Update for New Implementations

As new implementations of the Itanium architecture are announced, we attempt to post appropriate updates on the support page for *Itanium Architecture for Programmers: Understanding 64-Bit Processors and EPIC Principles* at `http://www.viika.com/itanium/`.

While we tried to write an architecture book that would be largely unaffected by future developments, Chapter 13 suggested why that goal might be difficult indeed.

This electronic appendix contains additional or revised information that should be read in conjunction with the book itself. Sections in this appendix are of the form U.4.5.1, etc., for ease of correlation with the text.

U.4.5.1 Itanium Cache Structures

When new implementations of the Itanium architecture were released in 2003 to 2006 (formerly code-named Madison), the *Intel Itanium 2 Processor Reference Manual for Software Development and Optimization* was revised. Certain previously published descriptions of the Itanium 2 implementations released in 2002 (formerly code-named McKinley) changed also. Table U.4–3 reflects these changes.

Table U.4–3 Characteristics of Itanium 2 Memory Hierarchy

Level	Total Size	Line Size	Type	Replace Policy	Write Policy	Load Latency Cycles*		
						Integer	Floating	Instruction
L1-I	16 KiB	64 B	4-way	LRU	n/a	n/a	n/a	1
L1-D	16 KiB	64 B	4-way	NRU	write-through	1	n/a	n/a
L2	256 KiB	128 B	8-way	NRU	write-back	5, 7, 9	6, 8, 10	7, 9, 11
L3	up to 9 MiB	128 B	up to 24-way	NRU	write-back	12–14+	21–23+	14+
Memory	up to 1 PiB					> 100	> 100	> 100

*See *Intel Itanium 2 Processor Reference Manual for Software Development and Optimization,* revised 2004, for details about implementation differences.

The dual-core Itanium 2 implementation first released in 2006, code name Montecito, has separate L2-I (1 MiB) and L2-D (256 KiB) caches and a unified L3 cache (12 MiB per core). Other characteristics of the caches are similar to entries in Table U.4–3, although certain special cases will perform better in practice.

U.13.3.1 Comparison to the Itanium 2 Processors

Table U.13–1 compares the first Itanium processor and two Itanium 2 processor implementations along several key dimensions. We will not define or discuss every entry in this table, but we wanted to present a diverse set of measures customarily used for comparisons in the industry, as well as a few Itanium-specific characteristics to contrast these implementations.

Table U.13–1 Selected Characteristics of Itanium Processor Implementations

	Itanium	Itanium 2	
Code name during development	Merced	McKinley	Madison
Year of market release	2001	2002	2003
Chip technology:			
speeds marketed	733, 800 MHz	0.9, 1.0 GHz	1.3 to 1.7 GHz
process feature size	180 nm (Al)	180 nm (Al)	130 nm (Cu)
transistor count	25×10^6	221×10^6	$>= 400 \times 10^6$
layers	6	6	6
die size			374 mm^2
operating voltage	1.6 V	1.5 V	1.3 V
power (including cache)	116–130 W	130 W	130 W*
cache bandwidth		32 GB/s	48 GB/s
Processor features:			
physical stacked registers	96	96	
RSE modes	only enforced lazy	only enforced lazy	
integer units	2 M and 2 I	4 M and 2 I	
memory units	2 load or store	2 load and 2 store	
parallel floating-point units	2	1	
Pipeline depth:			
integer	10	8	
floating-point	12	10	
Memory support:			
physical address bits	44	50	
virtual address bits	54	64	
data bus width	64 bits	128 bits	
maximum page size	256 MiB	4 GiB	
System bus:			

	Itanium	Itanium 2	
speed	133 MHz	200 MHz	up to 667 MHz
width	64 bits	128 bits	
bandwidth	2.1 GB/s	3.2 GB/s	up to 10.6 GB/s
Instruction differences:			
`brl` instruction	emulated in op sys	implemented	
`brp` instruction	useful	affects instruction prefetching	
`chk.a` or `chk.s` handling	needs the op sys	usually done by hardware	
L3 cache location	off-chip, in-package	on-chip	

*Lower for some chips with less cache and lower speeds.

The dual-core Itanium 2 implementation first released in 2006, code name Montecito, has a feature size of 90 nm, 1.7×10^9 transistors, 1.6 GHz processor speed, up to 533 MHz front-side system bus speed, and 104 W power dissipation. This new implementation introduces a small number of new instructions (see *Dual-Core Update to the Intel Itanium 2 Processor Reference Manual for Software Development and Optimization*).

Other new features include a capability to support up to two threads per core (temporal hyperthreading), and some support for virtualization technology (hosting of multiple operating systems on one multi-processor computer installation).

