

Energy-Recovery CMOS for Highly Pipelined DSP Designs

W.C. Athas, W-C Liu, L.“J.” Svensson

USC/Information Sciences Institute, 4676 Admiralty Way, Marina del Rey, CA, 90292, USA

{athas,weichuan,svensson}@isi.edu

Abstract

We compare the frequency-versus-power dissipation performance of two energy-recovery CMOS implementations to that of a conventional, supply-voltage-scaled design. The application is a small but complete DSP function. All three designs are based on the same high-level organization and conform to the same I/O specification. SPICE simulations indicate that an energy-recovery design which requires only a small degree of modification to the conventional design offers more than a two-fold reduction in power across a wide range of operating frequencies.

1. Introduction

The pursuit of low-power applications that exploit the theory of energy recovery [1] has focussed upon two means for viably delivering and recovering CMOS circuit energies: stepwise charging and resonant charging. The key advantages of stepwise charging over resonant charging have been modularity, low-cost integration, and asynchronous operation. A practical application of stepwise charging is described elsewhere in these proceedings [2].

Conversely, the potential advantage of resonant charging is succinctly demonstrated in the graph of Figure 1. The provably best way to charge and discharge a capacitance C through a resistance R is via constant-current charging. Figure 1 compares this idealized best case to that of an idealized linear voltage ramp applied to the input side of the resistance. In both cases, the delay is expressed as the time it takes the voltage level at the load C to reach 95% of its swing. This is the well-known $3RC$ metric when charging from a dc voltage source. Constant-current charging can accomplish the same output swing more rapidly and with less dissipation than constant-voltage charging. The caveat is that the input voltage must exceed that of the maximum output voltage swing.

Also plotted in this graph are *laboratory* measurements from a stepwise-charged pad-driver implementation ($T \approx 10$ ns, $RC \approx 3$ ns) and a resonant clock driver, called the blip

circuit ($RC = 10$ ns) [3]. The latter is an all-resonant circuit topology which generates the clock timing signals depicted in Figure 2c. Unlike the sinusoidal waveforms (Figure 2a) typical of the more well-known resonant topologies, such as the Colpitts and Hartley oscillators, the two-phase outputs of the blip circuit have only small regions of overlap. The waveforms bear important similarities to the well-known two-phase clocking paradigm for digital systems (Figure 2b).

Intriguingly, the practical implementation of the blip circuit outperforms the idealized linear ramp case for the faster switching times. Also, like constant-current charging, it can drive the output in less than $3RC$ if one is willing to allow the input voltage swing to exceed the maximum allowed swing on the output.

The design challenge is how to exploit the efficient charging and discharging of capacitive nodes afforded by circuits such as the blip circuit in real CMOS systems. In this paper, we investigate exploiting the energy-recovery capability of the blip circuit for the specialized problem domain of highly-pipelined, regular designs. There are several reasons for selecting this problem domain: the design space is relatively small; there is a clear computational performance trade-off between latency and throughput; and the resulting designs can be clocked at very high frequencies using standard supply voltages, or, alternatively, operated at low to very low power levels for reduced frequencies and voltages.

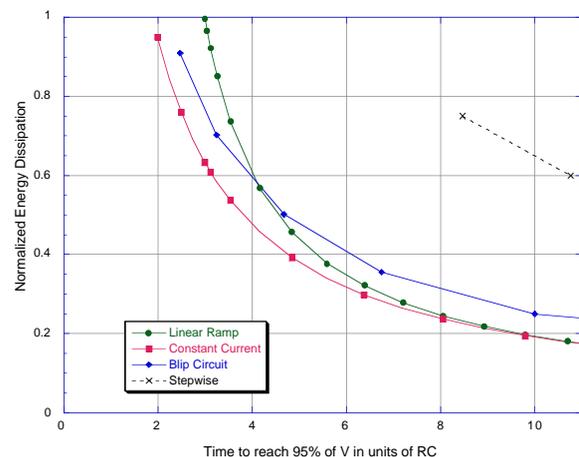


Figure 1. Energy recovery versus transition time: idealized cases and actual measurements.

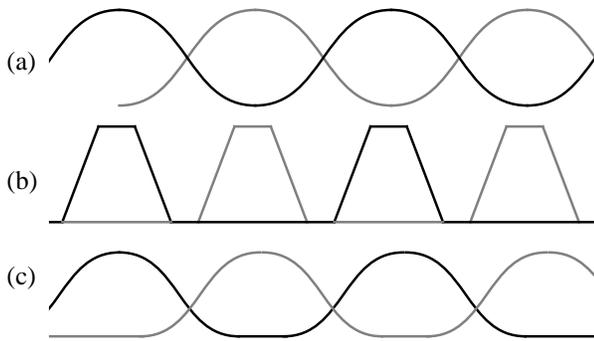


Figure 2. (a) complementary sine waves; (b) two-phase non-overlapping; two-phase, *almost* non-overlapping.

