Neural Networks 24 (2011) 1136–1142

Contents lists available at SciVerse ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Scaling-efficient in-situ training of CMOL CrossNet classifiers

Jung Hoon Lee *

Department of Physics and Astronomy, Stony Brook University, Stony Brook, NY, USA

ARTICLE INFO

Article history:
Received 24 December 2010
Received in revised form 15 June 2011
Accepted 17 June 2011

Keywords:
Hybrid CMOS/nanoelectronic
Neuromorphic
Defect-tolerant classifier

ABSTRACT

CMOL CrossNets, hybrid CMOS/nanoelectronic neuromorphic circuits, may open up exciting opportunities to build artificial intelligence similar to the brain. However, limited functionality of nanodevices used in CMOL circuits causes significant challenges to train CrossNets with the usual algorithms. In order to overcome these challenges, we developed an in-situ variety of the error backpropagation method for supervised training of CrossNet-based pattern classifiers. Although this algorithm successfully trained CrossNets to perform simple benchmark classification tasks in Proben1, we found that it did not scale up to larger problems such as the MNIST dataset. Therefore, we propose an alternative in-situ method, combining training with the hidden layer build-up. Simulated results suggest that our new in-situ approach is appropriate to train CrossNets to perform classification on practical problems.

1. Introduction

The development of hybrid CMOS/nanoelectronic integrated circuits (DeHon & Likharev, 2005; Heath, Kuekes, Snider, & Williams, 1998; Kuekes, Snider, & Williams, 2005; Likharev, 2008; Strukov & Likharev, 2005; Türel, Lee, Ma, & Likharev, 2004), particularly those of the so-called CMOL (CMOS/MOLecular) variety (Likharev, 2008; Strukov & Likharev, 2005), provides an opportunity to advance not only digital microelectronics (DeHon & Likharev, 2005; Heath et al., 1998; Kuekes et al., 2005), but also mixed-signal neuromorphic networks (Lee & Likharev, 2005, 2006, 2007; Ma & Likharev, 2007; Strukov & Likharev, 2005; Türel et al., 2004). In neuromorphic networks, named CrossNets, a CMOS subsystem is used to build neural cell bodies ("somas"), whereas a nanoelectronic subsystem (a nanowire crossbar with simple two-terminal devices at every crosspoint) implements synaptic communications between somas (Fig. 1).

The crosspoint devices, playing the role of elementary binary synapses (in Fig. 1, shown with circles), function as latching switches. In other words, they have two-internal states: an ON state with finite conductance, and an OFF state in which the current through the device is negligibly small. Recently, reproducible fabrication of such devices based on metal oxides (Chen et al., 2005) and chalcogenide materials (Chen et al., 2006) has been reported. Taken together with the fast progress of nanowire crossbar fabrication (Green et al., 2007; Jung et al., 2006), such advances lead us to believe that the experimental demonstration of CMOL circuits is imminent (for a recent review, see Likharev, 2008). Indeed, Xia et al. (2009) were able to fabricate small-scale hybrid circuits consisting of a CMOS subsystem and a nanowire grid. In CrossNets, voltages applied by pre-synaptic cells through the "axonic" nanowires induce currents through all ON-state latching switches. Passive summation of these currents in each "dendritic" nanowire creates the necessary input for each post-synaptic soma:

$$y_i = \sum_{j=1}^{M} w_{ij} x_j. \tag{1}$$

The operating speed of CrossNets depends mostly on how fast nanowires can be charged. Our earlier estimates of charging time constant predicted that the conduction delay of signals between CMOS cells is of an order of magnitude of 100 ns when a cell is connected to 1000 pre-synaptic cells (for detailed estimates, refer to Lee & Likharev, 2007; Türel et al., 2004). This delay is three orders of magnitude shorter than the conduction delay of AMPA synapse with the rise time constant being ∼0.5 ms. Consequently, CrossNets could outperform the mammalian cortex at least in terms of the signal propagation speed. Gao and Hammerstrom (2007) explored four different potential hardware implementations of both non-spiking and spiking neural networks. For non-spiking neural networks, their analysis suggested that the CrossNets could be comparable to biological systems in terms of densities and speeds. When it comes to spiking neural networks, CrossNets are denser, faster and more power-economic than the digital CMOS implementation unless the connectivity occurs sparsely. Providing sparse 0.1% connectivity, they found CrossNets are advantageous due to lower power consumption. The more recent study of Zaveri and Hammerstrom (2010) indicated that CrossNets provide the best performance/cost ratio for cortex-scale hardware.

* Correspondence to: Center for BioDynamics Boston University, 111 Cummington Street, Boston, MA 02215 USA. Tel.: +1 603 545 2886.
E-mail address: giscard88@gmail.com.

0893-6080/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2011.06.015

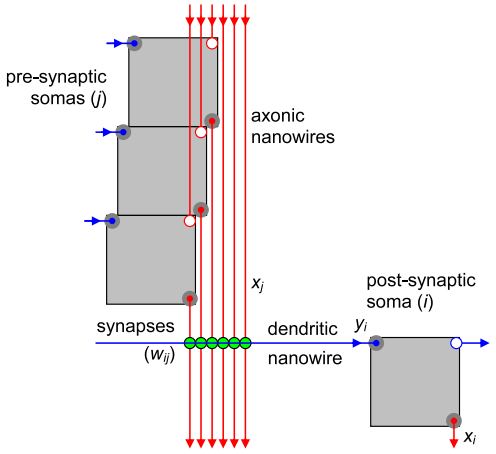

Fig. 1. *The simplest (feedforward, binary-synapse) CrossNet*. The axonic nanowire (shown as red) and dendritic nanowires (shown as blue) are physically similar. For clarity, only a fraction of M the elementary synapses (latching switches) contributing to one post-synaptic signal is shown (by green circles). The devices, in fact, form a continuous 2D array covering the chip area including CMOS-implemented somatic cells (shown as gray). Open points indicate the open-circuit terminations of axonic and dendritic lines. Due to these terminations, somas do not communicate directly (but only via synapses); they also limit the nanowire segment lengths and hence the cell connectivity M. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

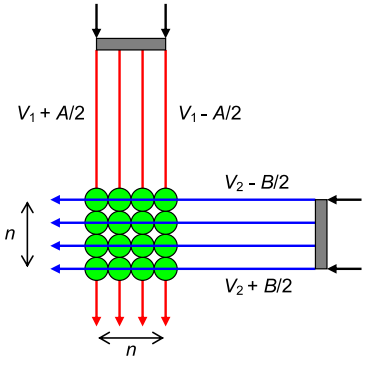

Fig. 2. *The multi-valued synapse of a CMOL CrossNet*. Gray rectangles show resistive strips connecting n nanowires. Manipulating applied voltages V_1 and V_2, the synaptic weight may be adjusted to any of the values $w = w_{\max}(l/n^2)$, where l is an arbitrary integer between 0 and n^2 (Lee & Likharev, 2007).

However, CMOL hardware imposes substantial limitations on CrossNet training procedures. Major challenges occur due to the following;

- Only binary synaptic communications are available through nanodevices (bi-stable latching switches), but most learning algorithms require continuous synaptic weights.
- The yield of nanodevice fabrication is low. In other words, the rates of defective devices would be high.

Our group has developed the training algorithms to overcome these challenges. Interestingly, their effects are not significant for Hopfield-type networks. The capacity of a CrossNet-based Hopfield network dropped not considerably (e.g., by ~30% at 99% fidelity) from that of a conventional Hopfield network with continuous synaptic weights (Türel et al., 2004), even though only binary synapses deliver intercommunications among (CMOS) somas. However, binary synapses imposed substantial penalty on fidelity of the feedforward network, Multi-Layer Perceptrons (MLP). We found that the degraded feedforward network performance could be restored if multi-valued synapses, arrays of $n \times n$ nanodevices shown in Fig. 2, are implemented on a CMOL chip (Lee & Likharev, 2007; Türel et al., 2004).

