

5-GHz 32-bit Integer Execution Core in 130-nm Dual- V_T CMOS

Sriram Vangal, *Member, IEEE*, Mark A. Anders, Nitin Borkar, Erik Seligman, Venkatesh Govindarajulu, Vasantha Erraguntla, *Member, IEEE*, Howard Wilson, Amaresh Pangal, Venkat Veeramachaneni, James W. Tschanz, Yibin Ye, Dinesh Somasekhar, *Member, IEEE*, Bradley A. Bloechel, *Associate Member, IEEE*, Gregory E. Dermer, Ram K. Krishnamurthy, *Member, IEEE*, K. Soumyanath, Sanu Mathew, Siva G. Narendra, Mircea R. Stan, *Senior Member, IEEE*, Scott Thompson, Vivek De, *Member, IEEE*, and Shekhar Borkar

Abstract—A 32-bit integer execution core containing a Han-Carlson arithmetic-logic unit (ALU), an 8-entry $\times 2$ ALU instruction scheduler loop and a 32-entry $\times 32$ -bit register file is described. In a 130 nm, six-metal, dual- V_T CMOS technology, the 2.3 mm² prototype contains 160 K transistors. Measurements demonstrate capability for 5-GHz single-cycle integer execution at 25 °C. Single-ended, leakage-tolerant dynamic scheme used in ALU and scheduler enables up to 9-wide ORs with 23% critical path speed improvement and 40% active leakage power reduction when compared to a conventional Kogge–Stone implementation. On-chip body-bias circuits provide additional performance improvement or leakage tolerance. Stack node preconditioning improves ALU performance by 10%. At 5 GHz, ALU power is 95 mW at 0.95 V and the register file consumes 172 mW at 1.37 V. The ALU performance is scalable to 6.5 GHz at 1.1 V and to 10 GHz at 1.7 V, 25 °C.

Index Terms—CMOS integrated circuits, integrated circuit design, logic design, microprocessors, very-high-speed integrated circuits.

I. INTRODUCTION

OUT-OF-ORDER execution engines of superscalar processors require: 1) wide instruction schedulers capable of scheduling back-to-back instructions into multiple arithmetic-logic units (ALUs) in the execution core; 2) fast ALUs capable of executing these instructions with single-cycle latency and throughput; and 3) leakage-tolerant register files capable of feeding the ALU units. A high-speed execution core is therefore essential to maximize processor performance [1]. In this paper, we describe key components of a integer execution core: a 32-bit ALU, an 8-entry $\times 2$ -ALU instruction scheduler and a 32-entry $\times 32$ -bit register file (RF), fabricated in 130 nm dual- V_T CMOS technology [2]. High-speed single-ended dynamic circuit techniques enable the evaluation of complex (up to 2×9 -way OR) logic operations while simultaneously

achieving: 1) high noise robustness; 2) low active leakage power dissipation; 3) maximum low- V_T usage; 4) simplified 2Φ 50% duty-cycle timing scheme with seamless scheduler/ALU interface time-borrowing; and 5) scalable performance up to 10 GHz, measured at 1.7 V, 25 °C. Stack node preconditioning enables further ALU performance improvement. In addition, the RF employs semi-dynamic flip-flops for increased speed and a static design for increased leakage tolerance. On-chip body-bias circuits are used to improve performance or reduce standby leakage power. The chip also supports full at-speed result capture and scan-out.

In the execution core, both the ALU and scheduler are organized as a loop [3], enabling single-cycle latency and throughput both for ALU operations and for resolving instruction dependencies and priorities (Fig. 1). The scheduler updates interinstruction dependency information each cycle, choosing the highest priority from among those instructions ready to execute. The chosen instruction controls ALU input selection as well as RF address. The 32-bit ALU executes add/subtract operations each cycle, allowing the previous results to be used directly in the following cycle. This architecture enables fast parallel out-of-order execution in superscalar microprocessors.

II. PROTOTYPE ARCHITECTURE

The core architecture contains: 1) three first in first out (FIFO) buffers, one FIFO for instructions and two FIFOs for data; 2) a tightly coupled RF-ALU-Scheduler loop; 3) a FIFO to capture output results (Fig. 2) [4]. The core executes instructions stored in the 32-bit wide, 4-deep circular instruction FIFO, operating at core speed. Data FIFOs (D0-D1) provide the desired operands. A central block forwards data and control signals to all units. RF-ALU instructions are single-cycle and can be scheduled back to back. A 416-bit long input scan chain feeds the data and control words. Results are captured at speed by a 32-bit wide, 4-deep result FIFO. Capture timing and interval are fully programmable via scan. Output results are scanned out using a 128-bit scan chain. The scan control block manages operations of all four FIFOs on chip. Separate power grids for ALU, RF and circuits in the rest of the core allow individual power measurement of each unit. In addition, RF and ALU units have on-chip body bias generator circuits to improve performance by applying 450 mV of forward body bias (FBB) to all pMOS devices during active operation. Body biases of high- V_T

Manuscript received March 15, 2002; revised June 10, 2002.

S. Vangal, M. A. Anders, N. Borkar, E. Seligman, V. Govindarajulu, V. Erraguntla, H. Wilson, A. Pangal, V. Veeramachaneni, J. W. Tschanz, Y. Ye, D. Somasekhar, B. A. Bloechel, G. E. Dermer, R. K. Krishnamurthy, K. Soumyanath, S. Mathew, S. G. Narendra, V. De, and S. Borkar are with the Intel Corporation, Circuits Research, Intel Labs, Hillsboro, OR 97124-6497 USA (e-mail: sriram.r.vangal@intel.com).

S. Thompson is with the Portland Technology Development, Intel Corporation, Hillsboro, OR 97124-6497 USA.

M. R. Stan is with the Department of Electrical Engineering, University of Virginia, Charlottesville, VA22904-4743 USA.

Digital Object Identifier 10.1109/JSSC.2002.803944

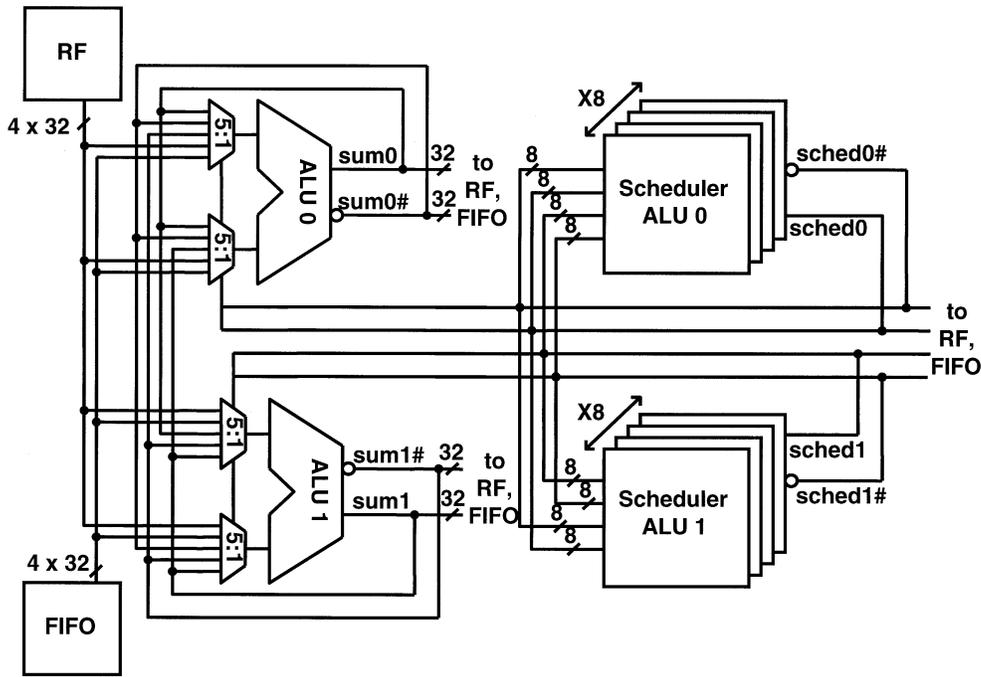


Fig. 1. Out-of-order execution core.