To make a direct comparison, we implemented a small bit-level-pipelined Finite Impulse Response (FIR) filter in three different ways. The control case was a conventional CMOS design based on two-phase, non-overlapping clock signal with fully-restoring logic and dynamic pipeline latches. Starting from this design, we then employed a modestly intrusive, energy-recovery method to recover the clock power: the dynamic latches were redesigned and a blip circuit was employed for the clock driver. The third design was the most aggressive in attempting energy recovery. It derived all of its operating power from its two clock lines using bootstrapped circuit techniques [4]. All three designs were carried out to the level of layout using the tool Magic and the MOSIS scalable CMOS design rules. They were simulated in HSPICE[®] from extracted layout. The CMOS technology was a 2 μm n-well Orbit process offered through MOSIS

2. The FIR Filter Experiment

The FIR filter is a simple digital-signal-processing function from which we expect our findings will reasonably generalize to larger examples. It has been extensively studied in the literature in a variety of implementations, including bit-level pipelining for low power [5,6]. We chose six bits and three taps so that prototypes could be inexpensively fabricated as small, 4 mm^2 test chips and so that we could reasonably simulate the entire filter in SPICE. The filter taps were 1/8, 3/4, and 1/8. All terms added to the output were shifted, sign-extended versions of the input, i.e., the middle 3/4 tap was implemented as 1/2 plus 1/4 [7].

The entire design required only two basic cells: a carry-save adder and a delay element. Carry-save adders were used to avoid a six-cycle delay in each adder. A five-cycle vector merge adder, built from the same full-adders and full-cycle delays as the rest of the chip, performed the final addition. The carry-out bits of the most-significant adder cells were generated (to not generate them would require another version of the cell), but would always be zero and were left unconnected.

The conventional, “control” design was based on fully restored, ratioed complementary gates for the logic starting from minimum-size geometry for the nFETs. The dynamic latches used for pipelining consisted of a transmission gate (unratioed) followed by a ratioed inverter. This arrangement required four separate clock signals to be distributed across the chip: the two phases and their complements.

The resonantly-clocked design was identical to the first, except that the transmission gate of the pipeline latch was replaced by an nFET pass gate and the clock swing was increased above V_{dd} to overcome the threshold drop due to the nFET. These modifications were made so that the resulting latches would be compatible with the blip circuit. The slight overlap in the two phases did not present a problem because the amplitude of the overlap was serendipitously always below the threshold voltage of the latch’s pass gate.

The third design was a fully clock-powered implementation. This version did not require a V_{dd} supply-voltage rail. It was an all nMOS design based on dual-rail logic and bootstrapped drivers. A block diagram for the dual-rail full adder is shown in Figure 3. The full adder consists of two And gates, two Exor gates, one half-cycle Delay, and a special Carry-Generate gate.

All logic gates were implemented using bootstrapped logic [4]. The schematic for the Exor gate is shown in Figure 4. The gate generates both the result and its complement. Two cross-coupled devices at the output of the dual-rail gate serve to clamp the undriven output to zero. The input is latched on the falling edge of the first (input) phase and the output is driven on the second (output) phase. All inputs must be driven on the input phase of the gate. Except for the And gate, cross-coupled output devices were used for all of the gates in the design. The And gate was implemented differently because it was used only to produce the g and k inputs for the Carry-Generate gate; those were not needed in both polarities. The And-gate output was not clamped when

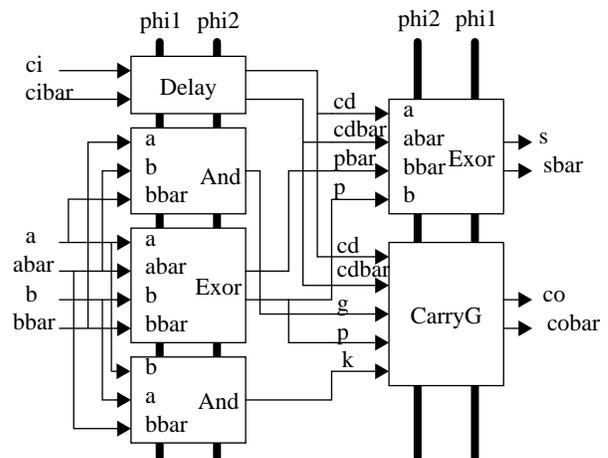


Figure 3. Block diagram of full adder for the bootstrapped, clock-powered version.

