Afterlife for Silicon: CMOL Circuit Architectures

Xialong Ma, Dmitri B. Strukov, Jung Hoon Lee, and Konstantin K. Likharev*
Stony Brook University, NY 11794-3800, U.S.A.
* Email: klikharev@notes.cc.sunysb.edu

Abstract — **This is a brief review of our recent work on architectures for the prospective hybrid CMOS/nanowire/ nanodevice ("CMOL") circuits including digital memories, reconfigurable Boolean-logic circuits, and mixed-signal neuromorphic networks. The basic idea of CMOL circuits is to combine the advantages of CMOS technology (including its flexibility and high fabrication yield) with the extremely high potential density of molecular-scale two-terminal nanodevices. Relatively large critical dimensions of CMOS components and the "bottom-up" approach to nanodevice fabrication may keep CMOL fabrication costs at affordable level. At the same time, the density of active devices in CMOL circuits may be as high as 10^{12} cm^2 and that they may provide an unparalleled information processing performance, up to 10^{20} operations per cm^2 per second, at manageable power consumption.**

I. Introduction

Recent results [1, 2], indicate that the current VLSI paradigm, based on a combination of lithographic patterning, CMOS circuits, and Boolean logic, can hardly be extended into a-few-nm region. The main reason is that at gate length below 10 nm, the sensitivity of parameters (most importantly, the gate voltage threshold) of silicon MOSFETs to inevitable fabrication spreads grows exponentially. As a result, the gate length should be controlled with a few-angstrom accuracy, far beyond even the long-term projections of the semiconductor industry [3]. (Similar problems are faced by the lithography-based single-electron devices [4].) Even if such accuracy could be technically implemented using sophisticated patterning technologies, this would send the fabrication facilities costs (growing exponentially even now) skyrocketing, and lead to the end of the Moore's Law some time during the next decade. This is why there is a rapidly growing consensus that the impending crisis of the microelectronics progress may be resolved only by a radical paradigm shift from the lithography-based fabrication to the "bottom-up" approach based on nanodevices with Nature-fixed size, e.g., specially designed molecules. Since the functionality of such nanodevices is relatively low [2], they necessarily should be used as an add-on to a CMOS subsystem. Several proposals of such hybrid CMOS/nanodevice circuits were put forward recently (for their review, see Ref. 5).

Our group is working on a particular circuit concept, dubbed CMOL [2, 6], for which the application prospects look best. As in several earlier proposals, nanodevices in CMOL circuits are formed (e.g., self-assembled) at each crosspoint of a "crossbar" array, consisting of two levels of nanowires (Fig. 1). However, in order to overcome the CMOS/nanodevice interface problems pertinent to earlier proposals, in CMOL circuits the interface is provided by pins that are distributed all over the circuit area, on the top of the CMOS stack. (Silicon-based technology necessary for fabrication of pins with nanometer-scale tips has been already developed in the context of field-emission arrays [7].) As Fig. 1c shows, pins of each type (reaching to either the lower or the upper nanowire level) are arranged into a square array with side $2\beta F_{\rm CMOS}$, where $F_{\rm CMOS}$ is the half-pitch of the CMOS subsystem, and β is a dimensionless factor larger than 1 that depends on the CMOS cell complexity. The nanowire crossbar is turned by angle $\alpha = \arcsin(F_{\rm nano}/\beta F_{\rm CMOS})$ relative to the CMOS pin array, where $F_{\rm nano}$ is the nanowiring half-pitch.

By activating two pairs of perpendicular CMOS lines, two pins (and two nanowires they contact) may be connected to CMOS data lines (Fig. 1b). As Fig. 1c illustrates, this approach allows a unique access to any nanodevice, even if $F_{\rm nano} \ll F_{\rm CMOS}$ - see Ref. 6 for a detailed discussion of this point. If the nanodevices have a sharp current threshold, like the usual diodes, such access allow to test each of them. Moreover, if the device may be switched between two internal states, e.g., as the single-electron latching switch [2, 8], each device may be switched into the desirable (ON or OFF) state by applying voltages $\pm V_{\rm W}$ to the selected nanowires, so that voltage $V = \pm 2V_{\rm W}$ applied to the selected nanodevice exceeds the corresponding switching threshold, while half-selected devices (with $V = \pm V_{\rm W}$) are not disturbed.

