A Dual-Core 64b UltraSPARC Microprocessor for Dense Server Applications

Toshinari Takayanagi, Jinuk Luke Shin, Bruce Petrick, Jeffrey Su and Ana Sonia Leon
Sun Microsystems, Inc.
430 N. Mary Ave.
Sunnyvale, CA 94085, U.S.A.
+1-408-720-4894
{toshinari.takayanagi, jinuk.shin, jeffrey.su, bruce.petrick, ana-sonia.leon}@sun.com

ABSTRACT

A processor core, previously implemented in a 0.25 μ m Al process, is redesigned for a 0.13 μ m Cu process to create a dual-core processor with 1MB integrated L2 cache, offering an efficient performance/power ratio for compute-dense server applications. Deep submicron circuit design challenges, including negative bias temperature instability (NBTI), leakage and coupling noise, and L2 cache implementation are discussed.

Categories and Subject Descriptors

B.7.1 [**Integrated Circuits**]: Types and Design Styles – Advanced technologies, Memory technologies, Microprocessors and microcomputers, Standard cells, VLSI; B.3.2 [**Memory Structures**]: Design Styles – Associative memories, Cache memories; B.8.1 [**Performance and Reliability**]: Reliability, Testing, and Fault-Tolerance; C.1.2 [**Processor Architectures**]: Multiple Data Stream Architectures (Microprocessors)

General Terms

Performance, Design, Reliability

Keywords

Multiprocessor, Dual-core, UltraSPARC, Dense server, Throughput computing, Deep submicron technology, cache, leakage, Negative Bias Temperature Instability, NBTI, Coupling noise, L2

1. INTRODUCTION

1.1 Background

For the past two decades, microprocessor performance has doubled every 18-24 months. The performance scaling has been sustained by increasing clock frequency and instruction-level parallelism (ILP). However, these two factors are both reaching the point of diminishing returns [1]. In addition, network computing based on today's pervasive use of the Internet has drastically changed the nature of application workloads. Network-based applications, such as online transaction processing, are rich in thread-level parallelism. These applications require high computing throughput to execute multiple threads/processes simultaneously rather than high single thread performance. Further, power consumption is critical for dense server environments in the data center.

1.2 Design Target

This 64b SPARC processor is designed for compute-dense systems such as rack-mount and blade servers for network computing [2][3]. The critical requirements for this type of applications are high computing throughput, high memory bandwidth, large addressing space, high reliability, low power and low cost. A short design cycle was also critical for this project.

2. ARCHITECTURE AND CHIP OVERVIEW

To address the above design targets, the optimum solution is to create an on-chip dual-core processor based on the UltraSPARC I/ II microarchitecture [4] with embedded 1MB L2 cache, DDR-1 memory controller and symmetric multiprocessor bus (JBus) interfaces (Figure 1). This core provides an efficient performance per watt having balanced hardware complexity with 4-issue superscalar, 9-stage pipeline and in-order execution/out-of-order completion. Predominantly static circuit design styles and the balanced H-tree clock distribution also contribute to achieving low power. Typical power dissipation at 1.2GHz and 1.3V is 23Watts, which is the lowest published figure for 64b server processors [2].

Parity bits are added to L1 cache data/tag and L2 tag arrays while L2 data arrays are protected by ECC for enterprise-class reliability. The memory controller supports up to 16GB of physical memory. JBus controllers allow low-cost multiprocessing systems with configurations of up to four-chips (eight threads) without any glue logic. The chip implements standard software interfaces to manage the multiple threads on a die in consistent ways, which Sun has developed to support its broad family of forthcoming Chip MultiThreading (CMT) designs.

DAC 2004, June 7–11, 2004, San Diego, California, USA
ACM 1-58113-828-8/04/0006.

