10.1 The Implementation of a 2-core Multi-Threaded Itanium®-Family Processor

Samuel Naffziger, Blaine Stackhouse, Tom Grutkowski

Intel, Fort Collins, CO

The next generation in the Itanium® processor family, code named Montecito is introduced. The processor has two dual-threaded cores integrated on die with 26.5MB of cache in a 90nm process with 7 layers of copper interconnect. The die is 21.5mm by 27.7mm and includes 1.72B transistors. With both cores at full frequency, it consumes 100W. The micro-architecture and circuit methodologies are leveraged from the prior Itanium2 processors[2]. Improvements include the integration of 2 cores on-die, each with a dedicated 12MB 3rd-level cache, a 1MB 2nd-level cache and dual-threading. Susceptibility to soft errors is also reduced and power efficiency improved through low-power techniques and active power management.

Multi-threading addresses the memory latency issue in large SMP systems. The Montecito style of multi-threading is referred to as temporal multi-threading (TMT) where the thread switch is triggered by long-latency events such as a 3rd-level cache miss. Duplicating all architectural state and some micro-architectural state creates the two logical processors that then share all the execution resources and the large caches. To the operating system, each Montecito die looks like 4 processors. The thread-switch logic and state is implemented without a cycle-time impact. At any given time, only one thread is active in the pipeline of each core and consumes all the physical execution resources while the caches continue to service both threads. The performance benefits of TMT on database workloads are estimated to range from 15 to 35% depending on memory latency. The integer and floating-point register files implement a duplicate set of registers in an area-efficient manner with a 15% area impact[2]. Other architectural state that must be duplicated requires dual-threaded latches (Fig. 10.1.5). The changes for TMT result in less than a 2% area growth in the core providing a very power- and area-efficient performance increase for SMP systems.

Soft errors caused by high-energy particles negatively affect system reliability resulting in customer downtime which is not acceptable for servers. Montecito provides several key enhancements over the previous Itanium processors. First, the parity protection of the 2nd-level data-cache tag structures is enhanced with ECC protection. This feature allows the transparent fixing of 86% of bits that would otherwise cause the processor to crash on a detected error. Second, the large L3 caches provide support for automatic disabling of individual cache lines that suffer from repeated errors. This is achieved by the combination of hardware and firmware support. Third, parity protection is added for the large register file structures (2 sets of 128 dual-threaded registers), with a low-overhead approach [2].

Power reduction is targeted as a primary design priority. At issue, is the fact that the last 3 generations of Itanium processors all had the same power consumption (130W) for a single-core processor despite process scaling. For Montecito to provide competitive database performance, 2 processor cores are integrated on-die and the cache is grown. Scaling indicates that this would consume nearly 300W in 90nm at the 2.0+GHz target frequency – clearly unacceptable. To reduce power while also providing a new level of power manageability, several capabilities are needed:
Power *in* should be managed in the processor, not just temperature (~Power *out*). Power *in* is a major driver of system and data center operating costs. Power *out* must be managed also as it affects heat-sink design and form-factor issues, but it is only half the problem.

The processor should be able to manage its power to a dynamically adjustable limit. This provides flexibility in system design and data-center management. The processor should also adapt its operating point for each power setting to extract the maximum *performance/watt* possible to maximize flexibility. This means maximizing frequency as a function of voltage.

In order to reduce power and provide the above capabilities, the Montecito processor applies the full range of traditional CMOS power-reduction techniques to the core design inherited from the previous generation. These include reducing leakage through long-L and high-V_t device insertion, gating-off clocks, and reducing FET width in over-designed circuits. The result is a power reduction of 28% for integer applications.

Montecito also targets the three main sources of power variability for removal: the difference between average application power and the maximum (1.4X at the chip level for Itanium2); the variability in leakage due to silicon processing which is at least 2X of a nominal 25% power component; and voltage variability due to regulator and package variations and transients that are ~10% of nominal.