Two advantages of CMOL circuits over other crossbar-type hybrids look most important. First, due to the uniformity of the nanowiring/nanodevice levels of CMOL, they do not need to be precisely aligned with each other and the underlying CMOS stack, thus allowing the use for nanowire formation of advanced patterning techniques [9], [10] which lack precise alignment. Second, CMOL circuits may work with two-terminal nanodevices (e.g., single-electron latching switches) whose fabrication and/or self-assembly is substantially less challenging than that for their three-terminal counterparts. Still, CMOL, similarly to all other nanodevice-based technologies, requires defect-tolerant circuit architectures, since the fabrication yield of such devices will hardly ever approach 100% as closely as that achieved for the semiconductor transistors.

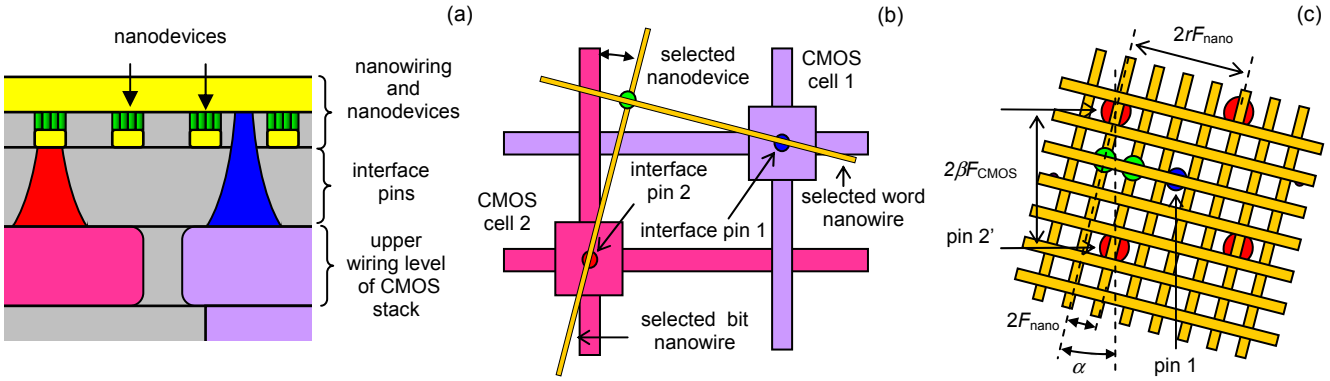

Fig. 1. Low-level structure of the generic CMOL circuit: (a) schematic side view; (b) the idea of addressing a particular nanodevice, and (c) zoom-in on several adjacent interface pins to show that any nanodevice may be addressed via the appropriate pin pair (e.g, pins 1 and 2 for the leftmost of the two shown devices, and pins 1 and 2' for the rightmost device). On panel (b), only the activated CMOS lines and nanowires are shown, while panel (c) shows only two devices. (In reality, similar nanodevices are formed at all nanowire crosspoints.) Also disguised on panel (c) are CMOS cells and wiring.

II. CMOL Memories

The most straightforward potential application of CMOL circuits are embedded memories and stand-alone memory chips, with their simple matrix structure. In such memories, each nanodevice would play the role of a single-bit memory cell, while the CMOS subsystem may be used for coding, decoding, line driving, sensing, and input/output functions.

