Neuromorphic CMOL Circuits

Konstantin K. Likharev
Stony Brook University
Stony Brook, NY 11794-3800, U.S.A.
klikharev@notes.cc.sunysb.edu

Abstract—**This is a brief review of the recent work on the development of neuromorphic architectures for future hybrid CMOS/nanowire/MOLecular ("CMOL") circuits. Such circuits may provide the first chance for the implementation of advanced information processing systems with areal density of (beyond 10^{12} active functions per cm^2) comparable to that of the human cerebral cortex, while operating at much higher speed (up to 10^{20} operations per second per cm^2), at acceptable power consumption. Our group has suggested a family of distributed crosspoint networks ("CrossNets") that are natural for implementation in CMOL technology, and has shown that such networks may be trained to perform at least the effective pattern recognition in the Hopfield mode. Work on CrossNet training to perform more complex tasks in under way.**

Neuromorphic networks, hybrid CMOS/nanowire/molecular circuits, single-electron devices, training, pattern recognition

I. INTRODUCTION: CMOL

VLSI circuits with sub-10-nm features would provide enormous benefits for all information technologies, including computing, networking, and signal processing. However, recent results [1, 2] indicate that the current VLSI paradigm based on CMOS technology can be hardly extended into this region: below 10 nm gate length the sensitivity of parameters (most importantly, the gate voltage threshold) of silicon field-effect transistors to inevitable fabrication spreads grows exponentially. This sensitivity will probably send the fabrication facilities costs (extremely high even now) skyrocketing, and lead to the end of Moore's Law some time during the next decade.

The main alternative nanodevice concept, single-electronics [2, 3], offers some potential advantages, e.g., a broader choice of possible materials for active devices. However, the critical dimension of single-electron transistors (the single-electron island size) for room-temperature operation should be below ~1 nm [3], far too small for the current and realistically envisioned lithographic techniques.

I believe the impending crisis of the microelectronics progress may only be resolved by a radical paradigm shift from the purely CMOS technology to "CMOL" hybrid circuits (Fig. 1) that would combine [2]:

- a level of advanced CMOS devices fabricated by the lithographic patterning,
- a few layers of parallel nanowire arrays formed, e.g., by nanoimprinting, and
- a level of molecular devices that would self-assemble on the nanowires from solution.

This work was supported in part by DOE and NSF.

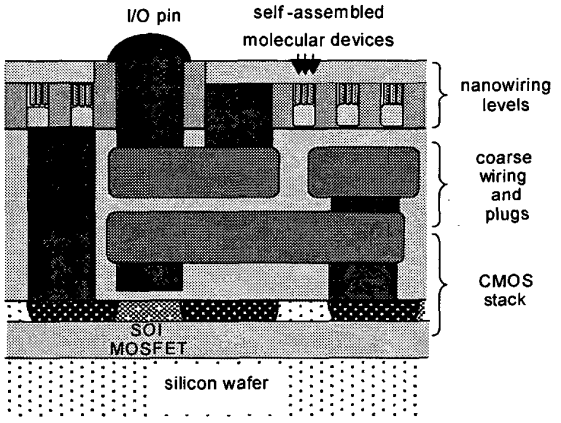

Figure 1. CMOL (hybrid CMOS/nanowire/MOLecular) circuit [2].

The CMOL concept allows a combination of the advantages of both its nanoscale components (e.g., reliability of CMOS circuits and minuscule footprint of molecular devices) and patterning techniques (a virtually complete flexibility of the usual lithography and potentially very low cost of nanoimprinting and chemically-directed self-assembly). This combination may enable an unparalleled density of CMOL circuits, beyond 10^{12} devices per cm^2, limited essentially only by quantum-mechanical tunneling between the adjacent nanowires.