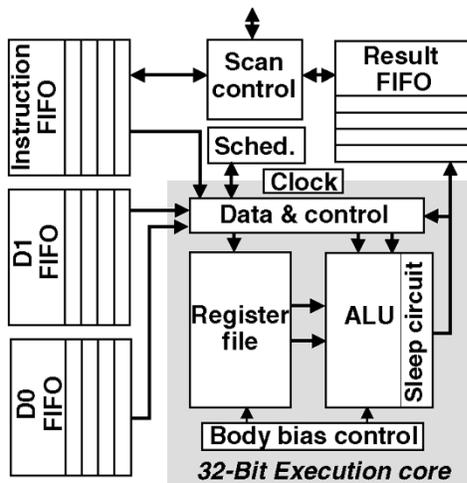


Fig. 2. Block diagram of integer execution core and test circuits.

and low- V_T devices can be controlled separately. On-chip FBB can be disabled; and forward, zero, or reverse body bias can be applied externally to all nMOS and pMOS devices to improve performance or reduce standby leakage power. The chip organization provides flexibility to characterize individual units or the complete core.

III. 8-ENTRY \times 2 INSTRUCTION SCHEDULER

The instruction scheduler is capable of scheduling dependent instructions to two 32-bit ALUs, choosing one of eight potentially ready instructions to execute in each ALU per cycle. An instruction is ready for execution if it is not dependent upon results of any other pending instructions and has not been scheduled in the previous cycle. The scheduler is organized into 16

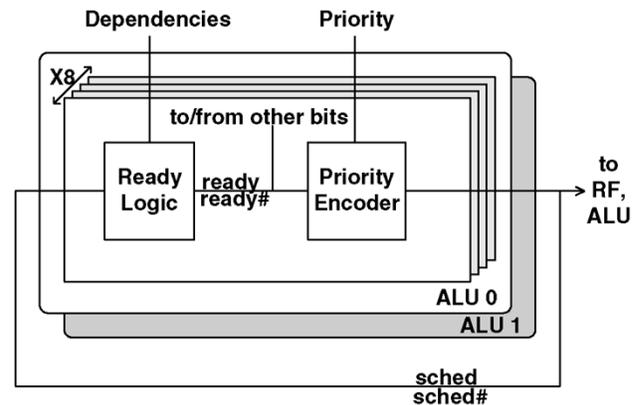


Fig. 3. Scheduler organization.

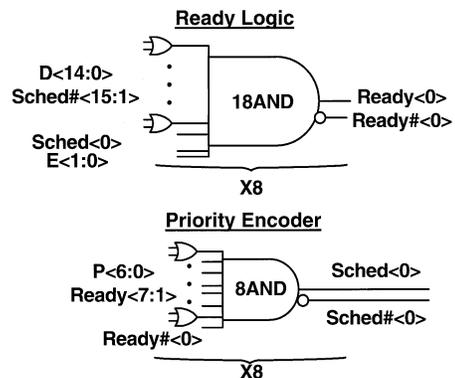


Fig. 4. Scheduler bitslice logic.

bit slices, with one ready logic evaluation and one priority encoder operation per bit slice (Fig. 3). The 15 dependencies for the 16 instructions currently in the pool, $D\langle 14 : 0 \rangle$, are evaluated and stored in a 1-bit \times 240-entry dependency matrix during

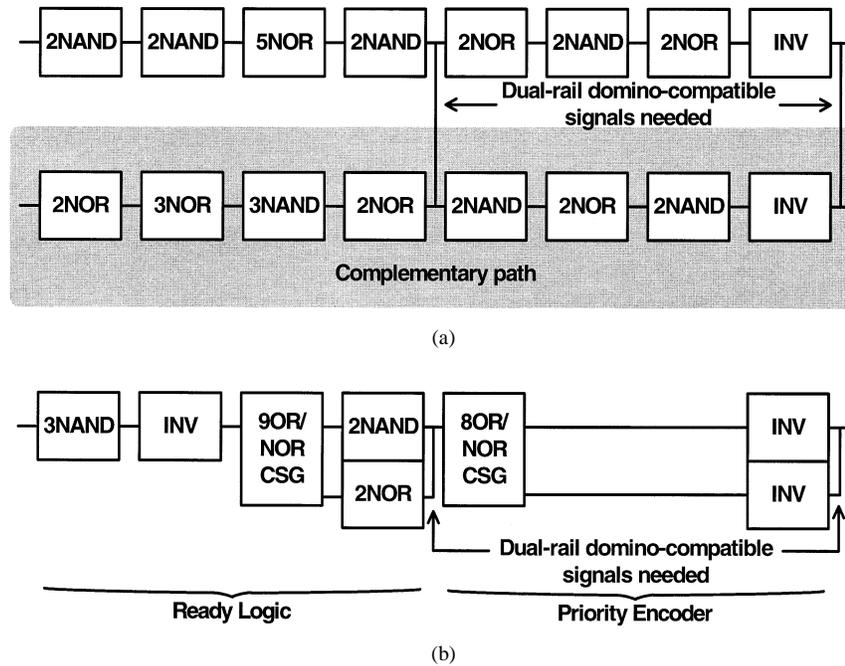


Fig. 5. Scheduler designs. (a) Dual-rail domino. (b) Single-rail CSG.

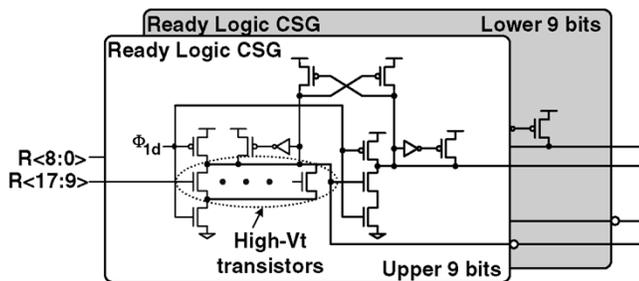


Fig. 6. Scheduler ready logic CSG circuit.

the previous cycle. The ready logic resolves dependencies between the 16 instructions in the pool and two external dependency signals ($E\langle 1 : 0 \rangle$), essentially requiring an 18-way AND operation (Fig. 4). An 8-way AND priority encoder then chooses from among the ready instructions using dynamically controlled priorities ($P\langle 6 : 0 \rangle$) and drives a 140- μm loopback bus into the ready logic and the shared ALU tri-state bus. The ready logic, using the priority encoder outputs from all other bit slices, determines if its instruction is dependent on any other instruction. The priority encoder, then, using the ready logic outputs only from the other 7 bit slices in its portion of the instruction queue, indicates if its instruction is the highest priority.

A domino implementation of the scheduler logic requires a fully dual-rail design, since both true and complementary domino-compatible inputs are required for both the ready logic as well as the priority encoder. An optimal dual-rail domino design requires 8 gate stages due to decreasing performance as evaluation stack heights are increased on the complementary path [Fig. 5(a)]. Fig. 5(b) shows the single-ended to domino-compatible complementary signal generator (CSG) based ready logic and priority encoder implementation, that eliminates the

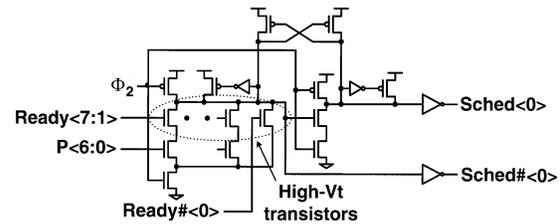


Fig. 7. Scheduler priority encoder CSG circuit.

wide AND paths and realizes the complete critical path with single-ended 2×9 -way (Fig. 6) and 8-way (Fig. 7) dynamic OR circuits, respectively. The CSG circuit enables domino-compatible dual-rail outputs but requires only a single-rail input. It contains two dynamic nodes, a traditional complementary dynamic node and a true dynamic node. Both nodes precharge using the same clock. During evaluation, one of these two nodes transitions low, causing the nonswitching node to be actively held by a pMOS device turned on by the evaluating node. These cross-coupled pMOS transistors provide additional noise immunity, allowing wider OR-gates than those possible when leakage is compensated only by a normal half-keeper. Dual- V_T optimization is conducted for high performance and to meet target noise margin constraints. High- V_T is used on the 9- and 8-way domino-OR nMOS pull down transistors and low- V_T is used for all other transistors. The complete scheduler path requires only 6 gate-stages, improving critical path performance by 23% over the corresponding dual-rail implementation. Furthermore, the single-ended design achieves 67% layout area reduction and 25% loopback interconnect length reduction due to eliminating 50% of the scheduler logic transistors, enabling a dense layout occupying 210 $\mu\text{m} \times 210 \mu\text{m}$. Total active leakage power dissipation is 50% lower than the dual-rail domino design.

