ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 423-428

Architectures for Nanoelectronic Neural Networks: New Results

Özgür Türel, Jung Hoon Lee, Xiaolong Ma and Konstantin K. Likharev[1]

Stony Brook University, Stony Brook, NY, 11794-3800

Abstract

Our group is developing artificial neural networks that may be implemented using hybrid semiconductor/molecular ("CMOL") circuits. Estimates show that such networks ("CrossNets") may eventually exceed the mammal brain in areal density, at much higher speed and acceptable power consumption. In this report, we demonstrate that CrossNets based on simple (two-terminal) molecular devices can work well in at least two modes: as Hopfield networks with high defect tolerance, as well as simple and multilayer perceptrons.

1. Introduction

Recent spectacular advances in molecular electronics (see, e.g., Refs. 1-3) give every hope for the electronic industry transfer, on a relatively short time scale - in 10 to 20 years - from a purely semiconductor-transistor ("CMOS") technology to hybrid semiconductor/nanowire/molecular ("CMOL") integrated circuits [4]. Such circuits may feature unprecedented density (up to 10^{12} active devices per cm^2), at acceptable fabrication costs. However, their molecular devices may have limited functionality (e.g., low voltage gain) and can hardly be assembled with a 100% yield. This is why they are more suitable for the implementation of naturally defect-tolerant artificial neural networks, than Boolean logic circuits. We have proposed [5-8] a family of ANN architectures called Distributed Crossbar Networks ("CrossNets") whose topology is uniquely suitable for CMOL implementation – see Fig. 1. CrossNet synapses are based on simple molecular devices (latching switches [5, 6]), while neural cell bodies (somas) may be implemented in the CMOS subsystem that physically underlies the molecular device level.

Our previous work was focused on CrossNets with three-terminal latching switches. In particular, we have shown [7, 8] how CrossNets of a specific ("InBar") variety, based on such switches, can be used as Hopfield networks, e.g., for fast restoration of corrupted images. The goal of this communication is to report three new important results. First, we proved that the Hopfield operation mode of CrossNets may be highly defect-tolerant, in some cases providing 99% fidelity with more than 80% fraction of bad devices - see Sec. 2. Second, the same functionality may be obtained with *two-terminal* switches (like those considered in our first work [5, 6]) that are much simpler for self-assembly than their three-terminal counterparts (Sec. 3). Finally (Sec. 4),

[1] Ozgur.Turel@stonybrook.edu, jlee@grad.physics.sunysb.edu, xiama@ic.sunysb.edu and klikharev@notes.cc.sunysb.edu

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 423-428

another CrossNets species ("FlossBar"), using a few switches per synapse, may be taught to function as either simple or multilayer feedforward perceptrons, with a very limited loss of performance in comparison with their continuous-weight prototypes.

