10.2 The Design and Implementation of a First-Generation CELL Processor

D. Pham[1], S. Asano[3], M. Bolliger[1], M. N. Day[1], H. P. Hofstee[1], C. Johns[1], J. Kahle[1], A. Kameyama[3], J. Keaty[1], Y. Masubuchi[3], M. Riley[1], D. Shippy[1], D. Stasiak[1], M. Suzuoki[2], M. Wang[1], J. Warnock[1], S. Weitzel[1], D. Wendel[1], T. Yamazaki[2], K. Yazawa[2]

[1]IBM, Austin, TX
[2]Sony, Tokyo, Japan
[3]Toshiba, Austin, TX

The implementation of a first-generation CELL processor that supports multiple operating systems including Linux consists of a 64b power processor element (PPE) and its L2 cache, multiple synergistic processor elements (SPE) [1] that each has its own local memory (LS) [2], a high-bandwidth internal element interconnect bus (EIB), two configurable non-coherent I/O interfaces, a memory interface controller (MIC), and a pervasive unit that supports extensive test, monitoring, and debug functions. The high level chip diagram is shown in Fig. 10.2.1. The key attributes include hardware content protection, virtualization and real-time support combined with extensive single-precision floating-point capability. By extending the Power architecture with SPE having coherent DMA access to system storage and with multi-operating-system resource-management, CELL supports concurrent real-time and conventional computing. With a dual-threaded PPE and 8 SPEs this implementation is capable of handling 10 simultaneous threads and over 128 outstanding memory requests.

Figure 10.2.7 shows the die micrograph with roughly 234M transistors from 17 physical entities and 580k repeaters and 1.4M nets implemented in 90nm SOI technology with 8 levels of copper interconnects and one local interconnect layer. At the center of the chip is the EIB composed of four 128b data rings plus a 64b tag operated at half the processor clock rate. The wires are arranged in groups of four, interleaved with GND and VDD shields twisted at the center to reduce coupling noise on the two unshielded wires. To ensure signal integrity, over 50% of global nets are engineered with 32k repeaters. The SoC uses 2965 C4s with four regions of different row-column pitches attached to a low-cost organic package. This structure supports 15 separate power domains on the chip, many of which overlap physically on the die. The processor element design, power and clock grids, global routing, and chip assembly support a modular design in a building-block-like construction.

The chip contains 3 distinct clock-distribution systems, each sourced by an independent PLL, to support processor, bus interface, and memory-interface requirements. The main high-frequency clock grid covers over 85% of the chip, delivering the clock signal to processors and miscellaneous circuits. Second and third clock grids, each operating at fractions of the main clock signal, are interleaved with the main clock-grid structure, creating multiple clock frequency islands within the chip. All clock grids are constructed on the lowest impedance final two layers of metal, and are supported by a matrix of over 850 individually tuned buffers. This enables control of the clock-arrival times and skews, especially on the main clock grid that supports regions of widely varying clock-load densities. High-frequency clock-signal distribution optimization and verification rely on wire simulation models that includes frequency-sensitive inductance and resistance phenomena. As shown in Fig. 10.2.2, final worst-case clock skew across chip is less than 12 ps.

Given a short cycle-time target, a significant amount of the power is consumed by latches, flipflops, and other clocked elements. However, the delay overhead imposed by standard flipflops is considerable. Therefore, a variety of latches and flipflops are developed to allow for both power and delay optimizations. The base clock components are shown in Fig. 10.2.3. In addition to test controls, the base block accepts a local clock-gate signal, with a small setup time relative to the falling global clock (cycle boundary). Input setup-and-hold times are specified against the falling clock edge, as a result of the built-in latching action of the base block. Local clocks are derived from the base block for a typical master-slave flip flop.

For timing-critical paths, a high-performance latch (HPL) is designed. It combines a wide MUX (up to 10-way), relying on a dynamic NOR gate, with a set-reset latch (Fig. 10.2.4). The dynamic NOR starts evaluating with the launch of the clock, and the input-data hold time is limited when all sel_b inputs are forced high after a fixed delay.

Dynamic circuits are used in several critical macros, in the arrays, and in PLAs. All dynamic macros are latch-bounded (macro-to-macro signals are static). Signals that feed the dynamic logic are usually launched from the master portion of a flipflop, and ANDed with *lclk* to provide a signal which resets to 0 every cycle when *lclk* is low. Dynamic logic is always followed by a set-reset latch similar to that used for the HPL shown earlier.

In addition, various rules are adopted to ensure a "correct-by-construction" design methodology. All circuits use a common set of clocking components to ensure uniformity across the design with no rotation of the components allowed. An extensive set of electrical and physical checks and audits are performed. Finally, a customized series of yield-related checking rules is employed to ensure manufacturability of the chip.

