20.3 Sub-500ps 64b ALUs in 0.18µm SOI/Bulk CMOS: Design & Scaling Trends

Sanu Mathew, Ram Krishnamurthy, Mark Anders, Rafael Rios, Kaizad Mistry, K. Soumyanath

Intel Corp., Hillsboro, OR

The requirements of high-throughput Internet servers necessitate the use of multiple ALUs in high-performance 64b execution cores. Consequently, each ALU demands a compact, energy-efficient 64b adder core with single-cycle latency. The resultant critical path, which is a balanced mix of interconnect, diffusion and gate loads, forms a representative test bed for evaluating competing circuit techniques and process technologies (bulk CMOS/SOI). This paper presents: (i)the design of an energy-efficient 64b ALU in 0.18µm bulk CMOS technology (ii)a direct port of this design to a comparable SOI technology and (iii)an SOI-optimal redesign of the adder core. Further, it describes design margining required for the SOI implementations and reports the results of shrinking the two architectures to 0.13µm Bulk/SOI. In both cases, a sophisticated SOI compact model that incorporates features to effectively model the SOI floating body effect is used.

The 64b ALU is based on a single-rail, radix-2, Han-Carlson (HC) adder core [1] with two wide multiplexer stages and a write-back bus (Figure 20.3.1). The HC adder core performs carry-merge on alternate bit slices. Consequently, it has 50% fewer carry-merge gates when compared to a conventional Kogge-Stone adder (at the cost of an additional final carry-merge stage). The reduced active leakage (21% smaller) and an efficient energy-delay characteristic make it suitable for the 64b bulk CMOS ALU (Figure 20.3.2a).

The single-rail implementation is enabled by use of a complementary function generator that delivers a domino-compatible Carry/Carry* from a single-ended carry signal (Figure 20.3.3a). The inputs to this circuit must be set up before the clock enters its evaluation phase, preventing time-borrowing. This 'non-time-borrowability' penalty is absorbed by placing the stage at the stretchable Φ_2 clock boundary. Ten stages of domino logic perform the PG-generation, Carry-merge and Sum-generation.

A single-ended XNOR circuit generates a domino-compatible partial sum output (Figure 20.3.3b). The partial sum output transitions low when one of the inputs (P_i/G_i) is asserted low subsequent to precharge. Potential noise glitches caused by skewed input arrival times are mitigated by the simultaneous evaluation of P_i and G_i.

The HC ALU operates at 482ps in 1.5V, 0.18µm bulk CMOS technology [2]. A direct port of this design to a comparable SOI technology [3] results in a 16% increase in performance (Figure 20.3.4a). The SOI and bulk processes [2][3] have matching $I_{OFF,DC}$ at room-temperature (Figure 20.3.2b), and compact model parameters are derived from experimental data. As expected, the maximum (35%) and minimum (2%) speedups are obtained in the diffusion capacitance dominated wide multiplexers and the load-dominated dynamic gates respectively (Figure 20.3.4b). These comparisons include layout (Figure 20.3.8) extracted parasitics, where the maximum interstage load is 750µm of interconnect, spanning 32 carry-merge bit slices. The 16% speedup is lower than that reported in [4]. This is attributable to the aggressive reduction of junction capacitances in the contending bulk CMOS process [3].

The floating body effects in SOI suggest a redesign of the adder core to SOI-amenable architecture using deeper stacks. However, the accompanying increase in fanout loads makes the Han-Carlson architecture unsuitable for such a redesign. A quaternary tree carry-select [5] adder core (Figure 20.3.5), enables use of deeper stacks, with a simultaneous reduction in P, G fanouts and interconnect loads (by 1/3 and 1/2 respectively). This characteristic results in an SOI-optimal realization of the ALU. The carry-merge tree is changed from a standard 2P-2N-2P-2N-2P-2N carry-merge to a 2P-4N-2P-3N chain, eliminating two stages of logic depth from the critical path. While the HC ALU in SOI is a direct port of bulk CMOS design to SOI, the redesigned Deeper Stack Carry Select (DSCS) ALU showcases the main advantages of SOI technology. The DSCS ALU provides a further 5% speedup (Figure 20.3.4a), raising the overall SOI performance improvement to 21%.

