

CrossNets: Neuromorphic Networks for Nanoelectronic Implementation

Özgür Türel and Konstantin Likharev

Stony Brook University, Stony Brook, NY 11794-3800, U.S.A.

Ozgur.Turel@sunysb.edu, klikharev@notes.cc.sunysb.edu

Abstract. Hybrid “CMOL” integrated circuits, incorporating advanced CMOS devices for neural cell bodies, nanowires as axons and dendrites, and single-molecule latching switches as synapses, may be used for the hardware implementation of extremely dense ($\sim 10^7$ cells and $\sim 10^{12}$ synapses per cm^2) neuromorphic networks, operating up to 10^6 times faster than their biological prototypes. We are exploring several “CrossNet” architectures that accommodate the limitations imposed by CMOL hardware and should allow effective training of the networks without a direct external access to individual synapses. CrossNet training in the Hopfield mode have been confirmed on a software model of the network.

1 Introduction

The recent demonstration of first single-molecule transistors [1-4] gives every hope for the development, within the next decade or so, of “CMOL” integrated circuits [5] with density beyond 10^{12} functions per cm^2 . Such a circuit (Fig. 1) would combine a level of advanced (e.g., 45-nm-node [6]) CMOS fabricated by the usual lithographic patterning, a few layers of parallel nanowire arrays formed, e.g., by nanoimprinting [7], and a level of molecular devices that would self-assemble on the wires from solution. CMOL circuits allow to combine advantages of their nanoscale components (e.g., reliability of CMOS circuits and miniscule footprint of molecular devices) and circumvent their drawbacks (e.g., low voltage gain of molecular devices).

Earlier we suggested [8] a simple 2-terminal device (latching switch) that allows single-molecule implementation and might be the basis of BiWAS (binary-weight, analog-signal) synapses in several network architectures sustaining the ultimate areal density of the synapses. The further study has indicated [9], however, that in order to allow effective training, 3-terminal devices, enabling Hebbian plasticity, may be necessary. The goal of this work is to demonstrate that dense “CrossNet” structures using the 3-terminal single-electron devices, may really allow effective training, at least in the Hopfield mode, without tutor’s access to individual synaptic weights.

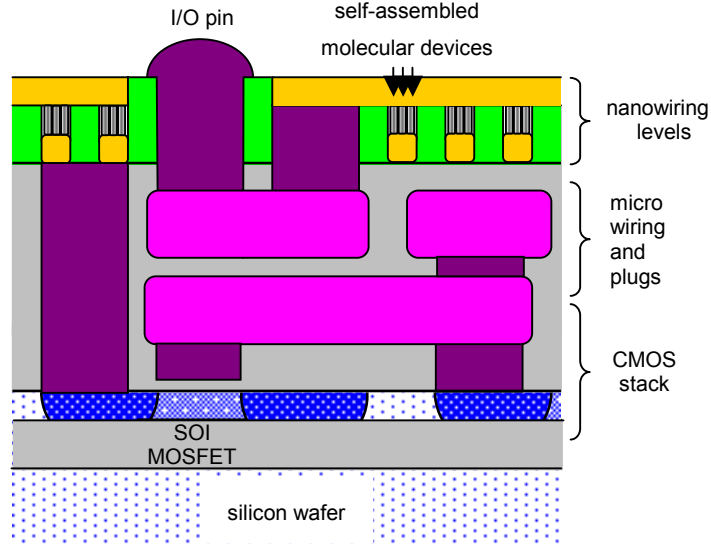


Fig. 1. The concept of CMOL (hybrid CMOS/nanowire/molecular) circuit [5]