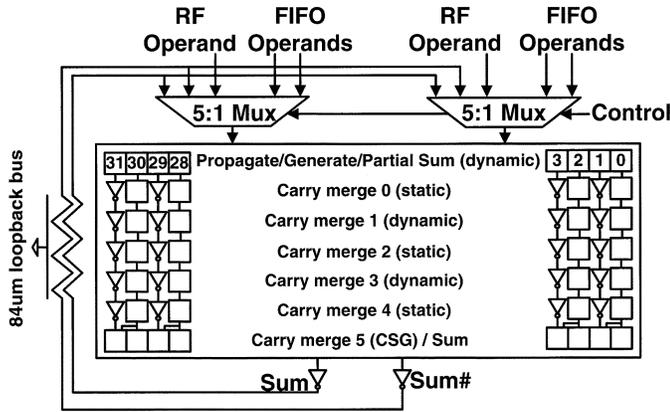


Fig. 8. 32-bit Han-Carlson ALU organization.

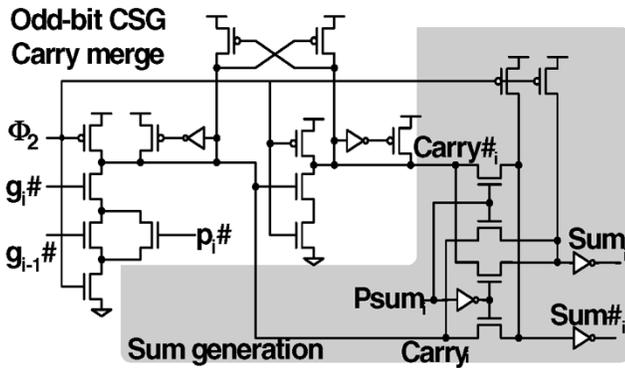


Fig. 9. ALU odd-bit CSG carry merge.

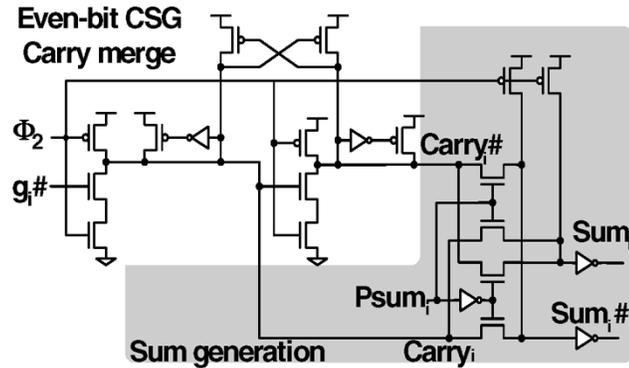


Fig. 10. ALU even-bit CSG carry merge.

IV. 32-BIT INTEGER ALU

The 32-bit ALU consists of a 5:1 source multiplexer, single-ended 32-bit dynamic adder core and an 84- μm differential ALU loopback bus (Fig. 8). The source multiplexer selects single-rail ALU operands from the true and complementary outputs of ALU loopback bus, 32-bit RF entries and external debug FIFO inputs. The sum/sum# adder outputs are driven onto the ALU loopback bus via a tristated bus driver. This organization enables single-cycle execution of add, subtract and accumulate instructions. The adder employs a radix-2 Han-Carlson architecture with carry-merge operation performed in both the dynamic and static stages of the domino gates. This results in a worst-case evaluation path of 3N-2P-2N-2P-2N-2P stacks, with

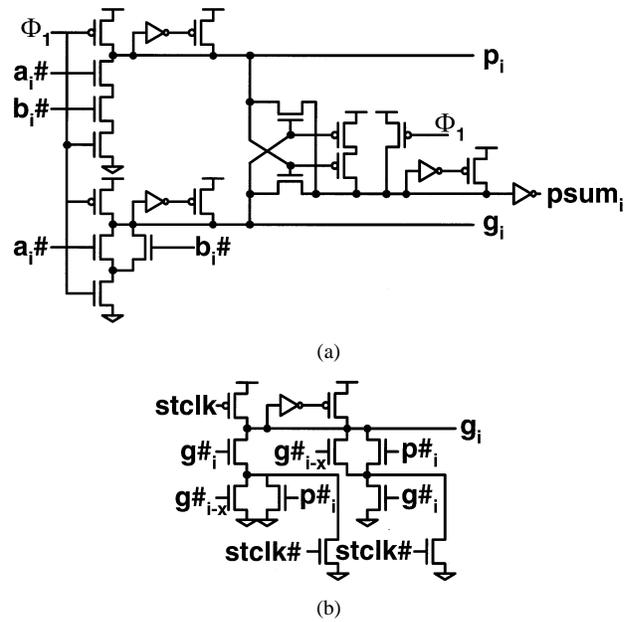
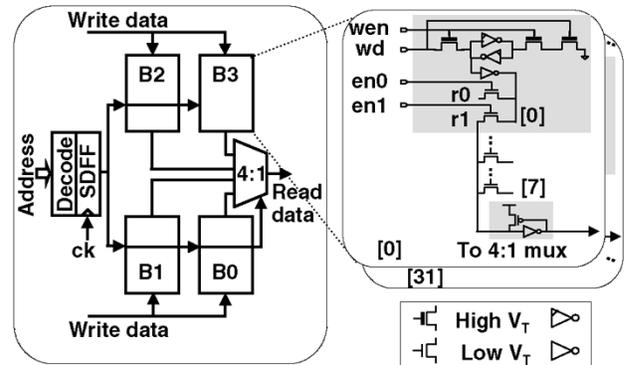


Fig. 11. ALU. (a) PG and partial sum circuit. (b) SNP carry merge stages.

Fig. 12. 32-entry \times 32-bit register file (RF).

initial P-G generation occurring in the first stage, followed by 5 stages of carry-merge logic. This implementation enables a 4-way carry-merge operation to be effected in two logic stages. Worst-case domino nMOS pull down is only 2-wide, allowing usage of performance-setting low- V_T transistors throughout the core while meeting noise immunity and active leakage power constraints. All dynamic nodes are fully shielded to minimize capacitive coupling noise. The Han-Carlson carry-merge tree skips odd carries (C_1, C_3, \dots, C_{31}) and generates 16 even carries (C_0, C_2, \dots, C_{30}) in 5 stages. An extra carry-merge logic stage is required to generate the missing odd carries at the end of the carry-merge tree. This logic is folded into a CSG and the output sum XORs to produce the dual-rail sum/sum# outputs for the odd bits in a single gate-stage, achieving a 10% delay reduction over the reference design in [5] (see Figs. 9 and 10). Unlike the scheduler, the CSG in the adder does not result in a gate stage reduction since the true and complementary paths were well balanced. Therefore this performance improvement is primarily due to wire length reductions throughout the carry merge tree from the elimination of the dual-rail path. The single-ended even carries also feed into a CSG with the output

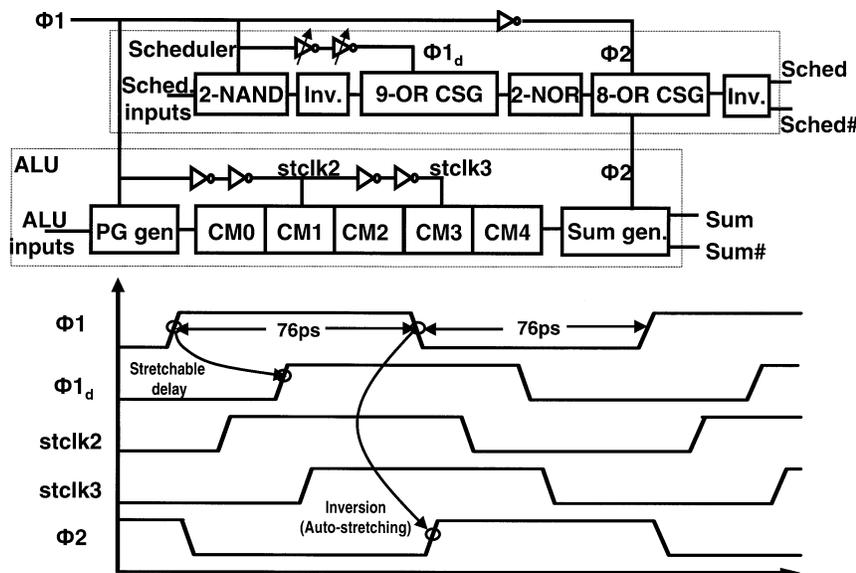


Fig. 13. Timing plan.

sum XORs folded-in to produce the dual-rail sum/sum# outputs for the even bits.

