How to reduce clock speed and increase performance

Oscar Manuel Lima Dias

Departamento de Informática, Universidade do Minho
4710 Braga, Portugal
odias@deb.uminho.pt

Abstract. Processor performance can be relatively estimated by multiplying the processor clock speed by the IPC, within a certain type of architecture. Normally performance is boosted by augmenting clock speed, as we have seen on the Pentium 4 series. But with Pentium M, clock speed was reduced, but performance was not. Pentium M architecture gave up clock speed up, decreasing the pipeline stages, relatively to Pentium 4, due to the later being more penalized with bubbles in the pipeline, and incorrect branch predictions. In fact, at similar clock frequency, Pentium III had a better performance than Pentium 4. Furthermore, Pentium M introduced some power-aware, performance-enhancer features like, micro-ops fusion, a dedicated stack manager, and an improved branch predictor. These new features, allied to its smaller pipeline, allowed it to work at lower frequencies, but often do better than its higher-frequency predecessor, the Mobile Pentium 4. In some occasions the Pentium M also performs competitively with the higher-frequency desktop Pentium 4 processors. It is physically impossible to provide data to a large number of pipelines, since the memory system is not powerful enough. Even with the existing pipelines there are a significant number of clocks missed, because data can not be transferred fast enough. To increase efficiency Intel has even decreased the pipeline stages to 14 in the new Intel Core architecture, giving up the clock speed escalation pursuit, proving that performance is not a clock speed side effect.

1. Introduction

A cluster like SeARCH, may have some heating issues, given that it has about 120 cpu's, on a confined space. Laptops also have the same setback. So the design for future clusters could be enhanced, using some features exploited to resolve laptops warm up problem.

Intel's approach to this problem as resulted in a processor with a reduced clock speed frequency and increased performance, the Intel Pentium M. The philosophy of its architects was the equilibrium between power consumption and performance, since lower power consumption means low temperature. Before Pentium M the mobile market was very small, thus desktop processors were just retrofitted to meet laptops power and thermal demands [1].

In section 2 Pentium M architecture will be discussed, in section 3 some performance studies will be presented and section 4 will present some conclusions.

2. Processor Architecture

Intel has been very careful about the disclosure of the pipeline allocated to Pentium M. It is supposed to be deeper than Pentium III 10 stages pipeline and smaller than the Pentium 4 twenty stages pipeline, since deeper pipeline consent higher clock frequencies and Pentium M aims to higher frequencies than old Pentium III. The number is probably closer to 10 than 20 stages, because with a longer pipeline, the CPU is less efficient at the same clock speed [2]. In fact, at similar clock frequency, Pentium III had a better performance than Pentium 4, as it will be seen in a study further on.

A shorter pipeline means fewer stages, since the instruction has to go trough less sections and each stage does less work (Figure 1). Fewer stages do not favour higher clock speed, since higher frequencies require more clock cycles to complete a stage, thus favouring a lower frequency. A shorter pipeline, combined with a lower clock speed, means a higher IPC (instruction per clock cycle), because the processor is able to complete more instructions each clock cycle.

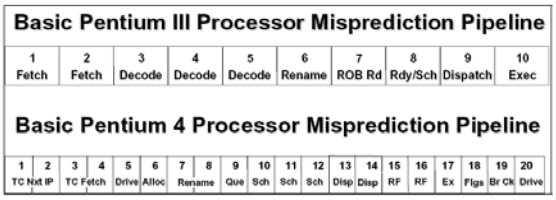

Figure 1. Pipeline structures, Pentium III and Pentium 4.

