Design and Demonstration of Micro-Electro-Mechanical Relay Multipliers

Hossein Fariborzi, Fred Chen
Vladimir Stojanović
Massachusetts Institute of Technology
Cambridge, MA
fariborz@mit.edu

Rhesa Nathanael, Jaeseok Jeon
Tsu-Jae King Liu
University of California
Berkeley, CA

***Abstract*— This paper describes the micro-architecture and circuit techniques for building multipliers with micro-electro-mechanical (MEM) relays. By optimizing the circuits and micro-architecture to suit relay device characteristics, the performance of the relay based multiplier is improved by a factor of ~8x over any known static CMOS-style implementation, and ~4x over CMOS pass-gate equivalent implementations. A 16-bit relay multiplier is shown to offer ~10x lower energy per operation at sub-10 MOPS throughputs when compared to an optimized CMOS multiplier at an equivalent 90 nm technology node. To demonstrate the viability of this technology, we experimentally demonstrate the operation of the primary multiplier building block: a full (7:3) compressor, built with 98 MEM-relays, which is the largest working MEM-relay circuit reported to date.**

I. Introduction

Due to their negligible leakage, MEM relays have recently been proposed as a solution to overcome the minimum energy limitations of CMOS circuits [1]. Although the mechanical movement makes relays slower than CMOS transistors, we have shown in adders and other basic circuit building blocks that appropriate circuit design strategies can enable relays to be more competitive at the circuit and system level [2, 3]. However, for MEM relays to be a practical replacement technology for CMOS, the performance gains must translate across the entire system. In this work we investigate the design of MEM relay based multiplier structures as they are commonly the most complex arithmetic blocks. We develop the multiplier micro-architecture and circuit techniques tailored to the relay device properties. These new micro-architectures are optimized around larger compressor circuits to minimize the mechanical delay. The larger pass-gate style compressor circuits are also optimized to provide single-mechanical delay operation and minimize the number of devices. The functionality of the optimized (7:3) compressor is experimentally demonstrated.

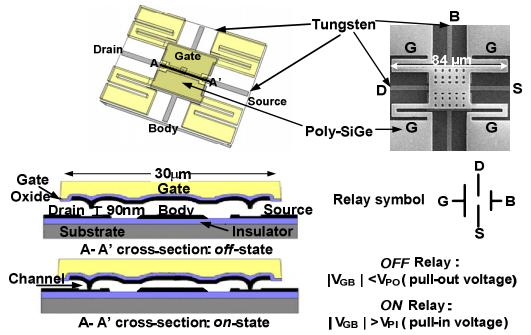

Figure 1. SEM, diagram, circuit symbol and operating states of the current MEM relay device

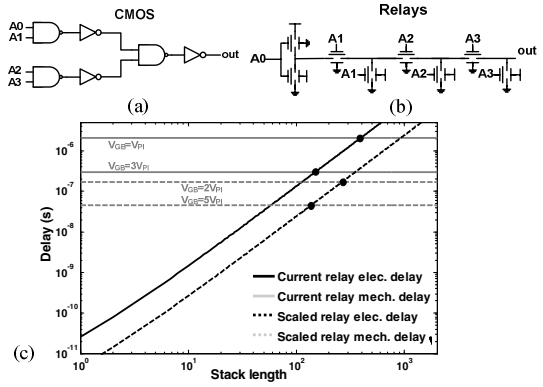

Figure 2. A 4-input AND built with CMOS (a) and MEM-relays (b). Electrical vs. mechanical delay for current (1 µm lithographic process) and scaled (90 nm lithographic process) relays (c)

II. MEM-Relay Operation and Circuit Design

The structure and operation of our 4-terminal MEM-relay [3] is shown in Fig. 1. When the gate-to-body ($|V_{GB}|$) voltage exceeds the *pull-in* voltage (V_{pi}) the channel connects the source and drain electrodes and allows current to flow. The current flow is disrupted by an air gap when $|V_{GB}|$ decreases below the *pull-out* voltage (V_{po}) and the relay de-actuates.