The P-G stage of the adder produces not only single-rail propagate and generate signals for the carry-merge tree, but also the partial sum, which is used in the final sum generation stage and therefore is not critical [Fig. 11(a)]. A dynamic pass-transistor XOR is used for the partial sum to reduce input loading. The inputs are set up before the $\Phi 1$ clock. Both sides of the pass-transistors in the XOR are precharged for robust glitch-free operation.

In addition to the above improvements, all intermediate stack nodes of the dynamic carry-merge stages are pre-discharged during precharge phase to minimize body effect, enabling best-case evaluate performance [6]. This stack node preconditioning is accomplished by adding small transistors to “precondition” the stack nodes of the gate during the precharge phase [Fig. 11(b)]. An nMOS transistor is added to the dynamic gate so that the stack node is pre-discharged to ground. A pMOS transistor is added to the static gate so that the stack node is precharged to V_{CC} . In order to minimize the charge-sharing noise inherent in this technique, the evaluation transistor stacks are split into two halves and transposed [6]. The technique provides a delay improvement of 10% in the ALU carry tree.

The Han-Carlson architecture with CSG usage enabled a single-rail ALU implementation with 50% fewer carry-merge gates and 40% less active leakage energy compared to a differential domino Kogge-Stone implementation [5]. Furthermore, with the Han-Carlson architecture, only alternate bits are propagated between consecutive carry-merge stages, resulting in a 50% reduction in inter-stage interconnect routing complexity compared to Kogge-Stone. This allowed a compact layout occupying $336 \mu\text{m} \times 84 \mu\text{m}$, with a worst-case inter-stage wire length of $168 \mu\text{m}$, contributing to further speed improvement.

V. 32-ENTRY \times 32-BIT REGISTER FILE

The RF unit is 32-entry by 32-bit with dual read ports and single write port (Fig. 12). The design is implemented as a large

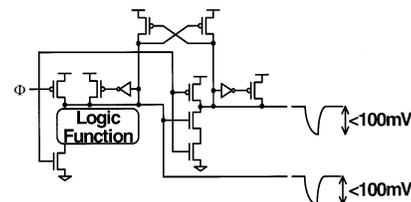


Fig. 14. CSG noise sensitivity.

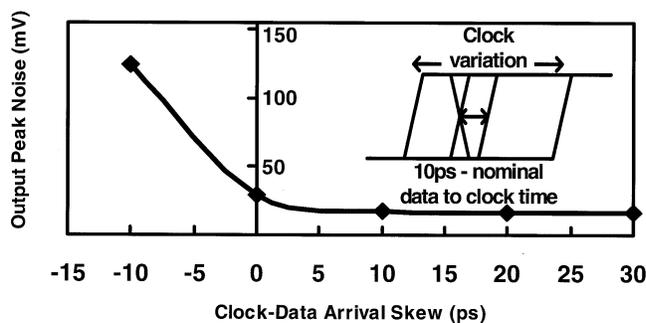


Fig. 15. CSG clock skew sensitivity.

signal memory array. A static design was chosen to reduce power and provide adequate robustness in the presence of large amounts of leakage. The RF design is organized in four identical 8-entry, 32-bit banks. For fast, single-cycle read operation, all four banks are simultaneously accessed and multiplexed to obtain the desired data. An 11-transistor, leakage-tolerant, dual- V_T optimized RF cell with 2-read/1-write ports is used. Reads and writes to two different locations in the RF occur simultaneously in a single clock cycle. To reduce the routing and area cost, the circuits for reading and writing registers are implemented in a single-ended fashion. Local bit lines are segmented to reduce bit-line capacitive loading and leakage. As a result, address decoding time, read access time, as well as robustness improve. RF read and write paths are dual- V_T optimized for best performance with minimum leakage. The RF RAM latch and access devices in the write path are made

high- V_T to reduce leakage power. Low- V_T devices are used everywhere else to improve critical read delay by 21% over a fully high- V_T design. For added noise immunity when reading a logic “1”, a half latch pulls the bit-line to rail. Using low- V_T allows reduced device sizes, providing a compact layout of $150 \mu\text{m} \times 340 \mu\text{m}$, where 83% of the total transistors in the design are low- V_T . A sparse body bias grid is routed over the entire unit.

VI. EXECUTION CORE TIMING

The execution core operates on a 50% duty-cycle 2Φ domino timing scheme, resulting in reduced circuit design and validation complexity (Fig. 13). Since the RF unit is implemented in static CMOS, it uses only the Φ_1 clock, while both the ALU and scheduler also use intermediate clocks. The Φ_2 clock is locally generated by inverting the incoming Φ_1 clock and triggers the CSG stages. Inputs to the CSG are setup before Φ_2 clock’s rising edge to minimize noise on the nonswitching output. This noise results because the true node of the CSG is poised to switch, with its input transitioning from high to low. In the case when the complementary node switches, the true node will have a glitch (Fig. 14). Peak output noise is limited to 100 mV for up to 30 ps of Φ_2 clock skew/jitter across process and temperature variations, meeting output noise constraints (Fig. 15). The dependence of the output noise glitch on clock skew, with positive clock-data arrival skew numbers indicating arrival of data before the clock, indicates the inherent robustness and leakage tolerance of the CSG. Even with simultaneous arrival of data and clock signals, the worst-case glitch is limited to 25 mV.

The scheduler’s ready logic CSG clock (Φ_{1d}) is a delayed version of Φ_1 clock, produced by an on-die programmable switched-capacitance delay cell to enable clock stretching for slow frequency debug. All Φ_1 clock boundaries use footed domino structures with embedded logic, enabling seamless time borrowing between the ALU, scheduler and register file interfaces without incurring an explicit skew/jitter penalty.

VII. BODY BIAS GENERATION AND DISTRIBUTION

On-chip body bias is used for the pMOS devices in the digital core of the chip. Fig. 16 shows the body bias generation and distribution details. A distributed bias generator architecture [7] was used to minimize variation of the body-to-source voltage due to global coupling and V_{CC} noise. A Central Bias Generator (CBG) uses a scaled bandgap circuit to generate a Process Voltage Temperature (PVT) insensitive 450-mV voltage with reference to V_{CCA} . This differential reference voltage is routed to 76 Local Bias Generators (LBG) distributed around the RF and ALU units in the execution core. Each LBG has a reference translation circuit that converts the 450 mV reference voltage to a voltage 450 mV below the local V_{CC} . This voltage is driven by a buffer stage and routed locally to the pMOS devices in the core to provide 450 mV of FBB during active operation. Local body bias routing tracks are placed adjacent to the local V_{CC} tracks to improve common-mode noise rejection and thus reduce noise-induced variations in the target 450 mV body bias to the pMOS devices. The voltage buffer and the local decoupling capacitor at the buffer output have been designed to min-

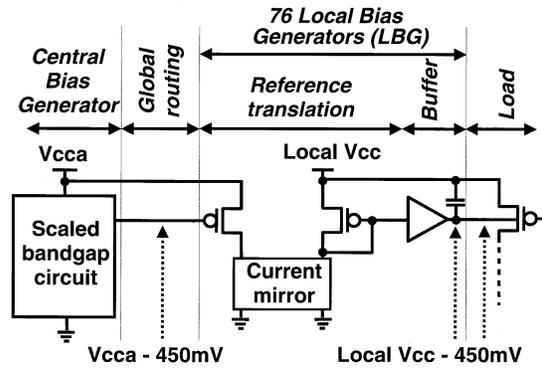


Fig. 16. Body bias generation and distribution.

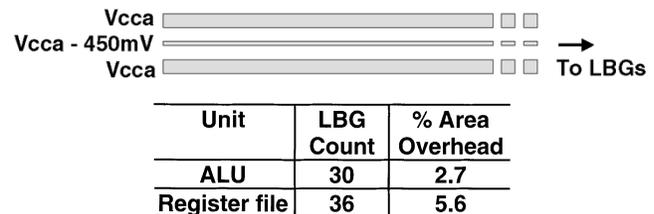


Fig. 17. Global body bias signal routing and biasing overhead.

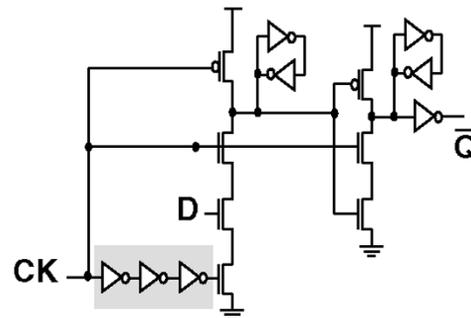


Fig. 18. Basic semi-dynamic flip-flop.

imize body bias variations induced by local coupling and V_{CC} noise.