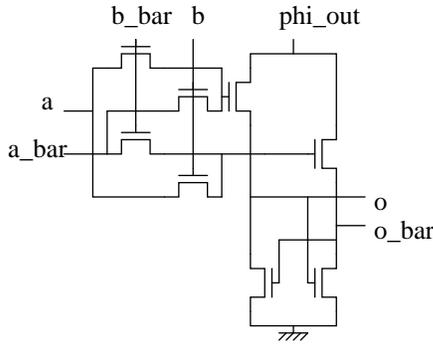


Figure 4. The bootstrapped Exor gate of the clock-powered version.

not driven, but a device controlled by the input phase ensured that the output did not accumulate stray charge when not driven for several cycles.

3. Measurements and Analysis

To determine the simulated frequency-versus-power dissipation performance, a digitized triangle wave was input to the filter. This choice of input waveform resulted in signal transitions in all of the filter's cells which served to reduce the effects of dual-rail versus signal-rail signal encoding and also to promote the relative importance of data versus clock dissipation.

Figure 5 contains a graph of the simulated frequency versus power performance of the three designs between 10 and 125 MHz, which was the top speed of the conventional filter at 5 V. The conventional design is plotted under two different sets of operating conditions. The first is when V_{dd} is held constant at 5 V and only the clock frequency is reduced. The second is when V_{dd} is decreased commensurately with the increase in clock period afforded by the lower operating frequency, i.e., supply-voltage scaling.

The clock-powered, bootstrapped version was simulated with non-overlapping clocks that had the shape of half sinusoids. All power returned by the clock source was assumed to have been recycled. The blip circuit was not used in the simulation, though this design had been fabricated and tested in the lab with the blip circuit. The top speed was 45 MHz with 4 V clock swings, compared to 100 MHz for the fully conventional design at $V_{dd} = 4$ V. The clock-powered design outperformed the frequency-scaled-only conventional version. However, reducing the supply voltage while decreasing the clock frequency (i.e., supply-voltage scaling) resulted in significantly lower dissipation for the conventional circuit except for the top two frequency simulations of the clock-powered design.

For the resonantly-clocked design, the overhead of the blip circuit was included in the simulations. The two power transistors M1 and M2 of the blip circuit, shown in Figure 6, were simulated from the same 2 μm process data as the filter.

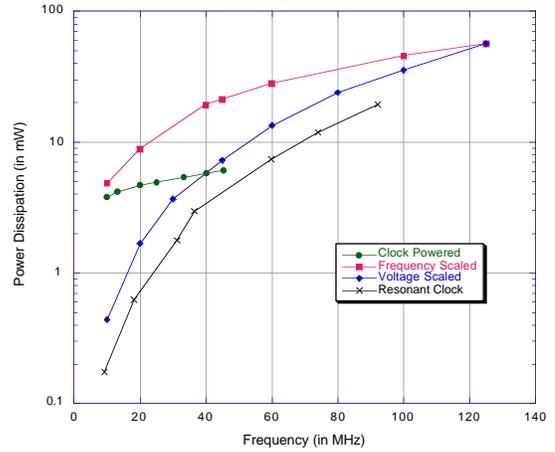


Figure 5. Clock frequency versus power dissipation for the four test cases.

The drawn channel width for each of these transistors was 1.2 mm. They did not contribute significantly to capacitive dissipation because they were themselves resonantly driven. Clock-line resistance was simulated at 3 Ω .

The clock frequency was controlled by selecting inductances for L1 and L2 of Figure 6. For a given frequency, the supply voltage and clock swing voltages were then minimized. The top frequency of this design was 92 MHz ($L1 = L2 = 360$ nH). As shown in Figure 5, this version exhibited significantly lower dissipation than the voltage-scaled conventional version. Furthermore, the relative advantage improved as the clock frequency was reduced. This is to be expected, since the Q of the resonant circuit increases as the frequency is reduced and the inductance is increased. Offsetting this improvement is that the power transistors M1 and M2 exhibit higher on-resistance as the clock swing is also decreased.

The contribution of clock power to total power for the conventional version ranged from 60–70%. In contrast, for the resonantly-clocked version, clock power was only 8–18% of total power. Input patterns that would generate less internal switching activity would further increase the relative benefit to the resonant clocking in the presence of constant switching activity for the clock lines.

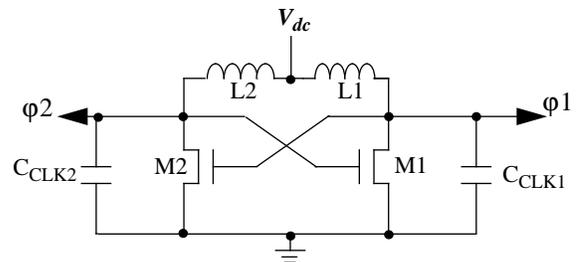


Figure 6. The resonant clock driver or "blip circuit."