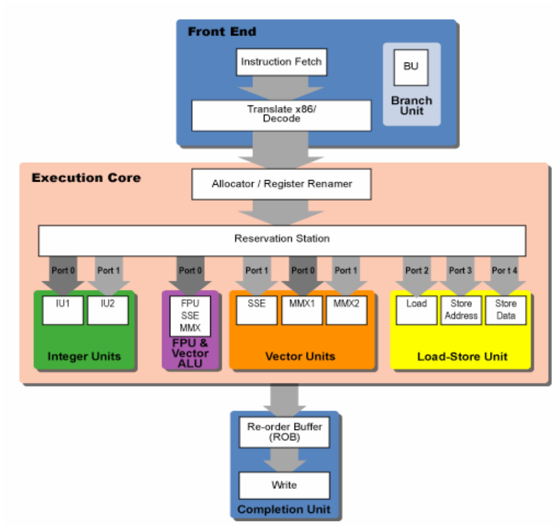

Figure 2. Pentium III internal architecture (in [8]).

A serious problem of a long pipeline is when an instruction stalls in order to wait for the execution of another instruction. That is called a bubble. A way to deal with bubbles is speculative execution, where microprocessors use branch prediction [4].

The IPC is the number of instructions completed for a clock cycle. Therefore, Pentium M microprocessor should belong to the *"Brainiacs"* family (high number of instructions per clock cycle and low clock speed), in opposition to Pentium 4 that belongs to the *"Speed Demons"* family (high clock frequency and low IPC) [3].

Clock speed was, for a long period of time, held responsible for performance enhancement, but with the dawn of this new microprocessor that premise was eradicated. Although, a higher clock speed permits Pentium 4 to have a high performance (*Performance = clock speed x IPC*), it increases the penalty when a branch is incorrectly predicted. The reason for that is that when such an error occurs it has to cleanup its pipeline, hence being so big it takes a lot time to do so. On the other hand, Pentium M, with its smaller pipeline, is not so penalized when the branch is incorrectly predicted. Though having a smaller pipeline than Pentium 4, Pentium M incorporates some branch predictor enhancements (Advanced Branch Prediction) that will be explained ahead.

Power aware, performance enhancement features

Since Pentium M has a slightly longer pipeline, than its P6 predecessor, some other changes were made in Pentium M in order to enhance its IPC, because a deeper pipeline implies a lower IPC. Most changes should have been made in the Front End, (Figure 2) like: micro-ops fusion, dedicated stack management and advanced branch prediction which includes a new branch predictor with an indirect branch prediction unit and a loop detector [8]. All of these features will be quickly explained next.

Advanced Branch prediction

When a branch is wrongly predicted performance declines. On a mobile CPU, an incorrect branch prediction does not only waste clock cycles, but also wastes battery power.

A method to reduce the penalty of a mistake on a branch prediction is the use of a trace cache, like in Pentium 4. A trace cache stores decoded micro-ops in their sequence of execution. This cache is called a trace cache because each line stores a snapshot, or trace, of the dynamic instruction stream (see [6] for more information about trace cache). Thus, on the event of a branch inaccurate prediction, the CPU can start, afterwards, in the pipeline. The problem with trace cache is large power consumption [7].

The advanced branch prediction in the Pentium M processor is based on the Pentium 4 processor branch predictor, but two other predictors were added, to distinguish special program flows: a Loop Detector and an Indirect Branch Predictor.

"The Loop Detector analyzes branches to see if they have loop behaviour. Loop behaviour is defined as moving in one direction (taken or not-taken) a fixed number of times interspersed with a single movement in the opposite direction" [1]. The loop detector works together with the advanced branch prediction unit and provides specialized data about the loops that exist in the program that is being executed [4, 7].

"The Indirect Branch Predictor chooses targets based on a global control flow history, much the same way a global branch predictor chooses the direction of conditional branches using global control flow history" [1].Unlike the branch predictor which stores information about if a branch was taken or not, the indirect branch prediction stores information about the destination of the branch. In fact the indirect branch prediction saves information about the favourite address of that particular branch [4].

Micro-ops fusion

Macro-instructions can be broken down into a sequence of one or more simple operations, called micro-operations, or micro-ops. A micro-op is the most basic operation of the microprocessor [1].

The micro-ops Fusion keeps the micro-ops united through most of the processor, but they are processed as non-fused operations, in order to preserve their single op properties, Figure 3. This process reduces the number of micro-ops, which helps to reduce the energy required to complete the instruction sequence, thereby minimizing power consumption [1, 5, and 7].

