The evolution of Itanium family

- Itanium : (2001) running at 800MHz on a 0,18micron process with a 4MBytes of L3 cache and 2,1GBytes/s of bandwidth

- Itanium2 : (2002) running at 1G Hz on a 0,18micron process with a 3MBytes of L3 cache and 6,4GBytes of bandwidth

- Itanium2 6M : (2004) running at 1,5G Hz on a 0,18micron process with a 6MBytes of L3 cache and 6,4GBytes of bandwidth
The Intel Itanium

Main features:

✓ Support for 64-bit addressing
✓ Reliability for mission critical applications
✓ Full IA-32 instruction set compatibility in hardware
✓ Scalability across wide range of operating systems and platforms

June 2007

The EPIC design

The Explicit Parallel Instruction Computing is employed by the Itanium processor to:

Shift the balance of responsibilities between software and hardware

Advantages:

• Powerful architectural semantics
• Global optimization by the software for scheduling scope → improvements of Instruction-Level Parallelism by the hardware

June 2007
The core pipeline

Three important aspects:

- **New level of parallelism**:
  - For enterprise and commercial codes: 6/8 instructions per clock
  - For scientific code: 12 parallels instructions
  - For digital content creation codes up to 20 parallels instructions

- **Dynamic Hardware**: dynamic features for run-time optimization

June 2007

Deep pipeline

The pipeline is based on 10 stages:

- QuickTime™ and a TIFF (Uncompressed) decompressor are needed to see this picture.

- ROT: instruction rotation
- REG: Register Read
- REN: register rename
- DET: Exception Detection
- WLD: word line decode

June 2007
How guarantee good performances?

- **Speculative fetches**: 16KByte, 4 way set-associative instruction cache is fully pipelined and can deliver 32 bytes of code.

- **Hierarchy of branch prediction**: the processor employs a hierarchy of branch prediction structures to deliver high-accuracy and low-penalty predictions across a wide spectrum of workloads.

- **Software-initiated prefetch**: instructions get prefetched from L2 cache into the IS Buffer.

June 2007

Predication and Speculation

To improve the performance in the Itanium were **eliminated** the branches and their misprediction penalties.

The processor can execute both branches simultaneously. Data can be loaded before the load command occurs in the instruction code → **Reduce memory latency**

June 2007
**Memory SubSystem**

The processor provides 3 levels of on package cache:

- At level 1 instruction and data caches are split in a 16KBytes, 4 way set associative and with a 32Byte line size

- At level 2 cache uses a 4-state **MESI** protocol for **multiprocessor coherence**

- At level 3 there is an on-package cache of 4MBytes that is 4-way set-associative and uses a 64Byte line-state. It communicates with the processor at core frequency of 800MHz using a 128-bit bus. Also at this level is used the MESI protocol.

**The Itanium 2 Processor**

**Main Features**

- 64-bit operations and register widths
- Parallel execution for instructions that don’t create data hazards

**Branch Prediction**

Two levels:
- L1, tightly coupled to the L1I cache.
- L2, provides the second storage level.

June 2007
Scoreboard and Hazards detection:
High performance in the face of L1D misses.

Cache: three levels
The L1D è 16-Kbyte, four way set-associative, physically addressed cache with a 64-byte line
The L2 è is a unified, 256-Kbyte, eight-way set-associative cache with a 128-byte line size
The L3 è is a unified, 3-Mbyte, 12-way set-associative cache with a 128-byte line size

System interface
operates at 200 MHz
Similar to that of the the Itanium (1) è scalability of several implementations

The Intel Itanium 6M

Main features:
• higher performance through increased frequency
• larger L3 cache è 6-Mbyte, 24-way set-associative on-die L3 cache

Architecture è similar to that of the previous implementation
• **Power reduction**
  - a 50 percent higher frequency
  - a 2 times larger L3 cache
  - 3.5 increase in transistor leakage

  Reduced the active power from 90 to 74 percent of the total.

• **Test, debug, and manufacturability features**
  - increasing the L3 cache repair from dual to quad redundancy
  - I/O loop-back feature enhancements

• **Cache**
  - L1 instruction and data caches are both 16-Kbyte, 4-way associative.
  - L2 cache is a unified 8-way, 256-Kbyte array;
  - L3 is a unified, 24-way, set associative, 6-Mbyte cache.

**June 2007**

---

**Thanks for the attention!**