4. Conclusions

In this experiment we have strived to equitably evaluate the frequency-versus-power performance of energy-recovery CMOS at the level of a small but complete CMOS system. We approached the design problem by starting with a resonant clock driver circuit which we have validated in laboratory testing for its energy efficiency and frequency range. Furthermore, the circuit generates waveforms that are akin to the two-phase clocking scheme that is well-known in digital design. The design problem that we chose was a bit-level pipelined FIR filter. This design required only two types of basic cells, which simplified the design effort considerably so that we were able to undertake a three-way comparison.

The technique of maximally pipelining the design, i.e., down to the bit level, offered more than an order-of-magnitude range in power versus frequency for the supply-voltage-scaled conventional design. The small fan-out and fan-in of each gate kept the capacitive loading small at the individual gate level.

The bootstrapped, clock-powered design attempted to maximize energy recovery by *tapping* the clock lines to power the data bits. The power dissipation due to the area overhead and higher voltage swing requirement of the clock-powered circuitry at the fine granularity of individual gates overwhelmed the power dissipation reduction which resulted from recovering the circuit energies. The sub-linearity of the frequency-versus-power curve suggests that most of the dissipation is due to causes which do not benefit from slower charging. The non-recoverable energy of the charge isolated at the gate nodes of the bootstrapped nFETs may be crucial to frequency-versus-power performance, as has been the case in other bootstrapped circuit experiments [8]. A second investigation of the bootstrapped circuit approach based on minimizing this effect is the subject of further research.

The resonantly-clocked design attempted to balance energy recovery with conventional techniques. A drawback to the bit-level pipelining approach is that the extensive application of pipeline latches incurs a very high capacitive loading of the clock lines. Clock power can quickly dominate the overall dissipation. Unless extraordinary measures are taken, the clocks will dissipate constant power regardless of the switching activity of the filter's data inputs. The resonant clock driver mitigates the clock power dissipation problem by recovering significant amounts of the capacitive clock line energy.

Starting with two-phase clocking and conventional CMOS logic, the only required change for compatibility in applying the blip circuit as a resonant clock driver is in the pipeline latches. The design change is to remove the pFETs whose gates normally connect to the complements of the clock phases. It is then desirable to compensate for the resulting threshold drop by increasing the clock swing voltage above V_{dd} .

From these small modifications, energy recovery is then feasible. There is some loss of maximum speed, e.g., 92 MHz versus 125 MHz in the FIR filter experiment. However, the energy-efficiency advantage improves as voltage and frequency are reduced because of the increased circuit Q for the blip circuit at lower frequencies. For applications where the chip must be run in a continuous mode, as is typical in many embedded DSP applications, this feature may be especially attractive.

Acknowledgments

The research described in this paper was supported by ARPA contracts DABT63-92-C0052 and DAAL01-95-K3528.

References

- [1] W. Athas, "Energy-Recovery CMOS," in J. Rabaey, M. Pedram (Eds) *Low-Power Design Methodologies*, Kluwer Academic Press, 1996.
- [2] J. Svensson, W. Athas, R. Wen, "A sub-CV2 pad driver with 10 ns transition time," *Proceedings of the International Symposium on Low-Power Electronics and Design*, Aug. 1996.
- [3] W.C. Athas, J. Svensson, N. Tzartzanis, "A Resonant Signal Driver For Two-phase, Almost-non-overlapping Clocks," *International Symposium on Circuits and Systems*, May 1996.
- [4] C. L. Seitz, A.H. Frey, S. Mattisson, S.D. Rabin, D.A. Speck, J.L.A. van de Snepscheut. Hot-clock NMOS. In *Proc. of the 1985 Chapel Hill Conference on VLSI*, pp. 1-17, 1985.
- [5] P.J. Duncan, S. Samy, R. Jain, "Low-power DSP using bit-level pipelined maximally-parallel architectures," *Symposium on Integrated Systems, Proceedings of the 1993 Symposium*, MIT Press, 1993.
- [6] C. Nagendra, R.M. Owens, M.J. Irwin, "Low power considerations in the design of pipelined FIR filters," *1995 IEEE Symposium on Low-Power Electronics*, Oct. 1995.
- [7] P. Yang and R. Jain. An FIR filter generator. In Brodersen (Ed), *Anatomy of a silicon compiler*, Chapter 18. Kluwer, 1992.
- [8] N. Tzartzanis, W. Athas, "Design and Analysis of a Low Power Energy-Recovery Adder," *Fifth Great Lakes Symposium on VLSI*, Mar. 1995.