

## 6.2 A Shared-Well Dual-Supply-Voltage 64-bit ALU

Yasuhisa Shimazaki<sup>1</sup>, Radu Zlatanovic<sup>2</sup>, Borivoje Nikoli<sup>2</sup>

<sup>1</sup>Hitachi, Tokyo Japan, now with SuperH, Tokyo Japan

<sup>2</sup>University of California, Berkeley CA

Power and power densities have constantly been increasing, which brings them to become primary constraints in present and future integrated circuit designs. A common goal is to use the design techniques that achieves the highest operating frequency with lowest power. A reduction in the supply voltage of a circuit decreases power dissipation, but degrades speed performance. The supply voltage can be selectively lowered by using a dual-supply technique [1], whereby a second, lower voltage can be supplied to non-critical timing paths without compromising performance. Additionally, this second voltage is employed to selectively reduce the power of gates that drive large switched capacitances with small impact on overall speed.

Figure 6.2.1a shows a conventional dual-supply layout where high supply ( $V_{DDH}$ ) and low supply ( $V_{DDL}$ ) are applied to two neighboring cells. The cells have to be placed in separate rows because of required well separation, resulting in an impractical layout for the datapath design. Figure 6.2.1b shows the circuit schematic and layout examples of a shared N-well dual-supply technique that is better suited to the datapaths. The power supply is split into  $V_{DDH}$  and  $V_{DDL}$  rails. The N-well is always tied to  $V_{DDH}$ , while the cells are supplied from either  $V_{DDH}$  or  $V_{DDL}$  by simple via placement. Both  $V_{DDH}$  and  $V_{DDL}$  cells can be placed in the same row, making this an area-efficient technique with no area overhead in a datapath. The main disadvantages of this method are reduced drive current of the PMOS transistors and issues with power routing. However, both are addressed through careful design.

In the shared N-well technique, the delay of  $V_{DDL}$  circuits is additionally increased due to negative back-biasing of the PMOS transistors. Figure 6.2.2 shows the simulated fanout-of-4 inverter (FO4-INV) delay and subthreshold current of a PMOS transistor. As  $V_{DDL}$  decreases, the delay increases, resulting in an 18% speed degradation at 1.2V, compared to a conventional, non-back-biased  $V_{DDL}$  circuit. Since this increase in delay significantly affects the performance of conventional CMOS logic, domino logic is preferred for this dual-supply approach. An additional benefit is that the PMOS leakage is reduced by two orders of magnitude. The second potential problem is in the increase of power rail resistances because of their reduced width. Although the dual supply technique reduces total power, the length of the rows between power straps has to be limited to avoid an increased IR drop. Added space between the two supply rails affects cell layout density for small cell heights. However, in datapaths, the cell height is usually determined by architectural and performance requirements; therefore, the datapath circuit cells are made tall enough to avoid any loss in density.

A block diagram of a 64b ALU module implemented in domino logic employing the proposed dual-supply technique is shown in Fig. 6.2.3. The ALU module, similar to [2,3], consists of an ALU, an output buffer and an input operand selector. The ALU can execute arithmetic (add/sub) and logic (and/or/xor) functions. The carry path is implemented in the  $V_{DDH}$  domain, while the partial sum generation and the logical unit are supplied from  $V_{DDL}$ . Carry signals are computed using a sparse radix-4 tree whose structure is shown in Fig. 6.2.4. Every fourth carry is calculated in the tree. While the full radix-4 tree suffers from a large number of complex carry-merge gates, the sparse implementation significantly reduces the gate and wire count while increasing the complexity of the sum computation. This sparse tree is par-

ticularly suitable for a dual-supply implementation, where the complex sum precompute gates are placed in the  $V_{DDL}$  domain. In this implementation they are in the critical path when  $V_{DDL}$  is lowered to 1.2V. While the  $V_{DDH}$  gates can freely drive  $V_{DDL}$  gates, returning to  $V_{DDH}$  domain requires level conversion. Domino level converters, similar to [4], are used in the sum selectors and the 9:1 multiplexers. Detailed circuit schematics of the output buffer and the 9:1 multiplexer are shown in Fig. 6.2.5. Since the forwarding interconnect is long with a high fanout load, the output buffer has a large power consumption. Lowering the supply on the buffer to 1.2V results in 56% energy reduction with 22% delay increase. However, this delay penalty corresponds to only 8% cycle time increase for the complete ALU module. The datapath is organized using cells with a pitch of 18 metal-1 tracks in a bit slice. Since the carry is computed only for every fourth bit, the sum precompute cells and buffers are placed in empty rows, resulting in a very dense layout.