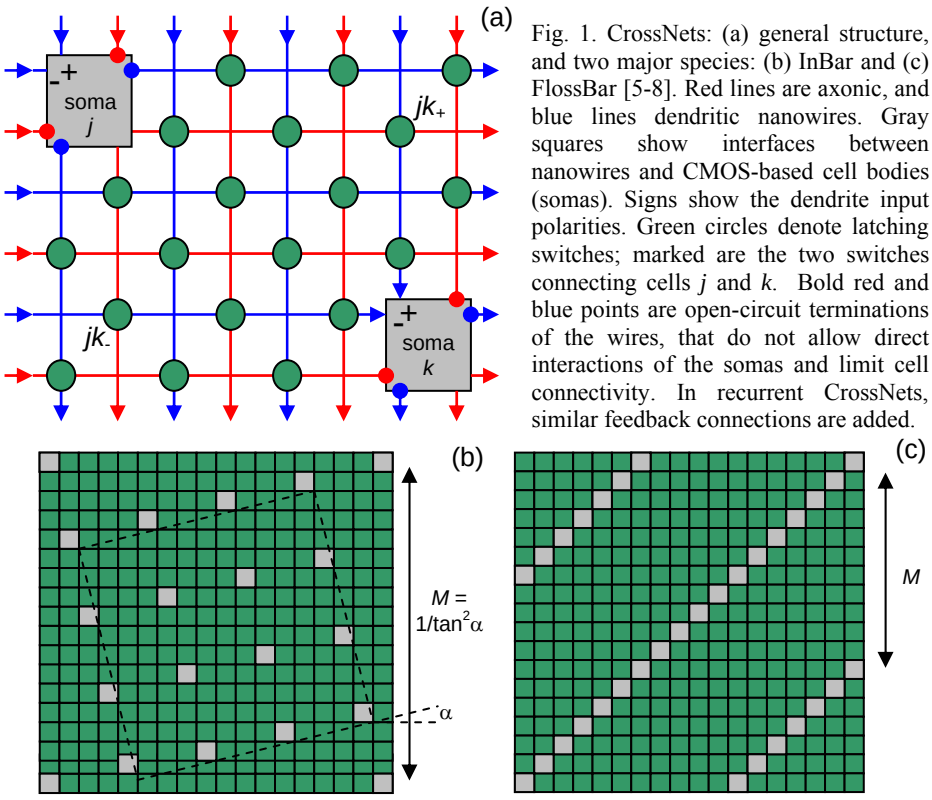

Fig. 1. CrossNets: (a) general structure, and two major species: (b) InBar and (c) FlossBar [5-8]. Red lines are axonic, and blue lines dendritic nanowires. Gray squares show interfaces between nanowires and CMOS-based cell bodies (somas). Signs show the dendrite input polarities. Green circles denote latching switches; marked are the two switches connecting cells j and k. Bold red and blue points are open-circuit terminations of the wires, that do not allow direct interactions of the somas and limit cell connectivity. In recurrent CrossNets, similar feedback connections are added.

2. Hopfield-Mode Defect Tolerance

A major challenge for CMOL CrossNet circuits is a lack of direct external access to individual molecular devices. As a result, special procedures are required for individual synaptic weight adjustment. Previously we had shown [7, 8] how this adjustment may be performed in recurrent InBar-type CrossNets. In the result of the adjustment, the CrossNet may function as a quasi-localized (finite-connectivity) Hopfield network with clipped symmetric weights. Our study have shown that, similarly to the fully-connected [9, 10] Hopfield networks, the network capacity loss due to clipping is rather marginal (~30% for 99% fidelity.)

In this work we have studied defect tolerance of this operation mode, using both the (approximate) analytical theory and numerical modeling. Figure 2 shows results obtained for a 3744-neuron InBar with connectivity parameter $M = 25$. (Each neuron is directly connected to $4M = 100$ other neurons, via two 4-switch synapses each way.) It is remarkable how resilient the network may be if the number of stored patterns P is not too close to $P_{max} \sim 0.4M$ [7]. For example, for $P = 3$ or 4, the network

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 423-428

functions reasonably well (with 99% fidelity) even in the case when almost 85% of switches are bad.

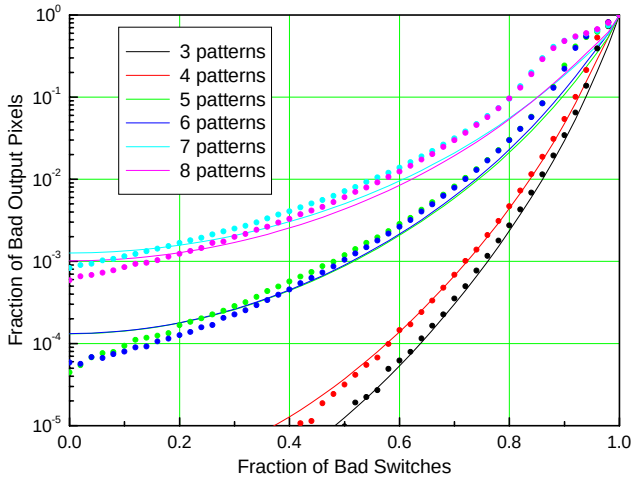

Fig. 2. The fraction of wrong output bits as a function of the fraction of disconnected latching switches. Lines show the results of an approximate analytical theory, while dots those of a numerical experiment.

3. Using Two-Terminal Devices

The main component of our previous designs [7, 8] was a three-terminal latching switch. The reason for that choice was that such switches serve as a basis of synapses with quasi-Hebbian dynamics. However, chemically-directed self-assembly of such molecules on prefabricated nanowires (the key step in CMOL circuit fabrication) with high yield presents a major technological challenge. The assembly may be much easier for two-terminal devices, possibly with the common third electrode [1-3] for global parameter adjustment.

In this work we show that CrossNets may indeed use two-terminal latching switches [5, 6] in architectures where the synaptic weight adjustment (training stage) may be separated from neural dynamics (operation stage). In fact, the ON/OFF switching rates of a two-terminal switch depend on both axonic and dendritic voltage [5-8]:

$$\Gamma_{\uparrow\downarrow} = \Gamma_0 \exp\{[\pm(V_a - V_d) - V_t]/T\}. \quad (1)$$