Measuring power *in* required the invention of several new technologies: An ammeter is integrated into the chip and package to continuously monitor the power consumed by the die[3]. This integration not only ensures good accuracy (~3%), but also dovetails with existing manufacturing test flows and eliminates the dependence on external components. Next, an autonomous microcontroller is implemented on-die to enable the processor to flexibly manage the voltage of the chip in response to changing power consumption and thermal conditions[3]. Finally, the core design has adaptive frequency to continuously varying voltage which in turn requires several capabilities: First, an asynchronous interface between cores and system bus, where one cycle of meta-stability resolution time is achieved using a high-gain resolver latch (Fig. 10.1.6). Second, a frequency synthesizer that is repeatable (manufacturing / test needs) by being as digital as possible, has a very low frequency-change latency (1 cycle) for adapting to voltage transients, enables fine-grained frequency changes (1/64th of T_{min}), and has good range and low jitter[5]. The final capability is a core that robustly operates across a broad range of frequencies and voltages. This required hold-time analysis at two voltage points and checks for low-voltage circuit operation as well as an active clock deskew system[4] that continuously maintains clock alignment within 10ps across each core as voltage and temperature vary.

Acknowledgements:
Many thanks to a talented and dedicated team from Intel in making this ambitious design a reality.

References:

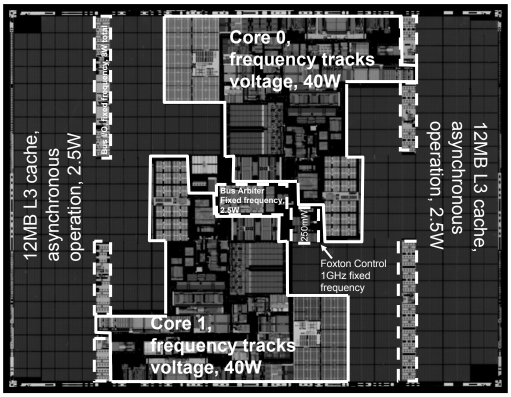
[1] S. Naffziger et al, "The Implementation of the Itanium2 Microprocessor," *IEEE J. of Solid-State Circuits*, vol. 37, no. 11, pp 1448-1460, Nov., 2002.

[2] E. Fetzer et al., "The Multi-Threaded, Parity-Protected, 128-Word Register Files on a Dual-Core Itanium®-Family Processor*," ISSCC Dig. Tech. Papers,* Paper 20.5*,* pp. 382-383, Feb., 2005.

[3] C. Poirier et al., "Power and Temperature Control on a 90nm Itanium®-Family Processor," *ISSCC Dig. Tech. Papers,* Paper 16.7, pp. 304-305, Feb., 2005.

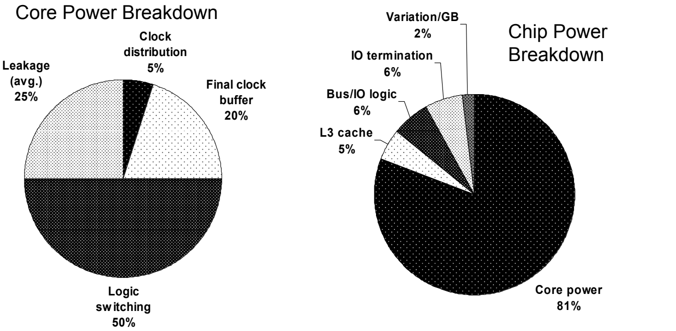
[4] E. Fetzer et al., "Clock Distribution on a Dual-core, Multi-threaded Itanium® Family Processor," *ISSCC Dig. Tech. Papers,* Paper 16.1, pp. 292-293, Feb., 2005.

[5] T. Fischer et al., "A 90nm Variable-Frequency Clock System for a Power-Managed Itanium®-Family Processor," *ISSCC Dig. Tech. Papers,* Paper 16.2, pp. 294-295, Feb., 2005.