A micrograph of the test chip is shown in Fig. 6.2.6. The chip uses a 1.8V, general-purpose 0.18 $\mu$ m 1P 5M CMOS process, with local interconnect technology. The chip includes 6 ALU modules, to simulate the loading conditions of a 6-issue integer execution unit, control circuitry, clock drivers and test circuitry. An additional capacitance is added to simulate the cache and register file load. The size of the ALU module is 200 x 760 $\mu$ m, while the overall chip size is 2mm x 1.5mm. With  $V_{DDH} = V_{DDL} = 1.8V$ , the chip operates at its nominal frequency of 1.16GHz, corresponding to 13 FO4-INV delays. Figure 6.2.7a summarizes the effect of the dual-supply operation on circuit speed and energy consumption. Single-supply operation is plotted as a reference where the supply is scaled down to meet the target delay. When the target delay is increased by 2.8%, total energy saving is 25.3% using dual supplies. A delay increase of 8.3% results in an energy savings of 33.3%. In comparison to a single reduced supply operation, the energy savings are 20.2% and 20.9% respectively. Leakage power is reduced by 42% at  $V_{DDL} = 1.2V$ . Figure 6.2.7b illustrates the effect of the negative back-biasing of PMOS transistors.

### Acknowledgements

This work was supported by Hitachi Ltd. and MARCO C2S2. The authors thank ST Microelectronics for test chip fabrication.

### References

- [1] K. Usami and M. Horowitz, "Clustered Voltage Scaling for Low-Power Design," *ISLPED*, pp. 3-8, Apr. 1995.
- [2] S. Mathew et al., "Sub-500ps 64b ALUs in 0.18 $\mu$ m SOI/Bulk CMOS: Design & Scaling Trends," *ISSCC Dig. Tech. Papers*, pp. 318-319, 2001.
- [3] E. Fetzter and T. Orton, "A Fully-Bypassed 6-Issue Integer Datapath and Register File on an Itanium Microprocessor," *ISSCC Dig. Tech. Papers*, pp. 418-419, 2002.
- [4] N. Tzartzanis et al., "A 34Word x 64b 10R/6W Write-Through Self-Timed Dual-Supply-Voltage Register File," *ISSCC Dig. Tech. Papers*, pp. 416-417, 2002.

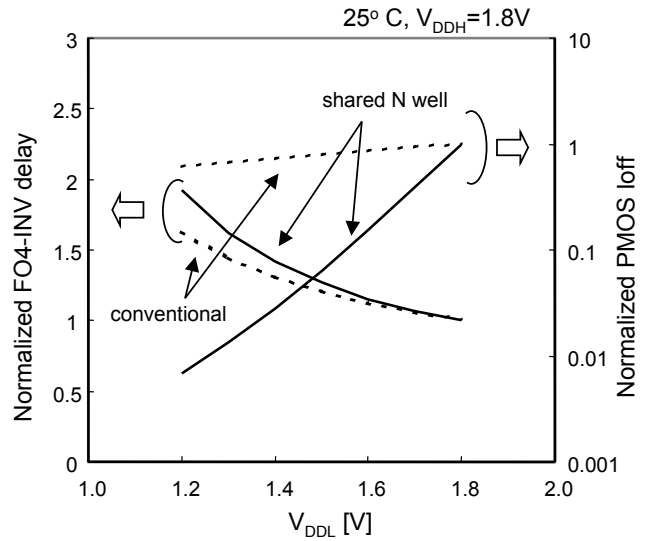
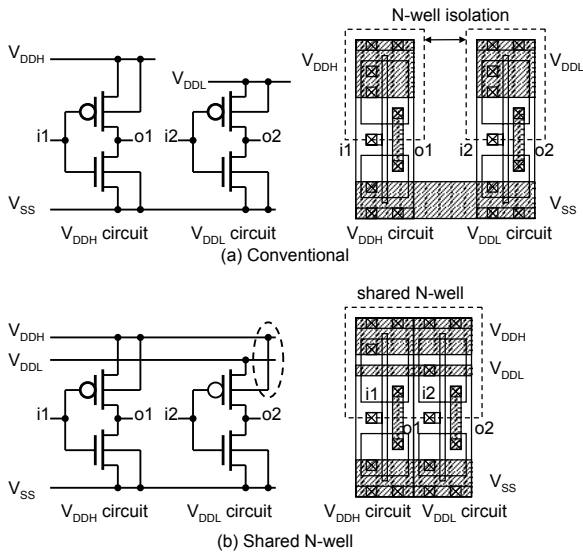


Figure 6.2.1: Dual-supply circuit schematic options and layout.

Figure 6.2.2: FO4-INV delay and PMOS leakage current.