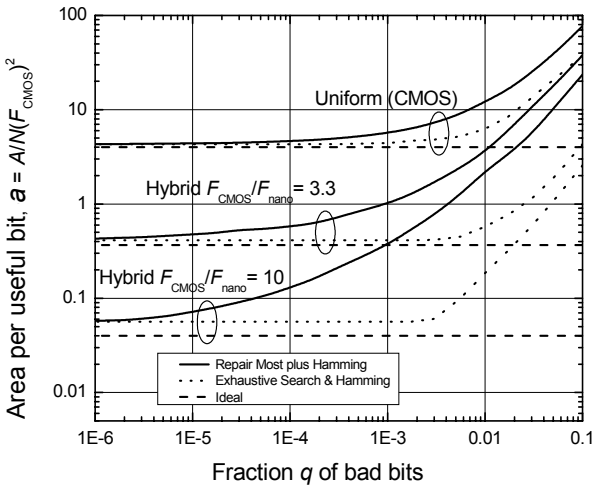

Fig. 2. The optimized area per useful bit as a function of single nanodevice yield, for hybrid (CMOL) and purely semiconductor memories.

We have carried out [11] a detailed analysis of CMOL memories, including the application of two major techniques for increasing their defect tolerance: the memory matrix reconfiguration (the replacement of several rows and columns, with the largest number of bad memory cells, for spare lines), and error correction (based on the Hamming codes). In particular, we have evaluated the additional memory area necessary to achieve a certain fixed yield. Figure 2 shows the optimized total chip area per useful bit, as a function of the nanodevice yield, for two values of the F_{CMOS}/F_{nano} ratio and two defect tolerance boost techniques. (Results for purely CMOS memories are also shown for comparison.) The results show that the array reconfiguration, especially applied in synergy with error correction, can increase the memory defect tolerance very substantially. However, in order to obtain an order-of-magnitude density advantage from the transfer to hybrid memories (such a goal seems natural for the introduction of a novel technology), the single bit device fraction has to be reduced to approximately 2% even if the exponentially long Exhaustive Search algorithm is used. For the simple Repair Most algorithm, this number drops to just 0.1%.

These results do not look overly optimistic, but this should not obscure the fact that when this yield threshold has been achieved, extremely impressive memories will become available. For example, the normalized cell area $a \equiv A/N(F_{CMOS})^2 = 0.4$ (Fig. 2) at $F_{CMOS} = 32$ nm means that a memory chip of a reasonable size (2×2 cm^2) can store about 1 terabit of data - crudely, one hundred Encyclopedia Britannica's.

III. CMOL FPGA: Boolean Logic Circuits

The reconfiguration is the most efficient technique for coping with defective nanodevices in hybrid circuits. This is why the most significant published proposals for the implementation of logic circuits using CMOL-like hybrid structures had been based on reconfigurable regular structures like the field-programmable gate arrays (FPGA). Before our recent work, two FPGA varieties had been analyzed, one based on look-up tables (LUT) and another one using programmable-logic arrays (PLA). Unfortunately, all these approaches run into substantial problems - see, e.g., Ref. 6 for their critical review.

Recently, we suggested [12] an alternative approach to Boolean logic circuits based on CMOL concept, that is close to the so-called cell-based FPGA [13]. In this approach an elementary CMOS cell includes two pass transistors and an inverter, and is connected to the nanowire/nanodevice subsystem via two pins. Disabling the

CMOS inverters allows to carry out the circuit reconfiguration via cell's pass transistors.

