THERE'S PLENTY OF ROOM AT THE TOP

Tsu-Jae King Liu[1], Urmita Sikder[1], Kimihiko Kato[2] and Vladimir Stojanovic[1]*
[1]University of California at Berkeley, Berkeley, California USA
[2]University of Tokyo, Tokyo Japan

ABSTRACT

The virtuous cycle of integrated-circuit technology advancement has been sustained for over 50 years, resulting in the proliferation of information technology (IT) with dramatic economic and social impact. Although there is still some "room at the bottom" today to manipulate and control matter at ever smaller scales, physics and economics limit the benefits of further transistor scaling. This paper discusses how the computational functionality and energy efficiency of microchips can be enhanced by integrating MEM switches on top of transistors, to sustain the IT revolution.

INTRODUCTION

Ever since Dr. Richard Feynman spoke of immense possibilities afforded by the miniaturization of electronic and mechanical devices down to the atomic scale [1], the semiconductor industry has steadily advanced planar process technology [2] to fabricate ever smaller transistors with high manufacturing yield. This advancement in integrated circuit (IC) technology over the past five decades has been driven primarily by economics [3]: as the density of transistors in the most advanced chip designs has roughly doubled with each new generation of technology due to higher-resolution patterning capability, the cost per transistor has decreased exponentially over time. (This is because the number of steps in the microchip fabrication process does not increase proportionately with the density of transistors.) Together with improvements in IC performance and functionality arising from transistor scaling and increased integration, this has led to the proliferation of information technology (IT), with dramatic economic and social impact. IT-using and IT-producing industries are estimated to generate nearly all economic productivity growth in the U.S. economy from 2010-2020 [4].

Today, techniques such as atomic layer deposition [5] and atomic layer etching [6] are used in high-volume IC manufacturing, and transistors down to 1 nm gate length have been experimentally demonstrated [7]. Although there is still some "room at the bottom" to manipulate and control matter at even smaller scales, the physics of transistor operation limit the benefits of further transistor scaling [8]. Therefore, alternative approaches eventually will be needed to improve chip performance and functionality at reasonable cost. Also, with the proliferation of mobile electronic devices and the emergence of the Internet of Things (IoT), power consumption has moved to the fore of challenges for future information-processing devices. Monolithic integration of micro-electro-mechanical structures (MEMS) with transistor-based electronic circuitry already has enabled products with enhanced functionality and/or improved energy efficiency. Examples are Texas Instruments' digital micromirror device [9] used in digital projectors, and resonators used for energy-efficient (high quality factor) radio-frequency signal processing and sensing applications [10]. This paper discusses how the computational functionality and energy efficiency of microchips can be further enhanced by integrating micro-electro-mechanical (MEM) switches on top of transistors, to sustain the IT revolution.

WHY MEM SWITCHES?

Complementary metal-oxide-semiconductor (CMOS) transistor technology is predominantly used for digital computing because it provides for lower static power consumption than other transistor technologies. In a digital logic circuit, a transistor functions simply as an electronic switch: in the ON state (when the magnitude of the voltage applied to the "gate" terminal exceeds a threshold value) it conducts current, causing the voltage between its conductive terminals (referred to as the "source" and "drain" in a MOS transistor) to be reduced to zero; in the OFF state it impedes current, allowing a large voltage to be sustained between its conductive terminals. Due to their physics of operation, transistors have non-zero OFF-state leakage current (I_{OFF}), which fundamentally limits the energy efficiency of CMOS-based digital computing [11].

In contrast to a transistor, a mechanical switch can have zero I_{OFF} because an air gap exists between its conductive terminals in the OFF state. For this reason, there has been renewed interest in mechanical switches for ultra-low-power computing [12, 13]. Much progress has been made in recent years to demonstrate MEM relay ICs [14] and to lower the operating voltage of MEM relays for lower active power consumption [15].

