5th SASTech
Iran | Mashhad
May 12 - 17 | 2011
5th Symposium on Advances in Science & Technology

A New Architecture for Efficient Implementation of Neuromorphic Networks based on Nanodevice

Abdalhossein Rezai, Semnan University, a_h_rezai@sun.semnan.ac.ir
Parviz Keshavarzi, Semnan University, pkeshavarzi@semnan.ac.ir
Reza Mahdiye, Semnan University, r_mahdiye@semnan.ac.ir
Paper Reference Number: 19
Name of the Presenter: Reza Mahdiye

Abstract

This paper presents a new architecture for efficient implementation of neural network in hybrid CMOS/Nano hardware system. In this new architecture, latching switches are used in order to determine synaptic weights and each synaptic weight is implemented by just one latching switch. Using this new architecture, not only the CrossNet can be trained dynamically but also the number of CMOS transistors is decreased significantly. The results show that the proposed architecture leads to the higher speed, lower power consumption, and higher tolerance in the network compared to similar networks using other architectures. Therefore, the proposed structure has a huge potential to become an efficient implementation of neuromorphic networks based on nanodevice. In addition, the best results were achieved by implementing the proposed architecture in MLP networks

Key words: Neuromorphic network, nanodevice, crossnets, synaptic weight, nanoelectronic.

1. Introduction

Biological neural networks perform complicated information processing tasks at higher speed than conventional computers based on conventional algorithms. This has inspired researchers to look into the way these networks function, and propose artificial networks that mimic their behavior (Bishop 1995, Akazawa and Ameniya 1997). Nowadays, ANNs are being widely used in a great variety of fields such as pattern recognition, signal processing, control systems etc. To take the advantage of inherent parallelism and applicability of neural networks, some hardware implementations have been designed and made in CMOS technology so far (Akazawa and Ameniya 1997) (Turel and Likharev 2003). However, most of CMOS implementations do not provide either the speed or the complexity of a human brain. Due to high density and low power dissipation characteristics, Nanoelectronics would be a favorable candidate for developing more efficient ANNs (Likharev 2003, Strukov and Likharev 2003). However pure molecular-size integrated circuits are hardly operative because of their stability (Strukov and Likharev 2003).

In recent years, a CMOS/Nano hybrid technology, called CMOL, has been introduced. It is shown schematically in figure 1 (Snider and Williams 2007). It can provide an efficient hardware platform to build a family of neuromorphic networks.

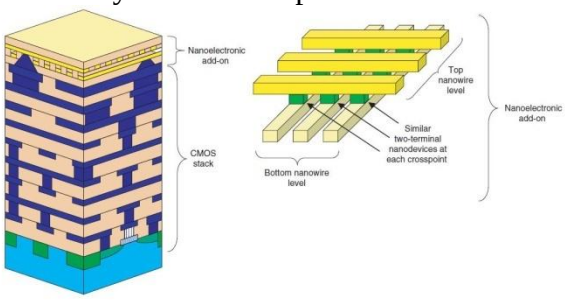

Fig. 1: Schematic of CMOL

In CMOL architecture, a crossbar of nanowires is placed on a CMOS subsystem with a rotation angle of α; which nanoswitches are placed in crosspoints of nanowires. Moreover, interfaces between the CMOS and the nanowire subsystem are provided by pins that are distributed all over the circuit area (figure 2) (Snider and Williams 2007, Folling, Turel and Likharev 2001).