Core 0, frequency tracks voltage, 40W
12MB L3 cache, asynchronous operation, 2.5W
12MB L3 cache, asynchronous operation, 2.5W
Bus IO, fixed frequency, 8W total
Bus Arbiter Fixed frequency 2.5W
250mW
Foxton Control 1GHz fixed frequency
Core 1, frequency tracks voltage, 40W

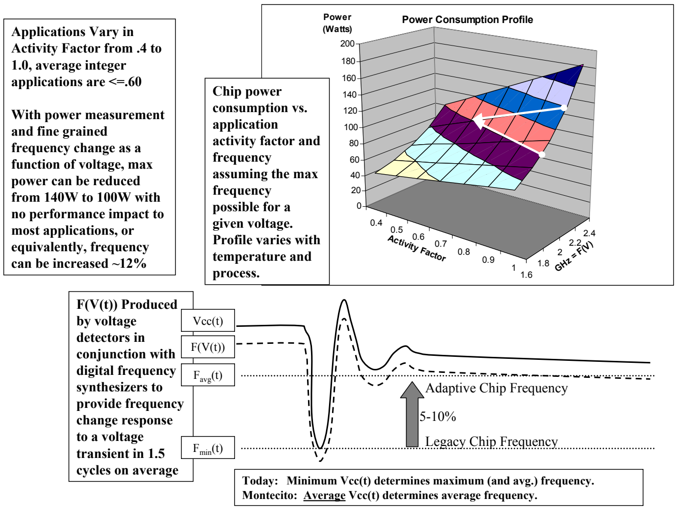
Domain	Voltage	Frequency	Power
Cores	.8 – 1.2V	Tracks voltage	80W total
Cache	.9-1V	Self-timed	5W total
IO term	1.2V	Bus	8W total
Fixed	1.15V	Fixed	6W total

Figure 10.1.1: Chip power consumption.



	FET count	low Vt
Core logic	57M	1.7%
Core caches	106.5M	0
L3 Cache	1550M	0
Bus logic & IO	6.7M	0.3%
Total	1.72G	.06%

Figure 10.1.2: Chip statistics at nominal operating point.


Applications Vary in Activity Factor from .4 to 1.0, average integer applications are <=.60

With power measurement and fine grained frequency change as a function of voltage, max power can be reduced from 140W to 100W with no performance impact to most applications, or equivalently, frequency can be increased ~12%

Chip power consumption vs. application activity factor and frequency assuming the max frequency possible for a given voltage. Profile varies with temperature and process.

F(V(t)) Produced by voltage detectors in conjunction with digital frequency synthesizers to provide frequency change response to a voltage transient in 1.5 cycles on average

Today: Minimum Vcc(t) determines maximum (and avg.) frequency.
Montecito: Average Vcc(t) determines average frequency.

Figure 10.1.3: Power reduction.

Feature	Frequency Gain @ 100W
Manage to application power vs. max power	12%
Adapt Vcc to optimal value for each part	5%
Manage junction temperature to the minimum possible	3%
Adapt frequency to Vcc to operate at frequency(V_{avg}) vs. frequency(V_{min})	7%
Optimize circuits for low switching and low leakage	11%
Cache power optimization (low V, asynch. etc.)	5%

With P ~ F^3:
$P = P_{orig}/(1+.12+.05+.03+.07+.11+.05)^3 = P_{orig}/2.92$

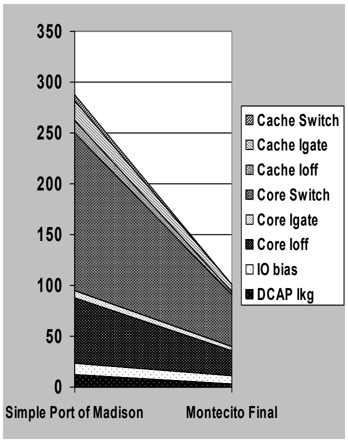


Figure 10.1.4: Power management.