Routing details of the global body bias signals are shown in Fig. 17. Global routing includes the PVT insensitive 450 mV reference voltage routed along with V_{CCA} tracks on both sides for proper shielding and adequate common-mode noise rejection. A digital control configures the LBG to apply forward or zero body bias the pMOS devices. An additional global control signal is used to disable the LBG for external body bias control. The ALU unit instantiates 30 LBGs with a 2.7% area overhead, while the RF unit uses 36 LBGs with a 5.6% area overhead. The dense layout of the register file results in increased area penalty.

VIII. FLIP-FLOPS

To enable 5-GHz operation, semi-dynamic flip-flops [8] are used for sequentials in the core. SDFP offers better clock-to-Q delay and clock skew tolerance than conventional static master-slave flops. SDFP (Fig. 18) has a dynamic master stage coupled to a pseudo-static slave stage. For best performance, all SDFPs were designed using 100% low- V_T devices. As is shown in the

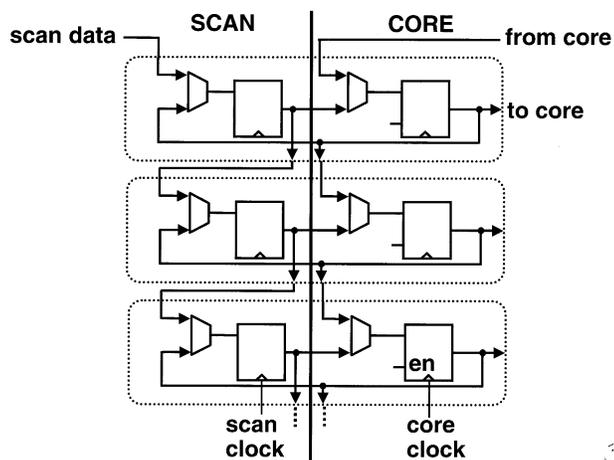


Fig. 19. FIFO cell and organization.

schematic, the flip-flops are implicitly pulsed. Pulsed flip-flops have several advantages over nonpulsed designs. One main benefit is that they allow time borrowing across cycle boundaries due to the fact that data can arrive coincident with, or even after, the clock edge. Thus negative setup time can be taken advantage of in the logic. Another benefit of negative setup time is that the flip-flop becomes less sensitive to jitter on the clock when the data arrives after clock. However, pulsed flip-flops have several important disadvantages. The worst-case hold time of this flip-flop can exceed clock-to-output delay because of pulse width variations across PVT conditions. Therefore, careful design is needed to avoid failures due to min-delay violations. All flip-flops used in the execution core were designed for an optimal energy-delay product.

IX. TEST CIRCUITS AND MEASUREMENTS

A. FIFO Design

Feeding the core at more than 5-GHz data rates and supporting at-speed results capture requires high-performance FIFOs. Hence, the core flip-flop in the FIFO cell is built using fast SDFF flops. The same cell is used in both the input and output FIFOs. Fig. 19 shows one column of the FIFO. The FIFO cell was designed to support both a low speed scan mode and a high-speed parallel FIFO mode. The output FIFO cell captures the 32-bit wide core data at-speed. The design allows easy transfer of this data between the core flop, operating at full speed and the scan flop, operating at a much lower speed. The scan clock can be run at arbitrary speeds and is only active during scan operations to save power.

The logic that enables at-speed capture is detailed in Fig. 20. First, the start and stop capture values are serially scanned in. The logic compares the start capture value to an internal 20-bit counter value and when equal, enables the start of result capture sequence. The logic then disables result capture sequence once the stop capture value is reached. The waveforms summarize the at-speed capture timing sequence. Once core execution starts, the logic asserts enable exactly after t_a core cycles and de-asserts enable exactly t_b core cycles past the assertion edge. The resulting enable signal is routed to the capture flip-flops in the output FIFO.

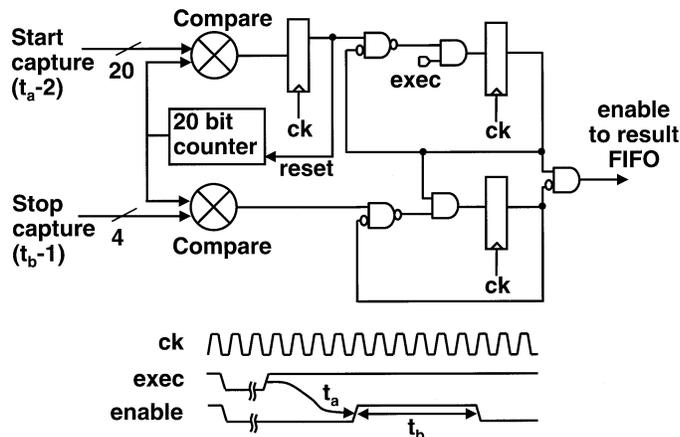


Fig. 20. At-speed capture logic.

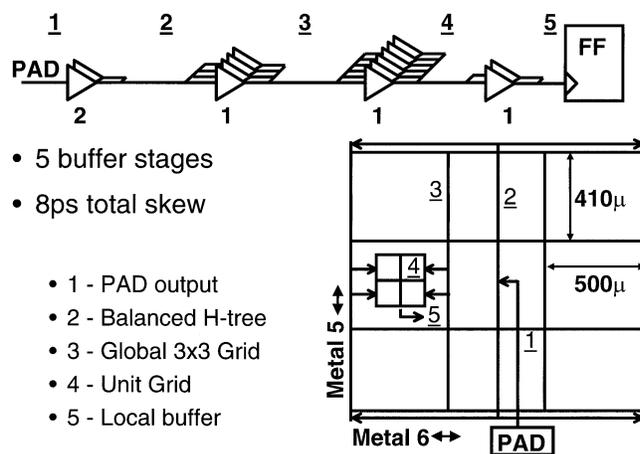


Fig. 21. Clock distribution.

B. On-Die Clock Distribution

The core clock distribution is shown in Fig. 21. There are a total of 5 stages of clock buffering from the pads to the clock inputs of the flip-flops in the execution core. First, there are two stages of buffering local to the pads used to drive the core clock to the center of the die. From the center, one more stage of buffering is added to drive a balanced H-tree to the four corners of the die. From the corners, another buffer stage is added to drive a symmetric, balanced, 3×3 global grid. Finally, the last stage of local buffering is added to all units. This last stage is sized according to the clock load of the particular unit. All clock buffers are composed of two CMOS inverters to minimize variations and use local decoupling capacitors to minimize jitter. The entire clock distribution uses upper-level metals (M6/M5) with V_{CC}/V_{SS} shielding for noise isolation and for symmetric current return paths. The core clock distribution network was simulated to have a maximum of 8 ps of total inter-unit skew and 2-ps worst-case skew between directly communicating units. Fig. 22 shows the core clock input circuit. An operational amplifier, located in the pads, converts differential sinusoidal clock inputs to a single-ended clock and forward it to buffers located at the center of the chip. The differential clocks are externally biased for duty cycle control, a feature needed for optimal operation of

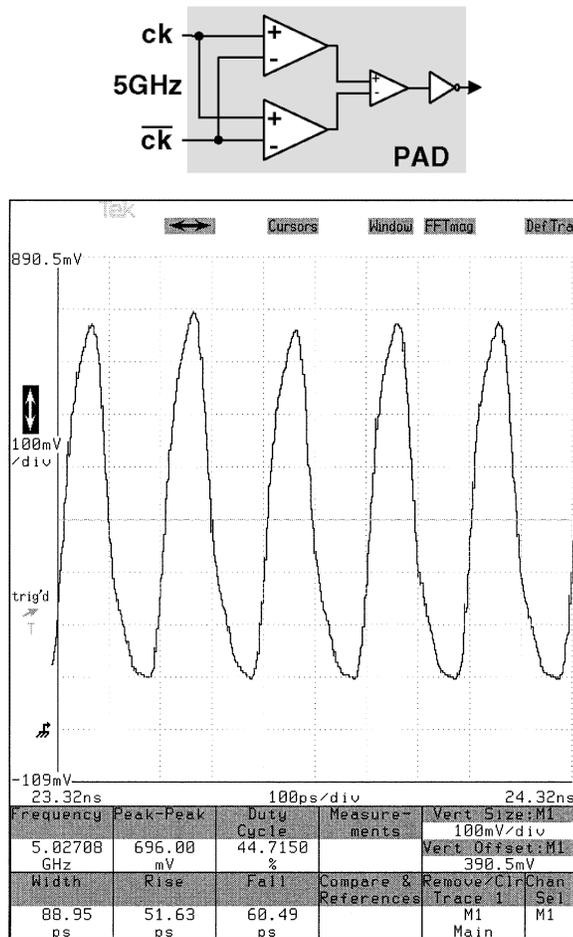


Fig. 22. Clock source and measured clock waveform.