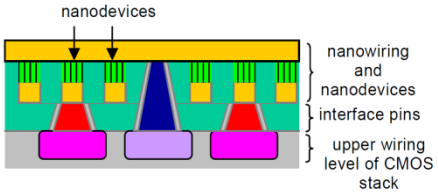

Fig. 2: Schematic side view of CMOL Interface

Since the CMOL hardware allows for very high density of nanoscale bistable devices (Turel and Likharev 2003), the use them as synapses is proposed in literatures (Likharev 2008). Hereafter Crossnets such as Inbar, Flossbar, Randbar are introduced and considered (Likharev 2008, Strukov and Likharev 2003). Thereafter, the implementation of some neural networks such as MLP and Hopfield is surveyed (Turel and Muckra 2003). Since nanoswitches have two states (On and Off), and the weights of network have a wide range, performance of these implemented networks was very low (Folling,Turel and Likharev 2001).
A solution that presented as yet, a square array of n×n nanoswitches is used for implementation of each weight. Since only ON-state switches will pass current, the total current collected in the dendritic strip will depend on the number of ON-state switches in this array, which will be arbitrary values between 0 and n^2. In this way, it is easy to extend bipolar values such as 0*; 1; 2 …;* n^2-1*;* n^2 using an array of $n \times n$ (Folling ,Turel and Likharev 2001).
The rest of this paper organized as follows: section 2 introduces the nanoswitch. Crossnet concepts and its types are described in Section 3. The proposed CMOL architecture for neural network is presented in Section 4. Section 5 discuss about the results. Finally conclusion is given in Section 6.

2.Nanoswitch

The structure of a single electron nanoswitch, which may be used as synapses, is shown in figure 3. This nanoswitch includes two parts: 1) A single electron transistor that connects vertical and horizontal nanowires; and 2) a trap (Strukov and Likharev 2003, Lee and Likharev 1999). Conductance of this device for small applied source–drain voltages might be very low (the 'Coulomb blockade' state), unless the blockade is lifted by an electric field generated by trap (Lee and Likharev 1999, Lee 2007).

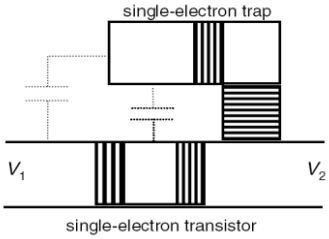

Fig. 3: Single Electron Transistor

If the voltage is low, the trap in equilibrium has no extra electrons and its total electric charge Q=ne is zero. As a result, the transistor remains in the Coulomb blockade state. Subsequently, input and output wires are essentially disconnected. If V is increased beyond a certain threshold, Vinj (which should be lower than the Coulomb blockade threshold voltage, V_t, of the transistor), one electron is injected into the trap: n→1. In this charge state the Coulomb blockade in the transistor is lifted keeping the wires connected by a finite varier resistance (Lee 2007). Figure 4 demonstrate V-I characteristic of nanoswitch.

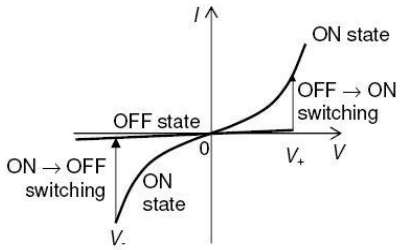

Fig. 4: I-V Curve of a nanoswitch

3. Crossnet

Uniquely structure of CMOL circuits makes them suitable for the implementation of more complex, mixed-signal information processing systems, including ultradense and ultrafast neuromorphic networks (Likharev 2003). For implementing neural network in CMOL, the neural cell bodies (somas) are implemented in CMOS subsection and the nanodevices with reconfigurable capabilities can be used as synapses. The dendrites and axons are implemented by interconnects or nanowires. Accordingly, somas can be connected to each other through the nanowires and nanoswitches. Such netwoks in CMOL called CrossNet (Turel and Muckra 2003, Turel and Likharev 2004). Figure 5 shows the general schematics of a CrossNet. The neuron somas (shown as gray squares) are essentially nonlinear amplifiers implemented as CMOS circuits. The amplifiers feed axons and are, in turn, fed by dendrites (dark lines) which are implemented as similar metallic nanowires (Turel and Likharev 2004).