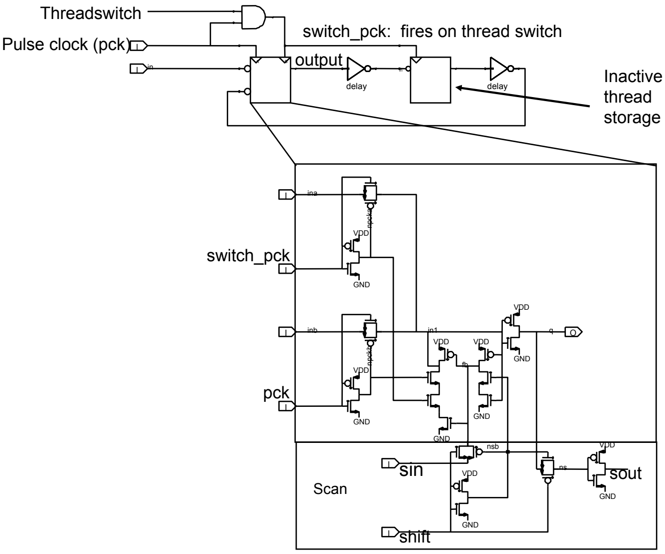


Figure 10.1.5: Dual threaded latch.

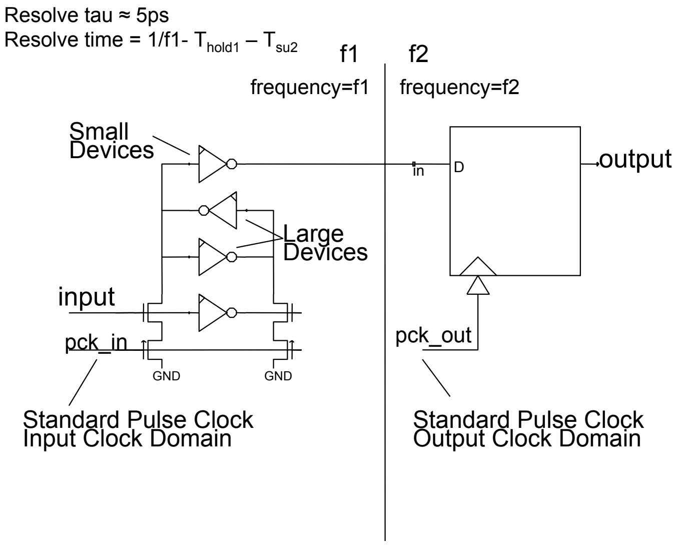

Resolve tau ≈ 5ps
Resolve time = 1/f1- T_{hold1} – T_{su2}

Figure 10.1.6: Clock domain crossing resolver latch.

Continued on Page 592

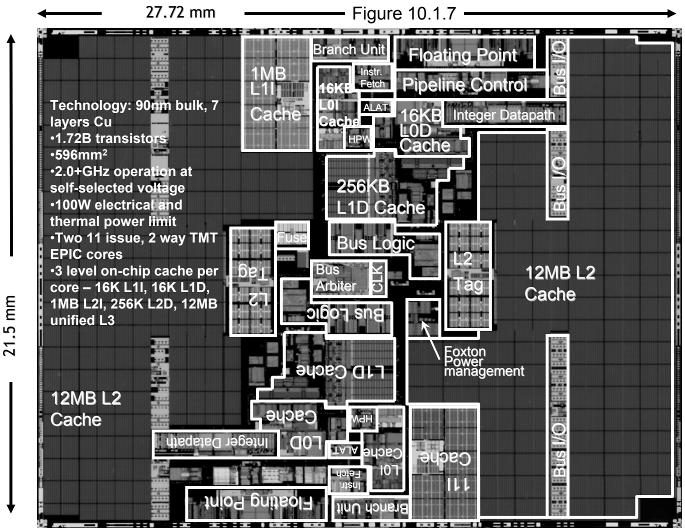


Figure 10.1.7: Die micrograph.