

Subthreshold MOS Implementation of Neural Networks with On-Chip Error Back-Propagation Learning

Yoon Kyung Choi and Soo-Young Lee

Computation and Neural Systems Laboratory, Department of Electrical Engineering

Korea Advanced Institute of Science and Technology

373-1 Kusong-dong, Yusong-gu, Taejon 305-701, Korea (South)

Tel: +82-42-869-3431 / Fax: +82-42-869-3410 / E-mail: sylee@ee.kaist.ac.kr

Abstract

Subthreshold analog circuits for MOS implementation of artificial neural networks are presented with on-chip learning capability. Each synapse circuits consist of a storage capacitor and 3 analog multiplier, i.e. one for signal feed-forward, one for outer-product synaptic weight adjustments, and one for error back-propagation. While all the 3 multipliers are used for error back-propagation learning, only the first 2 multipliers are used for Hebbian learning. Each neuron circuits are composed of a sigmoid circuit and a sigmoid derivative circuit, which show near ideal sigmoid characteristics and provide external gain-control capability. All the circuits incorporate modular architecture, and are designed to increase numbers of neurons and layers with multiple chips. Also the subthreshold operation provides low power consumption and large scale implementation.

Introduction

Although neural networks is capable of solving complicated pattern recognition and adaptive control problems, special hardwares are required to fully utilize its inherent massive parallelism. There have been two approaches on neural hardware developments, with and without on-chip learning capability. The first approach usually assumes that adaptive learning has been done by other hardwares, probably by conventional von Neumann computers. However the training of neural networks for very complicated applications with large number of neurons, where neural networks has potential advantages over conventional algorithms and special hardwares are worthwhile to build, requires enormous computational capability and there exists no other hardware available. There have been approaches to use conventional hardwares for calculation of the synaptic weight adjustments while the neural hardwares perform signal feed-forward and possibly error back-propagation. However computational requirement for the calculation of synaptic weight adjustment is of the same order with that of signal feed-forward or error back-propagation, and the neural hardwares can not improve the overall training speed much. Neural hardwares with on-chip learning capability is essential for practical large-scale implementations.

Recently several attempts have been made to put on-chip learning capability on the neural hardwares. Both digital and analog circuits have been developed. For high density and speed we use analog circuits. However, unlike other developments [1], subthreshold operation on MOS circuits is utilized to provide low power consumption and higher density synapses. Also outer-product learning algorithm is chosen for popular error back-propagation and Hebbian learning rules. Without time-multiplexing implementable number of neurons on a chip is always limited by numbers of pads and package pins, and modular architecture is required to provide capability to make larger systems with multiple chips. In this paper we present a modular analog neuro-chip with subthreshold operation and outer-product learning.

System Architecture and Circuits

Two basic components of our neuro-chip is shown in Fig.1. As shown in Fig.1(a), each synapse is composed of a voltage storage device and 3 analog multipliers, one for signal feed-forward, one for outer-products for calculation of synaptic weight adjustments, and one for error back-propagation. For neural network models with error minimizing learning algorithms such as (multilayer) Perceptron the outer products are done between input neural activation and output errors. For Hebbian neural networks with unsupervised learning

algorithms the outer products are done between neural activations of both input and output layers. Each neuron in Fig.1(b) consists of a sigmoid circuit, sigmoid derivative circuit, and an analog multiplier. The multiplier is used to get multiplication of the back-propagated error and sigmoid derivatives.

In synaptic circuits signals are coded as differential currents, which offer high dynamic range and common-mode noise immunity. The analog multipliers are based on four-quadrant Gilbert multipliers with differential input and output currents. [2,3] However synaptic weight is represented as node voltage and adjusted by currents during learning. Capacitors are selected as synaptic weight storage device for easy fabrication. Charge loss due to leakage currents may be considered as a part of continuous learning, or should be compensated by periodic refreshing.

