

3.8 Over One Million TPCC with a 45nm 6-Core Xeon® CPU

Ravi Kuppaswamy, Shankar R. Sawant, Srikanth Balasubramanian, Pradeep Kaushik, Narayanan Natarajan, Jeffrey D. Gilbert

Intel, Bangalore, India

This paper describes the 6-core Xeon® 7400 series processor family, code-name Dunnington, designed for a broad range of highly power efficient servers. The processor consists of three dual-core 45nm Core™ processors [1] and a shared inclusive 16MB L3 cache (LLC) integrated on a monolithic 503mm² die. The system interface is FSB based with the I/Os incorporated into the center of the die. Figure 3.8.1 shows the chip floorplan. The core-to-FSB connection is replaced with an on-die low-latency uncore interface. The uncore arbitrates among core, LLC, and external bus requests. The processor has 1.9B transistors and is implemented in 45nm CMOS using high-κ metal-gate transistors and nine copper interconnect layers [2]. The maximum thermal design power is 130W.

The processor runs at up to 2.66GHz core frequency and power levels down to 50W, including the first 6-core x86-compatible 65W version, which translates to just under 11W per core. A low-leakage version of the process, achieved by V_T tuning and L_e targeting, reduces leakage by 3× beyond the standard 45nm process, without affecting the intrinsic operating V_{MIN} of the circuits [2]. Traditional power-reduction methods like voltage scaling are largely limited by the large-die V_{MIN} constraints [3]. The low-leakage process delivers power reduction inline with a V_{CC} reduction of about 100mV. The product V_{MIN} is constrained by the large on-die LLC cache, followed by the smaller caches and then the register files in the cores. The power-consumption distribution, shown in Fig. 3.8.2, highlights the reduced leakage power in comparison to the active power components. The power-delivery solution is composed of three power planes. The primary digital-power plane supports the cores, the uncore and the LLC; with separate power planes for the I/O, fuse and thermal circuits; and PLLs. The high-frequency noise performance is optimized to 8% noise at nominal voltage for the digital power plane.

To actively manage processor thermals, a pair of sensors are embedded in each core: proc-hot sensor and catastrophic trip sensor, as shown in Fig. 3.8.3. The uncore arbitration logic collates all thermal sensor trip and proc-hot data to reduce the temperature by thermal-management 1 and 2 protocols (TM1 and TM2). TM1 involves modulating the clock duty cycle (off for 5000 cycles) while TM2 includes dropping the voltage and frequency. A trip of any of the catastrophic sensors leads to power down of the chip.

The processor has three levels of caches: 32KB each of data and instruction cache in each core, 3MBs of non-inclusive mid-level cache (MLC) for each core pair, and 16MB of inclusive LLC. There is a total of 25MB of cache integrated with the 6-cores on a monolithic die with a single primary power plane. Physically, the 16MB data cache is organized in 4MB blocks, each containing 4096 sets and 16 ways. The lower-cache-size products are supported by reducing the number of ways to 12 or 8. A 0.3816μm² cell is used in the data array. The data cache is protected by single-error correct double-error detect (SECDED) ECC, performed inline with fixed single-cycle latency. Both row and column redundancy is included. During an LLC access, only 1/16 of a 1MB slice is active, reducing both active and leakage power. The 1.5MB tag cache is built using a 0.54μm² cell and also protected by a SECDED scheme. Distributed sleep transistors create virtual islands that are put into a sleep or shut-off mode, to reduce leakage power in unused and disabled portions of the cache. The sleep modes allow supply-voltage reduction limited by the retention voltage of the SRAM.

There are 8 PLLs on a single die with a cascaded pair serving its respective core or uncore, as shown in Fig. 3.8.4. PLLs are fed from a single external bus clock (xxCLK: 266MHz) C4-bump pair with balanced distribution. The processor has three clock domains: quad-pumped FSB clock (ZCLK: 1067MHz); high-frequency clock (GCLK) supporting the core; and a half-frequency clock that supports most of the uncore logic and the LLC. In the cascaded configuration, the I/O PLL receives the xxCLK and synthesizes the 4× frequency ZCLK. The core PLL generates the high-frequency core clock required by the core logic. All the core and uncore PLLs are constrained to run at the same bus ratio. To combat large die variation, independent post-silicon controls are provided to adjust the supply derived from on-die regulators for each PLL. Each core has an independent de-skewing scheme [1] to have tight control on the skew. The uncore network, spanning over 30mm achieves, less than 15ps of global skew for less than 3W of power at the highest product frequency. Given the large die size and points-of-divergence that are four PLLs apart, the skew at the core-uncore boundary can be large. The core-uncore communication is enabled through a rate-matched FIFO with programmable pointer separation allowing for post-silicon optimization to achieve higher throughput. 9σ jitter across core and uncore clocks is measured to be about 100ps, enabling point-of-separation programming at 1 GCLK.