2 Hebbian Synapse

Figure 2a shows schematics of our 3-terminal latching switch. It is essentially a combination of two well known components [5]: a single-electron transistor connecting two nanowires (modeling an axon and a dendrite, respectively), and a single-electron trap. The device and physics of the device operation are essentially the same as for the 2-terminal switch [8, 9], besides that now the signal applied to the trapping island comes equally from two sources: an axon coming from the source cell j and an axon leaving the target neural cell j' . This is why the net voltage $V = V_j + V_{j'} = (V_0/2) \times (\pm x_j \pm x_{j'})$ depends on activity $x_j(t)$ of both cells. (We accept such normalization that $|x_j|_{\max} = 1$). Due to the random character of single-electron tunneling, V only determines the rates of electron injection into the trap (Γ_{\uparrow}), opening the transistor; and its ejection from the trap (Γ_{\downarrow}), closing the transistor. The rates, in turn determines the dynamics of probability p to have the transistor open:

$$dp/dt = \Gamma_{\uparrow}(1 - p) - \Gamma_{\downarrow}p. \quad (1)$$

The theory of single-electron tunneling shows that, in a good approximation, the rates may be presented as

$$\Gamma_{\uparrow\downarrow} = \Gamma_0 \exp\{\pm(V - S)/T\}, \quad (2)$$

where Γ_0 , S , and T are constants depending on physical parameters of the synapse. (The last parameter is the effective temperature expressed in voltage units.) Despite the random character of the switching, the strong nonlinearity of Eq. (2) allows to

limit the degree of device fuzziness. In fact, solving Eqs. (1), (2) for the case when signals are applied for sufficiently long time, we get

$$p = \frac{1}{2} [1 + \tanh(V - S)/T], \quad (3)$$

so that if $V_0 - S, S \gg T$ and $|x_j| \approx 1$, we get $p \approx 1$ for $\pm \text{sgn}(x_j) \pm \text{sgn}(x_{j'}) = 2$, and $p \approx 0$ for any other combination of signal signs. Let us connect in parallel 4 switches with all possible sign combinations, sending each output current to cell j' with the same polarity as that of the feedback axonic voltage $V_{j'}$. Then we get a composite synapse with an almost deterministic weight that satisfies the “clipped Hebb rule” [13] $w_{jj'} \propto \text{sgn}(x_j x_{j'})$.