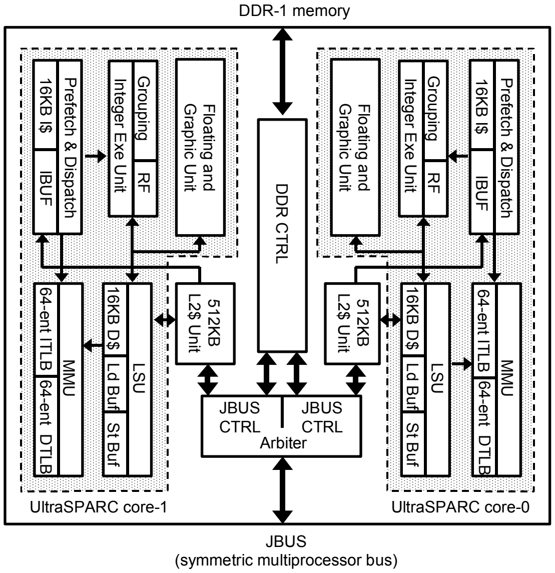

Figure 1. Chip block diagram

The chip is fabricated in Texas Instruments' advanced 0.13µm CMOS process with 7 layers of Cu and a low-k dielectric. The transistor count is 80M, of which 72M is SRAM. The 206mm^2 die is packaged in a 959-pin ceramic µPGA. The chip interface is made pin-compatible with UltraSPARC IIIi processor [5] to effectively reuse the existing system resources. The core and the chip features are summarized in Table 1 together with the comparison with the original UltraSPARC I processor.

Table 1. Chip feature summary

	UltraSPARC I (1995)	Dual-Core UltraSPARC (2003)
Architecture	64b SPARC-V9 with multimedia instruction extensions	• Dual core with CMP features (core disable/parking/error reporting) • Core is based on UltraSPARC I / II microarchitecture • DDR-1 memory controller • JBUS (SMP bus) interfaces • E*Star mode support • L1 caches are parity protected
Pipeline	• 4-issue superscalar (2 integer ops, 2 FPU/graphics ops / cycle) • 9-stage pipeline • In-order execution / out-of-order completion	
L1 cache / MMU	• 16KB 2-way I-cache • 2KB next field RAM • 16KB direct D-cache • 64-entry full-associative I-TLB and D-TLB	
L2 cache	Off-chip	• On-chip unified 512KB 4-way set-associative cache for each core • ECC protected • 4-cycle latency to core / 9-cycle load-use latency • 2-cycle data throughput
Process	0.5µm CMOS / 4 layers Al (Transported to 0.35µm process UltraSPARC II in 1996, 0.25µm in 1998)	0.13µm CMOS / 7 layers Cu
Voltage	3.3V	1.3V
Transistors	5.2M	80M total, 72M for SRAM
Die size	315mm^2	206mm^2 , 29 mm^2 per core
Clock freq.	167MHz	1.2GHz
Power	30W	23W (typical), 5W per core

The core circuits were originally implemented in 0.5µm/3.3V and last implemented in 0.25µm/2.1V technology. Redesigning these circuits for 0.13mm/1.3V technology raised deep submicron design challenges facing the semiconductor industry today, including negative bias temperature instability (NBTI), leakage and coupling noise. The project goals also required achieving a short on-chip L2 cache latency with ECC. These challenges are discussed in detail in the next sections.

3. DEEP SUBMICRON CIRCUIT DESIGN CHALLENGES

3.1 NBTI

NBTI is the aging effect that decreases PMOS current mainly due to Vt shift over silicon lifetime [9][10]. This Vt shift is strongly dependent on gate-source bias and temperature but barely dependent on drain voltage. The damaging mechanism is related to the hole trapping in the gate oxide and interface state generation. The NBTI impact on the circuits includes speed degradation, increased delay variation, shifted PMOS/NMOS drive current ratio, decreased Vdd/Vt headroom and Vt mismatch. Many circuits were modified to enhance margins for NBTI.

Particularly, current-mode latch sense amps used for L1 caches and TLBs (Figure. 2)[6] were highly affected, degrading the total sense delay by 42% (Figure. 3-c). The cross-coupled PMOS pair, M3 and M4, which act as low input impedance devices during equilibration [7][8] and positive feedback while sensing, were unevenly affected due to unequal 0 or 1 read rate. This situation is particularly worse for TLB matchline sense amp as each entry of the TLB will read out the “miss” data most of the time while one of the other 63 entries provides the “hit” signal. This could cause a worst case Vt mismatch of about 50mV between the PMOS pair, requiring a longer signal development time to overcome the offset. The PMOS Vt shift also attenuated the gain.