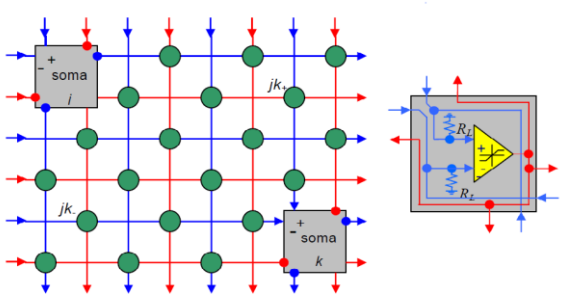

Fig. 5: Schematic of Crossnet

Both axonic and dendritic nanowires have open-circuit terminations on one end, excluding the direct connections between any two neurons located on the same row or column. Due to these terminations, neurons can communicate only through the synapses. CrossNet properties depend, to a certain extent, on the way somas are distributed over the array. Two different neuron distribution patterns, InBar and FlossBar, are shown in figure 6-a and 6-b, respectively (Turel and Muckra 2003, Turel and Likharev 2004).

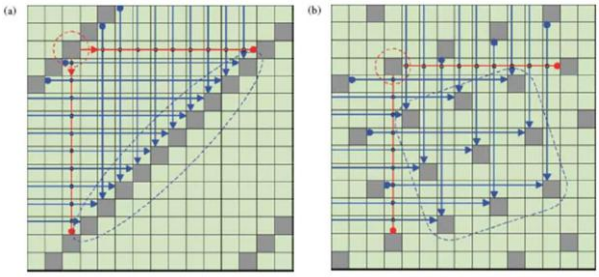

Fig.6: Crossnet types a) Flossbar b) Inbar

The main points in CrossNets are the somas which connected through nanoswitches. If the nanoswitch is in its ON state, some current flows into the latter wire, and charge it. Since such currents are injected into each dendritic wire through several open synapses, their addition provides a natural passive summation of signals from the corresponding somas, required for all neuromorphic networks:

$$x_i = \sum \omega_{ik} . y_k \qquad (1)$$

If the weights of network have two values (one and zero), by using a nanoswitch and assigning each value to one mode (Off mode for zero and ON mode for one), the weights could be implemented simply and the network works correctly. In this case, each nanoswitch is equivalent for any weight. Nevertheless if the weights be in a wide range, the performance would be decreased extremely (Likharev 2003, Turel 2007).

Lee and likharev proposed an array of n×n nanoswitches for implementing the weights (Lee and Likharev 1999). Figure 7 shows this idea of multi-valued synapses. *n* nanowires are attached to a metallic strip, which serves as an effective interface between the CMOS cell and nanowires. According to this method, range of the weights is divided into n^2 levels and the weights are quantized to these levels as each level denotes number of nanoswitches that

should be ON. Note that, the information loss at clipped synapse may affect the performance of networks more seriously (Turel 2007).

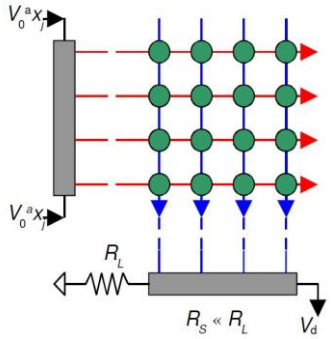

Fig. 7: A composite synapse with n^2 +1 weight values

4. The proposed Architecture

In this paper we use nonlinear I-V Characteristic of nanodevice to control nanodevice current in order to resolve the information loss at clipped synapse. We propose using nonlinear I-V curve of nanoswitch by controlling its voltage. Besides, we can implement synapses of network. Table 1 shows the values that extracted from nonlinear I-V curve of nanodevice (Likharev 2003).

V(v)	-0.4	-0.3	-0.2	-0.1	0	0.1	0.2	0.3	0.4
I(μA)	-0.2	-0.1	-0.044	-0.02	0	0.02	0.043	0.118	0.2

Table 1: Values of voltage and current of nanoswitch

As it is shown in figure 4 and table 1, current of the nanodevice depends on the voltage drop across it. We use nanodevice based uniquely and nonlinear characteristic to determine synaptic weight by one nanoswitch. The proposed architecture is shown in figure 8.

In the proposed architecture, Axons and dendrites are implemented as nanowires, and synapses at cross-points between axonic and denderitic nanowires. Neural cell bodies (somas) which implemented in previous architectures in the CMOS subsystem, in the proposed architecture implemented in nano and CMOS subsystem simultaneously.