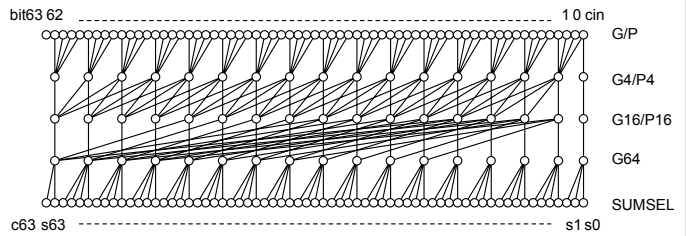
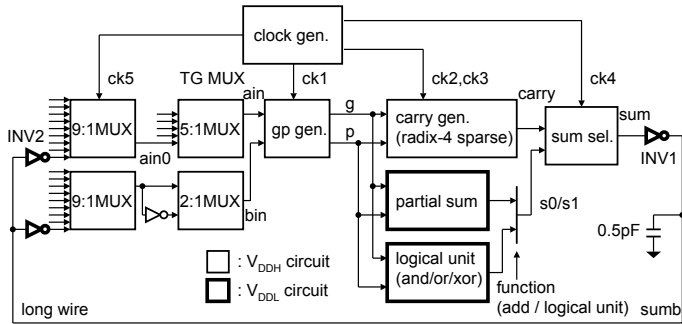


Figure 6.2.3: Block diagram of a 64-bit ALU.

Figure 6.2.4: Sparse radix-4 carry tree.

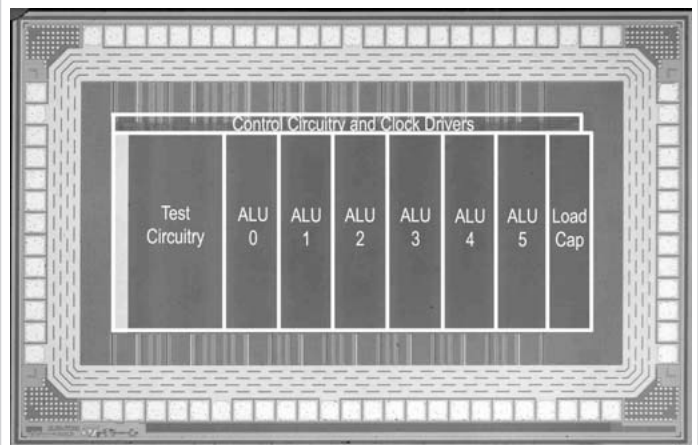
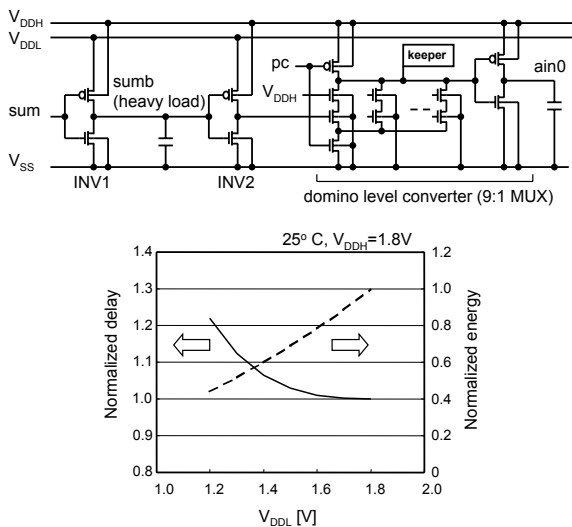


Figure 6.2.5: Output buffer and domino level converter.

Figure 6.2.6: Chip micrograph.

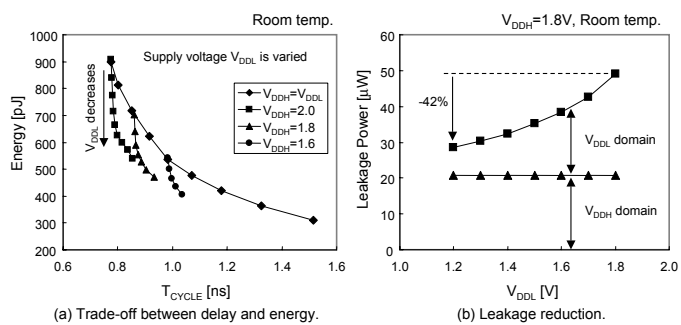


Figure 6.2.7: Measured results.