Challenges for micro-mechanical computing

Purely relay-based ICs will always operate more slowly than purely transistor-based ICs, since the turn-ON delay of a MEM switch (>1 ns) is orders of magnitude longer than that of a transistor (< 1 ps) [13]. Also, the endurance of MEM switches generally is much worse than that of transistors [16]. Another challenge for mechanical switches is that they are difficult to scale down to physical size as small as that of modern CMOS transistors. Electro-mechanical relay miniaturization is constrained by the requirement of low actuation voltage and by manufacturing process limitations which set a lower limit on air-gap thickness. A sufficiently large actuation area is needed to generate sufficient electrostatic force to overcome the spring restoring force of the movable structure in order to turn ON the switch; in turn, the spring restoring force must be larger than the contact adhesive force in order for a switch to turn OFF when the applied gate voltage is reduced to zero [17]. To minimize the layout area of a relay, its movable element should be a vertically oriented beam (Fig. 1).

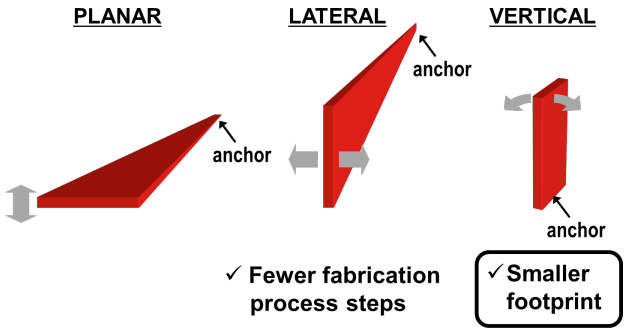

Figure 1: Options for MEM switch structural orientation.

MEM SWITCHES ON TOP OF CMOS

The density of metallic wires used for interconnecting transistors has steadily increased together with the density of transistors, as IC technology has advanced over time. Today there are more than a dozen "back end of line" (BEOL) interconnect layers in the most advanced CMOS manufacturing process, with air gaps incorporated in-between the wires to reduce undesirable capacitive coupling [19]. These trends (increasing number of wiring layers, reduced spacing between wires within a layer, and the incorporation of air gaps) can be leveraged to implement vertically oriented MEM switches in BEOL layers at relatively low incremental cost. Such monolithic integration with underlying CMOS circuitry can provide for ultra-low-power, compact digital logic and memory circuits [18].

In consideration of the challenges mentioned above, MEM switches in a hybrid CMOS+MEMS IC ideally should only be cold-switched and should not be required to cycle ON/OFF a large number ($>10^{10}$) of times. Also, the speed of circuit operation should not be limited by mechanical switching delay. Each of these operating requirements can be met by using MEM switches as non-volatile memory (NVM) elements rather than as digital-logic switches. In order for a MEM switch to retain its contacting state when the power supply is shut off, the movable structure must be designed so that its spring restoring force is smaller than the contact adhesive force.

An advantage of MEM switches over other NVM devices is that they can achieve a larger (ideally infinite) difference in current flow between two charge-conduction states – contact *vs*. non-contact states, for a mechanical switch – representing one binary digit of information. Additionally, it is possible to achieve perfectly complementary connections to a pair of conducting electrodes, using a single-pole/double-throw (SPDT) switch design as shown in Fig. 2. Notably, since no direct current flow is necessary to set and reset the state of this switch (*i.e.* electrodes D0 and D1 can be electrically floating during a set/reset operation), the energy needed to program it is orders of magnitude lower than that required for other types of NVM devices. In 7 nm-generation CMOS technology, the minimum pitch of line/space features in a metal layer is 40 nm, *i.e.* the minimum spacing between metal electrodes is 20 nm [20]. Three-dimensional (3-D) device simulations [21] indicate that a compact BEOL NV-MEM switch implemented in this technology could be set with less than 10 aJ [18] *vs.* more than 1 fJ for other emerging NVM devices [22].

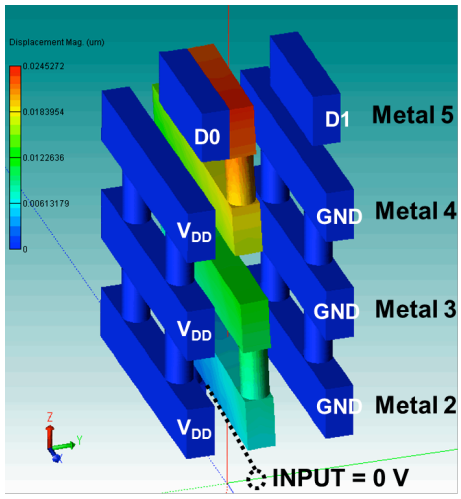

Figure 2: Illustration of a SPDT MEM switch implemented using multiple BEOL metal and via layers in a CMOS manufacturing process. The movable electrode (a serpentine beam) is anchored using the first layer of metal (Metal 1, not shown), and is actuated by electrostatic force resulting from voltages applied to the beam and the actuation electrodes on either side of it. The color scale indicates physical displacement.