This SoC presented new challenges in the thermal design of the chip. The higher-heat flux from smaller hot-spots hindered spreading of the heat across the silicon substrate. Extensive thermal analysis carried out early in the design cycle ensures that the maximum junction temperature as well as the average temperature of the die, are within the design specifications. Various workloads are simulated for each component and power maps are constructed. From these maps, a matrix of small power sources is created, for use with package and heat-sink models. Thermal models are then created and used to simulate both steady-state and transient thermal behavior. This data is analyzed to improve the design and floorplan of the chip and also provide feedback for improved thermal-sensor design (Fig. 10.2.5).

Due to local heating caused by individual processing units, sophisticated local thermal-sensing strategies and and thermal-control mechanisms are used to allow an aggressive low-cost thermal design. The processor contains a linear sensor and 10 local digital thermal sensors. The linear sensor is essentially a diode connected to two external I/Os, used to measure the die's global temperature of the die and to adjust the system cooling. The digital thermal sensors provide for early warning of any temperature increase and for thermal protection.

In conclusion, special circuit techniques, rules for modularity and reuse, customized clocking structures, and unique power- and thermal-management concepts are applied to optimize the design. Correct operation is observed in the lab on first-pass silicon at frequencies well over 4GHz as shown in Fig. 10.2.6.

Acknowledgements:
The authors gratefully acknowledge the many contributions from the entire Sony-Toshiba-IBM team who worked tirelessly side-by-side on the design of this processor.

References:
[1] B. Flachs et al., "A Streaming Processor Unit for a CELL Processor," *ISSCC Dig. Tech. Papers*, Paper 7.4, pp. 134-135, Feb., 2005.
[2] T. Asano et al., "A 4.8GHz Fully Pipelined Embedded SRAM in the Streaming Processor of a CELL Processor," *ISSCC Dig. Tech. Papers*, Paper 26.7, pp. 486-487, Feb., 2005.

 0-7803-8904-2/05/$20.00 ©2005 IEEE.

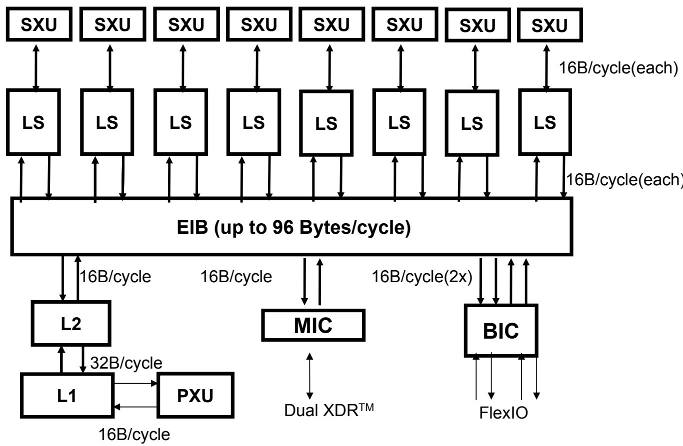

Figure 10.2.1: Processor high-level diagram.

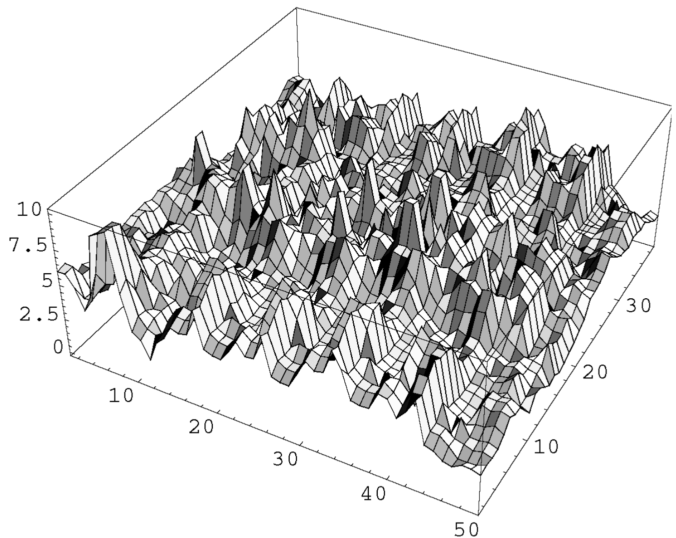

Figure 10.2.2: Clock-skew map.

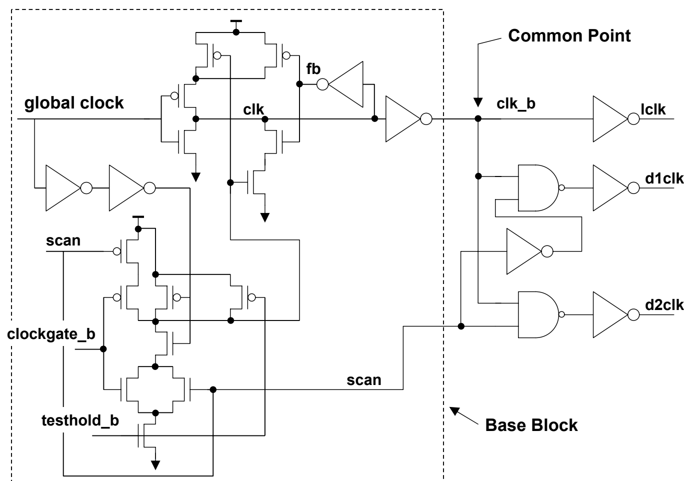

Figure 10.2.3: Local clock generation.

