Digital Systems Architecture

Intel Pentium 4 Processor

Jehanzeb Hameed
Sylvester Onyishi
The Quote

“You say it best when you say nothing at all”
Boyzone, Ronan Keaton…??
Outline

• Introduction
• Instruction Set Architecture (ISA)
• Pentium 4 micro-architecture
• Pipelining
• Dynamic Scheduling
• Branch Prediction
• Memory System
Introduction

• Intel markets it as a desktop processor.
• Intel Pentium 4 processor is the latest IA-32 processor equipped with the full set of IA-32 SIMD operations
• New micro-architecture which is called “NetBurst” by Intel
## Some Boring Facts

<table>
<thead>
<tr>
<th>Intel Processor</th>
<th>Date Introduced</th>
<th>Micro-architecture</th>
<th>Clock Frequency at Introduction</th>
<th>Transistors per Die</th>
<th>Register Sizes¹</th>
<th>System Bus Bandwidth</th>
<th>Max. Extern. Addr. Space</th>
<th>On-die Caches²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pentium III processor³</td>
<td>1999</td>
<td>P6</td>
<td>700 MHz</td>
<td>28 M</td>
<td>GP: 32</td>
<td>Up to 1.06 GB/s</td>
<td>64 GB</td>
<td>32KB L1; 256KB L2</td>
</tr>
<tr>
<td>Pentium 4 processor</td>
<td>2000</td>
<td>Intel NetBurst micro-architecture</td>
<td>1.50 GHz</td>
<td>42 M</td>
<td>GP: 32</td>
<td>3.2 GB/s</td>
<td>64 GB</td>
<td>12K µop Execution Trace Cache; 8KB L1; 256KB L2</td>
</tr>
</tbody>
</table>
Clock Rates!!

Relative Frequencies of Intel’s Processors (Cores)
Instruction Set Architecture

• Extension of Pentium 3 ISA (called IA-32)
• Streaming SIMD Extensions 2 (SSE2): Extends the SIMD capabilities of Intel MMX technology
  – 144 new instructions that perform 128-bit SIMD integer arithmetic operations
  – 128-bit SIMD double-precision floating-point (FP) operations.
But what is the P3 ISA?

- The IA-32 ISA used in the Pentium 4 is the same base ISA that was used by Intel in the first generation 80x86 processor, the 8086, in 1978.
- IA-32 is a CISC-like, general purpose register architecture.
- P4 decodes IA-32 instructions into basic operations called uops.
The NetBurst Micro-Architecture

• Goals
  – Execute legacy IA-32 and SIMD applications
  – Operate at high clock rates that will scale easily in the near future.
Main Features

- Hyper-Pipelined Technology
- Advanced Dynamic Execution
- Rapid Execution Engine
- Execution Trace Cache
- 533 MHz Front Side Bus
- Advanced Transfer Cache
- Streaming SIMD Extensions 2 (SSE2) instructions
- Hyper-Threading Technology
The "Hyper" Pipeline

### Basic Pentium III Processor Misprediction Pipeline

<table>
<thead>
<tr>
<th></th>
<th>Fetch</th>
<th>Fetch</th>
<th>Decode</th>
<th>Decode</th>
<th>Decode</th>
<th>Rename</th>
<th>ROB Rd</th>
<th>Rdy/Sch</th>
<th>Dispatch</th>
<th>Exec</th>
</tr>
</thead>
</table>

### Basic Pentium 4 Processor Misprediction Pipeline

<table>
<thead>
<tr>
<th></th>
<th>TC Nxt IP</th>
<th>TC Fetch</th>
<th>Drive</th>
<th>Alloc</th>
<th>Rename</th>
<th>Que</th>
<th>Sch</th>
<th>Sch</th>
<th>Disp</th>
<th>Disp</th>
<th>RF</th>
<th>RF</th>
<th>Ex</th>
<th>Flgs</th>
<th>Br Ck</th>
<th>Drive</th>
</tr>
</thead>
</table>
Wire Delay

