A&QT-R 2002 (THETA 13)
2002 IEEE-TTTC - International Conference on Automation, Quality and Testing, Robotics
May 23 – 25, 2002, Cluj-Napoca, Romania

SIMULATION MODELS OF PIPELINING IN INTEL PENTIUM PROCESSORS

Elena Zaharieva-Stojanova

Technical University of Gabrovo, Bulgaria
4 H.Dimitar str., 5300 Gabrovo, Bulgaria,
phone +359 66 223 529, e-mail Zaharieva@tugab.bg

Abstract: In this paper some simulation models of the Intel Pentium processors' structure and working are represented. These models could be used for creation of a software simulator needed to show the Intel Pentium superscalar architecture. The objective of the new simulator creation is to show the base concepts of the Intel Pentium processors functionality by means of short assembler programs.

In the next stages of its development IA-32 Intel processor architecture introduces Intel Pentium processor, P6 (Pentium Pro, Pentium II, Xeon, Intel Seleron, Pentium III) and Intel Pentium 4 NetBurst architecture. The most important features of these architectures are given.

This paper also represents the model of Pentium processor architecture and the base parts of the software simulator showing the main concepts of superscalar implementation. It not necessary to simulate exactly all work processes in the Pentium processor. The objective is creation of a simulator, which is able to show the pipelining in a superscalar architecture using a real existing architecture as an example. It will be applied in the hi-school education to show the functionality of superscalar architecture.

Key words: pipelining**,** superscalar architecture, IA-32 architecture, Intel Pentium processor, and simulators.

1. INTRODUCTION

The development of up-to-date processors' architecture is clously related with instruction-level parallelism and superscalar concepts. The key points in superscalar execution are: to fetch more than one instruction at a time from the cache; to decide when instruction are independent and thus can be executed simultaneously; and to execute more than one instruction at one time [2].

The Intel Processors are among of most used processors in the computer systems. Developing IA-32 architecture, Intel Corporation introduces superscalar technique in the Pentium processor. The Intel Pentium is the first processor with superscalar architecture. The next steps of IA-32 architecture development are coming of P6 and NetBurst processor architectures. [5]

The simulation is a frequently used technique in computer architecture development. The software simulators could be used as a tool for studying these architectures and optimisation processes.

In this paper simulation models of the Intel Pentium processors' structure and functionality are represented. The existing demo programs, for example DynExec, show these principals just as overall picture. The objective of new simulator creation is to show the base concepts of the Intel Pentium processors working by means of short

assembler programs. This simulator could be used also for the evaluation of source code efficiency. It will be applied in the hi-school education to show how superscalar architecture works.

2. DEVELOPMENT OF SUPERSCALAR ARCHITECTURE IN INTEL PENTIUM PROCESSORS

2.1.The Intel Pentium processor architecture

The Intel Pentium processor is the first Intel processor with a superscalar architecture. It has two execution pipelines to achieve superscalar performance. Two 5-stages pipelines, known as U and V, together can execute two instructions per clock. In comparison with 80486, the on-chip first-level cache was doubled, with 8 KB devoted to code, and another 8 KB devoted to data. Branch prediction with an on-chip branch table was added to increase performance in looping constructs. The Pentium processor also implements 8-stages pipeline for float-point instructions [5].

The U pipeline is known as a main pipeline. It can execute all 80x86 and Pentium instructions. The V pipeline is used just for "simple" instructions [6]. The U and V pipelines have 5 stages: Instruction Fetch (IF), Decode1 (D1), Decode2 (D2), Execution (EX), and Write Back (WB) [2].

The stage IF fetches two instructions from the first-level (L1) cache. If there is an JUMP or CALL instruction (JMP, Jcc, and CALL), the branch prediction is activated. Second stage D1 consists of two parallel decoders. The Pentium processor determines whether two consequent instructions (K and K+1) can execute together. If it is possible, instruction K is loaded into U pipeline and instruction K+1 is loaded into V pipeline as shown on fig.1. The instructions K and K+1 are coupled. If it is not possible to execute instructions K and K+1 together, instruction K is loaded into U pipeline and instruction K+1 waits for the next instruction (K+2).