In optimized relay-based circuits all mechanical movement should happen simultaneously to minimize the impact of the relatively slow mechanical delay [2], thus favoring a tailored pass-gate design. Figure 2(a,b) illustrates the difference between CMOS and MEM-relay logic design styles. A straightforward substitution of CMOS transistors with relays in a standard CMOS 4-input AND logic circuit would result in 4 mechanical delays as each signal hitting a gate triggers an additional mechanical delay. The optimized relay design shown in Fig. 2(b) incurs only one mechanical delay since all mechanical movements happen at the same time. Thus, given a logic function, the design strategy is to stack as many MEM-relays in series as possible. The upper bound on the number of MEM-relay devices in a stack is reached when the electrical and mechanical delays are equal. Figure 2(c) compares the electrical delay of a device stack with the mechanical delay, for our 1 µm and 90 nm MEM-relays [3]. The mechanical delay is obtained for a reasonable range of V_{GB} overdrives and shows that this design approach is extendable to hundreds of series devices and consequently encompasses most practical logic functions.

III. MEM-Relay Multiplier Design

A. Design Overview

The main opportunity for innovation in relay multiplier design comes from the logic for the partial product matrix reduction. Figures 3 and 4 show 6-bit examples of the two

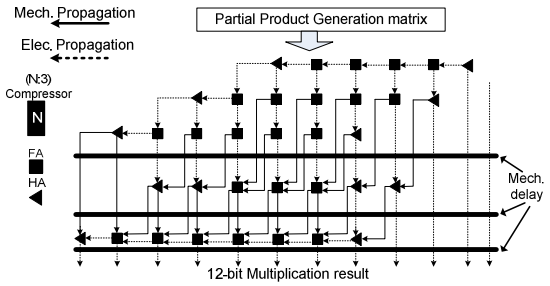

Figure 3. A 6-bit relay multiplier built with full and half adders

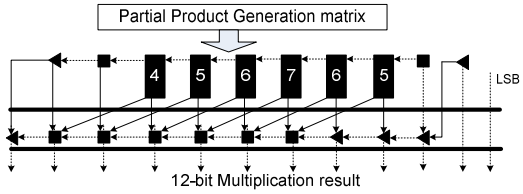

Figure 4. A 6-bit relay multiplier built with (7:3) and smaller compressors

primary micro-architectures explored in this work. Figure 3 shows a multiplier composed of (3:2) compressor (full-adders) and half adder circuits shown in Fig. 5. In both relay-based adder cells, there is one source/drain input whose output delay is only the electrical delay of the relay. The electrical propagation path allows for stacking (3:2) compressors and half adders without additional mechanical delays. Figure 4 shows the second micro-architecture which uses a higher compression ratio to decrease the total number of reduction steps. In each reduction step, various (N:3) compressors are stacked in such a way as to avoid paths with mechanical delays. The largest block is the (7:3) compressor with six gate inputs and one source/drain input. To illustrate the impact of using higher ratio compressors, Figs. 3 and 4 show the design of a simple 6-bit multiplier with both approaches, where the multiplier using (N:3) compressors reduces the number of mechanical delays from 4 to 3 (including one mechanical delay for partial product generation).

B. Partial Product Generation

As expected, using more complex compressors with larger radix introduces an area/delay tradeoff. A technique that can potentially benefit both area and delay is the modified radix-4 Booth encoding which reduces the total number of partial products by half [4]. The corresponding relay-based partial product generation circuit is shown in Fig. 6. Although it adds one mechanical delay to the partial product generation step, compared to a simple AND

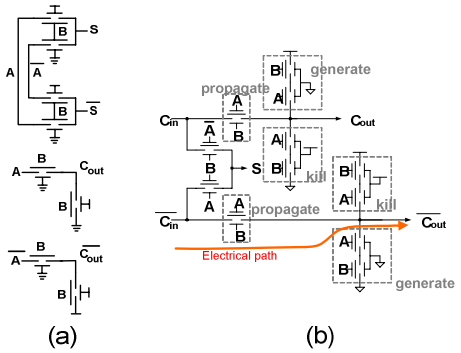

Figure 5. Implementation of (a) half adder and (b) full adder [2]

