Battle against Intel CPU Monopoly Moves from Lower Costs to Faster IA-32 Processors.

Giovana Mendes

Dep. Informática, Universidade do Minho
4710-057 Braga, Portugal
gimendes@uol.com.br

Abstract. The battle among IA-32 processor's manufacturers goes on, with Intel launching the new Pentium 4, AMD with the Athlon, and, a little more modest, Cyrix with its C3 Samuel 2. This communication aims to analyse the novelties in these launchings, namely in multimedia technologies - SSE2 and 3Dnow - and at the micro-architecture level - Netburst and Quantispeed. Remarks on the cost/performance ratio close the communication.

1 Introduction

CPUs have gone through many changes through the past years since Intel came out with the first microprocessor. IBM chose Intel's 8088 processor for the CPU of the first personal computer. This choice by IBM is what made Intel the supposed leader of the CPU market. Intel remains the leader of microprocessor development. They usually come out with the new ideas first. Then companies such as AMD and Cyrix come in with their versions, usually with some minor improvements and slightly faster.

Nowadays Intel does not lead alone the market, it has as main competitor AMD that is side to side in the releases of new processors.

The manufacturers throw their new processors betting in the innovation of the technologies and architecture, for not only to get the market but also to impress their competitors. At the same time, they use a great marketing strategy, showing that they have adapted today's technology. Along this communication we will show what really exists in each manufacturer technologies, namely the multimedia features and new architectures.

2 Architectures

2.1 Intel Pentium 4

The Intel® NetBurst™ Micro-Architecture. The design goals of Intel NetBurst micro-architecture are: (a) to execute both the legacy IA-32 code and applications based on single-instruction, multiple-data (SIMD) technology at high processing rates; (b) to operate at high clock rates, and to scale to higher performance and clock rates in the future. To accomplish these design goals, the Intel NetBurst micro-architecture has many advanced features and improvements over the Pentium Pro processor microarchitecture.

The major design consideration of the Intel NetBurst micro-architecture to enable high performance and highly scalable clock rates are as follows:

It uses a deeply pipelined design to enable high clock rates with different parts of the chip running at different clock rates, some faster and some slower than the nominally-quoted clock frequency of the processor. The Intel NetBurst micro-architecture allows the Pentium 4 processor to achieve significantly higher clock rates as compared with the Pentium III processor. These clock rates will achieve well above 1 GHz.

Its pipeline provides high performance by optimizing for the common case of frequently executed instructions. This means that the most frequently executed instructions in common circumstances (such as a cache hit) are efficiently decoded and executed with short latencies, such that frequently found code sequences are processed with high throughput.

It employs many techniques to hide stall penalties. Among these are parallel execution, buffering, and speculation. Furthermore, the Intel NetBurst micro-architecture executes instructions dynamically and outororder, so the time it takes to execute each individual instruction is not always deterministic. Performance of a particular code sequence may vary depending on the state the machine was in when that code sequence was entered.

The pipeline of the Intel NetBurst micro-architecture contain three sections:

- the in-order issue front end
- the out-of-order superscalar execution core
- the in-order retirement unit.

The front end supplies instructions in program order to the out-of-order core. It fetches and decodes IA-32 instructions. The decoded IA-32 instructions are translated into micro-operations (µops). The front end's primary job is to feed a continuous stream of µops to the execution core in original program order.

The core can then issue multiple µops per cycle, and aggressively reorder µops so that those µops, whose inputs are ready and have execution resources available, can execute as soon as possible. The retirement section ensures that the results of execution of the µops are processed according to original program order and that the proper architectural states are updated.