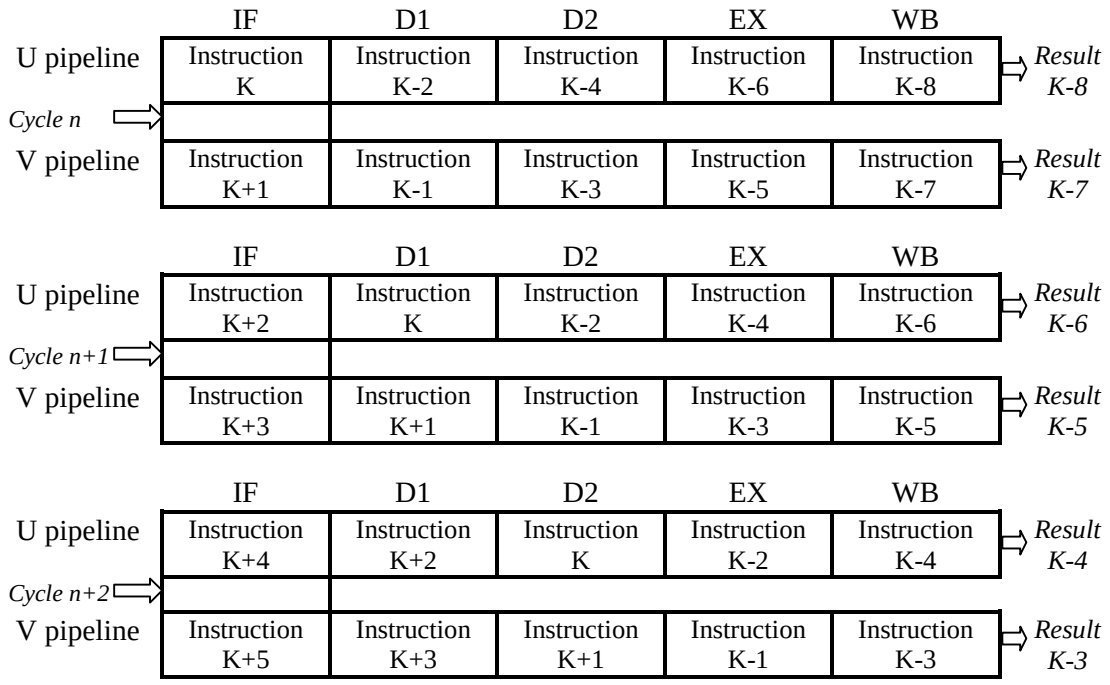

Fig. 1. Pipelining in the Intel Pentium processor

The stage D2 determines data addresses. The instructions run to this stage and they leave it together. The forth pipeline stage is Execution. The instructions run to this stage together but it is not necessary to leave it together. In this case, the instruction in U pipeline has to leave this stage first. The last stage is WB, where the results are written into registers and the flags are changed.

The results from the branch prediction could be controlled at the stage EX (for JMP and CALL instructions), or at the stage WB (for conditional branches).

2.2. The P6 family micro-architecture

The next step of the Intel superscalar architecture development is the P6 family of processors. This processor family was based on a new superscalar micro-architecture that established new performance standards. Members of the P6 processor family are: Intel Pentium Pro, Intel Pentium II, Xeon, Intel Celeron, Intel Pentium III processors.

The micro-architecture pipeline of the P6 processor is divided into four sections: the 1st level and 2nd level caches, the front end, the out -of-order execution core, and the retire section. Instructions and data are supplied to these units through the bus interface unit. The second-level (L2) cache was called Advanced Transfer Cache [3].