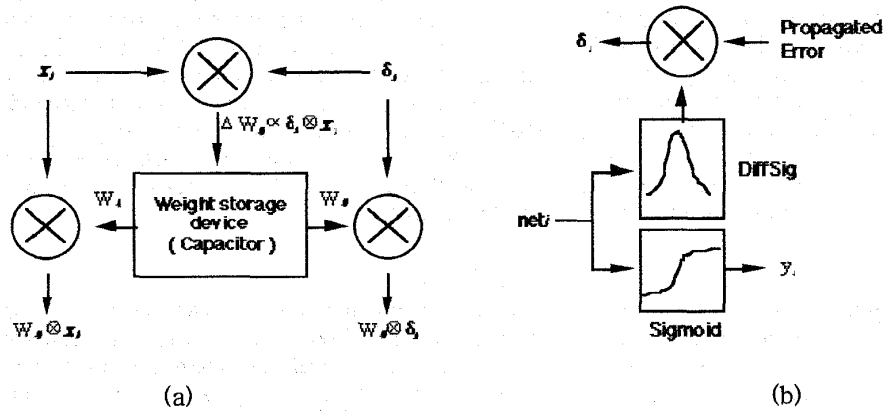


Fig.1 Basic components of the neuro-chip. (a) synapse; (b) neuron

Current outputs from the synapses are summed at the input node of corresponding neuron circuits. The range of this summed currents depends upon number of connected neurons through synapses, and need to be scaled to a fixed value for modular architectures with possibly variable number of neurons by multiple chips. A scaling circuit is shown in Fig.2. The differential output currents of each synapse satisfies $i_j^+ + i_j^- = I_b$ and $i_j^+ - i_j^- = W_j x_j$. By the principle of translinear array normalizer [4] I_j^+ and I_j^- in Fig.2 are given as

$$I_j^+ = \frac{i_j^+}{\sum_j (i_j^+ + i_j^-)} I_{const} = \frac{i_j^+}{NI_b} I_{const}, \quad I_j^- = \frac{i_j^-}{NI_b} I_{const}. \quad (1)$$

The scaled net output value $I_{net}^+ - I_{net}^-$ now becomes $\sum_j W_j x_j \frac{I_{const}}{NI_b}$.

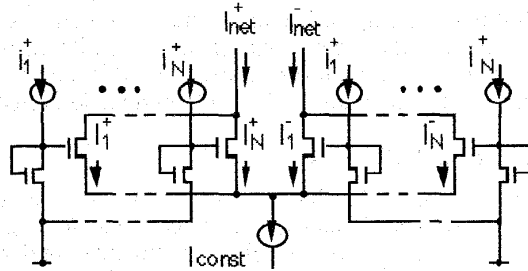


Fig.2 Scaling circuits

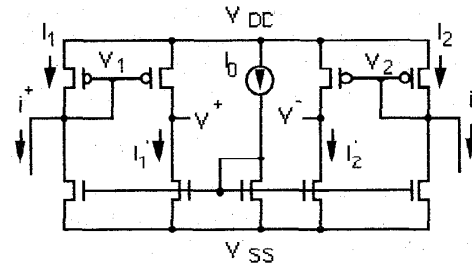


Fig.3 Current-to-voltage converter

Since the inputs of the neuron circuits in Fig.1(b) are differential voltages, the current outputs of the scaling circuits should be converted to differential voltages. As shown in Fig.3, provided $I_o \gg i^+$ and $I_o \gg i^-$ were satisfied, the output differential voltage becomes

$$V^+ - V^- = \frac{1 + \lambda_p V_{DD}}{I_o(\lambda_p + \lambda_n)} (i^+ - i^-). \quad (2)$$

Unlike the other circuits this converter circuit is biased to saturation region, and the currents magnitude conditions are easily satisfied. As shown in Eq.(2) the converted voltage is inversely proportional to the current I_o , and provides a way to effectively control gain of the sigmoid function.

The circuits for the sigmoid function and its derivative are shown in Figs. 4 and 5, respectively. Both circuits provide hyperbolic tangent and its derivative up to the first order approximation [2,5], and their characteristics are shown in Figs. 6 and 7. Several lines in Fig.6 show sigmoid functions with different gains controlled by I_o .