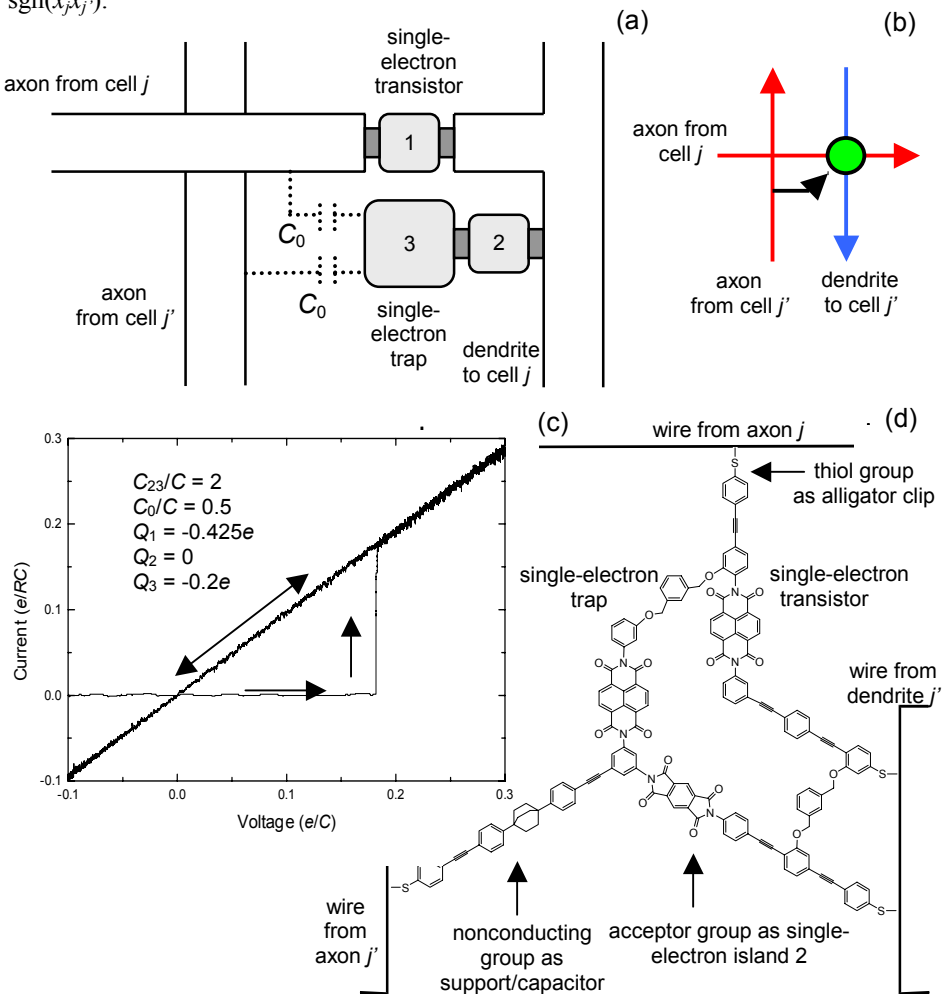


Fig. 2. Three-terminal single-electron latching switch: (a) schematics [9], (b) circuit notation used in Fig. 3, (c) result of simulation [8], and (d) possible molecular implementation [10]. C and R are the capacitance and resistance of each tunnel junction (shown dark-gray in panel a); Q_i are the background charges [5] of the single-electron islands (light gray)

3 CrossNets

We have suggested a family [9] of Distributed Crossbar Networks (“CrossNets”) in which neural cell bodies (implemented within the fire rate model, see gray cells in Fig. 3) are inserted sparsely into large, uniform 2D arrays of synapses clustered into identical “plaquettes” (green cells). Each pair of sufficiently close cells are connected in two directions, via two composite (4-switch) Hebbian synapses each way. Various CrossNets differ only by the gray cell insertion rule. In particular, in InBar (Fig. 3) the cells sit on a square lattice inclined relative to that of synaptic plaquettes. The incline enables each cell to be connected with $4M = 4/\tan^2\alpha$ other cells.

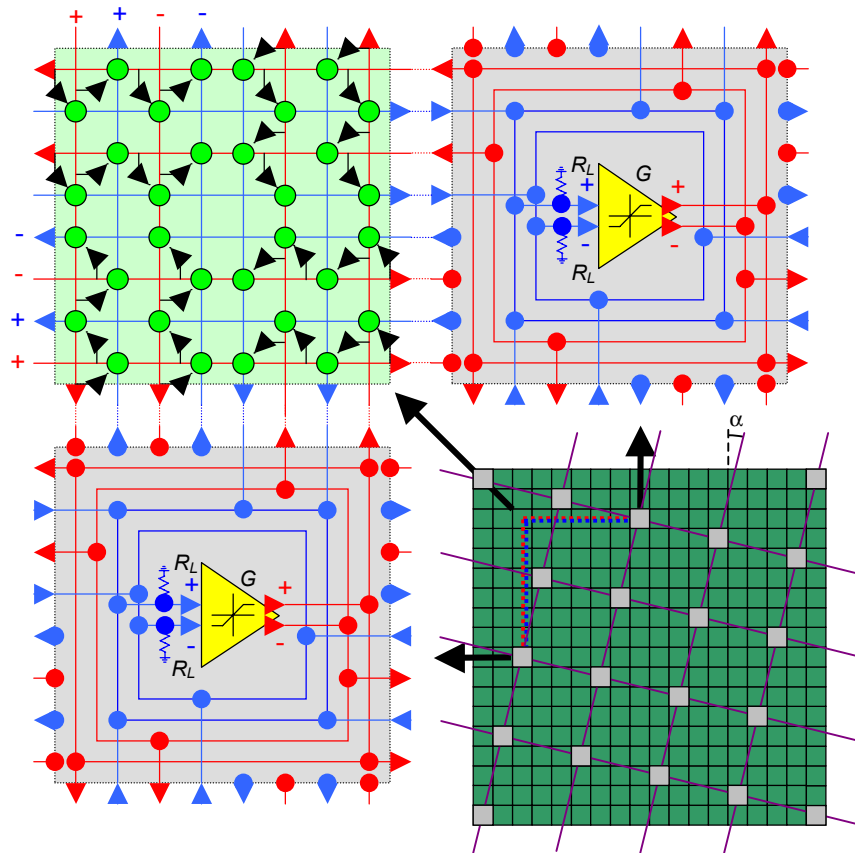


Fig. 3. CrossNet: synaptic plaquettes (green cells) and somas (gray cells) and their connection with axonic (red) and dendritic (blue) wires. Solid points on the somatic cell borders denote open-circuit terminations of the dendritic and axonic wires; due to these terminations the neural cells in the same row or column of the InBar do not interact. The bottom right panel shows the most explored CrossNet architecture, the Inclined Crossbar (“InBar”). In that panel, dashed red and blue lines show the particular connections show in the rest of the figure, while this violet lines indicate the square lattice formed by the somatic cells in InBar. These lines may be used for input/output of signals from individual cell (but not an individual synapse)