This micro-architecture is a three-way superscalar, pipelined architecture. It means that using parallel processing techniques, the processor is able on average to execute of three instructions per clock cycle. To handle this level of instruction throughput, the P6 processor family uses a decoupled, 12-stage superpipeline that supports out-of-order instruction execution. Three instruction decode units worked in parallel to decode object code into smaller operations called *micro-ops*. These micro-ops are fed into an instruction pool, and can be executed out of order by the five parallel execution units: two integer, two FPU and one memory interface unit. The Retirement Unit retires completed micro-ops in their original program order, taking account of any branches [3].

P6 processor architecture also introduced the concept of ***dynamic execution*** to reach out-of-order execution in a superscalar implementation [5]. Dynamic execution incorporates three data-processing concepts:

- ***Deep branch prediction*** - the P6 processor family implements highly optimised branch prediction algorithm to predict the direction of the instruction stream through multiple levels of branches, procedure calls, and returns.
- ***Dynamic data flow analysis*** - it involves real-time analysis of the flow of data through the processor to determine data and register dependencies and to detect opportunities for out-of-order instruction execution.
- ***Speculative execution*** - it refers to the processor's ability to execute instructions that lie beyond a conditional branch that has not yet been resolved, and ultimately to commit the results in the order of the original instruction stream.

2.3. The Intel Pentium 4 NetBurst architecture

The Intel Pentium 4 processor is the first based on the Intel NetBurst micro-architecture. This architecture provides the following important features: *Rapid execution engine* (ALUs run at twice the processor frequency and basic integer operations executes in 1/2 processor clock tick); *Hyper pipelined technology*; *Advanced dynamic execution; Advanced branch prediction algorithm; New cache subsystem; Full-speed, unified 8-way 2nd-Level on-die Advance Transfer Cache* [5].

The NetBurst micro-architecture pipeline is made up of three sections [4]:

- ***The Front End Pipeline*** **-** It performs several basic functions: Prefetch IA-32 instructions that are likely to be executed; Fetch instructions that have not already been prefetched. Decode IA-32 instructions into micro-operations. Generate microcode for complex instructions and special-purpose code. Deliver decoded instructions from the execution trace cache. Predict branches using highly advanced algorithm.
- ***The Out-of-order Core*** **-** This feature enables the processor to reorder instructions so that if one µop is delayed while waiting for data or a contended execution resource, other µops that are later in program order may proceed around it. This implies that when one portion of the pipeline experiences a delay, that delay may be covered by other operations executing in parallel or by the execution of µops which were previously queued up in a buffer.
- ***Retirement*** **-** The retirement section receives the results of the executed µops from the execution core and processes the results so that the proper architectural state is updated according to the original program order. For semantically correct execution, the results of IA-32 instructions must be committed in original program order before it is retired. Exceptions may be raised as instructions retired. Thus, exceptions cannot occur speculatively, they occur in the correct order, and the machine can be correctly restarted after an exception.

3. DESIGN OF SIMULATION MODELS FOR SUPERSCALAR PENTIUM ARCHITECTURE

According to the overview, IA-32 processor architecture development can be divided into three stages:

- superscalar architecture of the Pentium processor;
- P6 architecture;
- NetBurst architecture.