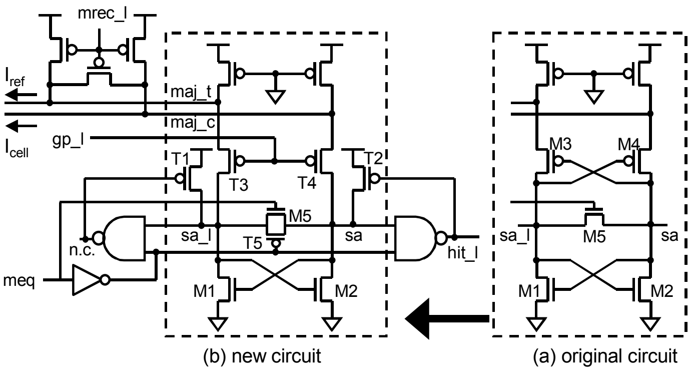

Figure 2. TLB matchline sense amplifier (simplified)

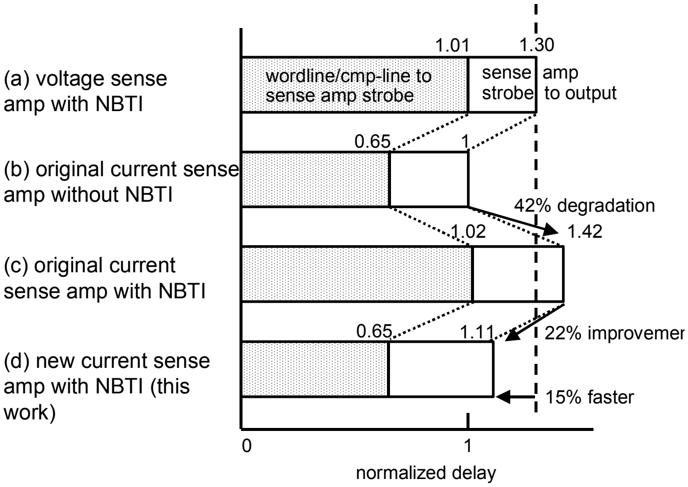

Figure 3. Sense amplifier delay comparison

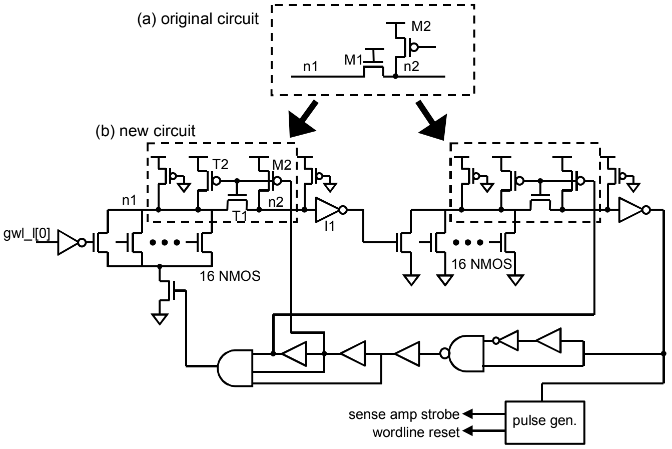

Figure 4. I-cache wordline detector

To cope with NBTI, the cross-coupled PMOS pair M3 and M4 are replaced by T3 and T4 whose gates are commonly biased at about 40% of Vdd during sensing. These act as low impedance devices being biased in saturation mode. As the gate bias is identical for T3 and T4, the Vt imbalance is minimized. In addition, PMOS T1 and T2 are added to speed up the low-to-high output transition. Although T1 and T2 can get an uneven Vt shift, it is not critical as they get activated after the significant part of the amplification is completed. These modifications improved the deteriorated sense delay by 22%, achieving 15% speedup over voltage sense amps (Figure. 3-d).

3.2 Leakage and Coupling Noise

In order to address leakage and noise issues, additional circuit modifications are required. The I-cache wordline detector in Figure 4 is one example. This circuit is a 256-inputs OR gate, which consists of two levels of 16-inputs self-resetting dynamic OR gates, to detect a wordline transition for sense amp strobe. The wired-NOR net, n1, is susceptible to leakage and noise, as 16 NMOSs are connected in parallel with a long wire. In the original circuit, n1 was precharged to Vdd-Vt for speed, however the noise margin of this circuit was reduced due to the lower supply voltage: a 100mV drop at n1 could cause the circuit to fail as M1 turns on easily to discharge node n2.