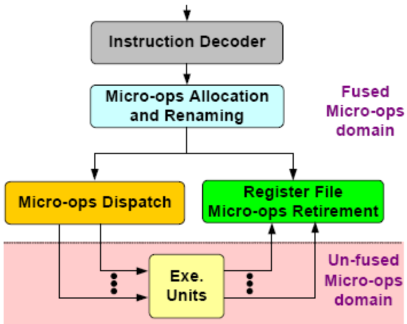

Figure 3. Fused Micro-ops Execution (in [1]).

One of the extra stages added by Intel to the Pentium M pipeline may be especially intended to deal with the micro-ops fusion [4].

Dedicated Stack Engine

Operations like *push* or *call* have an inner process, called a side effect, which updates de %esp register. Normally this update is done by including a micro-op to de instruction to add or subtract a direct value to the register, and then updating it. The new Pentium M uses a dedicated stack engine. This unit keeps track of the side effects and update the ESP_D register, where:

$$ESP_P := ESP_0 + ESP_D$$

ESP_0 is the momentous ESP value; ESP_D is the delta ESP that registers the variations of the ESP value during an instruction, and ESP_P is the ESP at a given moment in time. Since the ESP updating micro-ops have been eliminated, the ALU's are free to forthcoming instructions to the processor, increasing execution bandwidth. [1]. The result is fewer micro-ops (about

5%) per task, providing a lower IPC, increasing performance and reducing power consumption per task (a similar decrease of 5%) [5, 8].

3. Performance Analysis

The performance analysis will be divided in two categories: a benchmark study where Pentium III is compared to Pentium 4 at the same clock frequency, on a *“Tom’s Hardware”* benchmarking study, described in [11]; and desktop Pentium 4 versus Pentium M, a *Dell* benchmark, described in [2]. Additional and detailed information about these benchmarks can be found in the respective articles.

Pentium III vs. Pentium 4

In this benchmark study “*Tom’s Hardware*” underclocked a Pentium 4 down to 1 GHz. This study tested 8 benchmarks, and the results are presented in the next figure.

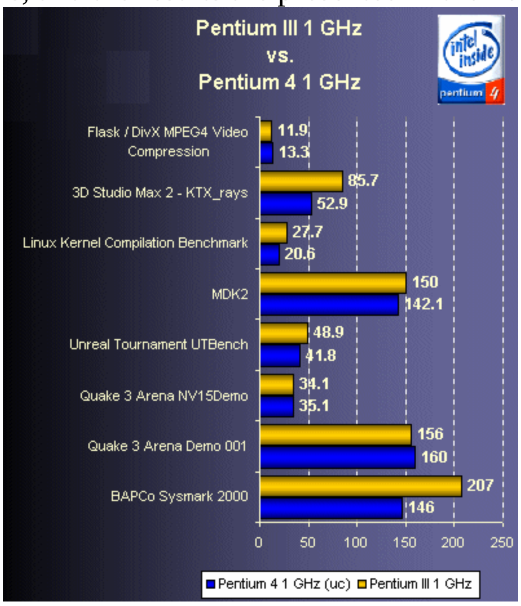

Figure 4. Comparison between Pentium III (yellow bar) and Pentium 4 (blue bar) (in [11]).

As it can be seen in Figure 4, Pentium III has a better performance in all tests, except the MPEG4 compression software, and Quake 3 arena. In SYSmark 2000 Pentium III is over 1.4 times superior than Pentium 4. In 3D Studio Max, Pentium III is 1.6 times superior than Pentium 4, every other benchmarks are more or less similar, with a slim advantage to Pentium III. With a smaller pipeline thus having smaller penalties for erroneous predictions or bubbles, Pentium III as a improved performance, clock for clock, than its “successor” Pentium 4, that is very penalized by incorrect predictions, despite its enhanced branch prediction, and trace cache.

Desktop Pentium 4 versus Pentium M

Dell Performance labs benchmarks are divided into three categories: office productivity, gaming, and CPU. The configurations were designed to isolate processor performance as much as possible by using identical hard drives and graphics cards.