Figure 1 illustrates a block diagram view of the major functional blocks associated with the Intel NetBurst micro-architecture pipeline. The paragraphs that follow Figure 3 provide an overview of each of the three sections in the pipeline. [2]

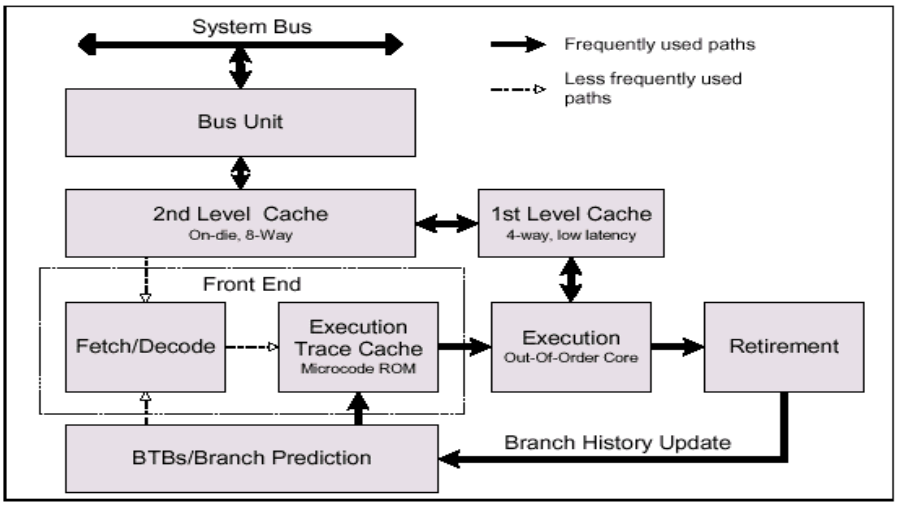

Fig.1. The Intel® NetBurst™ Micro-Architecture

2.2 AMD Athlon

QuantiSpeed™ Architecture for Enhanced Performance. A list of its main features follows:

- Nine-issue, superscalar, fully pipelined micro-architecture

At the heart of QuantiSpeed architecture is a nine- issue, superscalar, fully pipelined core.

This provides more pathways to feed application instructions into the execution engines of the core, simply allowing the processor to complete more work in a given clock cycle . The delicate balance between the depth of the pathways (pipelines) and operating frequency of the processor produces high levels of performance. Longer pipelines alone translate into lower IPC, but high operating frequencies. However, shorter pipelines alone result in increased IPC, but lower operating frequencies. The AMD Athlon XP processor is designed to maintain a balanced approach between pipeline depth and processor frequency to provide extraordinary levels of overall processor performance.

- Superscalar, fully pipelined Floating Point Unit (FPU)

QuantiSpeed architecture features a superscalar, fully pipelined FPU, which completes more floating point operations per clock cycle than competitive x86 processors and permits high operating frequencies. The end result is the most powerful x86 FPU available today. The AMD Athlon XP processor has ample computing power to tackle the most computation intensive software applications.

- Hardware data prefetch

Prefetching instructions from system memory to the processor's level 1 instruction cache is a common technique used to increase the processor's work throughput (IPC) and therefore overall performance. This feature of QuantiSpeed architecture prefetches data from system memory to the processor's level 1 data cache, which reduces the time it takes to feed the processor critical data, increasing work throughput. As a result, application performance is automatically enhanced when using the AMD Athlon XP processor with QuantiSpeed architecture.

- Exclusive and speculative Translation Look-aside Buffers (TLBs)

The TLB structures in QuantiSpeed architecture keep the maps to critical data close to the processor. This is designed to prevent the processor from stalling or waiting when future data is requested. These TLB structures are now larger, exclusive between caches, and speculative. Larger TLB's give the AMD Athlon XP processor access to additional data maps. The exclusive nature of these structures removes the duplication of information, freeing- up more space in the level 2 cache for other useful data to be used by the processor.[4]

2.3 Cyrix - C3 Samuel 2

The VIA C3 Samuel 2 processor architecture seems simple; instructions are issued, executed, and retired in order, only one instruction can be issued per clock, and most data cache misses stall the pipeline. However, the design is actually highly optimized and highly tuned to achieve high performance in the targeted environment. Some of the significant features providing this performance are:

- High internal clock frequency. The 12-stage pipeline facilitates high MHz
- Extensive features to further minimize bus stalls. These include:
    - Full memory type range registers (MTRRs)
    - A non-stalling write-allocate implementation
    - Implementation of the "cache lock" feature
    - Non-blocking out-of-order retirement of pipeline stores
    - Implementation of x86 prefetch instruction (3DNow!)
    - Implicit speculative data prefetch into D-cache
    - Aggressive implicit instruction prefetch into I-cache

- Highly asynchronous execution with extensive queuing to allow fetching, decoding and translating, executing, and data movement to proceed in parallel

- High-performance bus implementation.
  The VIA C3 Samuel 2 processor (socket 370) compatible bus implementation includes the following performance enhancements:
  - No stalls on snoops
  - Up to 8 transactions can be initiated by the processor
  - Aggressive write pipelining
  - Smart bus allocation prioritization
  - 100MHz and 133MHz bus operation
- Good performance for highly used instructions.
  Heavily used instructions—including complex functions such as protect-mode segment-register loads and string instructions—are executed fast.
  In particular, the pipeline is arranged to provide one-clock execution of the heavily used register– memory and memory–register forms of x86 instructions. Many instructions require fewer pipeline clocks than on comparable processors.[1]

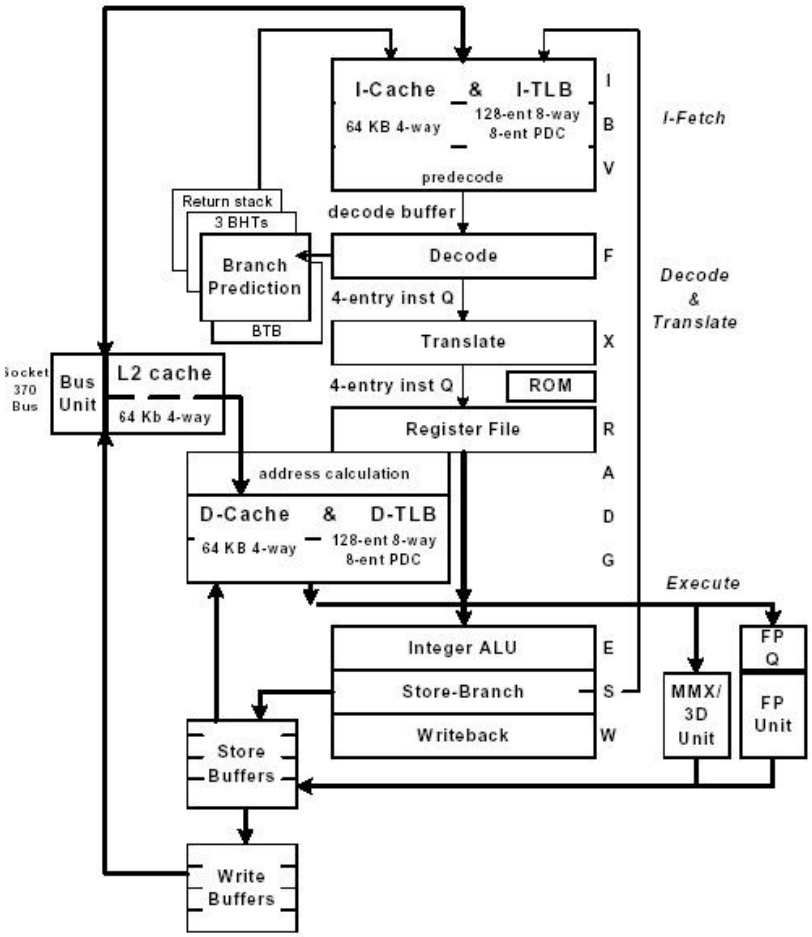

Fig.2. - The VIA C3 Samuel 2 Processor Pipeline Structure

3 Multimedia

3.1 Intel Pentium 4

Streaming SIMD Extensions 2 Instructions. The most important positive feature of this processor is the use of the second-generation instructions called SSE2 (Streaming SIMD Extensions 2).