In the new circuit, n1 and n2 are both precharged to Vdd with NMOS T1 between them. The gate of T1 is high during the evaluation and low during the precharge. During the evaluation, T1 acts as a noise decoupler since the voltage drop at n1 does not propagate to n2 unless it is large enough to turn on T1. In addition, T1 decouples n2 from the large capacitance on n1 during the precharge, speeding up the reset path. Compared to a conventional domino gate, this circuit achieves similar speed while improving the noise margin by 35%. The wordline detection slowed down by 9% from the original circuit, but the cache access time is not impacted since the extra delay is absorbed in the sense amp strobe buffering stage.

3.3 Intra-die Process Variation

Since clock skew does not scale proportionally to gate delay due to intra-die process variation, a significant number of new hold-time violations are created (Figure 5). These needed to be fixed with minimal physical design changes. New flops that have larger output delays while keeping the same footprint are created for the most frequently used flops. The flop depicted in Figure 7 utilizes the scan slave path for normal output as TG1 is kept on in normal operation mode with se=0 and thus sclk=1, achieving additional three-stage gate delay without increasing the flop size. Together with other types of new flops with negative hold-time, 75% of the 26,300 core level hold-time violations were fixed simply by replacing the flops. An automated flow is developed to identify and replace the flops to fix the hold-time violation without affecting the critical paths.