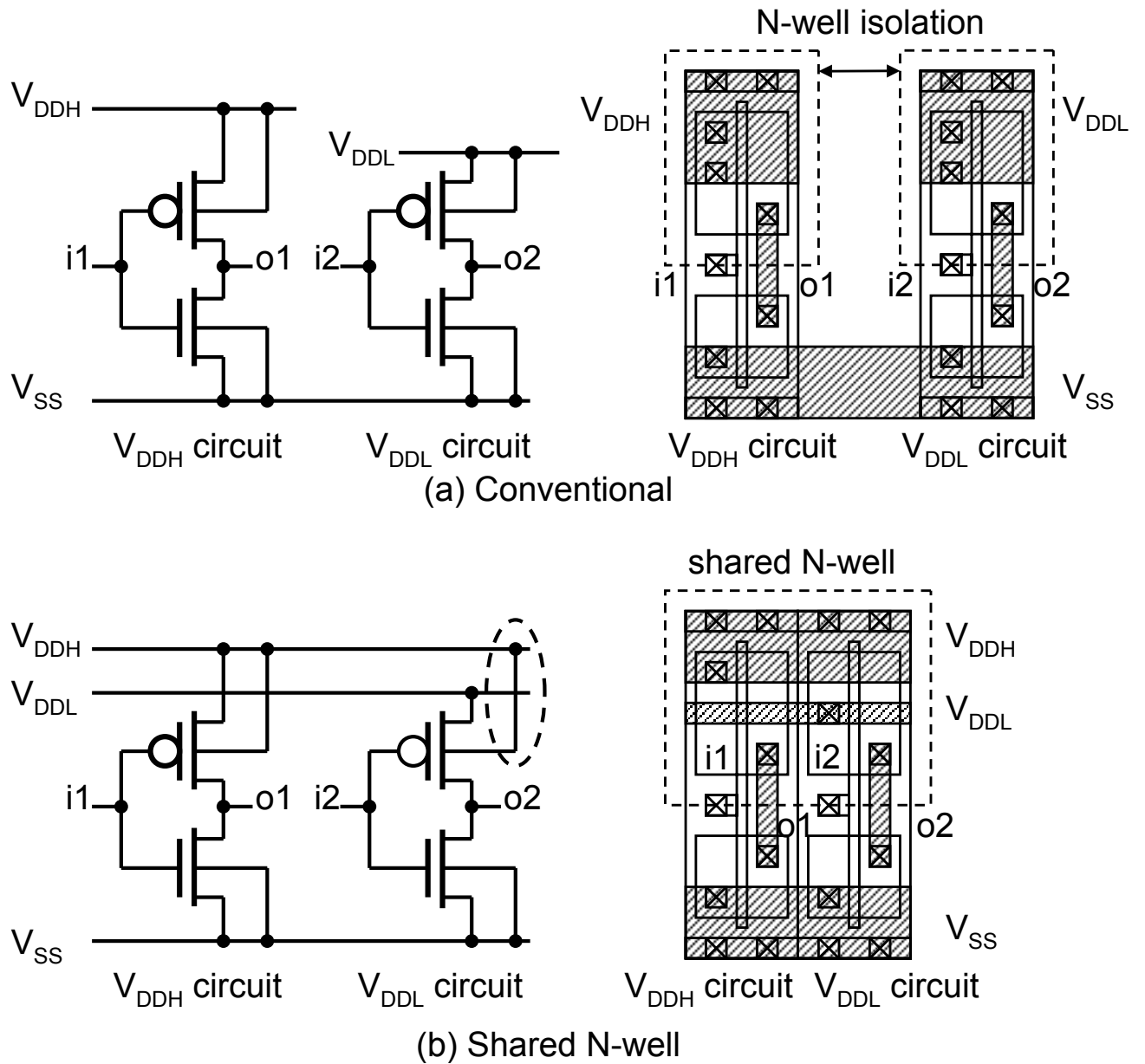


Figure 6.2.1: Dual-supply circuit schematic options and layout.