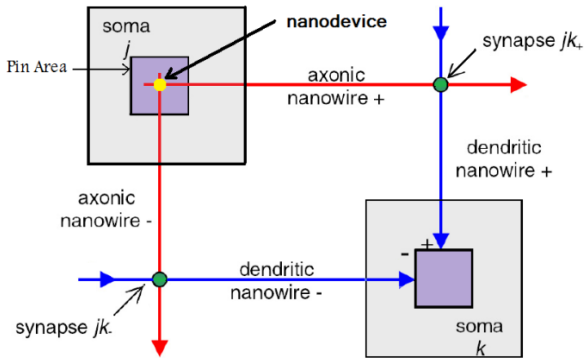

Fig. 8: Proposed archtecture

The proposed architecture is replacing a nanoswitch to determination synaptic weights, whereas the synaptic weights are implemented by nanodevices and each synaptic weight implemented by one latching switch. Therefore the CrossNet can be trained dynamically and also number of transistor in CMOS can be decreased.
Since the synaptic weights can be determined by nanoswitches, current and proportional voltage drop across the nanoswitch are depending on each other. Consequently implementation of synaptic weight in the proposed architecture is as follows: in first according to the weight, current of nanoswitch is determined. After that, according to the table 1 nanoswitch voltage is specified and finally specified voltage is dropped across the nanoswitch.
The proposed architecture is appropriate for implementation of MLP networks. In next section we compare the performance of MLP crossnet in our proposed architecture and architecture in (Turel and Likharev 2004).

5. Evaluation

In order to evaluate the performance of the proposed architecture, we analysis speed, power consumption and defect tolerance. The components that affect the performance of the circuit include the nanoswitches, the nanowire, and the pin-to-nanowire contact (pins interface CMOS and nanowire).
According to (Gao and Hammerstrom 2007), the time delay from the input pin to the output pin through the nanowires and nanoswitches as

$$\tau = (2R_{con} + 1.5R_{wire} + R_{on} / D)C_{wire} \quad (2)$$

Where R_{con} is the pin-to-nanowire contact resistance; R_{wire} is the ON mode nanowire resistance; R_{on} is the nanoswitch resistance; D is the number of nanoswitches which connected as parallel and C_{wire} is the nanowire capacitance.
According to (2), since Rcon and R_{on} in the proposed architecture is less than multi-value weights, the delay time of the proposed architecture is less than architectures that used multi-value weights such as (Likharev 2003, Turel 2007). So the speed of the proposed architecture is better than other architectures.
In addition, according to (Gao and Hammerstrom 2007), For CMOL crossbar arrays, the static power consumption include both the working power and the leakage power. A working "ON" power is due to the "ON" nanoswitches, and is given by

$$P_{on} = \alpha\beta NMV^2 / (2R_{con} + R_{wire} + R_{on} / D) \quad (3)$$

Where α is the average probability that the driving voltage to the input nanonwire is high (voltage on the nanoswitches is over V_t); β is the probability that the nanoswitches are "ON"; and N and M are the horizontal and vertical nanowire counts, respectively.
The leakage power is given by:

$$P_{leakage} = \alpha(1-\alpha)NMV^2 / (2R_{con} + R_{wire} + R_{off} / D) \quad (4)$$

According to (4) because N and M in proposed architecture is one, the power consumption of the proposed architecture is more less than architectures that used multi-value weights such as (Likharev 2003, Turel 2007).

In addition, we have studied the defect tolerance of this architecture, using MNIST data. Figure 9 shows the the wrong pixel fraction for a 3744-neuron InBar as a function of *q*. When the number of stored images is smaller than the capacity of the InBar, its performance is only gradually degraded until 80% defects.