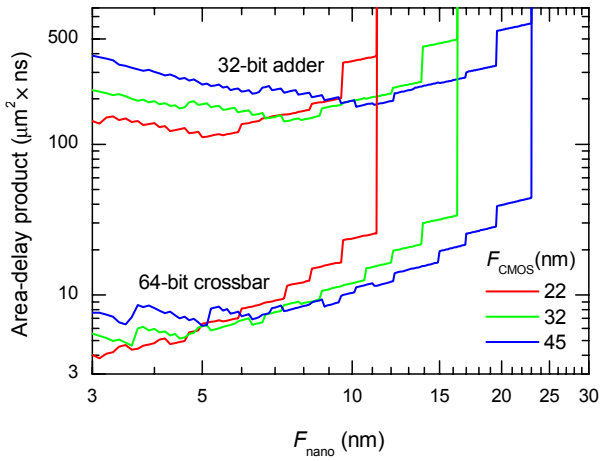

Fig. 3. Optimized CMOL FPGA area-delay product $A\tau$ of the two CMOL FPGA circuits as a function of nanowire half-pitch for three ITRS long-term CMOS technology nodes. The (formal) jump of the $A\tau$ product to infinity at some $(F\text{nano})_{max}$ reflects the fact that circuit mapping on the CMOL fabric may only be implemented for Fnano below this value. The finite sharp jumps are due to the discrete changes of angle α.

We have developed a simple, two-step approach to CMOL FPGA configuration, in which the desired circuit is first mapped on the apparently perfect (defect-free) CMOL fabric, and then is reconfigured around defective components using a simple algorithm [12]. The Monte Carlo simulation (so far only for the "stack-on-open"-type defects which are expected to dominate in CMOL circuits) has shown that even this simple configuration procedure may ensure very high defect tolerance. For example, the reconfiguration of a 32-bit Kogge-Stone adder, mapped on the CMOL fabric with realistic values of parameters, may allow to achieve the 99% circuit yield (sufficient for a ~90% yield of properly organized VLSI chips), with as may as 22% of defective devices, while the defect tolerance of another key circuit, a fully-connected 64-bit crossbar switch, is about 25%.

Our most striking result was that such high defect tolerance may coexist with high density and performance, at acceptable power consumption. For example, our estimates have shown [12] that for the total power of 200 W/cm^2 (planned by the ITRS for the long-term CMOS technology nodes [3]), an optimization may bring the logic delay of the 32-bit Kogge-Stone adder down to just 1.9 ns, at the total area of 110 μm^2, i.e. provide an area-delay product of 150 ns-μm^2, for realistic values F_{CMOS} = 32 nm and F_{nano} = 8 nm (Fig. 3). This result compares very favorably with the estimated 70,000 ns-μm^2 (with 1.7 ns delay and 39,000 μm^2 area) for a fully CMOS FPGA implementation of the same circuit (with the same F_{CMOS}).

IV. CMOL CrossNets: Neuromorphic Networks

One more possible application of CMOL circuits is neuromorphic networks (see, e.g., Ref. 14). We have explored a specific architecture of such networks, called Distributed Crossbar Networks ("CrossNets") [15, 16], which are uniquely suitable for CMOL implementation. In each CrossNet, relatively sparse neural cell bodies ("somas") are implemented in the CMOS subsystem, while the much denser latching switches are used as elementary synapses. The mutually perpendicular nanowires of the CMOL crossbar naturally implement the axons and dendrites which carry signals between the cells, allowing one cell to be directly connected to a virtually unlimited number M of other cells. Due to this parallelism, CrossNets can be spectacularly resilient. For example, a Hopfield-type CrossNet, operating at 50% capacity, may provide 99% fidelity with as many as 85% of bad nanodevices (Fig. 4).

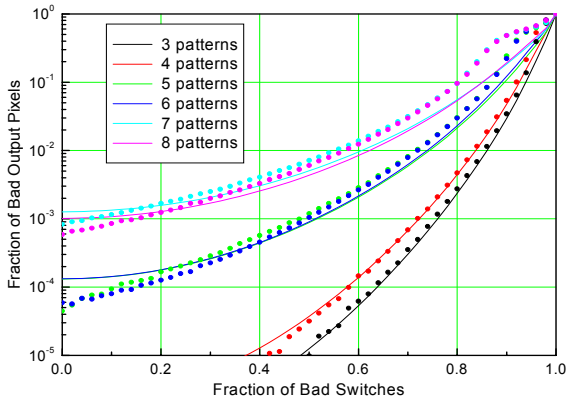

Fig. 4. Defect tolerance of a recurrent InBar with connectivity parameter M = 25, operating in the quasi-Hopfield mode. Lines show the results of an approximate analytical theory, while dots those of numerical experiments.