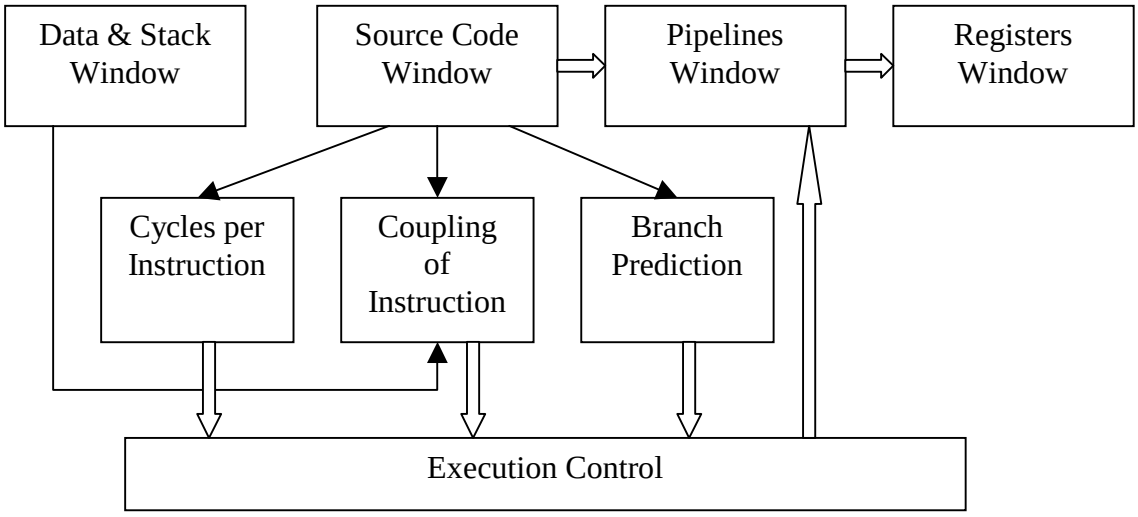

Fig. 2. Simulation Models of Intel Pentium Architecture

Including of the three types of architectures in the future software simulator would illustrate the whole Intel Pentium IA-32 architecture development. Furthermore, this approach has the possibility of comparison of different types of micro-architecture.

In this paper the model of the first Pentium processor will be introduced. Fig. 2 shows the simulation models of this processor. The Graphics User Interface (GUI) consists of the following forms:

- ***Source Code Window*** - it shows the source code and the currently executed instructions in the code segment. For example:

```
PUSHFD
POP EAX
MOV EBX,EAX
MOV ECX,EAX
XOR EAX, mask
PUSH EAX
POPFD
…
```

- ***Data & Stack Window*** - it represents the contents of the current data segment and the stack segment.
- ***Pipelines Window*** - this is a graphics representation of two pipelines work for the next 5 -10 cycles as shown in fig. 3. It is necessary to show the Stalls in pipelines, too.
- ***Register Window*** - it shows the contents of general-purpose registers: EAX, EBX, ECX, EDX, EBP, ESI, EDI.

