ASSOCIATIVE MEMORY IMPLEMENTATION WITH ARTIFICIAL NEURAL NETWORKS

Santosh Saraf[1], A. M. Bhavikatti[2]

[1]*Research scholar, Electronics and Communication Engineering, JJTU, Rajasthan, India*
[2]*Professor, Electronics and Communication Engineering, BKIT, Bhalki, Karnataka, India*

Abstract

The first description of ANN integrated circuit implements a continuous time analog circuit for AM. The design used a 22 x 22 matrix with 20,000 transistors, averaging 40 transistors per node to implement a Hopfield AM network. The design faced a scalability challenge at higher levels of integration. The paper advocates handling larger problems by a collection of smaller networks or hierarchical solutions, while predicting, "Significantly different connection technologies" as essential for success in larger systems.

Keywords: *Associative Memory (AM),CMOS (Complementary metal oxide semiconductors),Artificial Neural Network (ANN), Bayesian Memory Module (BMM), Field Programmable Gate Arrays (FPGA).*

----------***----------

1. INTRODUCTION

Learning is the way we acquire knowledge about the world around us, and it is through this process of knowledge acquisition, that the environment alerts our behavioral responses. Learning allows us to store and retain knowledge, it builds our memories. In a neurobiological context, memory refers to the relatively enduring neural alterations induced by the interactions of an organism with its environment. Without such a change, there is no memory .The memory must be useful and accessible to the nerves system that influences the future behavior. Memory and learning reintricately connected. When a particular activity pattern is learned, it is stored in the brain, where it can be recalled later when required. Learning encodes the information. A system learns a pattern if the system encodes the pattern in its structure and it changes as the system learns the information. So learning involves change that can be represented in memory for future behavior.

2. RELATED WORKS

The seminal work by Sage and Withers built AMs using discretetime analog technology for high-speed computation in combination with analog nonvolatile storage for synaptic weights. The network demonstrated was a 9x9 Hopfield [8] associative memory network. The issue with the design was that although the synaptic weights could dynamically adapt, there were only three possible states to the weights (1, 0, -1). Thus, the network could demonstrate learning for a very few specific computations only. The message from this study was that a continuous range weights would be a desirable feature for the synapses. In an attempt to achieve high resolution synaptic weights, Schwartz and Howard proposed representing each weight as a difference in voltage between two capacitors. With the additional circuitry for sense-amplifiers, a 32 x 32 matrix with 75,000 transistors averaged 70 transistors per neural node. The high-level integration required scaling of the components to nano-scale levels and further simplification of the node design.

Holler[7] proposedto use floating gate technology for the representation of synaptic weights to achieve higher synapse density, but the design has electrically programmable static weights, and the dynamics of input presentation has no bearing on the real-time network associations. A mix of 8 x 8 matrix of digitally stored weights gate the inhibitory/excitatory pulse stream from 4 x 4 input layer. The pulse stream generation, integration and modulation results in much lower densities (140 transistors per neural node) than the aforementioned designs

Among biological applications, Lyon et. al. [10], implemented an electronic analog equivalent for the human cochlea (inner-ear). The design uses CMOS transconductance amplifiers circuits, follower-integrator circuits and second-order filter circuits to emulate perceptron machines. The authors see inherent deficiencies with digital threshold logic and emphasize the need for high-density analog learning based implementations for more precise biological equivalence.