Here V_t is the switching threshold, while $T \ll V_t$ is the effective temperature expressed in voltage units. During the operation stage, low input resistance R_L of somatic cells may be used to reduce V_d well below V_a, thus preventing the undesirable anti-Hebbian readjustment of synaptic weights [5-8]. On the other hand, in the training mode the somatic cells (implemented with flexible CMOS circuits) may be readily re-configured to apply voltages proportional to the cell activity x_j to all nanowires, axons and dendrites alike. For example, for switches in the synapse connecting cells j and k, we may make $V_a = V_0 x_j$, $V_d = -V_0 x_k$ ($x_{j,k} = \pm 1$). At the appropriate choice of V_0 ($T \ll V_t \lesssim V_0$), only one switch of the two ($jk_\pm$ in Fig. 1) will be in the ON state, thus ensuring that the synaptic weight follows the clipped Hebb rule:

$$w_{jk} = \mathrm{sgn}\,(x_j \times x_k). \quad (2)$$

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 423-428

In particular, this property allows one to write desirable weight values into all synapses of any CrossNet with ordered geometric positions of somatic cells, e.g., InBar or FlossBar (Fig. 1b,c), from outside, row by row. For that, fixed "write enable" signals $x_j = V_0$ are sent to the horizontal wire of one switch row, "write disable" signals $x_j = -V_0$ to all other horizontal wires, while the signals $x_k = \pm V_0$, with the sign corresponding to the desirable w_{jk}, are applied to all vertical wires. (For example, in the Hopfield mode $w_{jk} = \mathrm{sgn}[\Sigma_p \xi_j^{(p)} \xi_k^{(p)}]$, where $\xi_j^{(p)}$ is the j-th pixel on the p-th pattern [10]). An elementary analysis shows that after this procedure all "selected" synapses (belonging to the selected row) will acquire the weights given by Eq. (2), while the weights of "deselected" synapses of all other rows (including those set earlier) would not be perturbed.

4. Multi-Valued Synapses

Some CrossNet species, e.g., feedforward FlossBars (Fig. 1a,c) may be used as multilayer perceptrons. Unfortunately, the information loss at synapse clipping may affect the performance of such feedforward networks as pattern classifiers more seriously than the Hopfield networks. For example, Fig. 3 shows the average error of a simple perceptron, induced by synapse clipping, i.e. rounding to the closest of L quantization levels. (We have got almost similar results for multilayer perceptrons, with effects of clipping slowly growing with the number of layers.) One can see that for binary synapses the error is above 20%, unacceptable for most applications. At the same time, an increase of the number of levels to, say, 32 makes the clipping-induced errors negligible (below 1%).

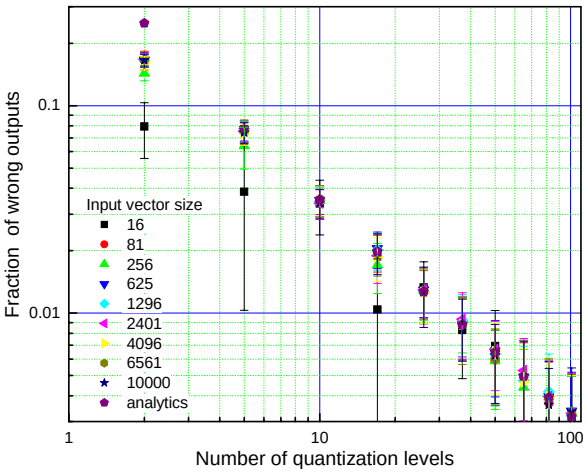

Fig. 3. Output error ε of a simple perceptron, induced by synaptic weight rounding to L discrete values. Each numerical result was obtained by averaging over 100 random input vectors. The results are described reasonably well by the simple formula $\varepsilon = 0.5 - \arctan(L-1)/\pi$, following from an approximate analytical treatment.

Such multi-valued synapses, with $L = 2n^2+1$, may be readily implemented by replacing each switch shown in Fig. 1 with a square array of $n \times n$ latching switches (Fig. 4). In the operation mode, all n axonic wires are fed with the same voltage, while the resulting currents flowing into n dendritic wires are just summed up. As a

result, the net output (post-synaptic) signal from two arrays is proportional to $w = (l_+ - l_-) /n^2$, where $l_\pm$ are the numbers of switches turned ON in each array.

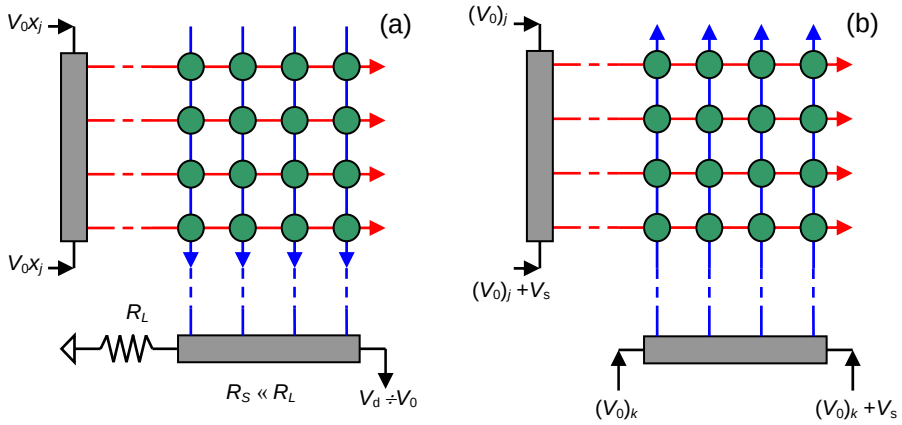

Fig. 4. A half of a composite synapse providing $L = n^2+1$ discrete levels of the synaptic weight, and its possible interfaces with somatic cell in (a) operation and (b) training mode. The dark-gray rectangles are resistive metallic strips (with total resistance $R_S \ll R_L$) serving as soma/nanowire interfaces.