While the forward-biased floating body in SOI contributes to the performance improvement over bulk, the body of an SOI device can also become reverse-biased during normal switching activity. Figure 20.3.6 shows the input sequence that sets up reverse body-bias in a 2-input static NOR gate. Starting with timestep T_1 (A=1 & B=0), the rising transition on B in T_2 is coupled onto the body of transistor A through the stack-node and bumps 'Body-A' to 1.5V. The subsequent events in T_3, T_4 and T_5 sets up 400mV of reverse body bias on transistor A. Consequently, the pMOS pull-up performance degrades, resulting in a 10% decline in SOI speedup between timesteps T_3 and T_6. This can cause a max-delay pushout, and produce new critical paths. Precharging the intermediate node to V_{cc} (with M_1 in Figure 20.3.6) reduces the speedup degradation to 2%. Since reverse-bias conditions are difficult to detect during timing analysis, a worst-case max-delay margin of 2% must be added, reducing the overall speedup of an SOI-optimal 64b ALU to 19%.

Figure 20.3.7 shows experimentally measured history effect data for a variety of gates, including chains of inverters, static 3-input NAND gates and transmission gates in a 0.18µm process. These gates cover the range of circuits that are present in this ALU design. Process model parameters that affect transient SOI behavior are fitted to measured data. Consequently, the simulated transient SOI effects correspond well with measured data. Although the history effect does not affect the overall speedup, timing analysis tools are required to budget a 11% min-delay race margin.

To quantify the scaling trend of ALU circuits in SOI, both ALU designs are ported to 0.13µm Bulk/SOI technologies. As in the case of the 0.18µm generation, $I_{OFF,DC}$ at room temperature for both simulation models are matched. Key compact model parameters related to the MOSFET and impact ionization are obtained from 0.13µm bulk data, while fitting techniques for SOI parasitic BJT/diode characteristics are unchanged from the 0.18µm fitting. The speedup with a direct port of the HC ALU, from a 0.13µm bulk technology to a comparable SOI process, reduces to 11% (Figure 20.3.4a). This correlates well with the device-level scaling trends forecast in Reference 3. In the case of the redesigned DSCS SOI ALU, the speedup over bulk falls from 21% in 0.18µm to 18% in 0.13µm. Margining for reverse-bias reduces the overall SOI speedup of both architectures to 9% and 16%, respectively.

Acknowledgments:
The authors thank S. Narendra and A. Keshavarzi for assistance with measured results. They also thank S. Borkar, F. Pollack, W. Holt, S. Rusu and G. Singer for encouragement and support.

References:
[1] T. Han et al.,"Fast area-efficient VLSI adders," Proc. 8th Symp. on Computer Arithmetic, Sept. 1987.
[2] T. Ghani et al., "100nm Gate Length High Performance/Low Power CMOS Transistor Structures," Tech. Dig. IEDM, Dec 1999
[3] K. Mistry et al., "Scalability Revisited: 100nm PD-SOI Transistors and Implications for 50nm Devices", Proc. Symp. On VLSI Tech., July 2000.
[4] D. Stasiak et al., "A 2nd Generation 440ps SOI 64b Adder", ISSCC Digest of Technical Papers, Feb. 2000.
[5] R. Woo et al. "A 670ps, 64b Dynamic Low Power Adder Design", Proc. ISCAS, May 2000.

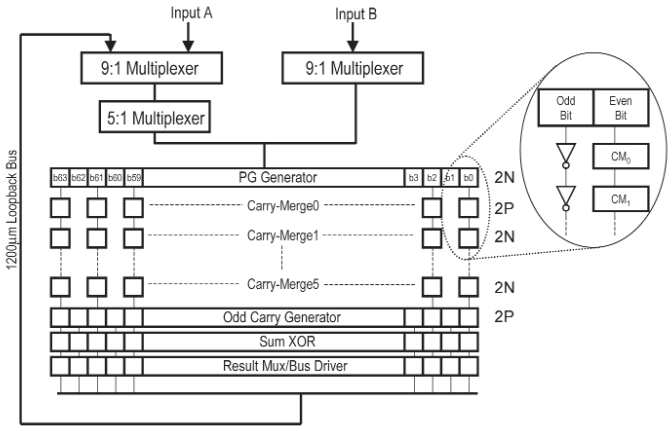


Figure 20.3.1: 64b ALU with Han-Carlson adder core.