TABLE I. VARIOUS MULTIPLIER DESIGN TRADE-OFFS

Multiplier		8x8		16x16		32x32		64x64	
		No Booth	Booth	No Booth	Booth	No Booth	Booth	No Booth	Booth
Logarithmic (row by row)	Mech. Delay	4	5	5	-	6	-	7	-
	Appx Relay Count	1.3k	1.5k	4k	-	14k	-	33k	-
(3:2) compressor	Mech. Delay	4	5	5	6	6	7	7	8
	Appx Relay Count	1.1k	1.4k	3.5k	3.4k	12k	9k	29k	19k
(7:3) & smaller compressor	Mech. Delay	4	4	4	5	5	5	5	6
	Appx Relay Count	1.4k	1.5k	5.5k	4.4k	26k	15k	57k	30k

network, it proves to be promising in reducing the overall complexity and delay for larger multipliers. Table I summarizes the trade-offs for a broad range of multipliers, indicating the performance of larger multipliers benefits even more from higher compression ratios, optimized compressor circuits and Booth encoding.

IV. MEM-RELAY MULTIPLIER COMPONENTS DESIGN

The logic function of an (N:3) compressor can be described as:

$$\overline{Y_2 Y_1 Y_0} = \sum_{i=0}^{N} A_i \qquad (1)$$

The corresponding circuit is comprised of 3 sub-circuits for generating Y_0, Y_1 and Y_2. The most important design consideration for this relay logic circuit is to include a shoot-through electrical path from an input to the output for all of the sub-circuits, which allows for stacking of compressors without additional mechanical delay penalty. In each circuit, these electrical paths are provided through A_0–$\overline{A_0}$ source/drain connections to Y_j–$\overline{Y_j}$.

According to (1), the LSB (Y_0) is a 7-input XOR gate. This sub-circuit can be built by cascading 6 two-input relay XOR gates (Fig. 7(b)). An alternative to this implementation is shown in Fig. 7(c), where the body terminal is used in the design to cut the total number of relays to half. However, for experimental robustness, the former design has been implemented in this work.

The implementations of the Y_1 and Y_2 sub-circuits are illustrated in Figs. 8 and 9, respectively. In both cases, a (5:3) compressor is used for illustration of different design steps. Figure 8(a) shows the propagation path logic for Y_1, where A_0 is passed to the output when $\sum_{i=1}^{4} A_i = 1$ and $\overline{A_0}$ is passed when $\sum_{i=1}^{4} A_i = 3$. Figure 8(b) shows the integration of "generate" and "kill" paths, which happen for the cases of $\sum_{i=1}^{4} A_i = 2$ and $\sum_{i=1}^{4} A_i = 0$ or 4, respectively. Using a similar method, the Y_1 sub-circuit of a full (7:3) compressor can be built, as shown in Fig. 8(c).

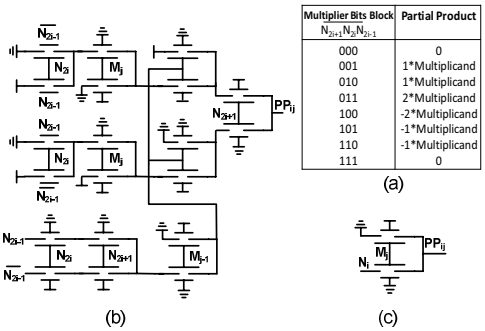

Figure. 6. (a) Booth encoding table, (b) Partial product generation circuit for Booth encoded generation, (c) Simple AND gate partial product generation

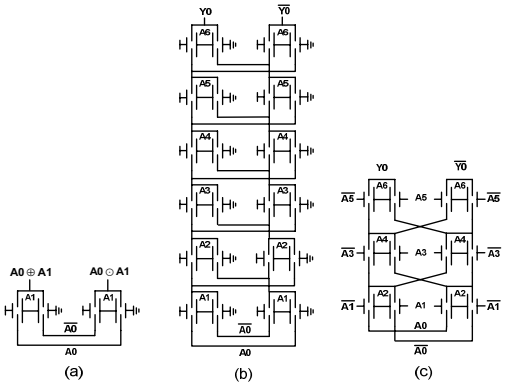

Figure 7. Implementation of (a) a two input XOR/XNOR gate and (b) the Y_0 sub-circuit (c) Y_0 sub-circuit with signals on body terminals

Symmetry in the Y_1 sub-circuit and fully complementary paths, enables sharing of intermediate nodes to implement $\overline{Y}_1$ by adding only 8 relays to the Y_1 design (Fig. 8(d)).