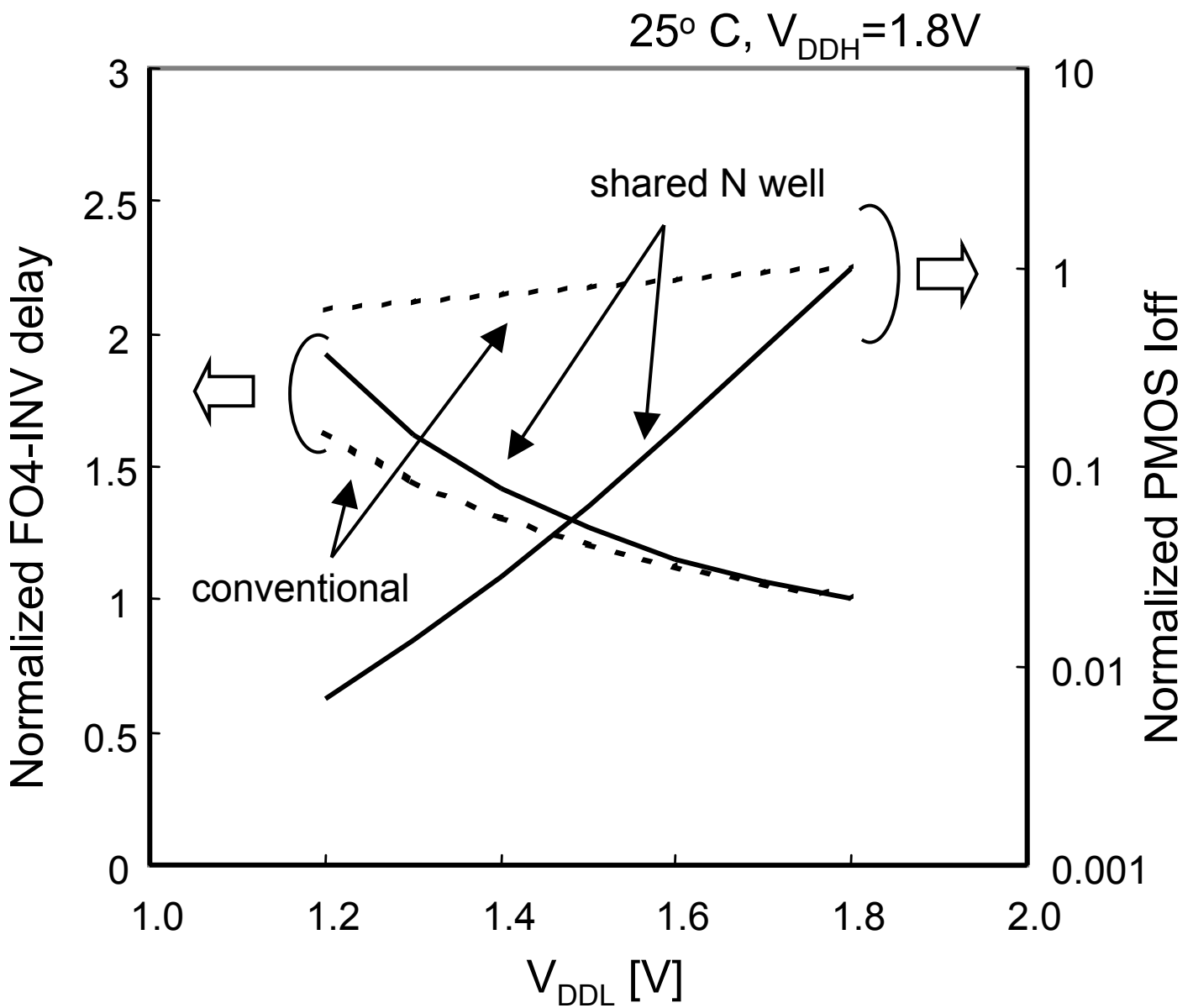


Figure 6.2.2: FO4-INV delay and PMOS leakage current.