This new instruction set supports new formats of packed data and increase the speed of manipulation of 128-bit SIMD integer operations.

Besides, this extension adds the ability to manipulate double precision floating point operations and several types of integer 128-bit operations. All new data types can be operated with in XMM registers.

There are some improvements that concern 68-bit SIMD integer instructions, which worked in Pentium II and Pentium III processors with 64-bit MMX-registers; in Willamette architecture they allow operating with XMM-registers. This goody will add some more flexibility when developing SIMD-code using both MMX and XMM-registers.

It's obvious, that these improvements will up the performance dramatically in such applications as audio and video coding/decoding, speech recognition, and will allow to obtain some increase in 3D-graphics. Scientific and engineering applications will see much of their speed boosts from the floating-point side of SSE2.

In addition to main SIMD instructions there brought in some new instructions allowing to control data caching. There is possible a preselection of data before they are wanted, and streaming data transfer from/to the registers without cache destroying.

Besides, SSE2 instructions are completely based on SSE, do not require OS support, and will probably prove to be the main ace of the Willamette processor.

3.2 AMD Athlon

3DNow! ™ Technology for 3D Operations. The 3DNow! Technology instructions are intended to open a major processing bottleneck in a 3D graphics application—floating-point operations. Today's 3D applications are facing limitations due to the fact that only one floating-point execution unit exists in the most advanced x86 processors. The front end of a typical 3D graphics software pipeline performs object physics, geometry transformations, clipping, and lighting calculations.

These computations are very floating-point intensive and often limit the features and functionality of a 3D application. The source of performance for the 3DNow! instructions originates from the single instruction multiple data (SIMD) implementation.

With SIMD, each instruction not only operates on two single-precision, floating-point operands, but the microarchitecture within the processor can execute up to two 3DNow! instructions per clock through two register execution pipelines, which allows for a total of four floating-point operations per clock. In addition, because the 3DNow! instructions use the same floating-point registers as the MMX™ technology instructions, task switching between MMX and 3DNow!operations is eliminated.

The 3DNow! technology instruction set contains 21 instructions that support SIMD floating-point operations and includes SIMD integer operations, data prefetching, and faster MMX-to-floating-point switching. To improve MPEG decoding, the 3DNow! instructions include a specific SIMD integer instruction created to facilitate pixel-motion compensation. Because media-based software typically operates on large data sets, the processor often needs to wait for this data to be transferred from main memory. The extra time involved with retrieving this data can be avoided by using the new 3DNow! instruc-

tion called PREFETCH. This instruction can ensure that data is in the level 1 cache when it is needed. To improve the time it takes to switch between MMX and x87 code, the 3DNow! instructions include the FEMMS (fast entry/exit multimedia state) instruction, which eliminates much of the overhead involved with the switch. The addition of 3DNow! Technology expands the capabilities of the AMD family of processors and enables a new generation of enriched user applications.[3]

3.3 Cyrix C3 Samuel 2

MMX UNIT. The VIA C3 Samuel 2 processor contains a separate execution unit for the MMX-compatible instructions.

MMX instructions proceed through the integer R, A, D and G stages. Thus, load-ALU x86 MMX instructions do not require an extra clock for the load. One MMX instruction can issue into the MMX unit every clock.

The MMX multiplier is fully pipelined and can start one non-dependent MMX multiply[-add] instruction (which consists of up to four separate multiplies) every clock.

Other MMX instructions execute in one clock. Multiplies followed by a dependent MMX instruction require two clocks.

Architecturally, the MMX registers are the same as the floating-point registers. However, there are actually two different register files (one in the FP-unit and one in the MMX units) that are kept synchronized by hardware.[1]

3DNOW! Unit. The VIA C3 Samuel 2 processor contains a separate execution unit for the 3DNow! instructions. These instructions are compatible with the AMD K6-II processor 3DNow! instructions and provide performance assists for graphics transformations via new SIMD single-precision floating-point capabilities.