Figures 9(a) and 9(b) show the implementation of propagation and kill/generate sections of the Y_2 sub-circuit for a (5:3) compressor using similar methods. Here $\sum_{i=1}^{4} A_i = 3$ propagates A_0, $\sum_{i=1}^{4} A_i > 3$ generates and $\sum_{i=1}^{4} A_i < 3$ kills the output. Fig. 9(c) shows the full Y_2 circuit for a (7:3) compressor. As can be seen in the compressor circuits, $\overline{A}_0$ is the only complementary signal required for the operation of the multiplier. In this design, that bit is created by $\overline{Y}_0$ and $\overline{Y}_1$ since the added area and energy overhead is relatively small due to the symmetry of those circuits. On the other hand, Y_2 lacks that symmetry, and hence implementation of $\overline{Y}_2$ requires twice the number of relays. As a result, the omission of $\overline{Y}_2$ reduces the total MEM relay count of a (7:3) compressor by 30.

In the proposed (7:3) compressor, the electrical path is provided through the source/drain inputs ($A_0/\overline{A}_0$), enabling compressor stacking. The resulting compressor uses only

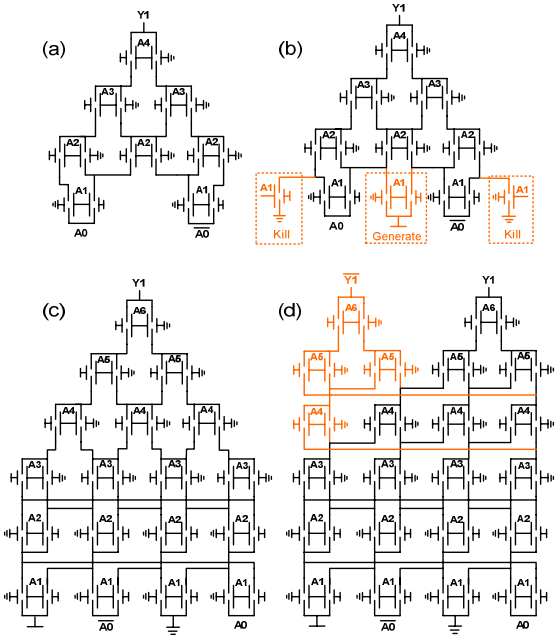

Figure 8. Steps toward implementation of the Y_1 sub-circuit of a complementary (7:3) compressor. (a) propagation paths for Y_1 of the (5:3) compressor, (b) Y_1 the (5:3) compressor, (c) Y_1 of the (7:3) compressor and (d) Y_1 of full complementary (7:3) compressor

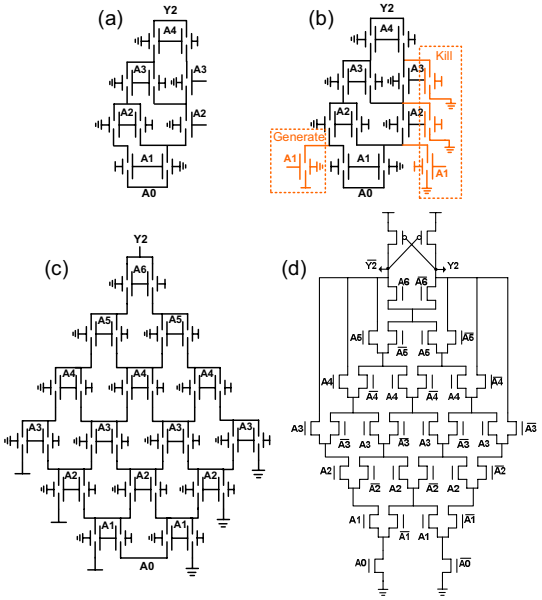

Figure 9. Steps toward implementation of the Y_2 sub-circuit of a (7:3) compressor: (a) propagation paths for Y_2 of the (5:3) compressor, (b) Y_2 of the (5:3) compressor, (c) Y_2 of the (7:3) compressor, (d) Y_2 of the (7:3) CMOS compressor [5]