For the single-molecule components, single-electron devices are the leading candidate, because (in contrast to field-effect transistors) their operation mechanism does not require high conductivity of the molecule-to-electrode interfaces. (The CMOS layer allows CMOL to circumvent one prominent drawback of single-electron transistors, their low voltage gain [3].) The recent spectacular demonstration of single-molecule single-electron transistors by several groups [4-6] offers every hope that first VLSI CMOL circuits will be implemented within the next 10-15 years, hopefully in time to preempt the impending Moore's Law crisis.

Any plans for practical applications of CMOL circuits should take into account the fact that the chemically-directed self-assembly will probably never give 100% of good single-molecule devices. Moreover, single-electron transistors (and probably any nanometer-scale electron devices) are vulnerable to single charged impurities in their nearest environment [2, 3]. As a result, any reasonable CMOL circuit architecture has to be defect-tolerant. Because of this, CMOL circuits will probably find their first applications in embedded and stand-alone

0-7803-7976-4/03/$17.00 ©2003 IEEE

memories, with molecules working as memory cells, and CMOS layer taking care of logic, decoder, and driver/sensor circuitry, because simple structure of memory matrices allows effective exclusion of bad cells [7, 8].

I believe, however, that CMOL circuits may be basis of much more interesting and complex systems, extending long beyond the usual digital number-crunching. The primary candidates for that are biologically-inspired neuromorphic architectures. A motivation for this choice comes, for example, from the well-known comparison of the performance of digital computers and biological neural systems for one of the basic simplest information processing tasks: image recognition (in more formal language, classification). A mammal's brain recognizes a complex visual image, with high fidelity, in approximately 100 milliseconds. Since the elementary process of neural cell-to-cell communication in the brain takes 10 to 30 milliseconds, the recognition is completed in just a few "ticks". In contrast, the fastest modern microprocessor operating at a few GHz and running the best commercially available software, would require minutes (i.e., of the order of 10^{11} clock periods) for an inferior classification of a similar image. The contrast is very striking indeed.

The goal of this paper is to give a brief review of our recent work [9-12] on the development of a family of neuromorphic architectures dubbed CrossNets (standing for distributed crossbar networks), that may reach high performance in advanced information processing tasks even with account of hardware limitations of CMOL circuits.

II. CrossNets

Figure 2 shows a possible molecular implementation [11] of the basic device of CrossNet circuits: a three-terminal latching switch [9-11]. This is essentially a combination of two well known single-electron devices, the transistor and the trap [3], where single-electron islands are implemented as diimide acceptor groups. The islands are connected by either oligo-ethynylenephenylene (OPE) bridges playing the role of tunnel junctions, or longer chains that do not conduct electrons, and just stabilize the geometric arrangement. The bridges and chains are terminated by thiol groups that serve as alligator clips that should allow the molecular self-assembly on gold nanowires.

The device is designed to operate in the following way: when the sum of voltages applied to nanowires 1 and 2 exceeds a certain threshold, an additional electron is injected from wire 3 into the trap island, and its electrostatic field opens the single-electron transistor connecting wires 1 and 3. This connection survives the reduction of the applied voltage for quite a while, because it takes time for the trapped electron to escape. As a result, the device function as a fan-in-two latching switch, incorporating a single-bit memory cell.

Because of the strong (exponential) dependence of the electron tunneling probability on the input signal amplitude, grouping of four such devices and using dual-rail presentation of signals (Fig. 3) the implementation of the "Hebbian" function that is very important for neuromorphic networks – see, e.g., Refs. 13, 14. Analysis shows [10, 11] that at large signal amplitude, current transferred through this group from input wires (show red) to output (blue) wire is proportional to the "synaptic weight" $w_{ij} \approx \mathrm{sgn}(y_i \times y_j)$, where y_i and y_j are the input signals. If the signal "activity" (average amplitude within a certain time window) drops below a certain value, the synaptic weights freeze, due to latching of the corresponding switches. Thus the composite device shown in Fig. 3 may play the role of the Hebbian synapse [13, 14] connecting "axonic" (red) and "dendritic (blue) wires. Physically, the wires are similar, and differ only by the way they are connected with CMOS-implemented neural cell bodies – "somas" –see below.