3DNow! instructions proceed through the integer R, A, D and G stages. Thus, load-ALU x86 3DNow! instructions do not require an extra clock for the load. One 3DNow! Instruction can issue into 3DNow! unit every clock.

The 3DNow! unit has two single-precision floating-point multipliers and two single-precision floatingpoint adders. Other functions such as conversions, reciprocal, and reciprocal square root are provided.

The multiplier and adder are fully pipelined and can start any non-dependent 3DNow! instruction every clock.[1]

Table 1. 3DNow versus SSE

Functionality	3DNow!™	SSE	Conclusion
SIMD floating point functionality (an AMD first)	21 (original 3DNow! Instructions)	~52	Comparable functionality: both technologies support 4 FP operations per clock cycle and deliver up to 4.0 GFLOPs at 1000 MHz. But 3DNow! Is simpler to implement. SSE has many more instructions because Intel's architecture requires MMX control functionality to be duplicated and requires two FP instructions per extension: one for SIMD and another for scalar operations.
MMX (integer) augmentation and data movement	19 new instructions	19	Comparable functionality: Both have instructions for cache and streaming controls.
DSP/communication extensions	5 new instructions	0	AMD functionality: AMD advances SIMD with DSP extensions for soft modems, soft ADSL, complex math, MP3, and Dolby Digital.
Total number of instructions	45	71	Advantage AMD: Enhanced 3DNow! Has greater functionality than SSE. When combined with the superior FP engine in the AMD Athlon processor, the result is the best FP/multimedia performance in x86 processors.

4 Conclusion

In the middle of this battle, we can conclude that the manufacturers are innovating in the release of their new processors, trying to get the Maximum of the functionalities. With the appraised technologies we can notice that AMD have their processors faster those from Intel, and still have the best cost benefit for the final user. Intel is increasing every day the clock speed of their processors, and in the case of Pentium 4 it increased the number of pipelines to 20 and with that it gets higher clock speeds, but also get slower performance for some current applications, because the instructions have to pass for more processing stages. Even in the case of the technology used for the multimedia (SSE2) we verified that it is very efficient but since the applications that used it, were specifically developed for that technology.

This way seems inevitable that we have to wait for the applications optimised for Pentium 4, because while those do not use the characteristics of SSE2 efficiently, the superiority of Pentium 4 won't be significant in relation to competition.

References

[1] Via Technologies Inc.: VIA C3 Samuel 2 Processor Datasheet, http://www.via.com.tw/jsp/en/viac3/VIAC3_S2datasheet.pdf , April 2001

[2] Intel, A Detailed Look Inside the Intel® NetBurst™ Micro-Architecture of the Intel Pentium®4 Processor, ftp://download.intel.com/pentium4/download/netburstdetail.pdf (November 2001)

[3] AMD: AMD Extensions to the 3DNow! and MMX Instruction Sets Manual, http://www.amd.com/usen/assets/content_type/white_papers_and_tech_docs/ 22466.pdf , March 2000

[4] AMD, Quantispeed Architecture, White Paper, http://athlonxp.amd.com/includes/content/whitePapers/quantiSpeed.pdf, September 7, 2001

[5] AMD: AMD Athlon XP Processor Model 6 Data Sheet, http://www.amd.com/usen/assets/content_type/white_papers_and_tech_docs/24332.pdf , December 2001

[6] Intel: Intel® Pentium® 4 Processor Optimization Reference Manual , http://developer.intel.com/design/pentium4/manuals/248966.htm , November 2000

[7] Via Technologies Inc.: VIA Cyrix III Processor Datasheet VIA Confidential, http://www.viatech.com/products/viacyrixIIIdatasheet_1_2.pdf , November ,2000

[8] Sandfile.org - http://www.sandpile.org/ia32/

[9] Joseph F Kovar: Computer Reseller News, Manhasset, November 22, 1999

[10] AMD vs Intel , PC World Magazine nº230 Pg.76-78 , December 2001