The product I/O is a point-to-point FSB design running at 1066MT/s. Unlike in conventional CPU designs, the I/O buffers are placed at the center of the die. Given the large die size, locating the I/O buffers at the ends of the die would create several internal timing-critical paths between the buffers and the bus controller logic. Resolving these paths without moving the buffer locations would require additional retiming flip-flops that increase I/O latency. Locating the I/Os close to the uncore logic and to each other improves the internal timing-critical paths, and enables the 1066MT/s without increasing latency. The central placement of the pads introduces challenges for the package design. Figure 3.8.5 shows the multiple bump pitches and pitch transitions that the package design supports for the center I/O location. A second tradeoff is the increased complexity in the package routes, which need to connect the I/O C4 bumps at the center of the die to the package pins. Figure 3.8.6 shows the routing, which ensures that there are no signal-integrity issues. The marginal increase in external timing is absorbed, without any impact, in the external platform timing budget.

Design-for-test (DFT) features include scan out, duty-cycle monitors (DCM), built-in self-test (BIST), low-yield analysis for cache arrays, I/O loopback test, and in-die process monitors. A key DFX feature is core symmetry for efficient re-use of high-volume manufacturing (HVM) content for test-time reduction. Core/cache recovery is used to enable partial good die as lower-power 4-core and 12M/8M cache products.

The processor delivers 1.2M tpmC (TPCC benchmark on transaction rate) on an 8-socket platform, and 684,508 tpmC for 4-socket servers.

Acknowledgment:

We gratefully acknowledge the work of the talented and dedicated Intel team that implemented this processor.

References:

- [1] V. George, S. Jahagirdar, C. Tong, et al. "Penryn: 45nm Next Generation Intel Core 2 Processor." *A-SSCC Dlg. Tech Papers*, pp. 14-17, Nov. 2007.
- [2] K. Mistry, C. Allen, C. Auth, et al., "A 45nm Logic Technology with High-κ+ Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193nm Dry Patterning, and 100% Pb-free Packaging", *IEDM Dlg. Tech Papers*, pp. 247-250, Dec. 2007.
- [3] K. Kuhn, C. Kenyon, A. Kornfeld, et al. "Managing Process Variation in Intel's 45 nm CMOS Technology." *Intel Technology Journal*, vol. 12, no. 2, pp. 93-109, Jun. 2008.

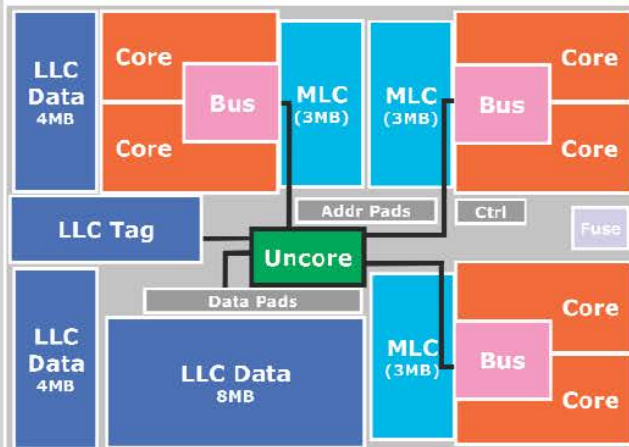


Figure 3.8.1: Chip floorplan.

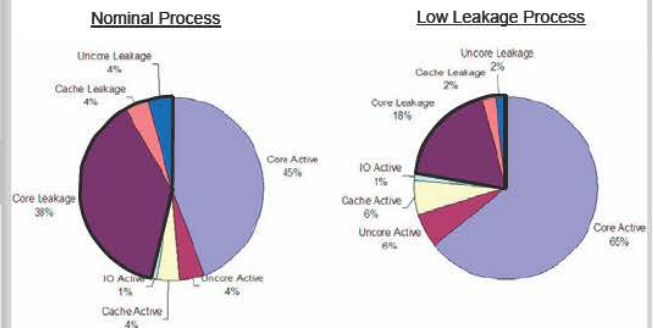


Figure 3.8.2: Low-leakage process enables 20% more active power for same TDP.