- Stages 5 and 20 used just to drive signal through the pipeline!!!
- Consequence of high clock rate
The Pipeline
Basic Block Diagram
The Front End

- Prefetches instructions that are likely to be executed
- Decodes instruction into \( \mu \)ops
- Generates \( \mu \)ops for complex instructions or special purpose code
- Predicts branches
The Front End

• Trace Cache
  – Primary Instruction Cache
  – Delivers upto 3 uops per cycle

Microcode ROM
  – Used for complex instructions which require more than 3 uops
Branch Prediction in P4

- Used when Trace Cache miss occurs
- Incorporates a BTB and Branch History Table
- The branch history table has 4K entries
- If no dynamic prediction is available, statically predict branches
  - Taken for backwards looping branches
  - Not taken for forward branches
The Execution Engine

• Out-of-Order Execution
• Designed to optimize performance by handling the most common operations in the most common context as fast as possible
• 126 $\mu$ops can in flight at once
Execution Engine – The Allocator

• Performs reordering of the instructions to exploit ILP
• Employs a Reorder Buffer (ROB) to track completion status of uops
• Also allocates one of the 128 floating point or integer registers to which the uop will write the result
Register Renaming

Pentium III

NetBurst

RF

ROB

Data

Status

Data

Status

Frontend RAT

Retirement RAT

EAX

EBX

ECX

EDX

ESI

EDI

ESP

EBP
Dynamic Scheduling

- Executes uops as soon as possible while maintaining dependencies
- Can dispatch up to 6 uops per cycle

Diagram:

- Exec Port 0:
  - ALU (Double Speed)
  - FP Move

- Exec Port 1:
  - ALU (Double Speed)
  - Integer Operation
  - FP execute

- Load Port:
  - Memory Load
  - All loads
  - LEA
  - SW prefetch

- Store Port:
  - Memory Store
  - Store Address

- Add/Sub Logic
  - Add/Sub

- Store Data Branches
  - Store Data Branches
  - FXCH

- FP/SSE Move
  - FP/SSE Move
  - FP/SSE Store

- FP/SSE-Add
  - FP/SSE-Add
  - FP/SSE-Mul
  - FP/SSE-Div
  - MMX
Pentium 4 is fun isn't it???

- Class about to end...but this is soooooooo interesting!!!
Retirement and Exceptions

- Can retire 3 μops per cycle
- Precise exceptions
- Reorder buffer to organize completed μops
- Also keeps track of branches and sends updated branch information to the BTB
Memory Subsystem

System Bus

Bus Unit

3rd Level Cache
Optional, Server Product Only

2nd Level Cache
5-Way

1st Level Cache
4-way

Front End

Fetch/Decode

Trace Cache
Microcode ROM

Execution
Out-Of-Order Core

Retirement

BTBs/Branch Prediction

Branch History Update

Frequently used paths

Less frequently used paths
Data Speculation

- The load scheduler to execution time, is longer than the load execution latency itself.
- The uop schedulers dispatch dependent operations before the parent load has finished executing.
- Assumes that the load will hit the L1 data cache.
- *Replay* logic tracks and re-executes instructions that use incorrect data.
400 MHz System Bus

• 3.2 Gbytes per second of bandwidth
• Quad pumps the 100 MHz bus
• Bus protocol allows 64 bytes access length
Hyper-threading

- Search for parallelism in “threads” instead of ILP
Ending Quote

“ No one will need more than 637 kb of memory for a personal computer “

Bill Gates
References

- The Microarchitecture of the Pentium 4 Processor, Glenn Hinton et al
- Intel Pentium 4 and NetBurst Micro-Architecture, Jason Waltman
- Inside the Pentium Pentium ® 4 Processor 4 Processor , Doug Carmean
- Intel Pentium 4 Processor, Steve Kelley and Zhijian Lu