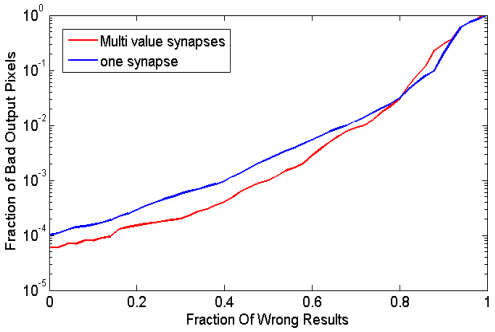

Fig. 9: Defect tolerance of a recurrent InBar

6. Conclusion

In this paper we proposed a new architecture for neural network implementation in CMOL nanoelectronic. Due to this, proposed architecture is introduced and after that special characteristic including speed, power consumption and defect tolerance was analyzed. According these results the proposed architecture in speed and power consumption is more efficient than other architectures.

References

Bishop, C., (1995), *Neural Networks for Pattern Recognition*. Oxford U.

Akazawa, M.and Ameniya, Y., (1997), "Boltzmann Machine Neuron Circuit Using Single-Electron Tunneling", *Appl.Phys. Lett*. vol. 70, pp. 670-673.

Ramacher, U., (1993), "Multiprocessor and Memory Architecture of the Neurocomputer synapse-1", *In World Congress on Neural Networks*, pp. 31-45.

Folling, S., Turel, O., Likharev K., (2001) "Single-electron latching switches as nanoscale synapses". *In Proc. 2001 IJCNN, Mount Royal* pp. 216–220.

O.Turel and K.K. Likharev., (2003) "Crossnets: Neuromorphic Networks for Nanoelectronic Implementation" *Lecture Notes in Computer Science*, NY, pp.753-760, 2003

Likharev, K.K., (2008) "Hybrid CMOS/nanoelectronic Circuit: Opportunities and Challengies" *J. nanoelectronic, Optoelectronic*, vol 3, pp. 203-230.

Likharev, K.K., (2003). "Nano and Giga Challenges in Microelectronics, chapter Electronics below 10 nm", . *Elsevier* pp. 27–68.

D.B. Strukov, D.B., and Likharev K.K,. (2005), *CMOL: Devices, circuits, and architectures*, Springer.

Snider G., and Williams, R., (2007) "Nano/CMOS architectures using a field-programmable nanowire interconnect," *Nanotechnol*, vol. 18, pp. 1–11.

Folling, S. , and Turel, O., and Likharev, K.K., (2001) "Single-electron latching switches as nanoscale synapses", *In Proc. IJCNN*, pages 216–220,.

Likharev, K.K., Mayr, A. and Muckra, I. and Turel. O. (2003) "Crossnets: High performance neuromorphic architectures for CMOL circuits" *Acad. Sci*, 1006:146-163.

Turel, O. and Muckra, I. and Likharev K.K.., (2003) " Possible Nanoelectronic Implementation of Neuromorphic Networks" *In Proc. IJCNN03*, pages 365-370.

O.Turel and K.K. Likharev., (2004). "Architecture for Nanoelectronic Implementation of Artificial Neural Network." *In Proc Elsevier Science*, pp 65-80.

Lee, J.H., and Likharev. K.K., (2006), "In situ training of CMOL crossnets" In *InProc. Of WCCI/IJCNN'06*, pp. 5026-5034.

See, and K. Likharev, (1999), "Single-Electron Devices and Their Applications" *Proc. of IEEE*, vol. 87, pp. 606-632.

Lee, J.H. (2007), "CMOL CrossNets as Defect-Tolerant Claassifiers" *PhD Thesis, Stony Brook Univ.* , pp.26-34.

Turel, O., (2007) "Devices and Circuits for Nanoelectronic Implementation of Artificial Neural Networks", *PhD Thesis, Stony Brook Un*, pp.29-32.

Masoumi, M., Raissi, F., Ahmadian, M. and Keshavarzi, P., (2006), "Design and evaluation of basic standard encryption algorithm modules using nanosized complementary metal-oxide-semiconductor-molecular circuits" *Nanotechnology*, 17(1):89–99.

Gao, C. and Hammerstrom D., (2007) "Cortical Models On to CMOL and CMOS-Architectures and Performance/Price " *IEEE Transactions on circuits and systems*, vol. 54, num. 11, pp 2502-2515.