

WP 3.6: A Bipolar Population Counter Using Wave Pipelining to Achieve 2.5x Normal Clock Frequency

Derek Wong, Giovanni De Micheli, Michael Flynn, Robert Huston*

Computer Systems Lab, Stanford Univ., Stanford/Trillium, Inc., San Jose, CA

Wave pipelining can boost the pipeline rate of a system without using additional registers or latches. In wave pipelining, multiple coherent waves of data are placed between storage elements by clocking the circuit faster than the propagation delay of the combinational logic (Figure 1). If all the propagation paths from the combinational circuit inputs to outputs have approximately the same delay, each wave propagates uniformly to the outputs without interfering with adjacent waves.

This bipolar LSI chip achieves 2.5 times the normal clock frequency without the use of additional storage elements. Related work is reported in References 1, and 3-7. Additional references are found in Reference 7.

The minimum clock period for a wave-pipelined circuit is bounded by the constraint: $t_{cp} > \Delta t_p + 2\Delta C + t_{sh} + t_{rf}$ [6,7], where: t_{cp} = clock period, Δt_p = maximum difference between longest and shortest path delays over worst-case design, process, and environment, ΔC = worst-case uncontrolled clock skew, t_{sh} = set-up plus hold time for edge-triggered registers, (for latches, t_{sh} = length of transparent period plus hold time), t_{rf} = worst-case rise or fall time (10% to 90% voltage swing) at the last logic stage.

By reducing Δt_p to a small fraction of the longest path delay, the clock period using wave pipelining can be made much smaller than that using the ordinary pipelining period. To make all the paths have approximately the same delay, special CAD algorithms and tools have been developed. These tools convert an ordinary design containing imbalanced delays using algorithms described elsewhere [5, 6, 7]. Fine tuning changes gate delays by adjusting gate drives, and rough tuning inserts a minimal number of delay elements along short paths that cannot be well-balanced by fine tuning alone. The methods are designed to lengthen short paths to approximately equal the length of the critical path(s). The critical path(s) are never lengthened using these methods.

The wave pipelining concept has been tested using a demonstration chip. The logic circuit performs a population counter function: the circuit takes 63 parallel inputs and produces the number of ones in that vector as a binary number. As shown in Figure 2, the design is split into two major sections. The first is a carry-save adder tree that takes 63 input lines and converts them into two 6b numbers. The adder tree is implemented using 3-2 counters. The second section is a 6b carry-lookahead adder.

The circuit is combinational logic with 21 gate levels and a nominal longest path delay of 8.5ns for the core logic plus 1ns for the output pin drivers. The path length difference after the design process is about 1.1ns, excluding the effects of differences in rising vs. falling delays and data-dependent delays. The total delay variation Δt_p includes these effects plus process and temperature variation within the chip. Rough tuning resulted in 6% added cell area.

The circuit is fabricated in a commercial BiCMOS process [2]. The minimum-sized npn transistors have the following parameters: $f_T = 13\text{GHz}$, $A_{EM} = 1 \times 2\mu\text{m}^2$, $C_{BE} = 6\text{fF}$, $C_{BC} = 6\text{fF}$, and $C_{CS} = 35\text{fF}$.

The circuit is implemented in single-level CML using a standard-cell technique. All the logic cells are OR/NOR gates using a single level of current switches. The gates are single-ended rather than differential and use a voltage swing of 500mV. The complete chip has 800 logic gates plus input and output buffers and voltage reference generators. The supply voltage is 5V. The nominal current consumption of the logic circuit, excluding I/Os and voltage generators, is 207mA.

Each resistance in each CML gate is implemented using a metal-programmable group of four resistors of value 2.5k, 1.5k, 0.5k and 0.5k Ω (Figure 3). Any combination of the four resistors can be wired in series to produce a resistor of any value between 0.5k and 5k in 0.5k Ω increments. This allows each gate current to be tuned without affecting the placement and routing of the overall circuit.

The number of input pads is reduced from 63 to 16 by wiring a few logic inputs to each pin, simplifying packaging and testing the prototype chip, while still allowing 2^{16} input patterns to be applied. A micrograph of the chip is shown in Figure 4. The core logic is 2.5x3.8mm², and the chip is about 4x6mm² total. Seventy-two chips were packaged from two wafers from the same fabrication process run. Twenty-six chips passed a 20000-vector functional test at 40MHz.

To test wave pipelining, a 40000-vector sequence is applied at vector rate up to 320MHz. The maximum wave pipelining frequencies for the 26 chips are as follows: 1 chip at 235MHz, 4 chips at 242MHz, 19 chips at 250MHz, and 2 chips at 258MHz. This is 2.4 to 2.65 times faster than the normal pipelining frequency of 97MHz. Figure 5 is an oscilloscope trace showing one input and one output pin during the test applied at 250MHz. Since the propagation delay is about 9.0ns for this part, more than two waves of data are stored within the combinational logic.

Tests show that wave pipelining works at the maximum rated frequency while the supply voltage is varied from 4 to 6V. High temperatures increase the maximum propagation delay by about 0.5ns and increase the worst-case difference between longest and shortest path delays by 0.25ns. Theoretically, this should cause the minimum clock period to increase by about 0.25ns. Figure 5 is a conceptual cell layout.