98 MEM-relays and results in a 5 mech. delay critical path for the 32-bit multiplier, while the static CMOS-style implementation of the same circuit would have 44 mech. delays. A direct MEM-relay translation of the CMOS pass-gate (7:3) compressor whose MSB sub-circuit is shown in Fig. 9(d) [5] requires two mechanical delays and would result in 19 mechanical delays in the critical path of a 32-bit multiplier. In the proposed design, the mechanical delay corresponding to the top PMOS in Fig. 9(d) is eliminated by direct implementation of kill/generate paths and hence the total mechanical delay is ~4x lower.

V. Energy/Delay Estimates of Scaled MEM Relay vs. CMOS Multipliers

In order to illustrate the potential gains over CMOS implementations, we benchmark the 16-bit relay multiplier built with our predictive scaled MEM relay parameters [3], against two different 16-bit CMOS multipliers. The first CMOS multiplier is designed using the Dadda tree algorithm [6] with Han-Carlson adders [7], and placed-and-routed using the Nangate 45 nm Standard Cell Library [8], resulting in a total area of 0.007 mm^2. The energy/delay curves are obtained by scaling the supply voltage between 0.7V-1.4V. The second CMOS multiplier employs optimally tiled compressor tree architecture (OTCT) with radix-4 Booth encoding and an arrival-profile aware completion adder [9]. This multiplier is built in a 90 nm CMOS technology and its total area is 0.03 mm^2. The energy/delay curves have been plotted for various operation voltages and frequencies reported in [9].

The MEM relay-based multiplier is built with 5610 relays in a projected 90 nm lithographic process, where each relay occupies 12 μm^2 resulting in a total multiplier area of approximately 0.087 mm^2. The energy/delay curves are shown for the operating voltage in the range of $2V_{pi}$ to $5V_{pi}$. The simulations are based on the MEM-relay model described in [10] and all parasitic and routing wires

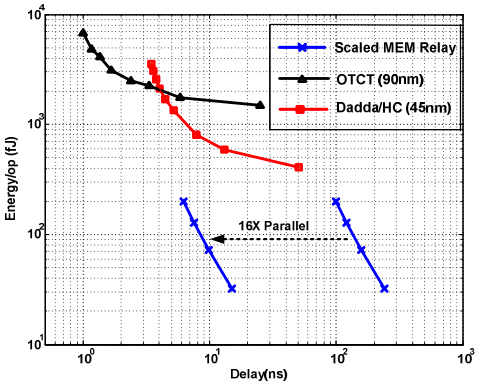

Figure 10. Energy/throughput comparison of CMOS and scaled MEM relay-based 16-bit multipliers

capacitances have been taken into account. The energy-delay trade-offs of both CMOS and MEM-relay 16-bit multipliers are shown in Fig. 10. The CMOS multipliers reach their minimum energy point for delays greater than 50 ns. As a result, the scaled MEM relay multiplier offers ~10x better energy-efficiency over CMOS OTCT and Dadda/Han-Carlson multipliers for sub-10 MOPS operation. As illustrated in Fig. 10, a further trade-off of increasing area for 16-parallel multiplications enables operation up to 100 MOPS, while preserving the energy-efficiency. The energy/op and delay of the same multiplier built with current 1µm MEM-relay and operating at 10V are approximately 10nJ and 20µs, 2 and 5 orders of magnitude more than the scaled relay design respectively, as the capacitances and V_{pi} are each approximately 2 orders of magnitude more than the scaled version. This shows the importance of reliable scaling of MEM-relays.

VI. The (7:3) Compressor Experimental Results

Reliability issues typically associated with MEM-relays such as contact oxidation, welding and charging can potentially prevent this technology from being used in larger circuit blocks. To illustrate the potential of this technology for building more complex circuits, we demonstrate a fully operational 98 relay (7:3) compressor block. The die photo and the operation of the fabricated compressor are shown in Figs. 11 and 12. A full-set of random vectors ranging from 0 to 127 is applied to the compressor, actuating the relay gates with V_{GB} voltage of ~10V and activating different propagate, generate and kill paths in the sub-circuits. The measured output code perfectly matches the expected value. The decline of output voltage in time is due to formation of native oxide on the surfaces of the electrodes of the MEM relays, which gradually increases their on-resistance. Before testing a circuit, the native oxide is intentionally broken by temporarily applying a larger source/drain voltage. Hermetic encapsulation of the chips would prevent contact oxidation and the need for an initial oxide breaking step.