Hammerstrom et al. [5] demonstrated one of the first custom digital ANN processor CNAPS. The CNAPS architecture, customized for ANN simulations, had significant performance vs. cost improvements over arrays of commercial microprocessors. The authors proposed that further speed-ups could be achieved by exploiting the high-speed memory structure and the inherent parallelism of field-programmable-gate-arrays (FPGAs). Along the lines of exploiting the FPGA

advantage, Changjian et al. [2] have demonstrated a best-match association using distributed representations on FPGA hardware. Similarly, Deshpande [4] implemented a Bayesian-memory (BM) module on a FPGA. The term BM is used by the author to describe a building block in the hierarchical design of an equivalent HTM model . Both these studies show that the performance of the FPGA based designs for associative memory models is dominated by the available chip area and the logic resources. Hammerstrom and Zaveri[6] analyzed the optimum use of such resources, and compared the performance vs. price trade-off for different architectures. They concluded that the mixed-signal CMOL design had the best performance-to-price ratio. The authors also suggested that "if in the future, nano devices/materials study provides robust solutions for implementing various analog functions using nanotechnology, this performance over price advantage is going to increase even further."

Digital implementations of ANN AM integrated circuits presents various trade-offs between silicon area and computation time. 16 x 16 pattern storage and recognition networks are implemented using multi-chip modules. All authors were convinced that their proposed architectures would benefit in terms of more complex computations, if digital devices scaled down another 1,000 fold from then existing 3–μm CMOS fabrication technology. Additionally, the proposed designs have only one stable state. More than one stable state exponentially increases the component count within each building block.

A fair insight into the algorithmic implementation of high-level AM algorithm including learning is achieved from the work. The design employed analog amplifiers to act as 'neurons', five-bit registers as synapses, and noise amplifiers for the simulated annealing. The paper highlighted several challenges:

- lack of an effective algorithm for learning in modular, hierarchical networks;
- necessity of modularity to manage connectivity;
- simplification of node design in addition to synaptic density;
- constraints, such as power dissipation and capacitive loading across the chips;
- at least 100 x 100 neurons interconnected by 1000 x 1000 synapse for the simplest of meaningful computation.

As Pao et al.set the rules for high-level AM models; the above experiments stressed the need for optimizing low-level AM designs that could be hierarchically integrated in densities not achievable even with the current 22nm nand-flash technologies. Additionally, Bailey et al.[1]assert the need for multiplexed interconnects for large scale ANN based AM system integration.

The implementation of building-blocks for high-level AMs memories is a challenging problem in artificial vision, image recognition, and other intelligent and adaptive computing areas. This challenge has previously been addressed in many different ways, for example by modeling artificial neural networks using traditional components such as resistors, capacitors, operational amplifiers, including voltage and current sources, as summarized above. However, the traditional approaches lack scalability. The other problem is that an AM building block unlearns as well if destructive input patterns are introduced.

3. PROPOSED WORK

This paper explores denser designs with novel nano-scale components that bypass the scalability hurdle with their inherently small form factors and with new properties. The device of choice for our investigation is a memristor.

3.1 Exploiting Memristors: Towards Nano-Scale Designs

Memristors were theoretically introduced by Leon O. Chua as passive two terminal devices in which the resistance is a function of the magnitude, polarity, and the duration of the applied voltage, and hence, a passive element with memory. In 1976 Chua and Sung [3] generalized memristors to a class of nonlinear dynamical systems or the memristive systems. Memristive systems were theoretically shown to be perfectly non-linear resistors showing hysteresis at lower frequencies and reducing to linear resistors at higher frequencies.

Memristance was observed in nano-scale solid state devices by Strukov et al.[11] and it was experimentally confirmed that resistance in memristive systems spans a continuous range of value. Memristance was quickly observed in several nano-scale systems, for example, oxide-vacancy-based titanium dioxide electrodes reported by Williams[13], electrochemically controlled polymeric memristors reported by Erokhin and Fontana , and magnetic memristors as reported byWang et al.[12].

The non-linear characteristic of a memristor is exemplified in Figure 1.

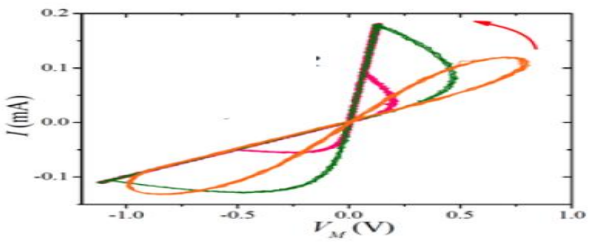

Fig 1: An example of the non-linear characteristic in the I–V plot of a memristor.