CMOS+MEMS FOR FUTURE ENERGY-EFFICIENT COMPUTING

Novel circuit architectures which leverage the aforementioned advantages of NV-MEM devices can provide for ultra-fast and energy-efficient computing. Two examples are given below.

Data Searching

Data searching is a basic operation used in computer networks and in the processing of large and complex data sets ("big data"). Conventionally, a data search operation utilizes a combination of a processor chip (CPU) and memory chip (DRAM). A single-chip solution for high-speed and energy-efficient data searching utilizing embedded NV-MEM devices was proposed recently [23]. Briefly, data strings are stored in an array of non-volatile memory cells, each comprising one access transistor and one NV-MEM switch (Fig. 3). The memory cell has a layout area equal to $8F^2$ (where F is the minimum half-pitch), which is suitable for high-density storage.

To find a matching data string stored within the memory array, two read operations are performed directly on the array, in parallel across all of the programmed cells. Circuit simulations indicate that the location of a data string can be found in less than 0.5 ns with less than 2.5 pJ. To put these numbers into perspective, for a die size of 76 mm^2 with F = 20 nm and 35% cell density (similar to that of DDR4 DRAM), the CMOS+MEMS chip would have a storage capacity of 8 Gb and would consume only 300 nJ to find a match on the whole chip. In comparison, it would take a combination of CPU and DRAM approximately 90 mJ and 80 ms for the same task. The relatively fast read speed and low power consumption make the proposed CMOS+NEMS solution well-suited for real-time data searching applications.

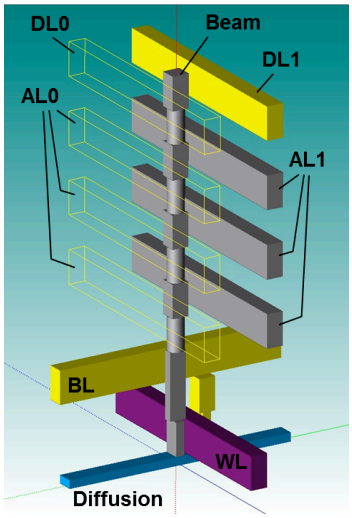

Figure 3: 3-D view of a NV-MEMory cell comprising one access transistor (formed in the silicon "Diffusion" layer) and one NV-MEM switch implemented in multiple layers of metal interconnects. To set the state of the cell, the beam is grounded (by driving the word-line high to turn on the transistor, thereby electrically connecting the beam to the bit-line that is driven low during the programming operation) and a voltage pulse is applied to one of the actuation electrodes AL0 or AL1 (shared with all of the cells in the same row of the memory array) to cause the beam to come into contact with one of the data lines DL0 or DL1 (which also are shared across the array).

Look-Up Tables (LUTs)

In the future, swarms of embedded electronic devices must be able to perform relatively simple real-time computations on sensor data with very high energy-efficiency for long battery life or self-powered operation. Lookup tables (Fig. 4), in which runtime computation is replaced by a simple memory-array indexing operation [24], are well suited for this application. A single-chip "in-memory computing" solution using embedded resistive memory (ReRAM) in a compact cross-point array architecture has been proposed and demonstrated [25]. MEM switches are advantageous for implementing the NVM array because they can be programmed with much less energy than required for other NVM devices, as noted above. Since LUTs generally do not need to be reprogrammed many times, the limited cycling endurance of MEM switches should not be an issue for this application. A NV-MEM-based LUT architecture and operating scheme recently was proposed to provide for improved speed and energy efficiency [26].

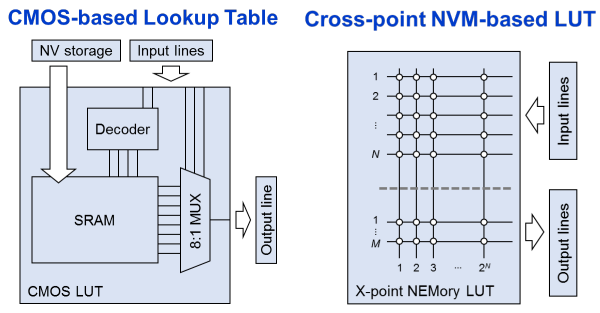

Figure 4: Comparison of LUT implementations.