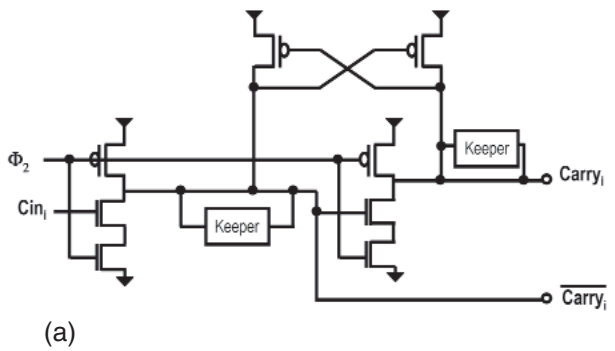
Adder Architecture	Energy/transition
Kogge-Stone	120pJ
Han-Carlson	68pJ

(a)

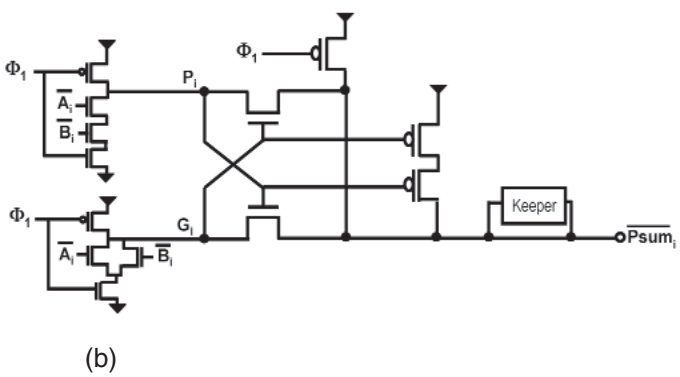
Technology	$I_{OFF\text{-}DC}$(nA/µm)	I_{DSAT}(µA/µm)
SOI-NMOS	3	1050
Bulk-NMOS	3	1040
SOI-PMOS	3	441
Bulk-PMOS	3	460

(b)

Figure 20.3.2: (a) Energy/transition of single rail 64b adders. (b) Process parameters: Vcc=1.5V, T=30°C, 0.18µm technology.



(a)



(b)

Figure 20.3.3: (a) Complementary function generator. (b) Partial sum generation circuit.

Technology (Adder Core Type)	ALU Delay (ps)	
	0.18µm Technology Vcc=1.5V	0.13µm Technology Vcc=1.2V
Bulk (HC)	482	351
SOI (HC) - Direct Port	403	312
SOI (DSCS) - Redesigned	380	286

(a)

Stage Type	Improvement over Bulk
Static Gates	12-15%
Dynamic Gates	2-9%
3:1 TG Mux	20%
5:1 TG Mux	23%
9:1 TG Mux	35%

(b)

Figure 20.3.4: (a) ALU simulation results at T=110°C. (b) Breakup of speedup obtained by direct port of HC ALU in 0.18µm technology.

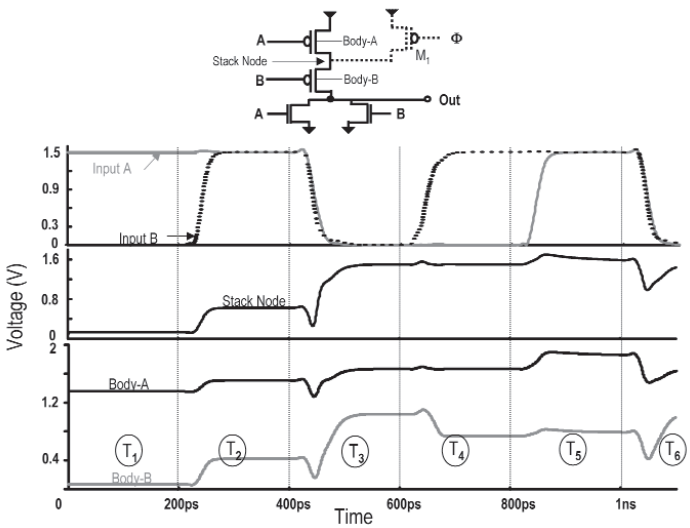


Figure 20.3.6: Waveforms showing reverse body-bias in 2-input NOR SOI gate.