CrossNet functionality is strongly dependent on the distribution of the somas over the axon/dendrite/synapse field (Fig. 5). For example, "FlossBars" (Fig. 5a) have a layered topology and hence can be used to implement multi-layer perceptron (MLPs) (see, e.g., Ref. 14), while "InBars" (Fig. 5b) have an "interleaved" structure which is natural for the implementation of recurrent networks [14].

CrossNet training faced several challenges including the binary character of the elementary synapse (latching switch) and a certain statistical uncertainty of its switching. In our recent work [17] we have proved that, despite these limitations, CrossNets can be taught, by at least two different methods, to perform virtually all the major functions demonstrated earlier with usual neural networks, including the corrupted pattern restoration in the recurrent quasi-Hopfield mode and pattern classification in the feedforward MLP mode [17, 18].

The importance of this result is in the CrossNet's potential unparalleled density and speed [15-17]: for realistic

parameters, the cell density may exceed that of cerebral cortex (above 10^7 cells per cm^2), while the average cell-to-cell communication delay may be as low as ~10 ns (i.e., about six orders of magnitude lower than that in the brain), at acceptable power. Even putting aside the exciting long-term prospects of creating high-speed artificial brain-like systems [17], CMOL CrossNet chips of modest size might be used for important present-day problems, e.g., online recognition of a person in a large crowd [18].

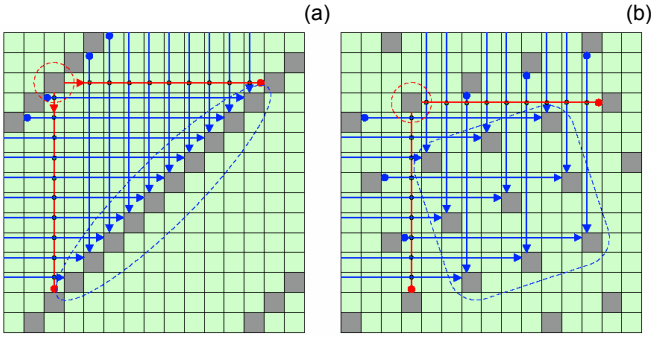

Fig. 5. Two particular CrossNet species: (a) FlossBar (shown for the connectivity parameter M equal to 10) and (b) InBar (for M = 9). For clarity, the figures show only the axons, dendrites, and synapses providing connections between one soma (indicated by the dashed red circle) and its direct signal recipients (inside the dashed blue lines), for the simplest (non-Hebbian feedforward) networks.

V. Conclusions

There is a chance for the development, within perhaps the next 10 to 20 years, of hybrid "CMOL" integrated circuits that will allow to extend Moore's Law to the few-nm range. Preliminary estimates show that such circuits could be used for several important applications, notably including terabit-scale memories, reconfigurable digital circuits with multi-teraflops-scale performance, and mixed-signal neuromorphic networks that may, for the first time, compete with biological neural systems in areal density, far exceeding them in speed, at acceptable power dissipation. The major challenges on the way toward practical CMOL circuits include the development of high-yield techniques for formation (e.g., chemically-directed molecular self-assembly) of single-electron latching switches, and even better architectures for digital and mixed-signal CMOL circuits.

Acknowledgements

Useful discussions of the issues considered in this paper with P. Adams, J. Barhen, V. Beiu, W. Chen, S. Das, J. Ellenbogen, X. Liu, J. Lukens, A. Mayr, V. Protopopescu, M. Reed, M. Stan, and Ö. Türel are gratefully acknowledged. The work has been supported in part by AFOSR, NSF, and MARCO via FENA Center.

References

[1] D. J. Frank, R. H. Dennard, E. Nowak, P. M. Solomon, Y. Taur, and H. S. P. Wong, "Device scaling limits of Si MOSFETs and their application dependencies," *Proc. IEEE,* vol. 89, no. 3, pp. 259-288, 2001.