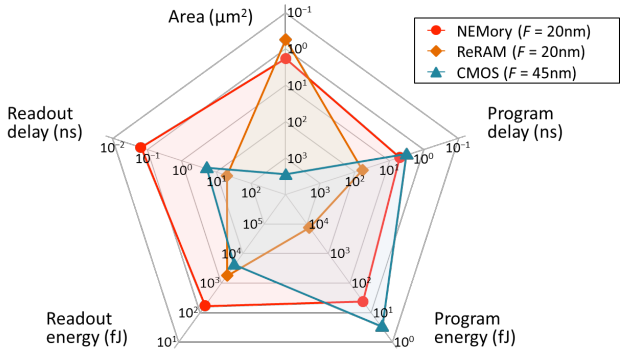

Figure 5: Radar plot comparing the performance characteristics of NV-MEM-based LUT against those of conventional CMOS-based LUT [24] and ReRAM-based in-memory computing [25]. Cell performance based on [27] and readout time of 20 ns are assumed for ReRAM.

The performance characteristics of a NV-MEM-based LUT are benchmarked against those of CMOS-based and ReRAM-based LUTs in Fig. 5. It can be seen from this chart that much lower readout energy and delay are remarkable advantages of the CMOS+MEMS single-chip solution.

CONCLUSION

Continual IC technology advancement is bringing the semiconductor industry ever closer to practical limits for increasing transistor density, but at the same time is facilitating the monolithic integration of MEM switches on top of transistors. In addition to diversifying functionality, CMOS+MEMS technology has tremendous potential for enhancing the computational functionality and energy efficiency of electronic devices, which will enable them to become truly pervasive. Therefore, we can look forward to a future where distributed information systems are used for sensing, computing, and communication to address the most pressing needs of our global society.

ACKNOWLEDGEMENTS

This work was supported in part by the National Science Foundation Center for Energy Efficient Electronics Science (NSF Award 0939514).

REFERENCES

[1] R. P. Feynman, "There's Plenty of Room at the Bottom," *Engineering and Science*, pp. 22-36, 1960.

[2] J. A. Hoerni, "Method of Manufacturing Semiconductor Devices" U.S. Patent 3025589, filed May 1, 1959.

[3] G. E. Moore, "Cramming more Components onto Integrated Circuits," *Electronics*, pp. 114-117, 1965.

[4] D. W. Jorgensen, M. S. Ho and J. D. Samuels, "Information technology and U.S. productivity growth: evidence from a prototype industry-level production account," *Journal of Productivity Analysis*, vol. 36, no. 2, pp. 159-175, 2011.

[5] S. M. George, "Atomic layer deposition: an overview," *Chemical Reviews*, vol. 110, no. 1, pp. 111-131, 2010.

[6] A. Agarwal and M. J. Kushner, "Plasma atomic layer etching using conventional plasma equipment," *Journal of Vacuum Science & Technology A*, vol. 27, pp. 37-50 , 2009.

[7] S. B. Desai, S. R. Madhvapathy, A. B. Sachid, J. P. Llinas, Q. Wang, G. H. Ahn, G. Pitner, M. J. Kim, J. Bokor, C. Hu, H.-S. P. Wong and A. Javey, "MoS_2 transistors with 1-nanometer gate lengths," *Science*, vol. 354, pp. 99-102, 2016.

[8] J. Wang and M. Lundstrom, "Does source-to-drain tunneling limit the ultimate scaling of MOSFETs?" *International Electron Devices Meeting Technical Digest*, pp. 707-710, 2002.

[9] L. J. Hornbeck, "Multi-level digital micromirror device," U. S. Patent 5583688, filed December 21, 1993.

[10] M. K. Zalalutdinov, J. D. Cross, J. W. Baldwin, B. R. Ilic, W. Zhou, B. H. Houston and J. M. Parpia, "CMOS-integrated RF MEMS resonators," *Journal of Microelectromechanical Systems*, vol. 19, no. 4, pp. 807-815, 2010.