VII. Conclusion

This work describes the micro-architectural and circuit techniques for design of MEM relay-based multipliers optimized over an area-delay trade-off space. Design analysis shows the performance benefits of using relay-tailored higher ratio compressors at the micro-architecture level as well as customized compressor circuits. The

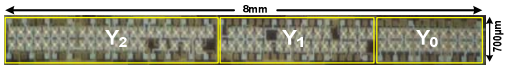

Figure 11. The (7:3) compressor die photo

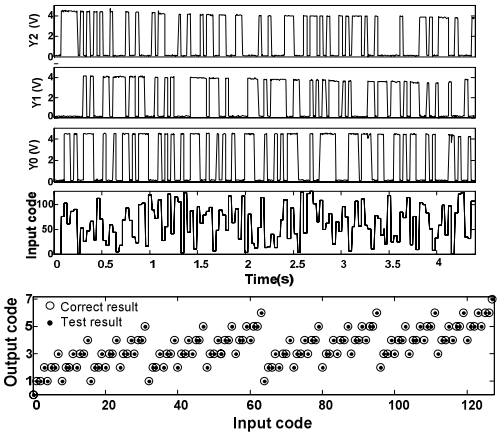

Figure 12. Experimental results of the (7:3) compressor

operation of the main building block of the MEM-relay based multiplier, the (7:3) compressor, is experimentally demonstrated. This circuit is the largest MEM-relay based circiut successfully tested to date. Simulation results of a 16-bit relay multiplier built in a scaled relay process predicts ~10x improvement in energy-efficiency over CMOS designs in the 10 MOPS performance range. The relative performance of the multiplier enhancements are in line with what was previously predicted by a MEM relay 32-bit adder [3], suggesting that complete VLSI systems (e.g. a microprocessor or an ASIC) would expect to see similar energy/performance improvements from adopting MEM relay technology.

Acknowledgments

We gratefully acknowledge the contributions of the whole DARPA NEMS team at UC Berkeley, UC Los Angeles and MIT, especially Matthew Spencer, Elad Alon, Hei Kam, Vincent Pott, Chengcheng Wang, Kevin Dwan and Dejan Marković. This work was supported in part by the DARPA NEMS program, the C2S2 and MSD FCRP centers, BWRC sponsors, MIT CICS, and NSF Infrastructure Grant No. 0403427.

References

[1] H. Kam, *et al,* "Design and reliability of a micro-relay technology for zero-standby-power digital logic applications," *IEDM* 2009.

[2] F. Chen, *et al*, "Integrated circuit design with NEM relays," *ICCAD*, 2008.

[3] M. Spencer *et al.,* "Demonstration of integrated micro-electro-mechanical relay circuits for VLSI applications," *JSSC* Jan. 2011.

[4] O. L. McSorley, "High speed arithmetic in binary computers," *Proc. IRE,* vol 49, pp 67-71, Jan. 1961

[5] P.J. Song, G. De Micheli, "Circuit and architecture trade-offs for high-speed multiplication ," *JSSC*, Sep 1991.

[6] L. Dadda, "Some schemes for parallel multipliers", *Alta Frequenza,* Vol. 34, No. 5, pp. 349--356, March 1965.

[7] T. D. Han and D. A. Carlson, "Fast area-efficient VLSI adders," *8th symposium on Computer Arithmetic*, May 1987.

[8] "Nangate 45nm Open Cell Library," www.nangate.com/openlibrary

[9] S.K. Hsu *et al.,* "A 110 GOPS/W 16-bit multiplier and reconfigurable PLA loop in 90-nm CMOS," *JSSC*, 2006.

[10] H. Kam, T.K. Liu, V. Stojanovic, , D. Markovic, E. Alon, "Design, optimization, and scaling of MEM selays for ultra-low-power digital logic," *Trans. Electron Devices,* Jan. 2011.