4 Hopfield Training

CrossNet training algorithms should take into account the following peculiarities of these networks:

- in CMOL-implemented networks, no external access to an individual synapse is possible (though individual somas may be accessed);
- CrossNets, by their very structure, are deeply recurrent.

These peculiarities do not allow the straightforward use of most known techniques of neural network training [11, 12], and new techniques should be developed.

We have started with the development of techniques of training InBar to operate as a Hopfield network. In the first, most straightforward method each pair $\{j, j'\}$ of cells is taught in turns. (Due to the network locality, cell pairs separated by large Manhattan distances, $r > M$, may be trained simultaneously.) For this, one of the cells of the pair

is fed with external signal $(V_0/P) \times \sum_p \xi_j^p \xi_{j'}^p$, where ξ_j^p is the j -th pixel of the p -

th image of the training set of P images, while the second cell is fed with positive signal V_0 , with $V_0 \gg T$. In this way, each of two synapses connecting the cell pair is exposed to training only once and, as described in Sec. 2 above, the probabilities of connection of its synapses are well saturated to provide virtually deterministic weights

$$w_{jj'} \approx \text{sgn} \sum_p \xi_j^p \xi_{j'}^p. \quad (4)$$

Such synaptic weights are known to work very well for fully connected Hopfield networks [11, 13]. Our numerical experiments have confirmed its efficiency for such localized networks as CrossNets. As an example, Fig. 4 shows the result of the restoration of one of three black-and-white images, being spoiled initially by flipping 40% of randomly selected pixels. In this case, the final restoration is perfect.

However, as the number of taught images P is increased, we see that the number of errors grow. For a ‘‘global’’ (fully connected) Hopfield network of N cells, the rule (4) is well known [13] to provide capacity $P_{\max} \approx 0.1N$, i.e. just ~30% below that for the continuous Hebb rule $w_{jj'} \propto \sum_p \xi_j^p \xi_{j'}^p$. In CrossNets, the synaptic connections only

extend to $4M \ll N$ cells. To our knowledge, no analytical theory had been developed for this case, but from what we knew about randomly diluted Hopfield networks (see, e.g., Ref. 14), we expected P_{\max} to be proportional to M . Our analytical estimate [15] has confirmed this expectation, giving $P_{\max} \approx [4/\pi f^2(m)]M$, where m is the average fraction of wrong bits in the restored image, and the function $f(m)$ is defined by equation: $m = \{1 - \text{erf}[f(m)]\}/2$. For $m = 1\%$, this formula gives $P_{\max} \approx 0.5M$, compatible with our numerical results.

In our second training method, the InBar matrix is considered to be partitioned to pixel panels with P cells each. Then each couple of cells from different pixels is fed sequentially by external signals according to the following rule: $V_{j,\pi} = V_0 \xi_\pi^p$, $V_{j+p,\pi'} = V_0 \xi_\pi^p$, where j is the cell number in the pixel, while π is the pixel number. In this way, each 4-group of synapses is again exposed to training only once and probabilities of

connection of its synapses are saturated to provide almost deterministic weights $w_{j,\pi}$, $j+p,\pi \approx \xi_\pi^j \xi_\pi^{j+p}$. The advantage of this method is that it does not require the external tutor system to provide multiplication of the taught signals. Our analysis and numerical experiments have shown, however, that although this method also works, it provides a lower network capacity: $P_{\max} \approx [2/f(m)]M^{1/2}$.

