Advanced Processor Architecture

Jin-Soo Kim (jinsookim@skku.edu)
Computer Systems Laboratory
Sungkyunkwan University
http://csl.skku.edu
Modern Microprocessors

More than just GHz

<table>
<thead>
<tr>
<th>CPU</th>
<th>Clock Speed</th>
<th>SPECint2000</th>
<th>SPECfp2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Athlon 64 FX-55</td>
<td>2.6GHz</td>
<td>1854</td>
<td>1782</td>
</tr>
<tr>
<td>Pentium 4 Extreme Edition</td>
<td>3.46GHz</td>
<td>1772</td>
<td>1724</td>
</tr>
<tr>
<td>Pentium 4 Prescott</td>
<td>3.8GHz</td>
<td>1671</td>
<td>1842</td>
</tr>
<tr>
<td>Opteron 150</td>
<td>2.4GHz</td>
<td>1655</td>
<td>1644</td>
</tr>
<tr>
<td>Itanium 2 9MB</td>
<td>1.6GHz</td>
<td>1590</td>
<td>2712</td>
</tr>
<tr>
<td>Pentium M 755</td>
<td>2.0GHz</td>
<td>1541</td>
<td>1088</td>
</tr>
<tr>
<td>POWER5</td>
<td>1.9GHz</td>
<td>1452</td>
<td>2702</td>
</tr>
<tr>
<td>SPARC64 V</td>
<td>1.89GHz</td>
<td>1345</td>
<td>1803</td>
</tr>
<tr>
<td>Athlon 64 3200+</td>
<td>2.2GHz</td>
<td>1080</td>
<td>1250</td>
</tr>
<tr>
<td>Alpha 21264C</td>
<td>1.25GHz</td>
<td>928</td>
<td>1019</td>
</tr>
</tbody>
</table>
Pipelining

- Sequential execution

- Pipelining (RISC)
Superpipelining

- **Superpipelining**
  - Subdivide each pipeline stage
  - Higher clock speed
Superscalar

- The execution stage has a bunch of different functional units.
- Execute multiple instructions in parallel
- Pentium: 2-way superscalar
Superpipelining + Superscalar

- 2-way: MIPS R5000
- 3-way: PowerPC G3/G4, Pentium Pro/II/III/M/4, Athlon
- 4-way: UltraSparc, MIPS R10000, PowerPC G4e, Alpha 21164 & 21264, Core 2 Duo
- 5-issue: PowerPC G5
Superpipelined Superscalar (2)

- Tackling instruction dependencies
  - Branch prediction + speculative execution
    - Mispredict penalty: 10 – 15 cycles in Pentium Pro/II/III
  - Instruction scheduling
    - In-order execution + compiler optimization
      » Rearrange the instructions at compile time.
      » Compiler can see further down the program than the hardware.
      » SuperSparc, HyperSparc, UltraSparc, Alpha 21064 & 21164
    - Out-of-order execution
      » Reorder instruction execution sequence in hardware at runtime
      » Register renaming reduces the dependency further.
      » MIPS R10000, Alpha 21264, POWER/PowerPC, Pentium Pro, Pentium 4, Core 2 Duo
Superpipelined Superscalar (3)

- ILP vs. Clock speed
Superpipelined Superscalar (4)

Core Microarchitecture

- 128 Entry ITLB
- 32 KB I-cache (8 way)
- Instruction Fetch Unit
- x86 Instruction Pre-Decode, Fetch Buffer
- Complex Decoder
- Simple Decoder
- Simple Decoder
- Simple Decoder
- ucode Sequencer
- 7+ Entry uop Buffer
- Register Alias Table and Allocator
- 96 Entry Reorder Buffer (ROB)
- 32 Entry Reservation Station
  - Port 0: 64 bit ALU, 128 bit SSE, ALU Shift Rotate
  - Port 1: 128 bit FADD, 128 bit FMUL, FDIV
  - Port 2: 128 bit SSE
  - Port 3: 64 bit ALU Branch
  - Port 4: Store Data
  - Port 5: Store Address, Load Address
- Internal Results Bus
- 32 KB dual ported D-cache (8 way)
- 256 Entry DTLB
- 4MB Shared L2 Cache (16 way)
- Retirement Register File (Program Visible State)
- Shared Bus Interface Unit

CSE2003: System Programming | Spring 2009 | Jin-Soo Kim (jinsookim@skku.edu)
VLIW

- **Very Long Instruction Word**
  - Compilers explicitly group instructions to be executed in parallel
  - Long instructions: often 128 bits or more
  - Easier to design CPUs: simpler decode/dispatch stage
  - Intel IA-64: Itanium I/II
SMT (1)

- Limitations in instruction-level parallelism
  - Instruction dependencies
  - Branch mispredictions
  - Cache misses
  - Interrupts and exceptions

- Simultaneous multithreading (SMT)
Intel Hyper-Threading Technology

- Utilizes thread-level parallelism
- Fill pipelines with the instructions from multiple threads running at the same time.
- An SMT processor appears as if it were multiple independent processors.
- Uses processor resources more effectively
Multi-core (1)

- **Multi-core**
  - Put two or more processor cores onto a single chip
  - Previously called CMP (Chip Multiprocessor)
  - Examples
    - AMD Opteron: dual-core (Apr. 2005)
    - AMD dual-core Athlon 64 X2: dual-core (May 2005)
    - Intel Core Duo, Core 2 Duo, Xeon: dual-core
    - Sun UltraSparc T1: eight-core, 32 threads (Nov. 2005)
    - Intel Core 2 Extreme Quad-core processor (Sep. 2006)
    - Intel Xeon X7460 Six-core processor (Sep. 2008)
    - Microsoft’s Xbox 360: triple-core PowerPC microprocessor
Multi-core (2)

Intel CPU Trends
(sources: Intel, Wikipedia, K. Olukotun)
Multi-core (3)

- **Memory wall**
  - CPU 55%/year, Memory 10%/year (1986 ~ 2000)
  - Caches show diminishing returns.

- **ILP wall**
  - Control dependency
  - Data dependency

- **Power wall**
  - Dynamic power $\propto$ Frequency$^3$
  - Static power $\propto$ Frequency
  - Total power $\propto$ The number of cores
Multi-core (4)

Performance

Power

Single-Core

Raise Clock (20%)

1.73x

1.13x

0.87x

Lower Clock (20%)

0.51x

Dual-Core

1.02x

More MIPS/watt

Source: Intel
Microprocessor Design

- **CISC**
  - Instructions take variable times to complete

- **RISC**
  - Simple instructions, optimized for speed
  - Same individual instruction latency
  - Greater throughput through instruction “overlap”

- **Superscalar**
  - Multiple instructions executing simultaneously

- **SMT**
  - Additional hardware resources (regs, PC, SP)
  - Each context gets processor for x cycles

- **VLIW**
  - “Superinstructions” grouped together
  - Decreased hardware control complexity
  - Need a good compiler

- **Multi-core**
  - Duplicate entire processors

---

Jin-Soo Kim (jinsookim@skku.edu)