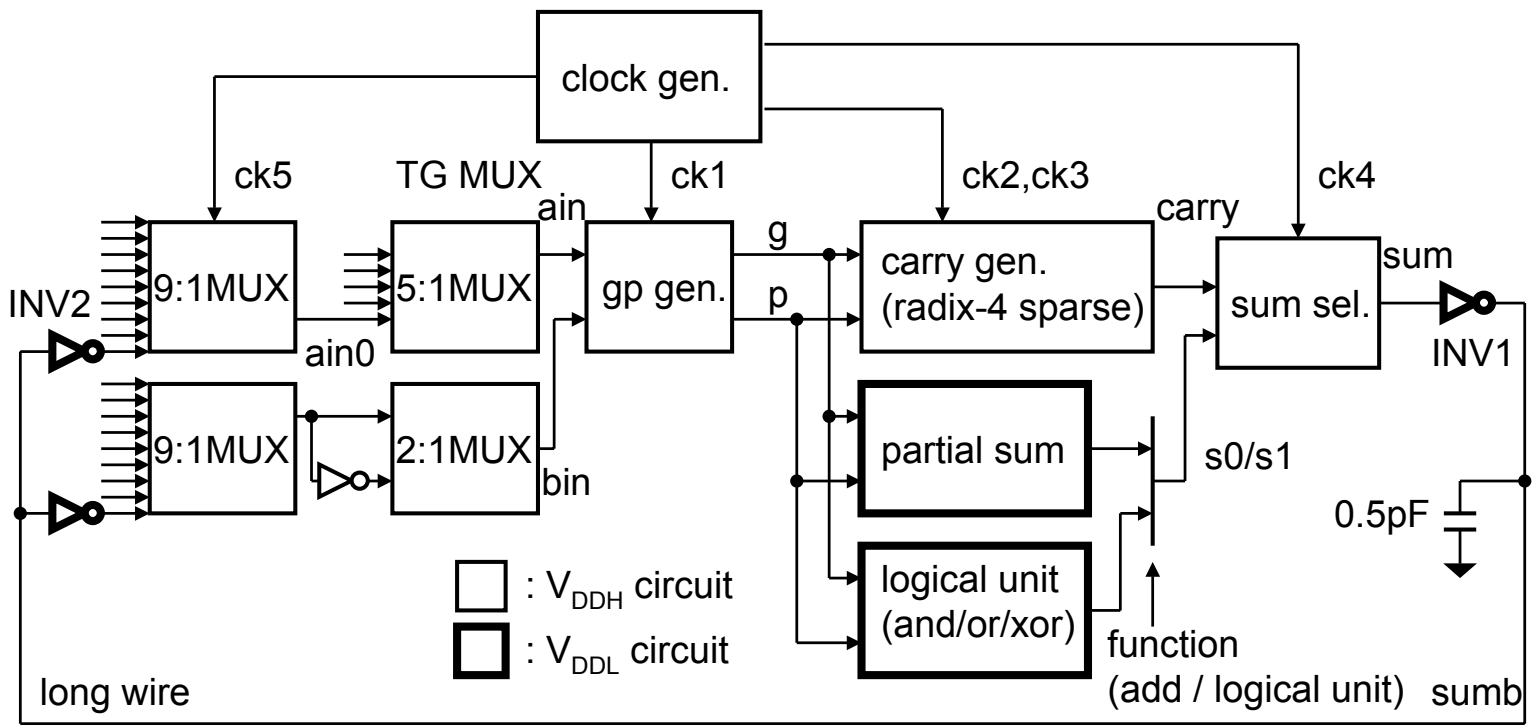
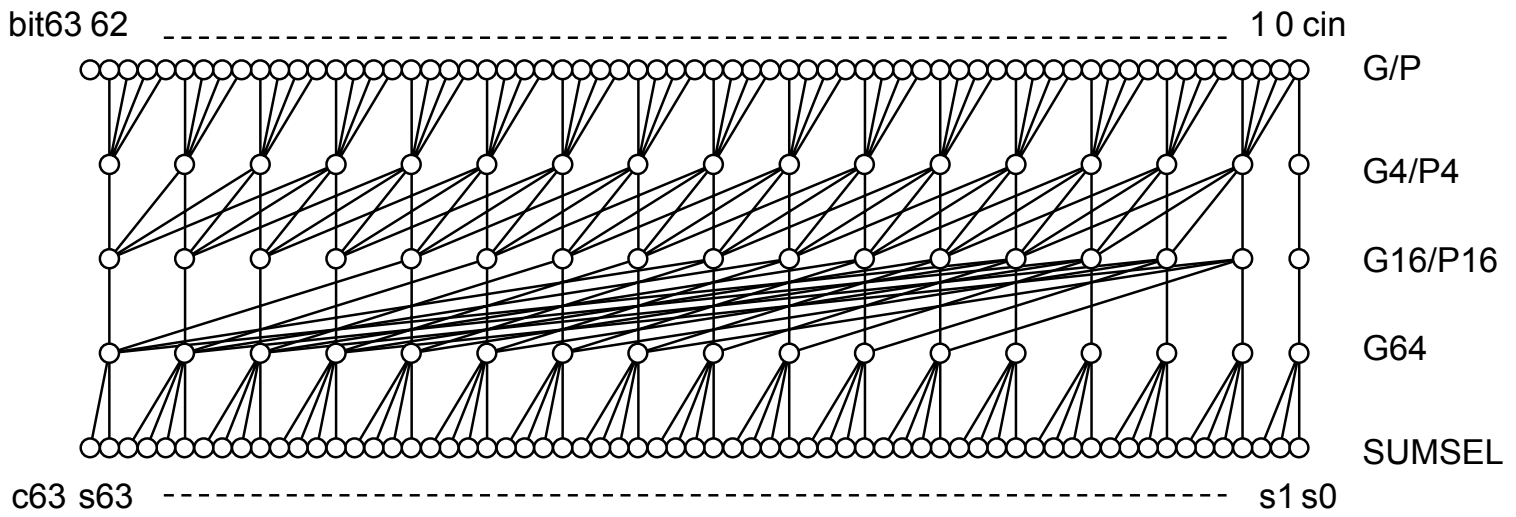


Figure 6.2.3: Block diagram of a 64-bit ALU.