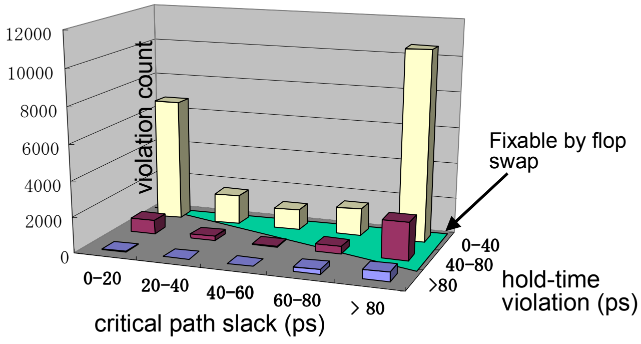

Figure 5. Hold-time violation statistic

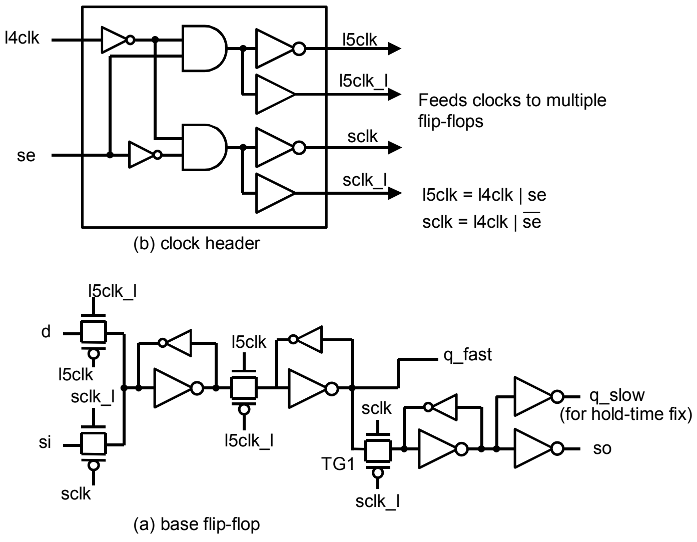

Figure 6. Scan flip-flop with delayed output

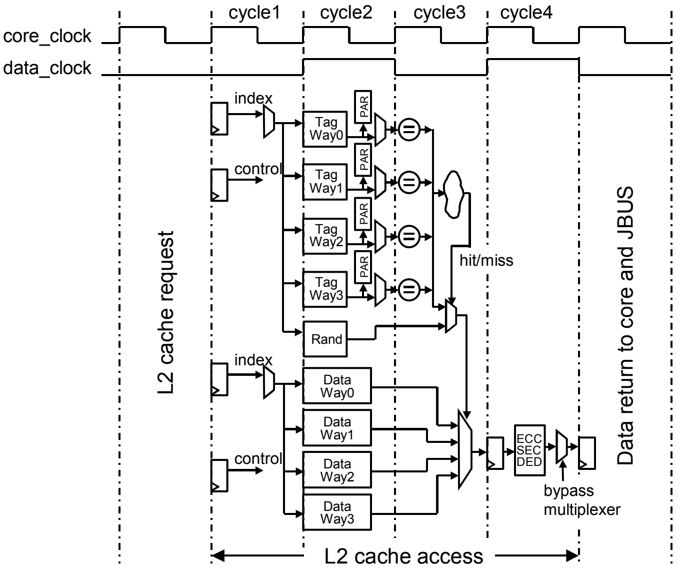

Figure 7. L2 cache pipeline diagram

4. L2 CACHE IMPLEMENTATION

The chip includes 512KB four way set-associative L2 cache per core. The data bus width is 128b and the line size is 64b. The data array has 8 ECC bits for 64b data with SEC/DED (single bit error correction/double bit error detection). The tag array is protected by parity bits.

Logic blocks, including way-selection, ECC and interfaces to the core and system bus, are optimally placed along with the dataflow in the center of the cluster and in the channel region between the SRAM arrays. Compared to a conventional approach where the logic is partitioned and placed outside the SRAM arrays, this approach reduces pipeline stages for communication between the arrays and logic blocks, minimizing the impact of long wire delays. This together with a multi-cycle clock scheme allows a low four-cycle latency from the L2 cache to the core including error correction. The L2 cache pipeline and floorplan are shown in Figure 7 and 8.

In cycle 1, addresses are dispatched from the core into each array. A full cycle is needed to send the addresses to the far end of the data arrays after multiplexing the addresses in the L2 control unit. In the next cycle, the data and tag arrays start the access simultaneously. In cycle 3, the address information from the tag array generates a way select signal before the data arrives at the waysel datapath. This cycle ends by registering a selected set of data in a datapath block located in the center of the cluster where the delay from all the data arrays is balanced. Cycle 2 and 3 are designed as multi-cycle paths for speed and power, eliminating the pipeline flops between the cycles, which minimizes flop and clock skew overhead. In the last cycle, ECC syndrome and correction is performed in the datapath located in the channel where the data is routed back to the core. The final data set is registered in the interface block located at the top of the L2 cluster.

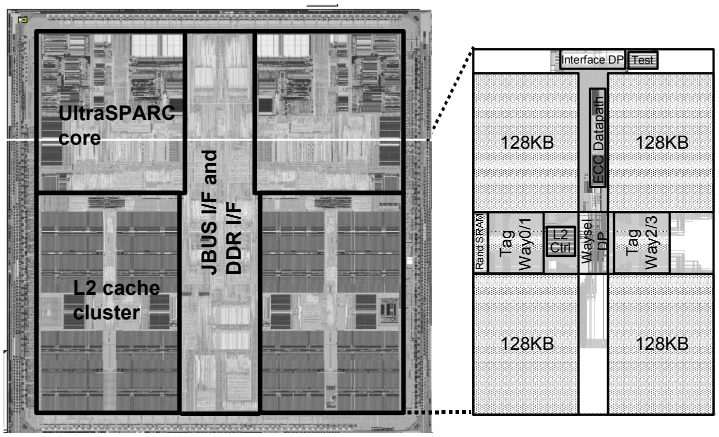

Figure 8. Chip micrograph and L2 cluster floorplan

5. INTEGRATION

Roughly 14% of the control and datapath blocks in the UltraSPARC core area are re-synthesized and place-and-routed, while other blocks are transported to the new technology with engineering change orders (ECOs) to achieve the faster floorplan and timing closure. The core level routing is redone to deal with the deep submicron interconnect issues. Trade-offs for the area, speed, coupling noise and electromigration (EM) are optimized by the use of various wire classes defined in terms of wire width, space, half-side shielded or both-side shielded. 2,400 repeaters are newly inserted at the core level and 12,900 repeaters are implemented at the chip level. Coupling noise is fully analyzed using an in-house CAD tool.

