IA-32

Nicholas FitzRoy-Dale

“At the forefront of the computer revolution” - Intel

“Difficult to explain and impossible to love” - Hennessy and Patterson

Datapoint 2200

- Released 1970
- 2K shift register main memory
- CPU: ~100 TTL components
- Instruction set implemented by Intel in the 8008

History

DP2200, 8008

8088, 8086, 81016, 80286

80386

i486

Pentium

Pentium Pro

Precursors to x86

Before IA-32: "x86". Rather dull.

IA-32. 32-bit addressing. Paging.

On-chip L1. Pipelining.

Superscalar! Branch prediction. SMP.

P6 microarchitecture. Dynamic execution. Speculation.

SMP.

Pentium II SIMD extensions.

Pentium III SIMD extensions #2.

Pentium 4 Netburst microarchitecture. SIMD extensions #3 and #4, virtualisation, 64-bit support

Pentium M microarchitecture.

SIMD extensions try #5.
Legacy

→ CISC ISA
→ Lack of registers
→ Four modes of operation
→ Segmentation
→ Superscalar
→ Complex addressing modes

Microarchitectures

→ New one “every two years”
→ Focus on P6

P6 microarchitecture

→ As RISC as possible
→ Register renaming
→ Superscalar
→ Out-of-order execution
→ Speculation
P6: As RISC as possible

- L1 I-cache
- Complex instruction decoder
- Simple instruction decoder 0
- Simple instruction decoder 1
- µ-op sequencer
- µ-op queue

Instruction decode

- Pentium 4: Trace cache
- Pentium M, Core: µ-op fusion
- ARM: Single instruction per clock
- IA-64: Two identical decoders (currently)

P6: “Register renaming”

- Performed by ROB
- 40 physical registers in RAT
- ARM: Lots of GPRs
- IA-64: Lots of GPRs and register rotation

Avoiding stalls

- Out-of-order execution
- Branch prediction
- Speculation
Out-of-order execution

P6: Superscalar

Superscalar: Netburst

Superscalar

→ ARM:
→ Varies per-core:
→ None
→ Separate ALU, MAC and LS (ARM11)
→ IA-64:
→ 2 I-units, 2 F-units, 3 B-units
→ But very few templates avoid “split issue”
P6: Branch prediction

- Branch target buffer: 512 entries
- Branch history and predicted address
- Mispredicts: 10-15 cycles
- Static prediction
- ~90% hit rate
- Pentium M: Loop detection

P6: Speculation

- Reservation station
- ROB
- Execution units
- ROB
- Write

Registers

<table>
<thead>
<tr>
<th>32</th>
<th>15</th>
<th>7</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>EAX</td>
<td>AH</td>
<td>AL</td>
<td></td>
</tr>
<tr>
<td>ECX</td>
<td>CH</td>
<td>CL</td>
<td></td>
</tr>
<tr>
<td>EDX</td>
<td>DH</td>
<td>DL</td>
<td></td>
</tr>
<tr>
<td>EBX</td>
<td>BH</td>
<td>BL</td>
<td></td>
</tr>
<tr>
<td>EBP</td>
<td>BP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESP</td>
<td>SP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESI</td>
<td>SI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EDI</td>
<td>DI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EIP</td>
<td>IP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EFLAGS</td>
<td>FLAGS</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Accumulator
- Count
- Data
- Base of data
- Base of stack
- Stack pointer
- String source idx
- String dest idx
- Instruction pointer
- CPU flags
- Code segment
- Stack segment
- Data segment
- Extra data segment
- Extra data segment 2
- Extra data segment 3

Floating-point

- FP stack
- FP status

Source: Hennessy and Patterson
**MMX**

<table>
<thead>
<tr>
<th>FPR</th>
<th>MMX GP register</th>
</tr>
</thead>
<tbody>
<tr>
<td>79</td>
<td>MM0</td>
</tr>
<tr>
<td>63</td>
<td>MM1</td>
</tr>
<tr>
<td>0</td>
<td>MM2</td>
</tr>
<tr>
<td></td>
<td>MM3</td>
</tr>
<tr>
<td></td>
<td>MM4</td>
</tr>
<tr>
<td></td>
<td>MM5</td>
</tr>
<tr>
<td></td>
<td>MM6</td>
</tr>
<tr>
<td></td>
<td>MM7</td>
</tr>
</tbody>
</table>

**SSE, SSE2, SSE3, SSE4**

<table>
<thead>
<tr>
<th>FPR</th>
<th>SSE GP register</th>
</tr>
</thead>
<tbody>
<tr>
<td>127</td>
<td>XMM0</td>
</tr>
<tr>
<td></td>
<td>XMM1</td>
</tr>
<tr>
<td></td>
<td>XMM2</td>
</tr>
<tr>
<td></td>
<td>XMM3</td>
</tr>
<tr>
<td></td>
<td>XMM4</td>
</tr>
<tr>
<td></td>
<td>XMM5</td>
</tr>
<tr>
<td></td>
<td>XMM6</td>
</tr>
<tr>
<td></td>
<td>XMM7</td>
</tr>
</tbody>
</table>

**IA-32 Instructions**

**Instruction counts**

![Instruction counts chart](chart.png)
**Instruction counts**

The code

```
other_value = localnode->from_values[i];
ldr r4, [r0, #12]
ldr r2, [r4, lr, lsl #2]
mov ecx,DWORD PTR [eax+12]
mov eax,DWORD PTR [ecx+ebx*4]
cur_value -= coeff * value;
muld f0, f2, f1
sufd f3, f3, f0
mulsd xmm1,xmm2
subsd xmm0,xmm1
```

```compute_nodes()```

**SIMD**

- **MMX**: Integer only. Renames the FP registers
- **SSE**: 8 new registers, each with 4 ints or 4 single-precision floats. Cache control.
- **SSE2**: Lots more data types
- **SSE3**: Horizontal operations
- **SSE4**: Unknown
Floating point

→ 8087: Stack architecture
→ P III, P4, Core: SSE, SSE2, SSE3, SSE4

Thread-level parallelism

→ SMP
→ Pentium Extreme Edition
→ Core Duo
→ Hyperthreading

SMP

→ P III, 4, Core: MESI (Athlon: MOESI)
→ Bus arbitration via dedicated lines
→ Interrupts: Whichever gets them

Hyperthreading

→ Replicate register renaming, architectural registers
→ Partition re-order buffer
→ Share caches and execution units
→ Relatively large performance gain
Performance

- “Speed demon” P4 was disappointing
- Focus on ILP / TLP
- Dual-core designs
- Hyperthreading

The future of IA-32

- EMT64
- Virtualisable
- Multicore

seuQnoitőőőőőőőőőőő