**Figure 6.2.4: Sparse radix-4 carry tree.**

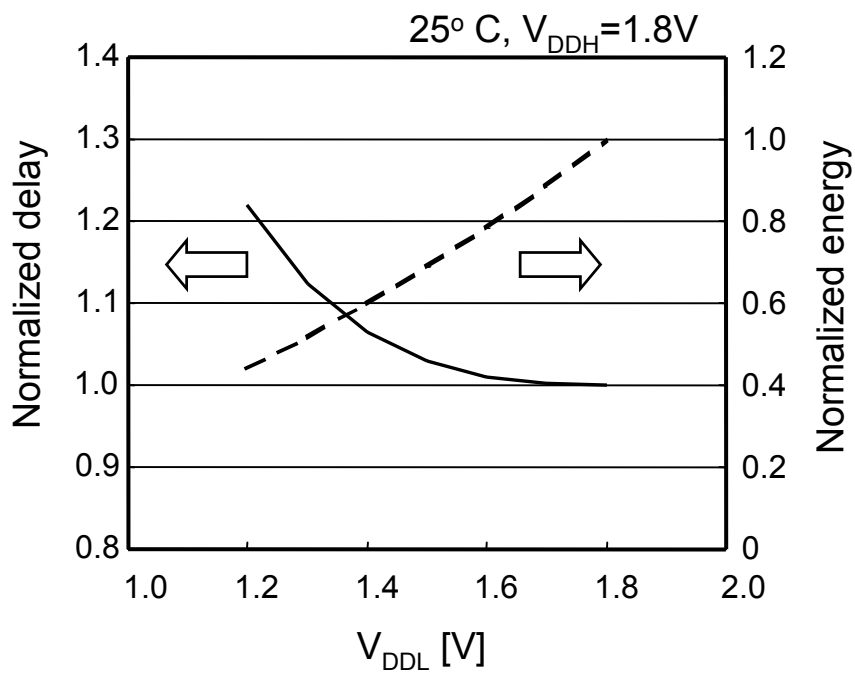
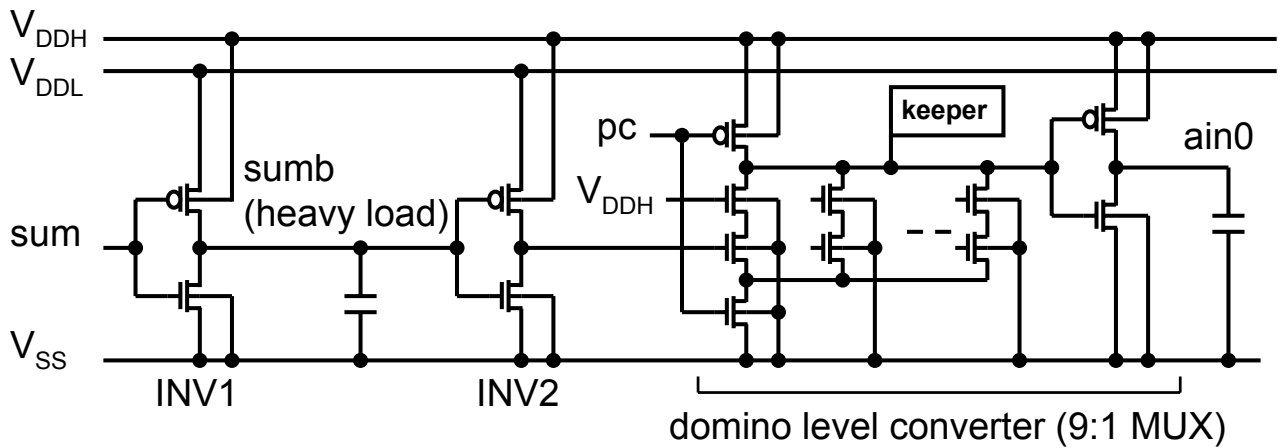


Figure 6.2.5: Output buffer and domino level converter.