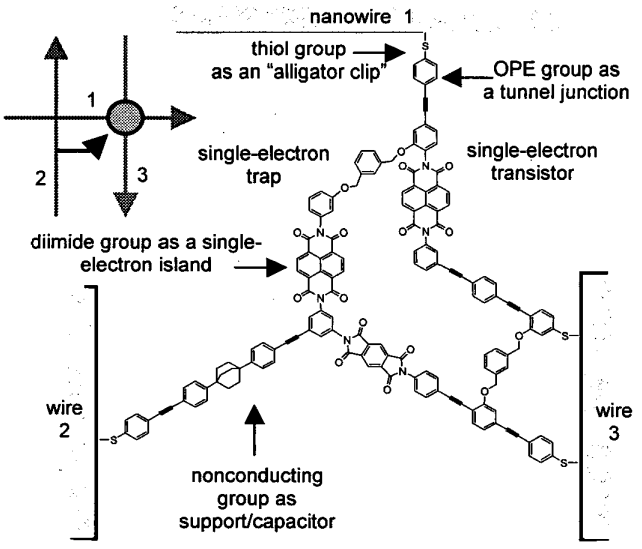

Figure 2. Possible molecular implementation of a three-terminal latching switch [11]. Circuit notation of the device is shown in the left upper corner.

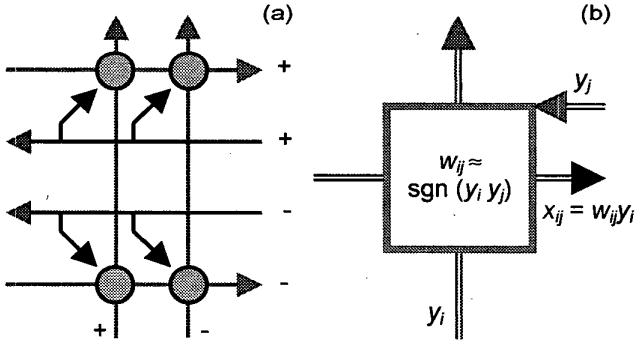

Figure 3. The group of four latching switches as a Hebbian synapse: (a) schematics, and (b) notation used in the following figures.

Figure 4 shows the general structure of CrossNets. These networks consist of synaptic tiles ("plaquettes") with 8 Hebbian synapses each, and sparsely located somatic (gray) cells, to be implemented in the CMOS layer of the CMOL circuit. In the simplest case, the gray cell may be just a nonlinear amplifier that performs the "activation" function $y_i = g(x_i)$ that is linear at small $|x_i|$, but saturates at $y_i = \mathrm{sgn}(x_i)$ at large $|x_i|$.

Various species of CrossNets [11, 12] differ by the way the rare somatic cells are imbedded into a large 2D array of

synaptic plaquettes. Figure 5 shows probably the most promising CrossNet type explored by our group: the so-called InBar. In this network, the somatic cells are located in the points of another square array, that is inclined by small angle α relative to the array of synaptic plaquettes. It is easy to check that synapses connect each somatic cell with $4M$ other cells, where the connectivity parameter M is defined by the incline: $M \equiv 1/\tan^2\alpha$. Of all explored CrossNets, InBar is especially convenient for the CMOL implementation, because each CMOS-based somatic cell may occupy a similar square-shaped area of the chip; the gray squares in Fig. 5 are not these cells per se, but rather the places of their interconnection with the molecule/nanowire layers – cf. Fig. 1.