Compared to an implementation using ordinary pipelining, a wave-pipelined circuit reduces the latency, area, and clock distribution required by pipeline registers or latches. Circuits having relatively few short paths, such as multiplier trees, require few padding elements during rough tuning while other structures, such as adders, require more.

Acknowledgements

This work was supported by the Center for Integrated Systems at Stanford, NSF Contract No. MIP88-22961, and NASA contract NAGW 419. The authors appreciate chip fabrication and design help by Philips/Sigmetics, chip testing by LTX/Trillium, and CAD tools from Mentor Graphics. G. De Micheli was supported by NSF, DEC, and AT&T under a PYI award. G. Bewick and many at Philips/Sigmetics and LTX/Trillium were helpful during chip design, simulation, fabrication, and testing.

References

- [1] Cotten, L., "Maximum Rate Pipelined Systems," AFIPS Spring Joint Computer Conference, pp. 581-586, 1969.
- [2] de Jong, J. L., et al, "Single-Polysilicon Layer Advanced Super High-speed Bi-CMOS Technology," IEEE Bipolar Circuits and Technology Meeting, Minneapolis, MN, pp. 182-185, Sept. 1989.
- [3] Gray, C. T., et al, "Theoretical and Practical Issues in CMOS Wave Pipelining," VLSI '91, Edinburgh, Aug. 1991.
- [4] Klass, F., and J. M. Mulder, "CMOS Implementation of Wave Pipelining," Technical Report 1-68340-44(1990)02, Delft University of Technology, Delft, Dec. 1990.
- [5] Wong, D., et al, "Inserting Active Delay Elements to Achieve Wave Pipelining," ICCAD '89, Santa Clara, pp. 270-273, Nov. 1989.
- [6] Wong, D., et al, "Designing High-Performance Digital Circuits Using Wave Pipelining," VLSI '89, Munich, pp. 241-252, Aug. 1989.
- [7] Wong, D., "Techniques for Designing High-Performance Digital Circuits Using Wave Pipelining," Ph.D. Dissertation and Computer Systems Laboratory Technical Report, Stanford University, Stanford, August 1991.

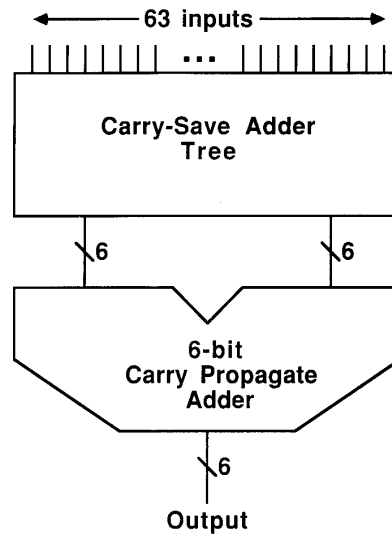


Figure 2: 63b population counter.

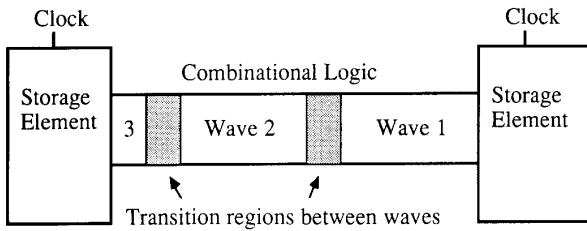


Figure 1: In wave pipelining, multiple coherent waves pass through a combinational logic pipeline.

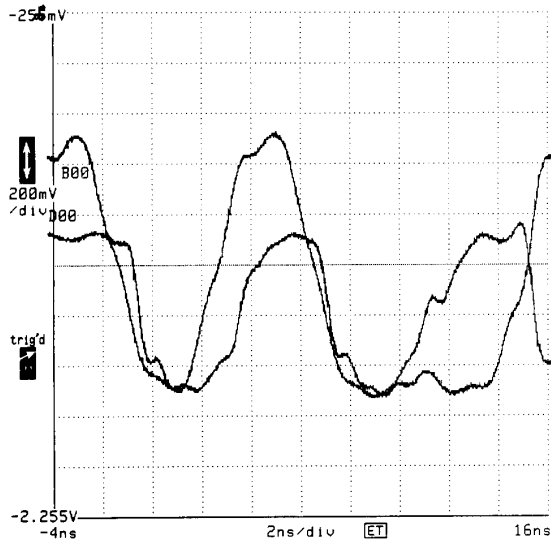


Figure 4: Oscilloscope trace of wave pipelining at 250MHz. B00 is input, D00 is output.

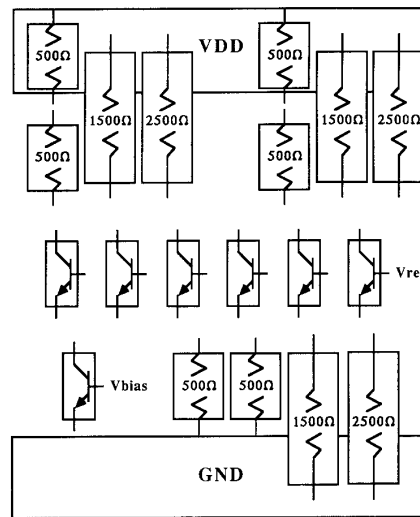


Figure 5: Conceptual cell layout.