In order to fix the desirable value of $l_\pm$ in each array during the weight adjustment mode, both vertical and horizontal wires are fed with graded voltages: $V_i = V_0 + V_s \times i/n$, here $0 < i < n$ is the wire number. This creates a gradient of the net voltage applied to switches and hence a domain of switches being turned on, with a boundary whose position depends on the average values V_0 of both horizontal and vertical voltages. (The boundary is inclined relative to the array edges, thus providing for a smoother $l(V_0)$ dependence.) A simple analysis, similar to that cited in Sec. 3 above, shows that the maximum possible value of the spread V_s equals $V_t/3$, and is achieved for the following values of V_0:

- write enable: $V_t/3$,
- write disable: $-V_t/3$,
- write: from $+2V_t/3$ (for $w = +1$) to $-2V_t/3$ (for $w = -1$).

The necessary voltages may be readily generated by CMOS circuitry of somatic cells, with the voltage gradient created, e.g., by simple resistive strips serving as contacts for axonic and dendritic nanowires – see Fig. 4.

5. Conclusions and Future Work

New results described in this work show that CrossNets with simple, two-terminal latching switched may work as both Hopfield networks and feedforward perceptrons, with very minor performance degradation caused by the switch conductance discreteness. For the former application, the ternary synaptic weights ($w = \{-1, 0, +1\}$) produced by just two latching switches (Fig. 1) are mostly sufficient. For the latter networks, especially used as classifiers, multi-valued synapses seem necessary. In this case the best set of continuous weights $-1 < w_{jk} < +1$ should be first generated by an

external tutor system (say, implemented on usual computers). After that, the necessary write voltage $(V_0)_j = V_a w_{jk}$ (with $V_a \approx 2V_t/3$) should be applied to dendritic lines, with the corresponding (k-th) axonic line write-enabled and other axonic lines write-disabled. While this training procedure would be limited in speed by the external tutor, the network speed in the operation mode may be extremely high. In fact, our estimates [6-8] show that CrossNets with a-few-nm nanowire pitch (limited by quantum tunneling) may generate output signals on the nanosecond scale, at acceptable power consumption.

As a result of the advances described in this communication, we believe that CrossNets may perform, at much higher speed, any function ever implemented with an artificial neural network. Moreover, since CMOL CrossNet circuits may reach very high integration scale (beyond 10^7 neurons per cm^2 even at the connectivity $\sim 10^4$), hierarchical systems based on such circuits [8] may be capable of performing much more intelligent tasks, possibly comparable with those typical for their biological prototypes. We are currently exploring this exciting opportunity.

References

[1] H. Park *et. al.*: Nanomechanical oscillation in a single C_{60} transistor. Nature, 407, 57-61 (2000).

[2] N. B. Zhitenev, H. Meng, and Z. Bao: Conductance of small molecular junctions. Phys. Rev. Lett., **88**, 226801 (2002).

[3] J. Park *et. al.*: Coulomb blockade and the Kondo effect in single-atom transistors", Nature, **417**, 722 (2002).

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

[5] S. Fölling, Ö. Türel and K. K. Likharev: Single-electron latching switches as nanoscale synapses. In: *Proceedings of the 2001 International Joint Conference on Neural Networks*, pp. 216-221.

[6] Ö. Türel and K. Likharev: CrossNets: Possible neuromorphic networks based on nanoscale components. Int. J. of Circuit Theory and Applications **31**, 37-53 (2003).

[7] Ö. Türel , I. Muckra and K. Likharev: Possible nanoelectronic implementation of neuromorphic networks. In: *Proceedings of the 2003 International Joint Conference on Neural Networks,* pp. 365-370.

[8] K. Likharev, A. Mayr, I. Muckra, and Ö. Türel: CrossNets: High-performance neuromorphic architectures for CMOL circuits. In: *Molecular Electronics III* (Annals of New York Acad. Sci., vol. 1006, 2003), pp. 146-163.

[9] J. L. van Hemmen and R. Kühn: Nonlinear neural networks. Phys. Rev. Lett., **57**, 913-916 (1986).

[10] J. Hertz, A. Krogh and R. G. Palmer, *Introduction to the Theory of Neural Computation* (Perseus, Cambridge, MA, 1991).