Figure re-drawn from [9]Soon after their discovery, memristor circuits were shown to perform universal logic operations, like material implication and useful arithmetic operation quite accurately. Memristor crossbar latch memory or RRAM have been demonstrated by Pascal et al. to achieve densities higher than DRAMs. However, a significant hurdle to realizing the potential of RRAM is the sneak path problem which occurs in larger passive arrays. Memristor-based circuits have been handdesigned to emulate biological responses like environment awareness in amoeba and may find application in pattern recognition.

4. CONCLUSIONS

In AM and ANN paradigms, memristors have been explored solely as synapses, effectively working as switches between 'on' and 'off' resistance values. We hypothesize that memristor network AMs will be more area-efficient than the traditional AMs if we exploit their:

- non-linear property for spatial association, and their
- Time-dependent property for temporal association.

Memristors, with their nano-scale form factors, continuous resistance range, non-linearity and time-dependency promise desirable circuits that can emulate low level AM model. Discovered only in 2008, these novel devices have no established design methodology yet as in the case of CMOS technology, which has developed over the past six decades. In the absence of design experience, we chose automated circuit discovery, employing stochastic search and evolutionary optimization as a tool for exploiting memristors design space.

REFERENCES

[1].Bailey J. and Hammerstrom D., Why VLSI implementations of associativeVLCNs require connection multiplexing , IEEE, International Conference on neural networks, pages 173–180 vol.2, 1988.

[2].Changjian G., Hammerstrom H., Zhu S. and Butts M., FPGA implementation of very large associative memories - scaling issues, In Ed. Amos Omondi, editor, FPGA Implementations of Neural Networks, Boston, MA, USA, 2003, Kluwer Academic Publishers

[3].Chua L. O.and Kang S. M. , Memristor devices and systems.Proceedings of the IEEE, 64(2) 209–223, 1976.

[4].Deshpande M., FPGA implementation & acceleration of building blocks for biologically inspired computational models. In M.S. AAT 1491185, Portland, OR, USA, 2011, Portland State University

[5].Hammerstrom D., Henry W., and Kuhn M. , The CNAPS architecture for neural network emulation, Parallel Digital Implementations of Neural Networks, pages 107–138, 1993.

[6].Hammerstrom D. and Zaveri M. S., Prospects for building cortex-scale.

[7]. Holler M., Tam S., Castro H. and Benson R., An electrically trainable artificial neural networks with 10240 'floating gates' synapses, in neural networks, 1989. IJCNN, International Joint conference on pages 191-196, vol. 2, 1989.

[8]. Hopfield J. J., Neural networks and physical systems with emergent collective computational abilities, Proceedings of the National Academy of Sciences, 79(8):2554–2558, 1982.

[9].Kantschik W. andBanzhaf W., Linear-graph GP - a new GP structure. In Proceedings of the 5th European Conference on Genetic Programming, EuroGP'02, pages 83–92, London, UK, 2002. Springer-Verlag.

[10]. Lyon R. F. and Mead C. ,An analog electronic cochlea, Acoustics, Speech and Signal Processing, IEEE Transactions on, 36(7):1119–1134, 1988.

[11].Strukov D B, Snider G S, Stewart D R and Williams R S ,The missing memristor found, Nature. Vol. 453(71910, pages 80-83 Nature publishing group, 2008

[12]. Wang F Z, Helian N, Wu S, Yang X, Guo Y, Lim G and Rashid M M , Delayed switching applied to memristor neural networks, Journal of applied physics, Vol. 11(7), pages 07E317-07E317. AIP, 2012.

[13]. Williams R, How we found the missing memristor, Spectrum, IEEE. Vol. 45 (12) pages 28-35. IEEE, 2008