The evaluated processors were:

- Intel Pentium M 2.13 GHz;
- Intel Pentium 4 3.8 GHz *HT*;

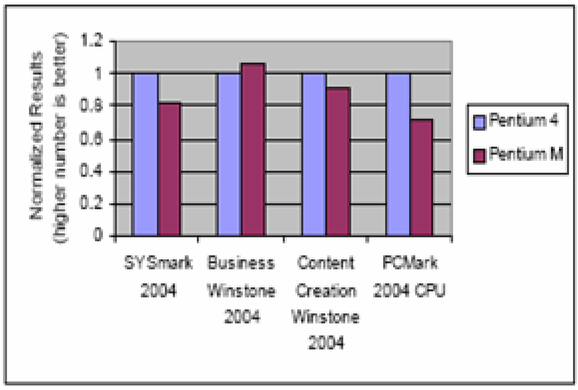

Figure 5. Office Productivity results (in [5]).

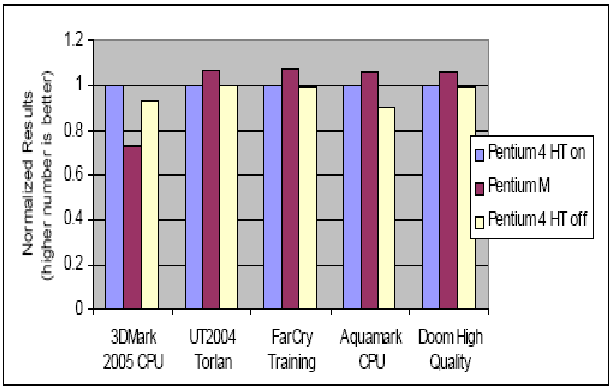

Figure 6. Gamming results (in [5]).

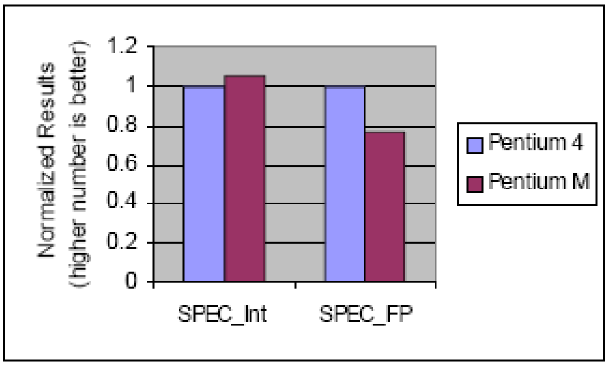

Figure 7. CPU results (in [5]).

Analysing Figure 5 we can see that Pentium 4 has the lead over Pentium M on the SYSmark 2004 and PC Mark CPU, because these applications are multithreaded and this Pentium 4 supports multi-threading. When multi-threading is disabled, the difference between the two processors decreases. But, on Business Winstone 2004, the larger cache of the Pentium M provides the advantage [5], however Content Creation Winstone 2004, gives advantage to Pentium 4 floating point unit.

The Pentium M exhibits very good gaming performance (Figure 6), as result of its larger cache and shorter pipeline. Only Aquamark and 3DMark05 CPU test, gaming benchmarks are designed to take advantage of multi-threading technology so the other benchmarks gain little performance with multi-threading.

SPEC CPU2000 (Figure 7) was the benchmark used to compare performance between processors. Pentium M cache, high IPC and shorter pipeline, gave it a little advantage on the integer benchmarks. The Pentium 4 was better on floating-point benchmarks due to its better floating-point unit, SSE performance, and higher clock speed [5].

4. Conclusions

The Intel Pentium M processor is the first microprocessor engineered purposely for the needs of mobile PCs. It offers performance while taking in consideration the thermal and energy constraints of mobile machines.

To have a huge pipeline depth, like in Pentium 4, though enhancing performance, has some constraints like high core temperature and high power consumption. It causes also, problems keeping the pipeline full, as it is physically impossible to provide data to a very large number of pipelines, given that the memory system is not powerful enough. Even with the existing pipelines there are a significant number of clocks missed, because data can not be transferred fast enough [10].