[2] K. K. Likharev, "Electronics below 10 nm," in *Nano and Giga Challenges in Microelectronics*. Amsterdam: Elsevier, 2003, pp. 27-68.

[3] *International Technology Roadmap for Semiconductors. 2003 Edition, 2004 Update,* Available online at http://public.itrs.net/.

[4] K. K. Likharev, "Single-electron devices and their applications," *Proc. IEEE,* vol. 87, no. 4, pp. 606-632, 1999.

[5] M. R. Stan, P. D. Franzon, S. C. Goldstein, J. C. Lach, and M. M. Ziegler, "Molecular electronics: From devices and interconnect to circuits and architecture," *Proc. IEEE,* vol. 91, no. 11, pp. 1940-1957, 2003.

[6] K. K. Likharev and D. B. Strukov, "CMOL: Devices, circuits, and architectures," in *Introducing Molecular Electronics,* G. Cuniberti, G. Fagas, and K. Richter, Eds. Berlin: Springer, 2005, to be published as Chapter 16.

[7] K. L. Jensen, "Field emitter arrays for plasma and microwave source applications," *Phys. Plasmas,* vol. 6, no. 5, pp. 2241-2253, 1999.

[8] S. Fölling, Ö. Türel, and K. K. Likharev, "Single-electron latching switches as nanoscale synapses," in *International Joint Conference on Neural Networks.* Mount Royal, NY: Int. Neural Network Soc., 2001, pp. 216-211.

[9] S. Zankovych, T. Hoffmann, J. Seekamp, J. U. Bruch, and C. M. S. Torres, "Nanoimprint lithography: Challenges and prospects," *Nanotechnology,* vol. 12, no. 2, pp. 91-95, 2001.

[10] S. R. J. Brueck, "There are no fundamental limits to optical lithography," in *International Trends in Applied Optics.* Bellingham, WA: SPIE Press, 2002, pp. 85-109.

[11] D. B. Strukov and K. K. Likharev, "Prospects for terabit-scale nanoelectronic memories," *Nanotechnology,* vol. 16, pp. 137-148, 2005.

[12] ——, "CMOL FPGA: A cell-based, reconfigurable architecture for hybrid digital circuits using two-terminal nanodevices," *Nanotechnology,* submitted for publication, preprint available at http://rsfq1.physics.sunysb.edu/figures//nano/FPGA05.pdf

[13] J. M. Rabaey, A. P. Chandrakasan, and B. Nikolic, *Digital Integrated Circuits: A Design Perspective,* 2nd ed. Upper Saddle River, NJ: Pearson Education, 2003.

[14] J. Hertz, R. G. Palmer, and A. S. Krogh, *Introduction to the Theory of Neural Computation.* Redwood City, CA: Addison-Wesley Pub. Co., 1991.

[15] Ö. Türel, J. H. Lee, X. L. Ma, and K. K. Likharev, "CrossNets: Possible neuromorphic networks based on nanoscale components," *Int. J. Circ. Theory App.,* vol. 31, no. 1, pp. 37-53, 2003.

[16] K. Likharev, A. Mayr, I. Muckra, and Ö. Türel, "CrossNets – High-performance neuromorphic architectures for CMOL circuits," *Ann. NY Acad. Sci.,* vol. 1006, pp. 146-163, 2003.

[17] Ö. Türel, J. H. Lee, X. L. Ma, and K. K. Likharev, "Neuromorphic architectures for nanoelectronic circuits," *Int. J. Circ. Theory App.,* vol. 32, no. 5, pp. 277-302, 2004.

[18] J. H. Lee and K. K. Likharev, "CMOL CrossNets as pattern classifiers," accepted for presentation at the *8th Int. Conf. on Artificial Neural Networks,* Barcelona, 2005, preprint available at http://rsfq1.physics.sunysb.edu/figures//nano/IWANN05.pdf.