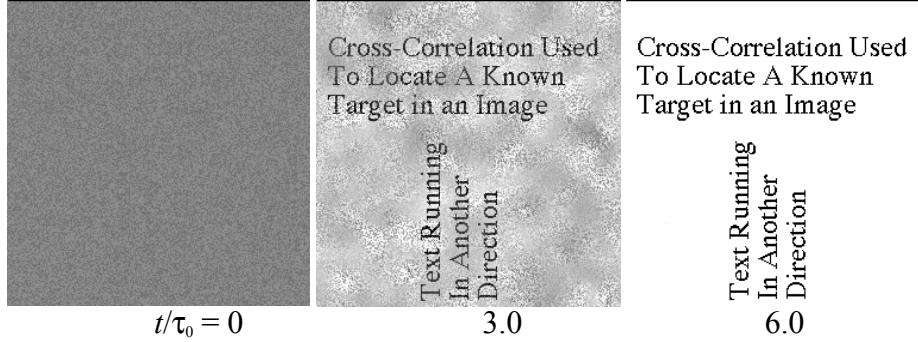


Fig. 4. The process of recall of one of three trained black-and-white images by an InBar-type CrossNet with 256×256 neural cells and connectivity $M = 64$. The initial image (left panel) was obtained from the trained image (identical to the one shown in the right panel) by flipping 40% of pixels. $\tau_0 = MR_L C_0$ is the effective time constant of intercell interaction, C_0 is the dendrite wire capacitance per one plaquette.

5 Runtime Training Prospects

The applied value of Hopfield networks is rather limited [11, 12]. Much more important would be the continuous (“runtime”) training of CrossNets as image classifiers. Our plans for such training are based on the following important feature of these networks.

Due to the signal sign symmetry of CrossNets (Fig. 3), if the latching switches are connected randomly, with probability $p = 1/2$, the average synaptic weight vanishes: $\langle w_{jj} \rangle = 0$. This means that in the absence of synaptic adaptation, the increase of the effective somatic gain $g \equiv GR_L/R$ (where G is the linear voltage gain of the somatic cell amplifier, and R_L is its load resistor – see Fig. 3), cannot lead to a global latch-up of all the cells in one of two possible saturated states. Simulations show that, similarly to the non-Hebbian case [8, 9], an increase of g above a certain threshold value $g_t \approx 1/\sqrt{M}$ leads to a self-excitation of the network. Near the threshold this excitation has a form of almost sinusoidal oscillations with a period of a few τ_0 , but at $g \gg g_t$ the activity is almost completely chaotic.

In order to train a CrossNet in runtime, we propose to set the initial shift S [see Eq. (2)] to zero. Then, according to Eq. (3), at $V = 0$ the synapse connection probability p settles to 0.5, and at sufficient gain $g > g_t$ the system goes to chaos. (If the somatic amplifiers are saturated at $V_s \lesssim T/\ln(1/\Gamma_0\tau)$, this activity as such does not generate

noticeable synaptic adaptation.) Now we insert external input signals to a subset of I ($N/M \sim I \ll N$) cells, and monitor activity of even smaller number of cells, $O \ll I$, as output signals. Most cells, not belonging to either input or output subsets, serve as a huge hidden “layer” (though this term is hardly applicable to our, deeply recurrent, networks).

As soon as the input/output combination is favorable, the external tutor increases parameter S (either globally or quasi-globally, in the vicinity of each output cell producing desirable output), so that the synaptic connection probability changes in accordance with the signal combination in this particular instant. Thus the chaotic activity should serve for driving the system through the phase space of possible signal states, enabling the tutor to pin down the favorable combinations. We hope that this method will be able to overcome the typical limitations of the firing rate models, with chaotic bursts serving for cell synchronization, similarly to spikes in biological neural systems and their integrate-and-fire models [11, 12]. The verification of this idea is our first priority for the near future.

6 Conclusions

Due to potentially low cost of chemically-directed self-assembly of single-molecule devices, CrossNets based on the CMOL technology may become the first artificial neuromorphic networks with the areal density comparable to that of the cerebral cortex, $\sim 10^7$ neurons per cm^2 , operating at much higher speed at acceptable power consumption P . (Specifically, the estimated constant τ_0 of intercell interaction is close to 20 ns, i.e. approximately 6 orders of magnitude smaller than the cortex, for the relatively high power $P = 100 \text{ W/cm}^2$. The power may be reduced with the proportional system slowdown: $P\tau_0 \approx \text{const.}$)

If created and trained to perform high-quality image classification and feature detection, these networks may create a viable market for CMOL circuits. In this case, large-scale ($\sim 30 \times 30 \text{ cm}^2$) CMOL circuits comparable in integration scale with the human cerebral cortex ($\sim 10^{10}$ cells at $\sim 10^{14}$ synapses) may become available. (In order to allow relatively rare but fast communications between its distant parts, such systems should have hierarchical organization including, as a minimum, flat CrossNet blocks connected by high-speed lines.)