[11] B. H. Calhoun, A. Wang and A. Chandrakasan, "Modeling and sizing for minimum energy operation in subthreshold circuits," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 9, pp. 1778-1786, 2005.

[12] K. Akarvardar, D. Elata, R. Parsa, G. C. Wan, K. Yoo, J. Provine, P. Peumans, R. T. Howe and H.-S. P Wong, "Design considerations for complementary nanoelectromechanical logic gates," *IEEE International Electron Devices Meeting Technical Digest*, pp. 299-302, 2007.

[13] F. Chen, H. Kam, D. Marković, T.-J. K. Liu, V. Stojanović and E. Alon, "Integrated circuit design with NEM relays," *IEEE/ACM International Conference on Computer-Aided Design*, pp. 750-757, 2008.

[14] M. Spencer, F. Chen, C. C. Wang, R. Nathanael, H. Fariborzi, A. Gupta, H. Kam, V. Pott, J. Jeon, T.-J. K. Liu, D. Marković, E. Alon and V. Stojanović, "Demonstration of integrated micro-electro-mechanical (MEM) switch circuits for VLSI applications," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 1, pp. 308-320, 2011.

[15] B. Osoba, B. Saha, L. Dougherty, J. Edgington, C. Qian, F. Niroui, J. H. Lang, V. Bulović, J. Wu and T.-J. K. Liu, "Sub-50 mV NEM relay operation enabled by self-assembled molecular coating," *IEEE International Electron Devices Meeting Technical Digest*, pp. 655-658, 2016.

[16] A. Peschot, C. Qian and T.-J. K. Liu, "Nanoelectromechanical switches for low-power digital computing," *Micromachines*, vol. 6, no. 8, pp. 1046-1065, 2015.

[17] C. Pawashe, K. Lin and K. J. Kuhn, "Scaling limits of electrostatic nanorelays," *IEEE Transactions on Electron Devices*, vol. 60, no. 9, pp. 2936–2942, 2013.

[18] N. Xu, J. Sun, I-R. Chen, L. Hutin, Y. Chen, J. Fujiki, C. Qian and T.-J. K. Liu, "Hybrid CMOS/BEOL-NEMS technology for ultra-low-power IC applications," *IEEE International Electron Devices Meeting Technical Digest*, pp. 677–680, 2014.

[19] S. Natarajan *et al*., "A 14nm logic technology featuring 2nd-generation FinFET, air-gapped interconnects, self-aligned double patterning and a 0.0588 μm^2 SRAM cell size," *IEEE International Electron Devices Meeting Technical Digest*, pp. 71-73, 2014.

[20] S.-Y. Wu *et al*., "A 7nm CMOS platform technology featuring 4th Generation FinFET transistors with a 0.027um^2 high density 6-T SRAM cell of mobile SoC applications," *IEEE International Electron Devices Meeting Technical Digest*, pp. 43-46, 2016.

[21] CoventorWare MEMS+ User Guide, 2013.

[22] Table ERD3, *International Technology Roadmap for Semiconductors*, 2011 Edition.

[23] K. Kato, V. Stojanović and T.-J. K. Liu, "Non-volatile nano-electro-mechanical memory for energy-efficient data searching," *IEEE Electron Device Letters*, vol. 37, no. 1, pp. 31-34, 2016.

[24] S. Paul and S. Bhunia, *Computing with Memory for Energy-Efficient Robust Systems* (New York: Springer), 2014.

[25] B. Chen, F. Cai, J. Zhou, W. Ma, P. Sheridan and W. D. Lu, "Efficient in-memory computing architecture based on crossbar arrays," *IEEE International Electron Devices Meeting Technical Digest*, pp. 459-462, 2015.

[26] K. Kato, V. Stojanović and T.-J. K. Liu, "Embedded nano-electro-mechanical memory for energy-efficient reconfigurable look-up tables," *IEEE Electron Device Letters*, vol. 37, no. 12, pp. 1563-1565, 2016.

[27] B. Govoreanu *et al*., "10×10 nm^2 Hf/HfO_x crossbar resistive RAM with excellent performance, reliability and low-energy operation," *IEEE International Electron Devices Meeting Technical Digest*, pp. 729-732, 2011.

CONTACT

*T.-J. K. Liu, tel: +1-510-642-2689; tsujae@berkeley.edu