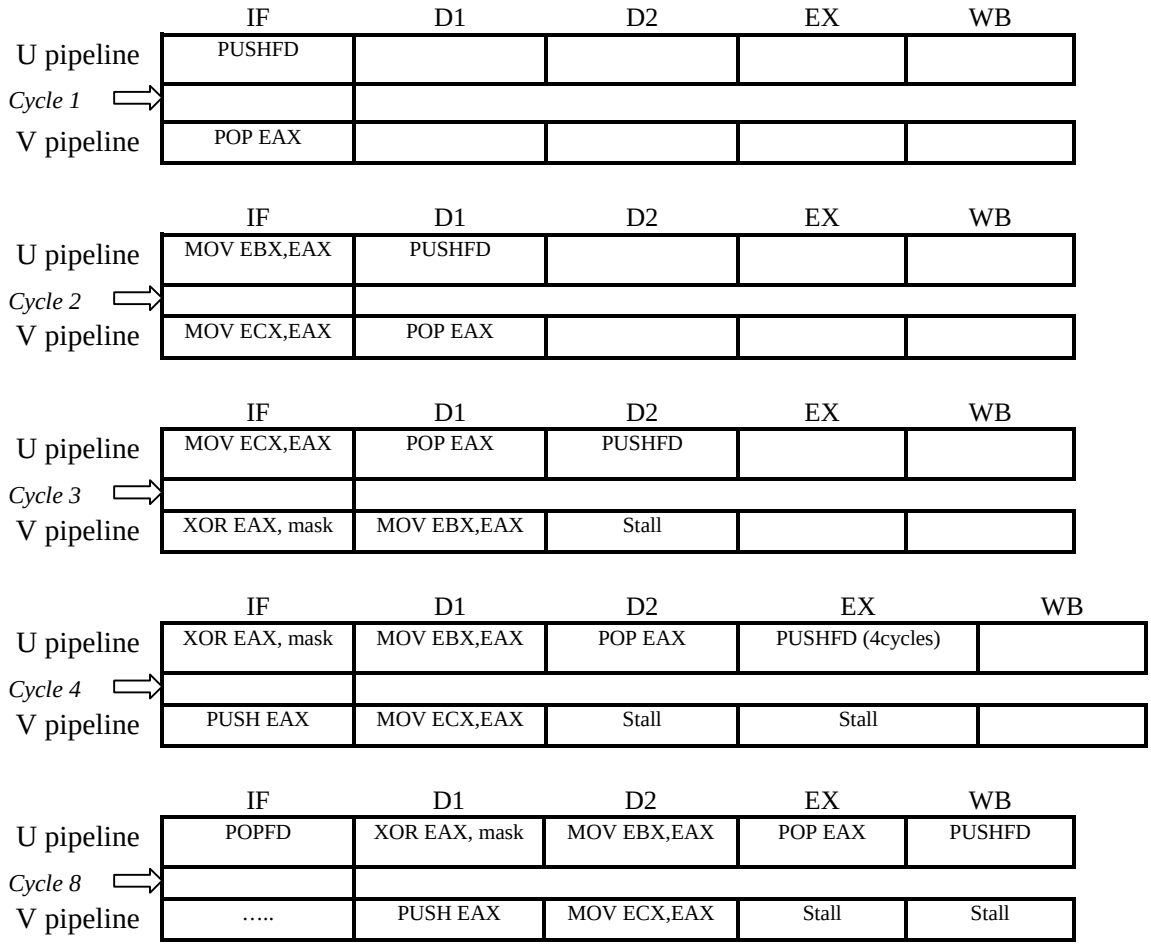

Fig. 3. Example of a Pipelines' Window

Except for the programs supported and controlled the GUI, the program system includes also the programs:

- ***Cycle per Instruction*** - it defines the number of the cycles needed for the execution of the current instruction.
- ***Coupling of Instruction*** - it defines whether two instructions can be coupled and executed simultaneously avoiding data hazard.
- ***Branch Prediction*** - this program simulates the functioning of branch prediction. If there is a jump, a branch, or a call instruction, branch prediction proceeds to predict if the branch shall be taken or not. Branch prediction logic goes through following states, as shown on the diagram:

highly non-taken branch -> lowly non-taken branch -> lowly taken branch -> highly taken branch

4. CONCLUSION

In this paper the simulation model of the of the Intel Pentium processor is represented. This model could be used for creation of a software simulator needed to show the Intel Pentium superscalar architecture.

This kind of simulator would be very useful in hi-school education to illustrate the pipelining in a superscalar architecture. Moreover, it shows a real existing often-used processor as Intel Pentium.

Simulating and showing all processes related with Pentium processor working at real-time is too hard task and it is not necessary for the objective of this research. The objective is creation of a simulator, which is able to show the pipelining in a superscalar architecture using a real existing architecture as an example.

The next step of this research is expanding the model with concepts of dynamic execution and hyper pipelined technology.

5. REFERENCES:

[1] Hinton G., D. Sager, M. Upton, D. Boggs, D.Carmean , A. Kyker, P. Roussel, (2001), The Micro architecture of the Pentium 4 Processor, *Intel Technology Journal Q1*.

[2] Hlavicka J., (1999), Computer Architecture, CVUT Publishing house.

[3] Keshava J., Vl. Pentkovski, (1999), Pentium III Processor Implementation Tradeoffs, *Intel Corp., Intel Technology Journal Q2*.

[4] A Detailed Look Inside the Intel NetBurst Micro-Architecture of the Intel Pentium 4 Processor, (2000), Intel Corporation.

[5] IA-32 Intel Architecture Software Developer's Manual, (2001), Intel Corporation.

[6] Pentium, (1998), NiSoft Ltd.