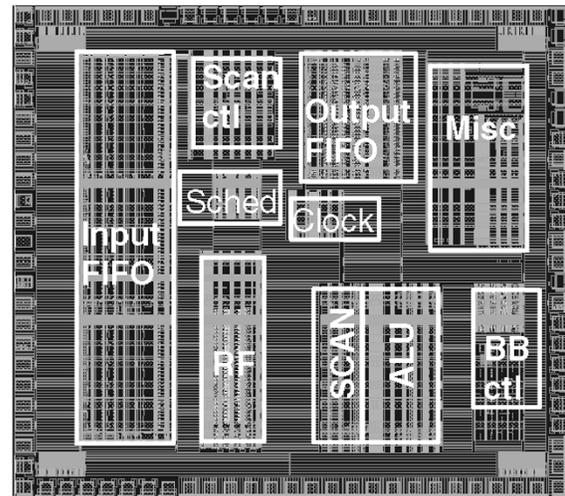
the domino ALU. Measured output clock waveform at 5 GHz from the output of a final clock buffer is also shown in Fig. 22.

C. Prototype Characteristics

Die micrograph and summary of chip characteristics are in Fig. 23. The blocks identified include the central clock drivers, register file, ALUs and instruction scheduler units, input and output FIFOs and the scan controller. The central bias generator circuits are part of the body bias control block. The 2.3-mm² fully custom design contains 160 000 transistors. There are 72 I/O pads along the die periphery, of which 30 are signal pads and 42 are power pads. Decoupling capacitors occupy approximately 20% of the total chip area.

D. Measurement Setup

The die was characterized on the wafer using a membrane probe card [9], length-matched to support differential clocks at speeds beyond 10 GHz (Fig. 24). A signal generator and a pulse inverting balun generated the differential clocks. An external power supply provides the DC bias for clock duty cycle control. A semiconductor parameter analyzer provides the external body bias supplies that individually control the biasing of nMOS and pMOS as well as high and low V_T devices for each unit. A PC running custom software is used to apply test vectors and observe results through the on-chip scan chain. The



Die Area	1.61 x 1.44 mm
Process	130nm CMOS
Interconnect	1 poly, 6 metal
Transistors	160K
Frequency	5GHz
Maximum V_{cc}	1.5V
Core Power	370mW @ 1.43V
Pad Count	72

Fig. 23. Die microphotograph and characteristics.

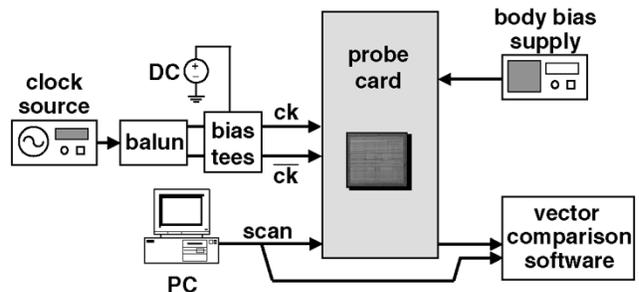


Fig. 24. Measurement setup.

membrane probe card used to characterize the design is shown in Fig. 25. The membrane probe consists of probe metallizations on a polyamide dielectric. Several 50- Ω microstrips on the polyamide connect the probe metallizations to semi-rigid coaxial lines for high-speed signals and to a supporting FR-4 probe card for lower speed signals.

X. MEASUREMENT RESULTS

Body bias improvement measurements showing frequency vs. supply voltage measurements of the ALU and RF are shown in Fig. 26. The domino ALU has better sensitivity in response to power supply increase when compared to the static register file design. At room temperature, 1.25 V and zero body bias, the ALU operates at 6.8 GHz. The RF frequency is 5.1 GHz at 1.43 V. Applying 450-mV FBB to both nMOS and pMOS transistors allows the target 5-GHz core frequency to be achieved at lower V_{CC} values for both ALU and RF. V_{CC} for 5-GHz operation is reduced from 1.05 to 0.95 V for the ALU, a 9.5% reduction and from 1.43 to 1.37 V for RF, a 4.2% reduction.

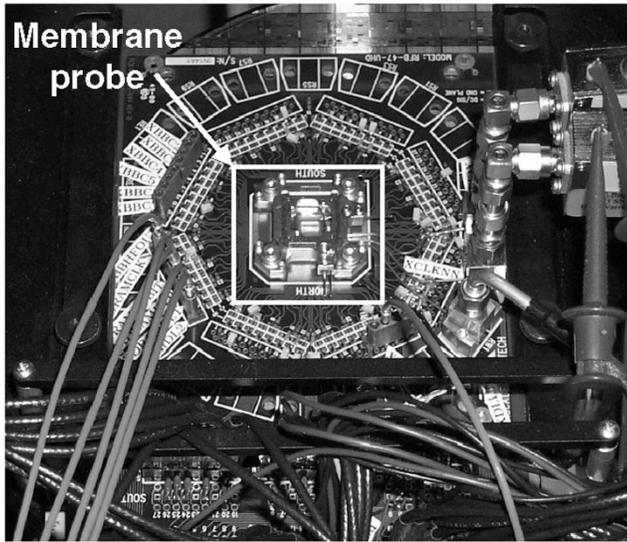
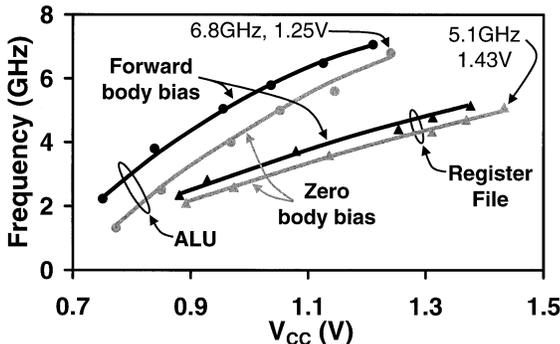
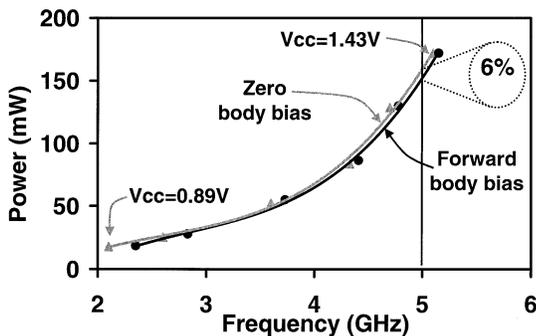


Fig. 25. Membrane probe card.



Vcc at 5GHz		
	ZBB	450mV FBB
ALU	1.05V	0.95V
RF	1.43V	1.37V

Fig. 26. ALU and register file frequency versus supply voltage.



5GHz	V _{cc} (V)	Power (mW)
ZBB	1.43	165
FBB	1.37	154

Fig. 27. Register file power versus frequency.

Power consumption of the RF as a function of frequency and with and without forward body bias is shown in Fig. 27. For this measurement, the power supply for the RF is varied from 0.89

	ALU	Scheduler
Area	50%	67%
Performance (Delay)	10%	23%
Active Leakage	40%	50%
Robustness	equal	equal

Fig. 28. Percentage improvement of single-rail CSG over dual-rail domino.

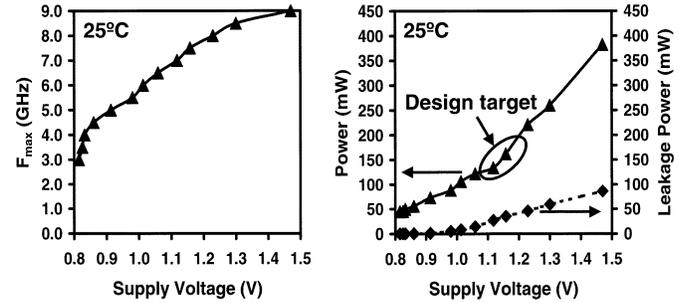


Fig. 29. ALU and instruction scheduler loop shmoo measurements.

to 1.43 V. At a target frequency of 5 GHz, with zero body bias and 1.43 V, the RF consumes 165 mW. The power consumption of the register file reduces by 6% to 154 mW on application of 450 mV of forward body bias.