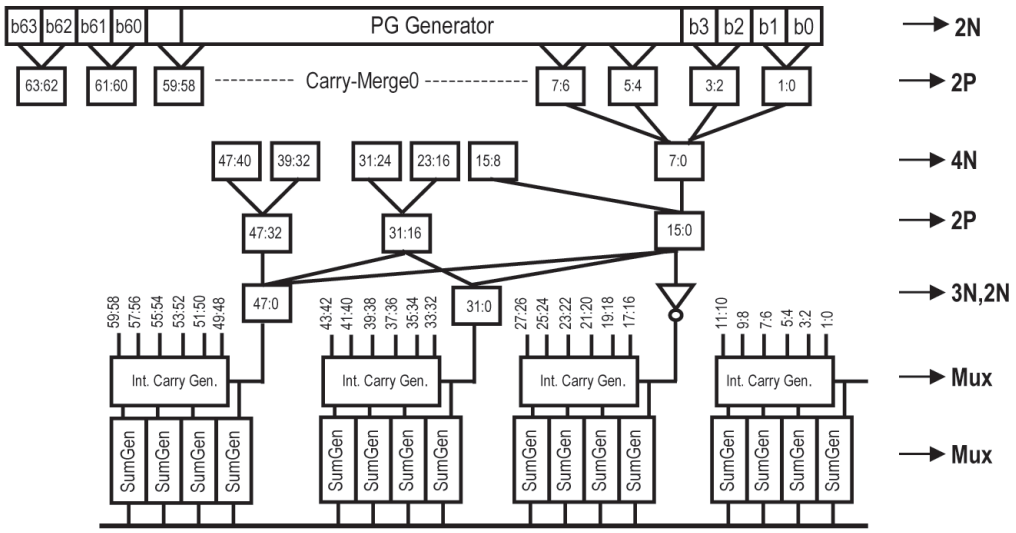


Figure 20.3.5: DSCS ALU adder core: SOI-optimal redesign.

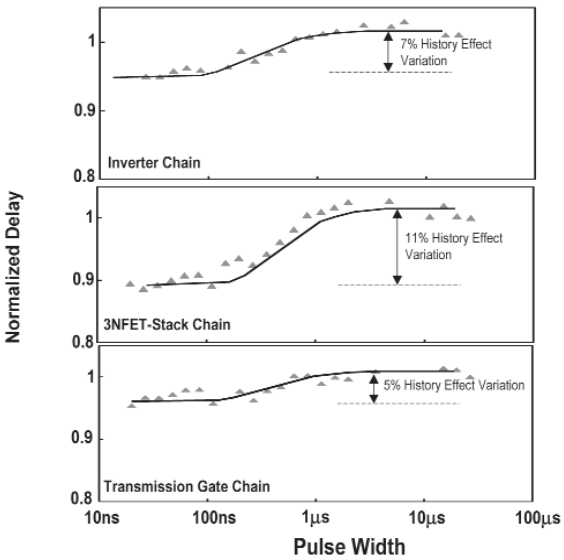


Figure 20.3.7: History effect measurements in 0.18µm SOI technology.

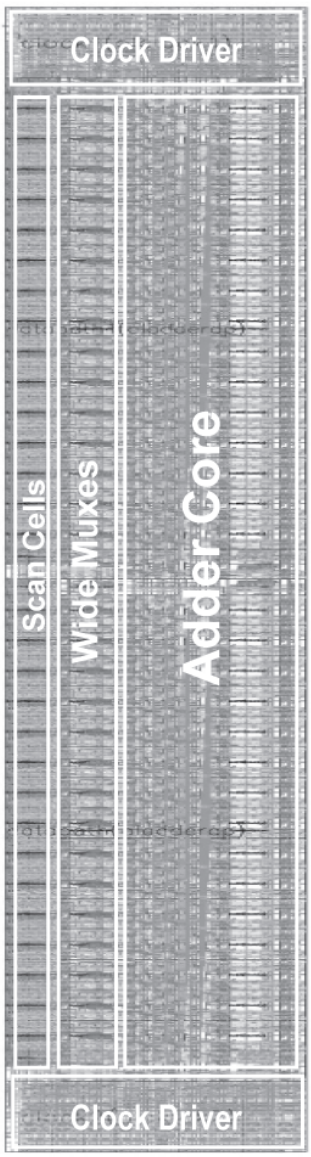


Figure 20.3.8: 64b ALU layout.