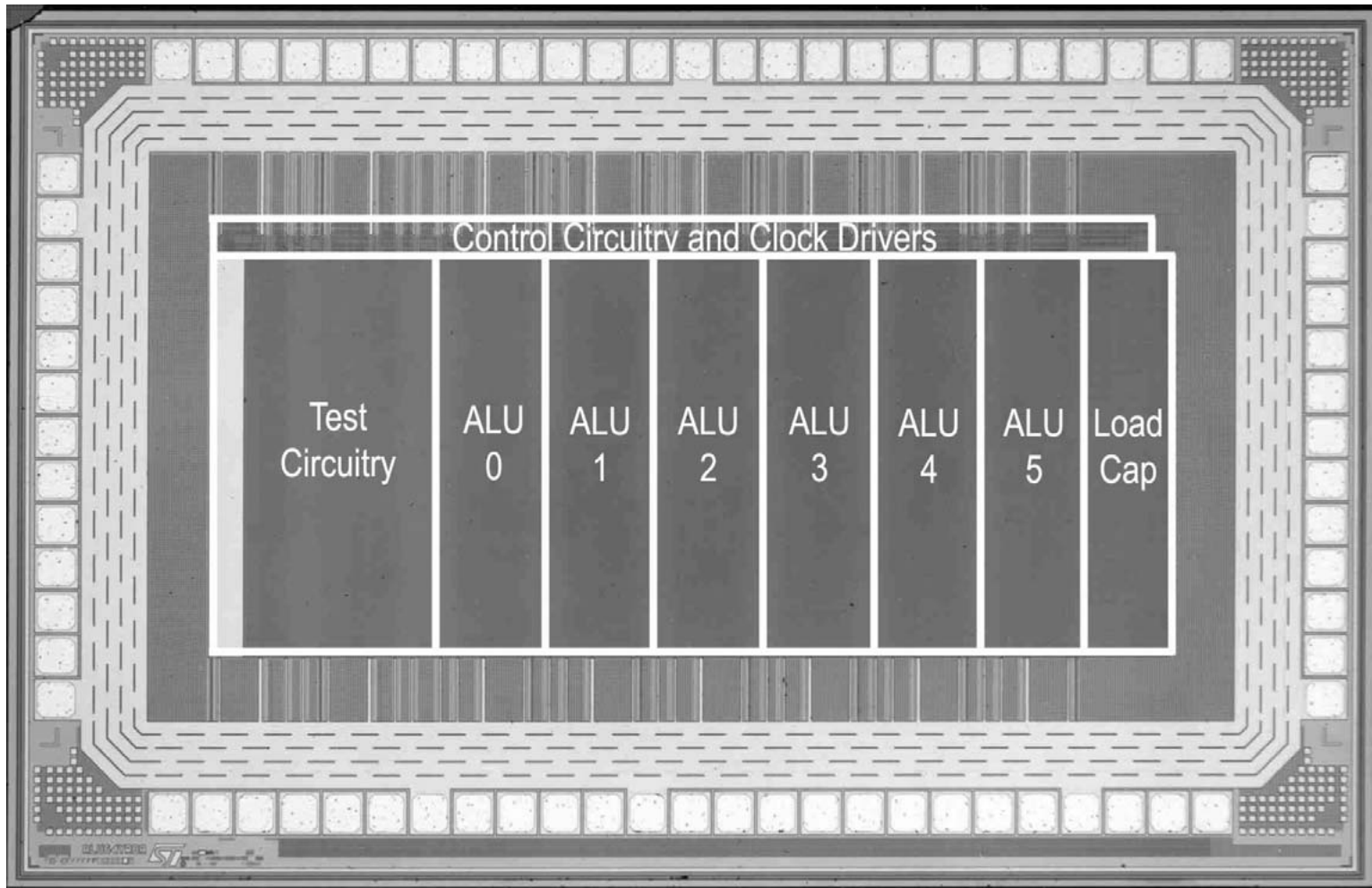
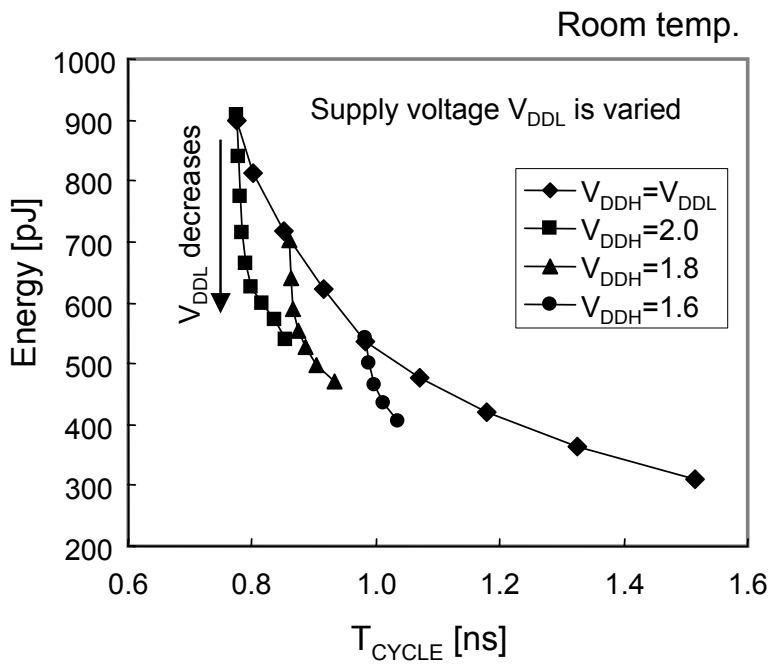
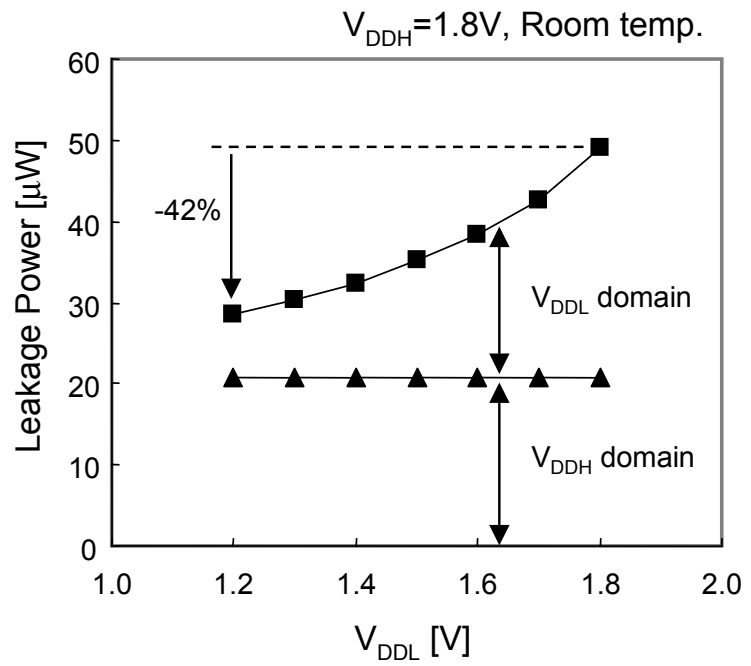


Figure 6.2.6: Chip micrograph.



(a) Trade-off between delay and energy.



(b) Leakage reduction.

**Figure 6.2.7: Measured results.**