At 5 GHz, the ALU dissipates 95 mW (1.05 V, 25 °C). At 6.5-GHz operation, the measured ALU and scheduler loop power increases to 120 mW with 15 mW of active leakage power. By increasing the voltage to 1.7 V, the ALU and scheduler loop frequency increases to 10 GHz. The advantages of the single-ended scheduler and ALU over dual rail schemes are summarized in Fig. 28. Area savings are 50% in the ALU since the dual-rail domino path has been eliminated. The scheduler savings are larger because the eliminated path consumed more than half the area. Both the ALU and instruction scheduler benefit from these area reductions as delay improvements, while the scheduler’s 23% delay improvement is also due to the reduction in gate stages. Active leakage is simultaneously reduced, as fewer transistors are needed to implement the logic. Fig. 29 shows the maximum frequency (F_{max}), switching power and active leakage versus supply voltage measurements.

XI. SUMMARY

The integer execution core consists of a 5-GHz 32-bit ALU, an 8-entry \times 2-ALU instruction scheduler and a 32-entry \times 32-bit leakage-tolerant register file, all fabricated in a 130-nm dual V_T CMOS process. At 5 GHz, the execution core dissipates 370 mW. The circuit innovations described enable simultaneous performance, area and leakage improvements in out-of-order execution engines of superscalar processors. The ALU and scheduler loop achieves 10-GHz operation at 1.7 V and 25 °C.

ACKNOWLEDGMENT

The authors thank all project members at Circuit Research Lab who contributed to this development; the Pyramid Probe Division of Cascade Microtech, Inc. for prompt and exceptional

membrane probe support; D. Sager, P. Madland and M. Milstein for discussions; K. Truong and K. Ikeda for their mask design expertise and, R. Hofsheier, F. Pollack and J. Rattner for their encouragement and support.

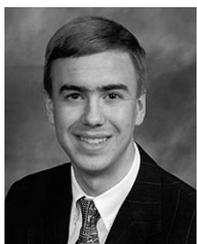
REFERENCES

- [1] D. Sagar *et al.*, "A 0.18 μm CMOS IA32 microprocessor with a 4 GHz integer execution unit," in *Proc. ISSCC Dig. Tech. Papers*, Feb. 2001, pp. 324–325.
- [2] S. Tyagi *et al.*, "A 130 nm generation logic technology featuring 70 nm transistors, dual V_T transistors and 6 layers of Cu interconnects," in *Proc. IEDM Tech. Dig.*, Dec. 2000, pp. 567–570.
- [3] M. Anders *et al.*, "A 6.5 GHz 130 nm single-ended dynamic ALU and instruction scheduler loop," in *Proc. ISSCC Dig. Tech. Papers*, Feb. 2002, pp. 410–411.
- [4] S. Vangal *et al.*, "A 5GHz 32 b integer execution core in 130 nm dual- V_T CMOS," in *Proc. ISSCC Dig. Tech. Papers*, Feb. 2002, pp. 412–413.
- [5] S. Mathew *et al.*, "Sub-500 ps 64b ALU's in 0.18 μm SOI/bulk CMOS: Design & scaling trends," in *Proc. ISSCC Dig. Tech. Papers*, Feb. 2001, pp. 318–319.
- [6] Y. Ye *et al.*, "Comparative delay, noise and energy of high-performance domino adders with stack node preconditioning," in *2000 Symp. on VLSI Circuits*, pp. 188–191.
- [7] S. Narendra *et al.*, "1.1 V 1 GHz communications router with on-chip body bias in 150 nm CMOS," in *Proc. ISSCC Dig. Tech. Papers*, Feb. 2002, pp. 270–271.
- [8] J. Tschanz *et al.*, "Comparative delay and energy of single edge-triggered & dual edge-triggered pulsed flip-flops for high-performance microprocessors," in *Proc. ISLPED '01*, pp. 147–151.
- [9] "Selecting, designing and using microwave pyramid probe [TM] cards," Cascade Microtech, Inc., Beaverton, OR, Application Note PYRPROAN-0397.



Sriram Vangal (S'90–M'98) received the B.S. degree from Bangalore University, India, in 1993, and the M.S. degree from University Of Nebraska, Lincoln, in 1995, both in electrical engineering.

He has been with Intel since 1995. He is currently a member of the Circuit Research Laboratories, Intel Laboratories, Hillsboro, OR, engaged in a variety of advanced prototype design activities. His research interests are in the area of low-power and high-speed circuits. He has 13 patents pending in these areas.



Mark A. Anders received the B.S. and M.S. degrees in 1998 and 1999, from the University of Illinois at Urbana-Champaign, both in electrical engineering.

Since graduation, he has been with Intel's Micro-processor Research Labs, Hillsboro, OR, where he is currently working on high-performance circuits research.



Nitin Borkar received the M.Sc. degree in physics from University of Bombay, Mumbai, India, in 1982, and the M.S.E.E. degree from Louisiana State University in 1985.

He joined Intel Corporation in 1986, where he worked on the design of the i960 family of Embedded microcontrollers. In 1990, he joined the i486DX2 microprocessor design team and led the design and the performance verification program. After successful completion of the i486DX2 development, Nitin worked on high-speed router

technology for the Teraflop machine. He now leads the prototype design team in Intel Labs, implementing and validating research ideas in the areas of in high performance—low power circuits and high speed signaling.



University of Phoenix.

Erik Seligman received B.A. degree in mathematics from Princeton University, Princeton, NJ, in 1991, and the M.S. degree in computer science from Carnegie Mellon University, Pittsburgh, PA, in 1993.

He has been with Intel for eight years. He is currently a CAD engineer in the Desktop Platforms Group, where he is working on formal equivalence verification of next-generation processor designs. His previous positions at Intel have included the Circuit Research Lab and the Strategic CAD Lab. In addition, he teaches mathematics part-time at the



Venkatesh Govindarajulu received the Bachelors degree in electronics engineering from Bangalore University, Bangalore, India, in 1993, and the Masters degree in computer engineering from Iowa State University, Ames, IA.

He has since worked in Intel®, in the micro-processor group on the Pentium®-II and Pentium®-III micro-processors, in the Circuit Research Labs on various advanced prototype designs and is currently a member of the XScale™ co-processor design team located in Austin, TX. He is engaged in both design methodologies and circuit design activities.



Vasantha Erraguntla received the Bachelors degree in electrical engineering from Osmania University, India, in 1989, and the Masters degree in computer engineering from University of Southwestern Louisiana, in 1991.

She joined Intel in 1991 and worked on the high-speed router technology for the Teraflop machine. She then joined Design Technology team validating performance verification tools for high-speed designs. For the last 5 years, she has been a part of the prototype design team in Intel Labs, implementing

and validating research ideas in the areas of in high performance & low power circuits and high speed signaling.



Howard Wilson was born in Chicago, IL, in 1957. He received the B.S. degree in electrical engineering from Southern Illinois University, Carbondale, in 1979.

From 1979 to 1984 he worked at Rockwell-Collins in Cedar Rapids, IA where he designed navigation equipment plus electronic flight display systems. From 1984 to 1991 he work at National Semiconductor in Santa Clara, CA designing telecom components for ISDN. With Intel since 1992, he is currently a member of the Circuits Research

Laboratory located in Hillsboro, OR, engaged in a variety of advanced prototype design activities.

Amaresh Pangal received the B.E. degree from University of Mysore, Mysore, India, in 1992, and the M.S. degree from Arizona State University, Tempe, in 1995.

He has been with Intel since 1995. His interests are in high-speed digital design and Network protocols. He has six patents pending in these areas.



Venkat Veeramachaneni received the B.E. degree in electrical engineering and the M.S. degree in physics from the Birla Institute of Technology and Science, Pilani, India, in 1997 and M.S. degree in electrical engineering from University of Virginia, Charlottesville, in 1999.

He has been with Intel Labs since 1999, where his work includes design of prototypes in the areas of low power high performance circuits and high speed signaling. He has authored or co-authored three papers and has two patents pending in these areas.



James W. Tschanz received the B.S. degree in computer engineering in 1997 and the M.S. degree in electrical engineering in 1999, both from the University of Illinois at Urbana-Champaign.

Since 1999 he has been a circuits researcher at Intel Laboratories, Hillsboro, OR. His research interests include low-power digital circuits, design techniques and methods for tolerating parameter variations. He is an adjunct faculty member at the Oregon Graduate Institute in Beaverton, OR, and has authored several papers and patents pending.