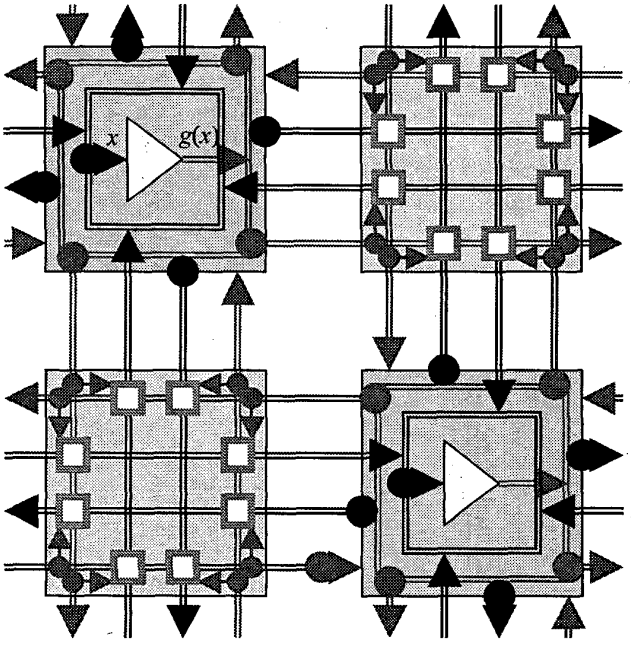

Figure 4. The general scheme of connection of two somatic (gray) cells in CrossNet arrays. Each soma feeds 4 dual-rail output (axonic) lines going in all 4 directions. If the corresponding synapse is connected ($w_{ij} \neq 0$), axonic signal makes a current contribution into each perpendicular "dendritic" (blue) line. The contribution may be either positive or negative, depending on the sign of the synaptic weight w_{ij}. Notice open-circuit terminations of axonic and dendritic lines at the borders of the somatic cells; due to these terminations these cells do not communicate directly.

The most important advantages of the CrossNets over other possible neuromorphic networks (see, e.g., Refs. 9, 15, 16) is that these networks allow to achieve large connectivity $M >> 1$ in quasi-2D CMOL circuits – see Fig. 1. This is very important, because connectivity determines the useful functionality of such networks [13, 14]. For example, in average connectivity of the cerebral cortex neurons is as high as $\sim 10^4$. On the other hand, it is well known that functionality of cellular-automata structures (with near-neighbor interactions and hence $M \sim 1$) is rather limited [15].

One more useful feature of CrossNets is that the physical length of wires connecting each pair of cells equals to their minimal Manhattan distance. This minimizes the interconnect capacitance and hence increases the circuit speed – for estimates, see below.

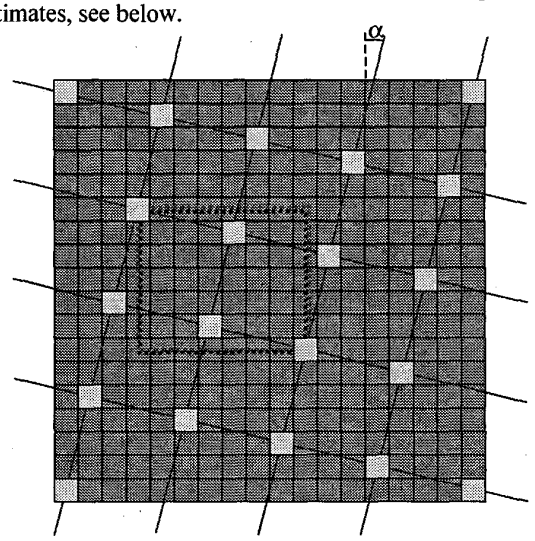

Figure 5. The most promising CrossNet geometry: the inclined distributed crossbar ("InBar"). Green squares are synaptic plaquettes, while gray squares are locations where the somatic cells contact nanowires of the molecular layer. Dotted red and blue lines indicate the interconnection path of a sample pair of somatic cells, while thin violet lines show the rectangular grid at which all the somas are located. The grid incline angle α determines the network connectivity: the number of somatic cells connected with an arbitrary cell is $4M = 4/\tan^2\alpha$

III. CrossNet Training

Neuromorphic networks do not require usual software, but need to be trained to perform their tasks. For CrossNets, the main challenge is that external access to individual synapse is impossible. Because of this, most well developed neural network training techniques cannot be directly applied.