Threads share the functional units (datapath pipelines), but each has its own architectural state information (i.e., registers) and thus appears to an operating system as a complete processor. This minimizes overhead of context changes at the operating system level. Since the time delay for thread switching is only 15 cycles (shorter than all cache hit intervals), it will be L3 cache misses that primarily trigger thread switching sooner than the expiration of a thread's time quantum as assigned by the operating system.

U.13.5IA-32 Compatibility

The Itanium and Itanium 2 processors listed in Table U–13–1 contain a dedicated unit to decode and expand complex IA-32 instructions into simpler operations that can be carried out as special cases in the standard Itanium execution units. This strategy resembles the way in which later Pentium processor implementations themselves break many of the complex IA-32 CISC instructions into simpler operations that are passed to a RISC-like core in the CPU for execution

With the IA-32 capability, an Itanium computer can run 32-bit operating systems. More importantly, 64-bit operating systems on Itanium-based computers can launch and run either 32-bit applications or native 64-bit applications. (According to information from

Hewlett-Packard Company, an Itanium 2 processor in IA-32 mode offers a performance level comparable to a 300-MHz Pentium Pro.)

Intel Corporation has also developed the IA-32 Execution Layer (IA-32 EL) software to emulate the full IA-32 instruction set. This layer plugs into an operating system and offers improved performance over the intrinsic hardware capability of Itanium processors. (According to information from Intel Corporation, an Itanium 2 processor with the IA-32 Execution Layer added to the operating system offers a performance level comparable to a Xeon processor having the same nominal clock speed.)

The dual-core Itanium 2 implementation first released in 2006, code name Montecito, has dropped the hardware IA-32 compatibility mode. It seems unlikely that any from future Itanium processor implementations would revive that capability, since software emulation has proven to be superior.

U.13.6 Determining Extensions and Implementation Version

Once an architecture acquires extensions, the need arises for compilers, operating systems, and system boot-up code to be able to determine the characteristics of a particular machine configuration.

The Itanium architecture provides a set of `cpuid` registers (Appendix D.8) that software can read. The `cpuid` register 3 has 8-bit fields specifying the architecture revision, the processor family, the processor model number, the processor revision number, and the highest implemented index number for the `cpuid` registers (at least 4). Bits <31:24> encode the processor family and bits <23:16> encode the model number. Table U.13–6 correlates these code values to the code names used during development of the various processor implementations and to a primary differentiating characteristic, the maximum amount of on-chip L3 cache.

Table U.13–6 Identification of Itanium Processors

Family	Model	Maximum L3 Cache	Code Name	Description
`0x07`	`0x00`		Merced	Itanium
`0x1f`	`0x00`	3 MiB	McKinley	Itanium 2
`0x1f`	`0x01`	6 MiB	Madison	
`0x1f`	`0x02`	9 MiB		
`0x20`	`0x00`		Montecito	Dual-core Itanium 2

Fields in `cpuid` register 4 indicate application-level features as a set of flag bits for presence (1) or absence (0) of each feature. For example, bit 0 indicates whether the processor implements the `brl` instruction (e.g., Itanium 2 processor) or the operating system must provide emulation when that opcode causes an exception (e.g., initial Itanium processor). Similarly, bit 2 indicates whether the 16-byte atomic operations (`ld16`, `st16`, and `cmp8xchg16`) are implemented.

REFERENCES

Baraz, Leonid, Tevi Devor, Orna Etzion, Shalom Goldenberg, Alex Skaletsky, Yun Wang, and Yigal Zemach, *IA-32 Execution Layer: a two-phase dynamic translator designed to support IA-32 applications on Itanium-based systems*, 2004 [cited 30 September 2006]. Available from http://www.intel.com/cd/ids/developer/asmo-na/eng/dc/itanium/93086.htm.

Intel Corporation, *Dual-Core Update to the Intel Itanium 2 Processor Reference Manual for Software Development and Optimization,* 2006.

Intel Corporation, *Intel Itanium 2 Processor Reference Manual for Software Development and Optimization,* revised 2004.

Intel Corporation, *Intel Itanium Processor Family Reference Guide: IA-32 Execution Layer,* 2004 [cited 30 May 2004]. Available from http://www.intel.com/design/Itanium/Downloads/254318.htm.

Microsoft Corporation, *Microsoft Windows Server 2003*, "The IA-32 Execution Layer 4.3 Software Driver," 2004 [cited 30 May 2004]. Available from http://www.microsoft.com/windowsserver2003/64bit/ipf/ia32el.mspx.

Stinson, Jason, and Stefan Rusu, *A 1.5GHz Third Generation Itanium Processor,* ISSCC-2003, 2003 [cited 2 July 2003]. Available from http://developer.intel.com/design/itanium2/download/14_4_slides_r31_nsn.htm.