To improve critical paths, low-Vt transistors are applied for 6% of the total transistors at the core, 3% at the chip level. Footprint compatible low-Vt library cells coupled with the automated cell replacement flow made this process easy.

```
0.500ns ...................** 2000.000Mhz
0.525ns .............******** 1904.762Mhz
0.550ns ..........*********** 1818.182Mhz
0.575ns ........************* 1739.130Mhz
0.600ns .......************** 1666.667Mhz
0.625ns ......*************** 1600.000Mhz
0.650ns .....**************** 1538.462Mhz
0.675ns .....**************** 1481.481Mhz
0.700ns .....**************** 1428.571Mhz
0.725ns ....***************** 1379.310Mhz
0.750ns ...****************** 1333.333Mhz
0.775ns ...****************** 1290.323Mhz
0.800ns ...****************** 1250.000Mhz
0.825ns ...****************** 1212.121Mhz
0.850ns ..******************* 1176.471Mhz
0.875ns ..******************* 1142.857Mhz
0.900ns ..******      ******* 1111.111Mhz
0.925ns .*******  PASS ******* 1081.081Mhz
0.950ns .*******      ******* 1052.632Mhz
0.975ns .******************** 1025.641Mhz
1.000ns .******************** 1000.000Mhz
1.025ns ********************* 975.610Mhz
1.050ns ********************* 952.381Mhz
1.075ns ********************* 930.233Mhz
1.100ns ********************* 909.091Mhz
1.125ns ********************* 888.889Mhz
1.150ns ********************* 869.565Mhz

        000011111111111111111
        .....................
        889900112233445566778
        050505050505050505050
        000000000000000000000
        VVVVVVVVVVVVVVVVVVVVV
```

Figure 9. Shmoo plot at 85°C

6. CONCLUSION

A highly integrated 64b dual-core microprocessor optimized for low-cost dense servers is successfully created. The target frequency of 1.2GHz at 1.3V, 85°C is achieved with comfortable margin (Figure 9), dissipating 23W with typical application workloads. The high performance/watt is achieved by reusing the power and area efficient UltraSPARC II core. The core circuits are successfully transported from 0.25µm process to 0.13µm process with high circuit margins coping with the new deep submicron design issues. The low-latency and high-capacity L2 cache is newly designed. The four-cycle core-to-L2 cache latency including ECC is among the best in the industry. The short design cycle was achieved by leveraging existing designs and effective design methodologies and flows.

7. ACKNOWLEDGMENTS

The authors acknowledge contributions from the entire Gemini development team and all those who supported the project both at Sun Microsystems and Texas Instruments. Special thanks to R. Heald and J. Kaku for their technical advice, and A. Kowalczyk, R. Raman, A. Strong and F. DeSantis for supporting the project.

8. REFERENCES

[1] M. Horowitz and W. Dally, “How Scaling Will Change Processor Architecture,” ISSCC Dig. Tech. Papers, Feb. 2004, pp. 132-133

[2] S. Kapil, "Gemini: A Power-efficient Chip Multi-Threaded UltraSPARC Processor," 15th Hot Chips Symposium, Aug. 2003

[3] T. Takayanagi et al., “A Dual-Core 64b UltraSPARC Microprocessor for Dense Server Applications,” ISSCC Dig. Tech. Papers, Feb. 2004, pp. 58-59

[4] Lev, L. A., et al, “A 64b microprocessor with multimedia support,” IEEE J. Solid State Circuits, vol. 30, no. 11, Nov. 1995, pp. 1227-1238

[5] G. K. Konstadinidis et al., “Implementation of a Third-Generation 1.1-GHz 64-bit Microprocessor,” IEEE J. Solid State Circuits, vol. 37, no. 11, Nov. 2002

[6] E. Anderson, “A 64-Entry 167MHz Fully-Associative TLB for a RISC Microprocessor,” ISSCC Dig. Tech. Papers, Feb. 1996, pp. 360-361

[7] E. Seevink et al., “Current-Mode Techniques for High-Speed VLSI Circuits with Application to Current Sense Amplifier for CMOS SRAM’s,” IEEE J. Solid State Circuits, vol. 26, no.4, April 1991, pp. 525-536

[8] N. Shibata, “Current Sense Amplifiers for Low-Voltage Memories”, IEICE Trans. Electron., vol. E79-C, no. 8, pp. 1120-30, Aug. 1996.

[9] V. Reddy et al., “Impact of Negative Bias Temperature Instability of Digital Circuit Reliability,” Reliability Physics Symposium, 2002, pp. 248-254

[10] N. Kimizuka et al: Symposium of VLSI Technology, (1999) p.73