Despite of this challenge, we have demonstrated [12] that at least one CrossNet species, InBar (Fig. 5), may be effectively trained to operate as a Hopfield network (essentially, an associative memory) [13, 14]. For this, one of the cells of each pair is fed with a strong external signal proportional to $\sum_p \xi_i^p \xi_j^p$, where ξ_i^p is the i-th pixel of the p-th image of the training set, while the second cell is fed with a positive signal of constant magnitude. In this way, each of two synapses connecting the cell pair (Fig. 4) acquires weight $w_{ij} \propto \mathrm{sgn} \sum_p \xi_i^p \xi_j^p$. This is exactly the so-called "clipped Hebb rule" that is known to work very well for fully connected Hopfield networks [13, 14]. Our estimates have shown that for a CrossNet trained by this method, the maximum number of stored patterns should scale as M.

Numerical experiments with InBar computer models have confirmed this result. An example of our results is shown in Fig. 6. In this case, an InBar had been taught three different images, and then has been able to restore a strongly corrupted version of any of these images, fed into it as an initial

condition. Remarkably, the recall process is perfect (no corrupted bits in the end) and very fast.

We have been certainly inspired by this first success, but practical applications of Hopfield networks are rather limited [13, 14]. Much more valuable would be to have CrossNets trained as patterns classifiers. This would open way to a large practical market of image classification and feature detection. Recently, we have developed [12] a plan for such training using the fact of chaotic excitation of CrossNets with sufficiently large somatic cell gain [9-11], and are currently working on its verification.

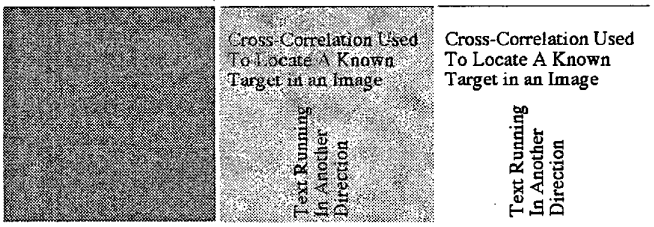

Figure 6. The process of recall of one of three trained black-and-white images by an InBar-type CrossNet with 256×256 neural cells and connectivity parameter $M = 64$. The initial image (left panel) was obtained from the trained image (identical to the one shown in the right panel) by flipping as many as 40% of pixels. $\tau_0 \approx RC_0$ is the effective time constant of intercell interaction. (R_0 is the resistance of the connected latching switch, while C_0 is the dendrite line capacitance per one synaptic plaquette.)

IV. Discussion: Possible CrossNet Performance

If the CrossNets may be indeed trained to perform advanced information processing, their performance may be very spectacular. Assuming the nanowire spacing ("half-pitch") of $F = 2$ nm, limited by wire-to-wire tunneling, for a CrossNet with connectivity $4M = 10^4$, the synaptic plaquette size would be 32×32 nm^2. (Notice that this corresponds to an areal density of ~10^{12} synapses and 10^8 somatic cells per cm^2, higher than that of the mammal's cerebral cortex.) Estimating the time scale τ_0 of neural cell interaction (that is dominated by the charging of the dendritic line capacitances through resistance R_0 of open single-electron transistors) and power dissipated by the circuit, we obtain [11, 12] that for high but acceptable power consumption of 100 W/cm^2, achieved at $R_0 \approx 10^{10}\Omega$ (a very realistic value for the devices like the one shown in Fig. 2), τ_0 is as small as ~20 ns. This speed is approximately 6 orders of magnitude higher than that of the cortical circuitry (and corresponds to ~10^{20} operations per sec per cm^2, the number well above ~3×10^{16} binary operations per second in Pentium 4 processor). Even scaling R_0 up by a factor of 100 to bring power consumption to a more comfortable level of 1 W/cm^2, would still leave CrossNet with 4 orders of magnitude of speed advantage over their biological prototypes. This is why we believe that even relatively small CrossNet CMOL chips may revolutionize the pattern classification field.