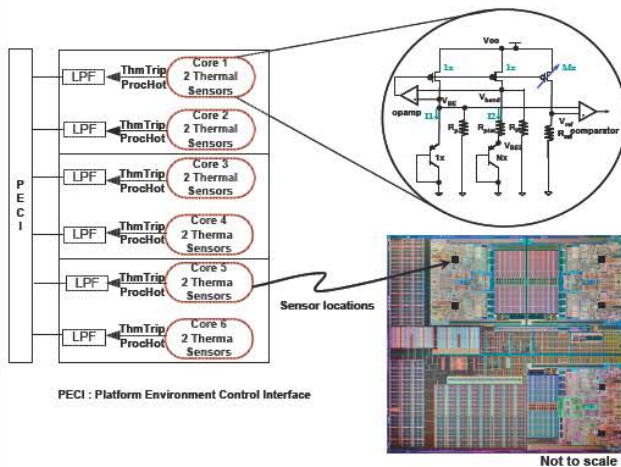


Figure 3.8.3: Thermal sensors; one pair per core.

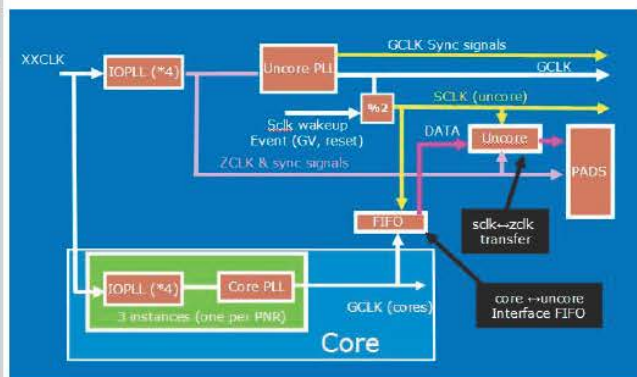


Figure 3.8.4: Clock architecture.

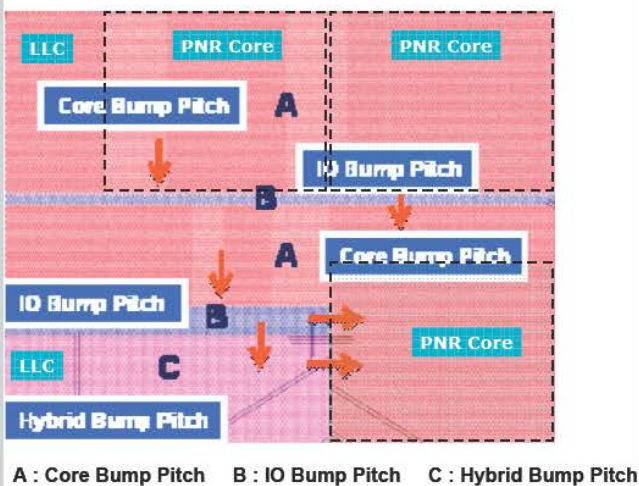


Figure 3.8.5: Bump pattern (arrows denote bump transition).

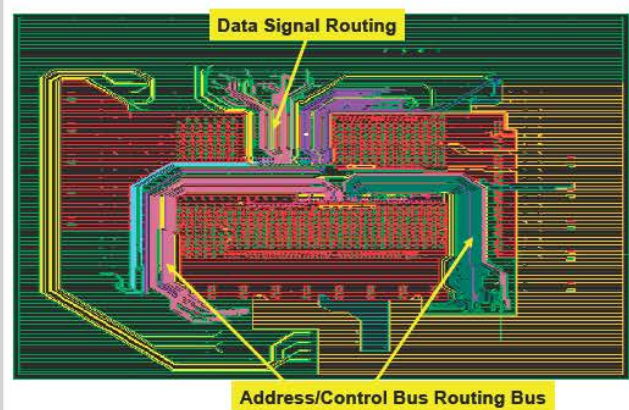


Figure 3.8.6: Package routing.

ISSCC 2009 PAPER CONTINUATIONS

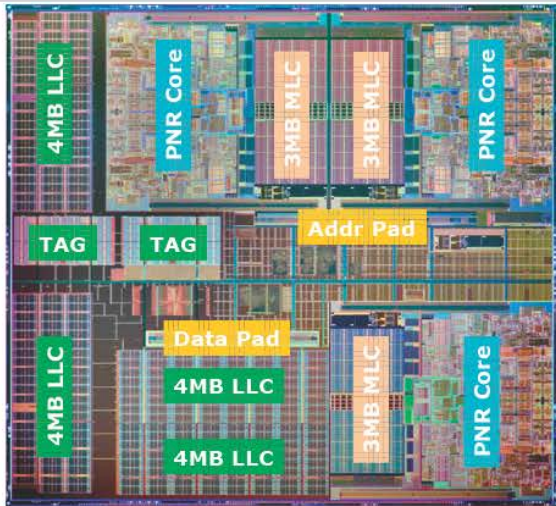


Figure 3.8.7: Die micrograph.