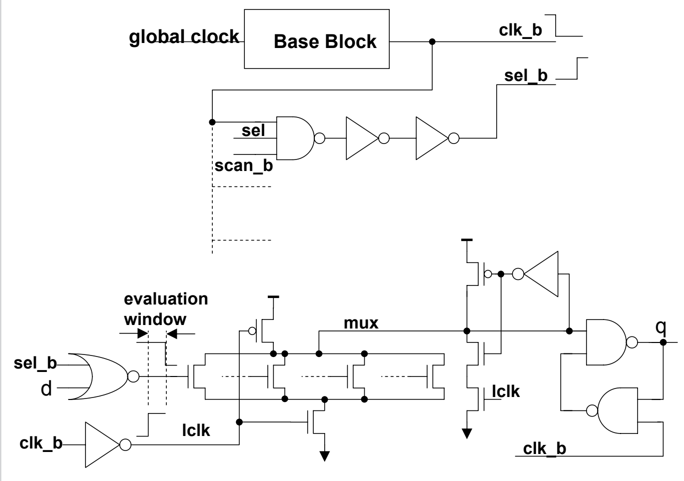

Figure 10.2.4: High-performance latch (HPL).

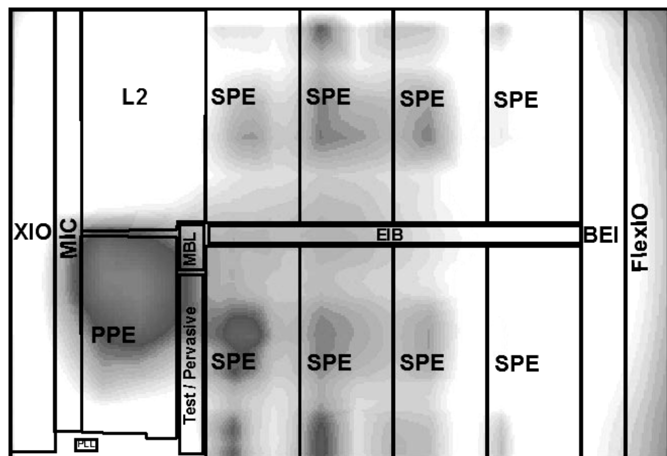

Figure 10.2.5: Die thermal map.

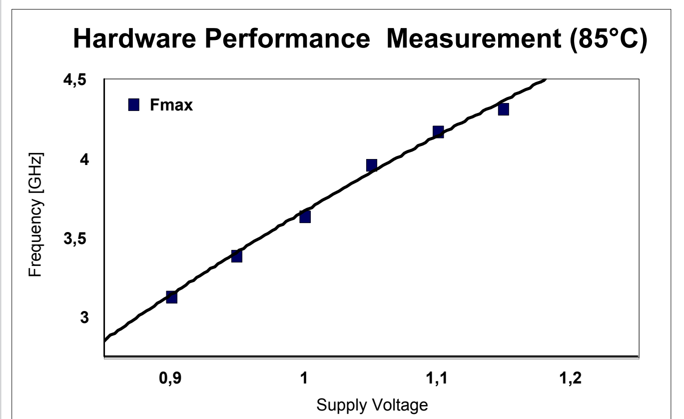

Figure 10.2.6: First-pass hardware measurement in the Lab.

Continued on Page 592

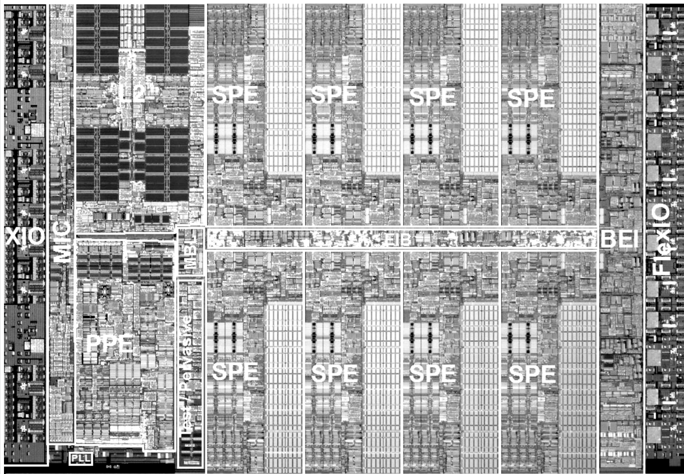

Figure 10.2.7: Die micrograph with high-level floorplan overlay.

 0-7803-8904-2/05/$20.00 ©2005 IEEE.