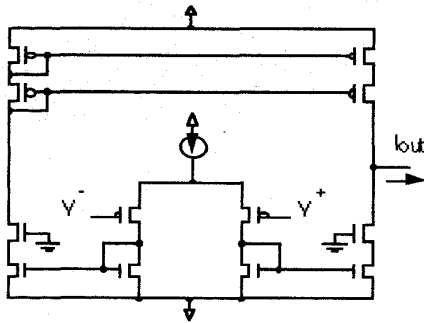


Fig.4 Circuit for the sigmoid function

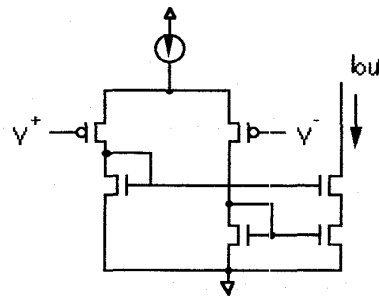


Fig.5 Circuit for derivatives of the sigmoid function

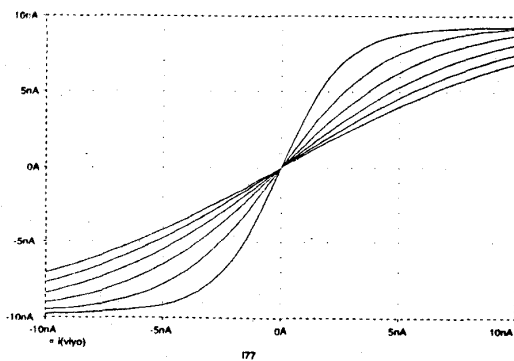


Fig.6 Characteristics of the circuit in Fig.4

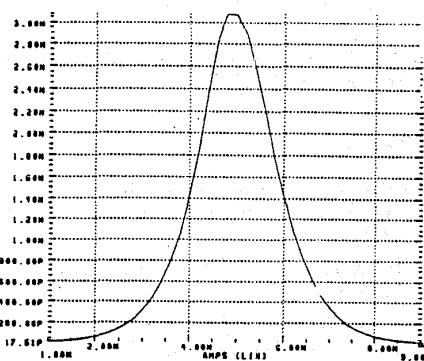


Fig.7 Characteristics of the circuit in Fig.5

By constructing a board with multiple chips one can easily increase number of neurons and number of layers. To train multilayer Perceptron one just apply input and target patterns. Signal feed-forward, error back-propagation, and weight adjustments all can be done asynchronously. For unsupervised learning models based on Hebbian learning rule only

the input patterns need to be applied. Although the circuits operate asynchronously, external synchronous control may be necessary to apply input (and target) patterns one by one.

Adaptive Learning Experiment

To show the learning capability the circuits are trained for the XOR pattern classifier. Transient SPICE analysis is used for this experiment, and the results are shown in Figs. 8 and 9. Evolution of the 3 synaptic weight values are plotted in Fig.8, and output error in Fig.9. Both figures clearly show adaptive learning of the circuits.

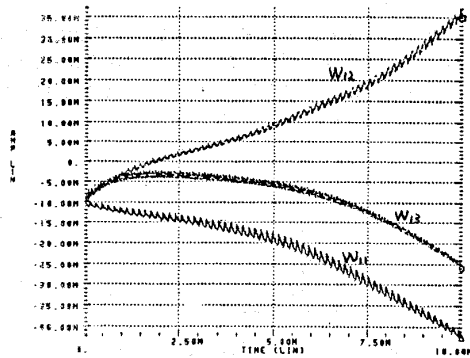


Fig.8 3 synaptic weights vs. learning epoch

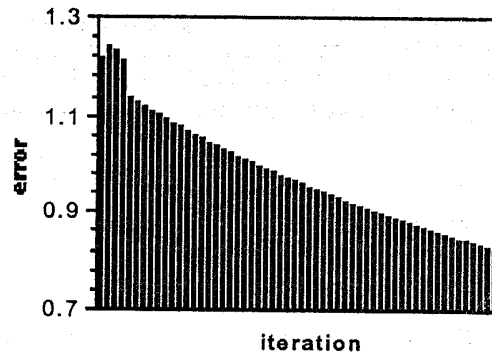


Fig.9 Error vs. learning epoch

Conclusion

In this paper we presented an on-chip learning circuit with error back-propagation or Hebbian learning. Due to subthreshold operation of the MOS circuits it consumes much lower power. Also its design is based on modular architecture with scaling circuits, which provides a capability to construct very complicated systems with multiple chips. The sigmoid function and its derivatives are implemented up to the first order approximation. Transient SPICE simulation for the XOR learning is presented, and results of the actual chip and board will be reported at the conference.

Acknowledgement: This research was supported by Korean Ministry of Science and Technology through KAIST as an Advanced Essential Technology Project.

References

- [1] T. Shima, T. Kimura, Y. Kamatani, T. Itakura, Y. Fujita, and T. Iida, "Neuro Chips with On-Chip Back-Propagation and/or Hebbian Learning," IEEE J. Solid State Circuits, vol.27, no.12, 1992.
- [2] C.A. Mead, *Analog VLSI and Neural Systems*, Addison-Wesley, Reading, 1989.
- [3] M.H. Cohen and A.G. Andreou, "Current-Mode Subthreshold MOS Implementation of the Herauld-Jutten Autoadaptive Network," IEEE J. Solid State Circuits, vol.27, no.5, 1992.
- [4] B. Gilbert, "A Monolithic 16-channel Analog Array Normalizer," IEEE J. Solid State Circuits, vol. SC-19, pp. 956-963, 1984.
- [5] T. Delbruck, "Bump Circuits for Computing Similarity and Dissimilarity of Analog Voltages," IJCNN, 1991.