CrossNet-based MLPs can be trained for classification by importing weights, previously calculated by an external tutor, into a CMOL circuit (Lee & Likharev, 2007; Likharev, 2008). This weight import procedure seems satisfactory for certain types of important tasks, for example, ultrafast recognition of a particular face in a large crowd (Lee & Likharev, 2005). However, for some large-scale applications, it may be impracticable to build and train such external tutor networks. Therefore, we developed (Lee & Likharev, 2006, 2007) a method for self-contained ("in-situ") training of CrossNet pattern classifiers. Although the method successfully trained CrossNets to perform simple classification tasks in Proben1 (Prechelt, 1994), its fidelity for a larger classification problem, the MNIST dataset (LeCun, Bottou, Bengio, & Haffner, 1998) of 60,000 handwritten-digits, has turned out to be quite poor (see Appendix).

The aim of this work is to suggest and explore an alternative method of in-situ training of CMOL CrossNets. In Section 2, for reader's convenience, we briefly summarize the basic idea of the in-situ synapse adaptation rule (Lee & Likharev, 2006), In Section 3, we discuss two forms of novel constructive algorithms developed for training CrossNets. In Section 4, we present simulation results, which suggest that CrossNets can be trained with in-situ algorithms to perform complex classification. The classification fidelity of CrossNets trained with 10,000 examples was comparable to the error rate 4.5% reported by LeCun et al. (1998), and it became even better with more training examples. In Section 5, we summarize our results and address future plans.

2. In-situ adaptation rule

Fig. 3 demonstrates our idea (Lee & Likharev, 2006) of in-situ adaptation of nanodevices at cross-points using the "stochastic multiplication" (in a different form, first discussed by Kondo and Sawada (1992)). Each resistive strip connecting n nanowires is biased by the voltage V_i, the output of a comparator fed by signal $\pm x_i$ and a random signal REF_i with a uniform probability distribution within the range $[0, x_{\max}]$. If the signal is larger than the current value of REF_i, the comparator's output is constant: $V_i \cong (V_{th}/2)\mathrm{sgn}(\pm x_i)$. If not, it is zero. V_{th}, the threshold voltage, triggers a nanodevice to switch its state from one to another (Likharev, 2008; Strukov & Likharev, 2005). As a result, the state of the nanodevice can be switched only when both comparators (C_1 and C_2) provide such outputs into the same nanodevice. The probability of the nonzero output is clearly proportional to x_i, and hence the probability of having nanodevices to be updated is

$$P_\Gamma = P_1 P_2 = \frac{x_1 x_2}{x_{\max}^2}. \tag{2}$$

In order to keep the weight change proportional to $\mathrm{sgn}(x_1 x_2)$, it is necessary to address each of 4 switch arrays of the composite synapse (Fig. 3) independently. The simplest way is to use a 4-step time division multiplexing procedure with a global shift $S(t) = \pm S_0$ added to all nanowires. In each step, the signs of global shift S_0 and x_i alternate according to the rule shown in the table inset in Fig. 3. We showed that if the period t of time division multiplexing (with random values of REF_i at every period) is much shorter than the characteristic time τ of nanodevice switching, the resulting average weight change follows a Hebbian-like rule;

$$\langle \Delta w \rangle = \eta x_1 x_2 \begin{cases} (w_{\max} - w) & \text{for } x_1 x_2 > 0, \\ (w_{\max} + w) & \text{for } x_1 x_2 < 0 \end{cases} \tag{3}$$

where $\eta = \Delta t/\tau$ is the effective learning rate.

As the error backpropagation training rule (Hertz, Krogh, & Palmer, 1991) is based on the weight change proportional to a product of two signals, this stochastic multiplication described by Eq. (3) may be used to implement a backpropagation-like