Equipped with broadband sensor/actuator interfaces, such hierarchical systems may be capable, after a period of initial supervised training, of further self-training in the process of interaction with environment (read self-evolution), with the speed several orders of magnitude higher than that of their biological prototypes. Needless to say, the development of such self-evolving systems would have a major impact on all information technologies and the society as a whole.

Acknowledgments

Fruitful discussions with P. Adams, J. Barhen, V. Protopopescu, and T. J. Sejnowski are gratefully acknowledged. I. Muckra has provided great help with network simulations. A. Mayr has kindly allowed the use of Fig. 2d before publication of Ref. 10. The work was supported in part by ARDA via ONR, DOE (both directly and via ORNL), and NSF. Most numerical calculations have been carried out on *Njal* computer cluster that was acquired with a grant from DoD's DURIP program via AFOSR.

References

1. Park, H. *et al.*: Nanomechanical Oscillations in a Single C₆₀ Transistor. *Nature* **407** (2000) 57-60
2. Zhitenev, N. B., Meng, H., Bao, Z.: Conductance of Small Molecular Junctions. *Phys. Rev. Lett.* **88** (2002) 226801 1-4
3. Park J. *et al.*: Coulomb Blockade and the Kondo Effect in Single-Atom Transistors. *Nature* **417** (2002) 722-725
4. Liang, W. J. *et al.*: Kondo Resonance in a Single-Molecule Transistor. *Nature* **417** (2002) 725-729
5. Likharev, K.: Electronics Below 10 nm. To be published in: Korin, A. (ed.): *Nano and Giga Challenges in Microelectronics*. Elsevier, Amsterdam (2003). Preprint available on the Web at <http://rsfq1.physics.sunysb.edu/~likharev/nano/NanoGiga036603.pdf>
6. International Technology Roadmap for Semiconductors, 2001 Edition, 2002 Update. Available on the Web at <http://public.itrs.net/Files/2001ITRS/Home.html>
7. Zankovych, S. *et al.*: Nanoimprint Lithography: Challenges and Prospects. *Nanotechnology* **12** (2001) 91-95
8. Fölling, S., Türel, Ö., Likharev, K. K.: Single-Electron Latching Switches as Nanoscale Synapses. In: *Proc. of the 2001 Int. Joint Conf. on Neural Networks*. Int. Neural Network Society, Mount Royal, NJ (2001) 216-221
9. Türel, Ö., Likharev, K. K.: CrossNets: Possible Neuromorphic Networks Based on Nanoscale Components. *Int. J. of Circuit Theory and Appl.* **31** (2003) 37-54
10. Likharev, K., Mayr, A., Muckra, I., Türel, Ö.: CrossNets: High-Performance Neuromorphic Architectures for CMOL Circuits. Report at the 6th Conf. on Molecular-Scale Electronics (Key West, FL, December 2002), to be published by the New York Acad. Sci. (2003)
11. Hertz J., Krogh A., Palmer R. G.: *Introduction to the Theory of Neural Computation*. Perseus, Cambridge, MA (1991)
12. Dayan, P., Abbott, L. F.: *Theoretical Neuroscience*. MIT Press, Cambridge, MA (2001)
13. van Hemmen, J. L., Kühn, R.: Nonlinear Neural Networks. *Phys. Rev. Lett.* **57** (1986) 913-916
14. Derrida, B., Gardner, E., Zippelius, A.: An Exactly Soluble Asymmetric Neural Network Model. *Europhys. Lett.* **4** (1987) 167-173
15. Türel, Ö., Muckra, I. Likharev, K. K.: Possible Nanoelectronic Implementation of Neuromorphic Networks, Accepted for presentation at the Int. Joint Conf. on Neural Networks (Portland, OR, July 2003), preprint available on the Web at <http://rsfq1.physics.sunysb.edu/~likharev/nano/IJCNN03.pdf>