Yibin Ye received M.S. and Ph.D. degrees in electrical engineering from Purdue University in 1994 and 1997, respectively.

He is currently with Circuit Research Lab, Intel Labs, Intel Corporation, Hillsboro, OR. His current research interests include high performance and low power circuit techniques, logic synthesis and optimization and algorithms in combinatorial optimization.



Dinesh Somasekhar (S'95–M'98) received the B.S.E.E. degree from Maharaja Sayajirao University, Baroda, India, the M.S.E.E. degree from Indian Institute of Science, Bangalore, India, and the Ph.D. degree from Purdue University, West Lafayette, IN, in 1989, 1991, and 1999, respectively.

From 1991 to 1994 he was an IC Design Engineer with Texas Instruments (TI), Bangalore, India, where he designed ASIC compiler memories and interface ICs. Since 1999, he has been a researcher in Microprocessor Research of Intel Labs, Hillsboro, OR.

Bradley A. Bloechel (M'95–A'96) received the A.A.S. degree in electronic engineering technology from Portland Community College, Portland, OR, in 1986.

He joined Intel Corp., Hillsboro, OR, in 1987 as a Graphics Design Technician for the iWarp project supporting the RFU and LLU design effort. In 1991, he transferred to Supercomputer Systems Division Component Technology, where he supported VLSI test/validation effort and extensive fixturing support for accurate high-speed test and measurement of the interconnect component used in the Tera ops computer project (Intel, DOE and Sandia). In 1995, he joined the Circuits Research Laboratory, Microcomputer Research Laboratory, where he is a Senior Lab Technician specializing in on-chip dc and high-speed I/O measurements and characterization.

Mr. Bloechel is a member of Phi Theta Kappa.



Gregory E. Dermer received the B.S. degree in electrical engineering from Indiana Institute of Technology, Fort Wayne, in 1977, and the M.S. degree in electrical and computer engineering from the University of Wisconsin, Madison, in 1983.

From 1979 to 1992, he held a variety of processor architecture, logic design and physical design positions at Cray Research, Inc., Nicolet Instrument Company, Astronautics Corporation of America, and Tandem Computers, Inc. In 1992, he joined the Intel Corporation's Supercomputer Systems Division.

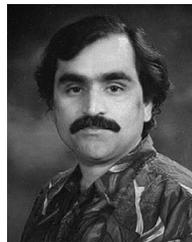
While there, he worked on clock system design and reliability modeling for the Intel ASCI Red supercomputer. For the past six years, he has worked in the circuits research area of Intel Labs, Hillsboro, OR, on physical design and measurements for high-speed interconnections.



Ram K. Krishnamurthy (S'92–M'98) received the B.E. degree in electrical engineering from Regional Engineering College, Trichy, India, in 1993 and the Ph.D. degree in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, in 1998. His Ph.D. research focused on low-power DSP circuit design.

Since graduation, he has been with Intel Corporation's Microprocessor Research Labs in Hillsboro, Oregon, where he is currently a Senior Staff Engineer and Manager of high-performance and low-voltage circuits research group. He is an adjunct faculty of Department of Electrical and Computer Engineering, Oregon State University, where he teaches VLSI System Design. He holds 16 patents issued, 40 patents pending, and has published over 35 papers in refereed journals and conferences.

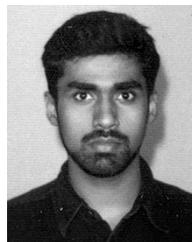
Dr. Krishnamurthy serves on the SRC ICSS Task Force and the program committees of the IEEE CICC, ASIC, and ISCAS conferences. He is the Technical Program Co-Chair for the 2003 IEEE International ASIC/SoC Conference.



K. Soumyanath received the B.E. degree in electronics and communication engineering from the Regional Engineering College Tiruchirappalli, India, in 1979, the M.S. degree in electrical communication engineering from the Indian Institute of Science, Bangalore, India, in 1985, and the Ph.D. degree in computer science from the University of Nebraska in 1993.

He was a faculty member at Tufts University, Medford MA until 1995 where he served as the director of the ARPA supported program in mixed signal IC design, for the Department of Defense. Since 1996 he has been at Intel Corporation where he is currently the Director of communications circuits research. He has published over 15 papers in VLSI and holds eight patents. In addition to CMOS circuits of all kinds, his research interests include classical Tamil poetry.

In 1998 Dr. Soumyanath served as the Chair for the Design Sciences task force for the Semiconductor Research Corporation and currently serves on the Technical Program Committee for ICCD.



Sanu Mathew received the Ph.D. degree in electrical engineering from the State University of New York at Buffalo in 1999. His dissertation focused on asynchronous circuit design.

He is currently part of the high-performance circuits research group at Intel Corporation's Microprocessor Research Labs, Hillsboro, OR.



Siva G. Narendra received the B.E. degree from Government College of Technology, Coimbatore, India, in 1992, the M.S. degree from Syracuse University, Syracuse, NY, in 1994 and the Ph.D. degree from Massachusetts Institute of Technology, Cambridge, in 2002.

He has been with Intel Laboratories since 1997, where his research areas include low voltage MOS analog and digital circuits and impact of MOS parameter variation on circuit design. He has authored or co-authored over 16 papers and has 15 issued and

27 pending patents in these areas. Dr. Narendra is an Adjunct Faculty with the Department of Electrical and Computer Engineering, Oregon State University, Corvallis.

Dr. Narendra is an Associate Editor for the IEEE TRANSACTIONS ON VLSI SYSTEMS and a Member of the Technical Program Committee of the 2002 International Symposium on Low Power Electronics and Design.



Mircea R. Stan (SM'94) received the Diploma in Electronics from the Polytechnic Institute of Bucharest, Romania, in 1984 and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Massachusetts at Amherst in 1994 and 1996, respectively.

Since 1996 he has been with the Electrical and Computer Engineering Department at the University of Virginia in Charlottesville, first as an assistant professor and since 2002 as an associate professor. He is teaching and doing research in the areas of

low-power VLSI, mixed-mode analog and digital circuits, computer arithmetic, embedded systems and nanocircuits. He has more than eight years of industrial experience as an R&D Engineer and has been a visiting faculty at IBM in 2000 and at Intel in 2002 and 1999.

In 1997 Dr. Stan has received the NSF CAREER Award for investigating low-power design techniques. He is a senior member of the IEEE and a member of ACM, Usenix. He is a member of Phi Kappa Phi and Sigma Xi.

Scott Thompson joined Intel in 1992 after completing his Ph.D., under Professor C. T. Sah at the University of Florida, on thin gate oxides. He has worked on transistor design and front-end process integration on Intel's 0.35, 0.25, 0.18, and 0.13 μm silicon process technology design for the Intel® Pentium® and the Pentium® II microprocessors. Scott is currently managing the development of Intel's 90 nm logic technology.



Vivek De (S'86-M'92) received the Ph.D. degree in Electrical Engineering from Rensselaer Polytechnic Institute, Troy, New York in 1992.

He is a Principal Engineer and Manager of Low Power Circuit Technology at Microprocessor Research of Intel Labs, Hillsboro, OR. He has authored 82 technical papers in refereed international conferences and journals and two book chapters on low power design. He has 23 issued patents and 45 more patents filed (pending).

Dr. De served as Technical Program Chair of 2001 ACM/IEEE International Symposium on Low Power Electronics and Design (ISLPED'01), General Chair of ISLPED'02 and Technical Program Chair of 2002 ACM Great Lakes Symposium on VLSI. He served on technical program committees of ARVLSI and ISQED conferences. He is the guest editor of a special issue on low power electronics for IEEE TRANSACTIONS ON VLSI SYSTEMS and an adjunct faculty at the Department of Elec and Computer Engineering at Oregon State University. He is the recipient of a best paper award at the 1996 IEEE International ASIC Conference in Portland, OR.



Shekhar Borkar received the B.Sc. and M.Sc. degrees in physics in 1977 and 1979, respectively, and the M.S.E.E. degree in 1981 from University of Notre Dame.

He joined Intel Corporation, where he worked on the design of the 8051 family of microcontrollers, high speed communication links for the iWarp multi-computer and Intel Supercomputers. He is an Intel Fellow and Director of Circuit Research in the Intel Labs, researching low power high performance circuits and high speed signaling. He is also an adjunct

faculty member of Oregon Graduate Institute and teaches Digital CMOS VLSI design.