This success would pave the way toward much more ambitious goals. It seems completely plausible that a cerebral-cortex-scale CrossNet-based system (with ~10^{10} neurons and 10^{14} synapses, that would require ~10×10 cm^2 silicon substrate) would be able, after a period of initial training by a dedicated external tutor, to learn directly from its interaction with the environment. In this case one can speak of a "self-evolving" system. If these expectations are confirmed, we may be able to revisit the initial dream of the neural network science of providing hardware means for reproducing the natural evolution of the cortex on a much faster time scale. Such evolution may lead to self-development of advanced features such as system self-awareness and reasoning. If a substantial success along these lines materializes, it will have a strong impact not only on information technology, but also on the society as a whole.

Acknowledgment

Substantial contributions into the work described in this review have been made by Stony Brook students S. Fölling, Ö. Türel, and I. Muckra. Useful discussions with P. Adams, J. Barhen, J. Lukens, A. Mayr, V. Protopopescu, and T. Sejnowski are also gratefully acknowledged.

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, pp. 259-288, March 2001.

[2] K. K. Likharev, "Electronics below 10 nm", in Nano and Giga Challenges in Microelectronics, J. Greer, A. Korkin and J. Labanowski, Eds. Amsterdam: Elsevier, in press; preprint available of the Web at http://rsfq1.physics.sunysb.edu/~likharev/nano/NanoGiga.pdf.

[3] K. K. Likharev, "Single-electron devices and their applications", Proc. IEEE, vol. 78, pp. 606-632, April1999.

[4] S. P. Gubin *et al.*, "Molecular clusters as building blocks for nanoelectronics: The first demonstration of cluster single-electron tunneling transistor at room temperature", Nanotechnology, vol. 13, pp. 185-194, April 2002.

[5] N. B. Zhitenev, H. Meng, and Z. Bao, "Conductance of small molecular junctions", Phys. Rev. Lett., vol. 88, pp. 226801 1-4, June 2002.

[6] J. Park *et al.*, "Coulomb blockade and the Kondo effect in single-atom transistors", Nature, vol 417, pp. 725-727, June 2002.

[7] C. P. Collier *et al.*, "Electronically configurable molecular-based logic gates", Science, vol. 285, pp. 391-394, July 1999.

[8] A. N. Korotkov, "Analysis of integrated single-electron memory operation", J. Appl. Phys., vol. 92, pp. 7291-7295, December 2002.

[9] S. Fölling, Ö. Türel, and K. Likharev, "Single-electron latching switches as nanoscale synapses", in Proc. of the 2001 Int. Joint Conf. on Neural Networks. Mount Royal, NJ: Int. Neural Network Society, 2001, pp. 216-221.

[10] Ö. Türel and K. K. Likharev, "CrossNets: Possible neuromorphic networks based on nanoscale components", Int. J. of Circuit Theory and Appl., vol. 31, pp. 34-54, January 2003.

[11] K. Likharev, A. Mayr, I. Muckra, and Ö. Türel, "CrossNets: High-performance neuromorphic architectures for CMOL circuits", in Proc. of 6th Conf. on Molecular-Scale Electronics, New York Acad. Sci., in press.

[12] Ö. Türel, I. Muckra, and K. Likharev, "Possible neuromorphic implementation of neuromorphic networks", in Proc. of 2003 Int. Joint Conf. on Neural Networks, in press.

[13] J. Hertz, A. Krogh, and R. G. Palmer, Introduction to the Theory of Neural Computation. Cambridge, MA: Perseus, 1991.

[14] S. Haykin, Neural Networks. Upper Saddle River, NJ: Prentice Hall, 1999.

[15] L. O. Chua and L. Yang, "Cellular neural networks ", IEEE Trans. on Circuits and Systems", vol. 35, pp. 1257-1290, October 1988.

[16] V. P. Roychowdhury, D. B. Janes, and S. Bandyopadhyay, "Nanoelectronic architecture for Boolean logic", Proc. IEEE, vol. 85, pp. 574-588, April 1997.