Pentium M was the first processor in which Intel decided to decrease the pipeline length relatively to their latest model, though keeping it deeper than Pentium III 10 stages. Pentium M deeper pipeline, relatively to P6 series, allows it to reach for slightly higher frequencies than Pentium III, (above 1.4GHz), but it will never reach the Pentium 4, 3.8GHz. Nor does it need, because, as it can be seen in "*Tom's Hardware*" benchmark tests, when a Pentium III and a Pentium 4 are working at the same clock rate, Pentium III has clearly the advantage.

Pentium M's pipeline will accommodate, as well, some performance enhancement features in order to increase the IPC like: advanced branch prediction, micro-operation fusion, and a dedicated stack engine. These unique features contributed to overcome mobile Pentium III and 4 [1]. Pentium M is also competitive on almost all benchmarks with a desktop Pentium 4. Gaming was the best category to Pentium M, were it outperformed his desktop counter partner. On all other tests Pentium M has proven to be very competitive, even though Pentium 4 was multithreaded and 1.67 GHz faster [5].

Within this point of view, it can be said that, Pentium M processor led the way on a new approach to performance enhancement on microprocessors. Intel has even decreased the pipeline stages to 14, in the new Intel Core architecture in order to increase performance [12]. Intel's new processors should include some of the power saving/performance enhancement features in order to reduce core temperature, and that would help "*supercomputers*" like the SeARCH Cluster. Intel's new approach to performance, giving up the clock speed escalation pursuit, reveals that performance can not be associated to clock frequency alone, but for a commitment between clock speed and IPC [11, 12].

References

[1] Gochman, Simcha; Ronen, Ronny; Anati, Ittai; Berkovits, Ariel; Kurts, Tsvika; Naveh, Alon; Saeed, Ali; Sperber, Zeev; Valentine, Robert C.; The Intel Pentium M Processor: Microarchitecture and Performance; Intel Technology Journal; volume 7; 2003;

[2] Pentium M review; www.cpuid.com/reviews/PentiumM/;

[3] Brainiac CPU; Wkipedia, The free Encyclopedia; http://en.wikipedia.org/wiki/Brainiac_CPU; September 2006;

[4] Stokes, John H.; The Pentium 4 and the G4e: an Architectural Comparisonz; http://arstechnica.com/articles/paedia/cpu/p4andg4e.ars/1; May 2001;

[5] Janakiraman, Deepak; Roesle, Chad; Performance Analysis of Intel Pentium M and Pentium 4 Processors; Dell White Paper; July 2005;

[6] Rotenberg, Eric; Bennett, Steve; Smith, James E.; Trace Cache: a Low Latency Approach to High Bandwidth Instruction Fetching; Proceedings of the 29th Annual International Symposium on Microarchitecture; Dec. 2-4; 1996; Paris, France;

[7] Shimpi, Anand Lal; Intel's Centrino CPU (Pentium-M): Revolutionizing the Mobile World; http://www.anandtech.com/showdoc.aspx?i=1800&p=1;

[8] Stokes, John H.; A Look at Centrino's Core: The Pentium M; http://arstechnica.com/articles/paedia/cpu/pentium-m.ars/1; February 2004;

[9] Pabst, Thomas; Intel's New Pentium 4 Processor; http://www.tomshardware.com/2000/11/20/intel/page21.html; November 2000;

[10] Karbo, Michael; PC arkitektur - i teori og praksis - english web version at http://www.karbosguide.com/books/pcarchitecture/chapter30.htm; IDG; 2002; Denmark;

[11] Gasior, Geoff; Intel's Pentium M 1.4GHz processor- More cache, less power; http://techreport.com/reviews/2003q3/pentiumm-1.4ghz/index.x?pg=1; August 2003;

[12] Torres, Gabriel; Inside Intel Core Microarchitecture; http://www.hardwaresecrets.com/article/313; April 2006;