2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

Spike-Timing Dependent Morphological Learning for a Neuron with Nonlinear Active Dendrites

Phyo Phyo San, Shaista Hussain and Arindam Basu
School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798
Email: arindam.basu@ntu.edu.sg, ppsan@ntu.edu.sg

***Abstract*— It has been shown earlier that simple abstraction of a neuron with nonlinear active dendrites and binary synapses has a higher computational power than a neuron with linearly summing dendrites. However, it has only been used to classify high dimensional binary patterns of mean spike rates. In this paper, a nonlinear dendritic (NLD) neuron equipped with binary synapses that is able to learn temporal features of spike input patterns is presented. Since the synapses are binary, learning happens through formation and elimination of connections between the inputs and the dendritic branches thus modifying the structure or "morphology" of the cell. A morphological learning algorithm inspired by the 'Tempotron'– a recently proposed temporal learning algorithm–is presented in this work. Experimental results indicate that our neuron with NLD with 1-bit synapses can obtain similar accuracy as a traditional Tempotron with 4-bit synapses in classifying a population of single spike latency patterns. Hence, the proposed method is better suited for robust hardware implementation in the presence of statistical variations.**

I. INTRODUCTION

Though the representation of stimulus by the neurons in our brain is a topic of much ongoing research and debate, it is widely believed that the timing of the action potentials or spikes fired by these neurons carries important information [1]. Spike latency codes i.e. delay in the spike time after stimulus presentation, have been suggested for tactile, olfactory and retinal systems [2]. They are also thought to offer significant advantages in terms of reducing power needed for communicating spikes as well as allowing rapid processing of inputs. Hence, neuromorphic engineers, who aim to mimic the brain's processing capabilities in silicon, have also been interested in spike timing based neural networks. Several analog CMOS integrated circuits operating in the sub-threshold regime have been designed in the past to implement somatic and synaptic functions [4], [5]. However, with the increase of statistical variations due to the constantly decreasing feature size of transistors, performance of silicon neural networks requiring accurate setting of a "weight" parameter become strongly compromised. This is also true for several nanometer scale non-CMOS devices (e.g. memristor or domain wall magnets) that have potential for use in neuromorphic applications. Hence, there is a strong need to develop algorithms and architectures that retain the performance of earlier systems but require low-resolution weights.

Financial Support from MOE through grant MOE2012-T2-2-054 is acknowledged.

In this paper, a hardware-friendly morphological learning rule that learned on nonlinear dendritic neuron (NLD) with spatiotemporal spike patterns is presented. Different from earlier work [3], [9] where mean rate encoded inputs were considered, spike-timing information is presented to the neuron with NLD in our case. Unlike the Tempotron learning rule [1] that requires weights with high resolution, the proposed network uses low-resolution integer weights and learns through modifying connections. This results in easier hardware implementation since a low-resolution integral weight of W can be implemented by activating a shared binary synapse W times through time multiplexing schemes like Address-Event Representation (AER) [6], [7].

The organization of this paper is as follows: in Section II, the architecture of nonlinear dendritic neurons and its morphological learning algorithm is introduced. In order to test the morphological learning's ability in recognizing spatiotemporal spike patterns, a classification task based on single spike latency patterns similar to the one in [1] is discussed in Section III. To show the effectiveness of the proposed method, several experiments are conducted with different number of spike input patterns in Section IV. Finally a conclusion is drawn in Section V.

II. METHOD

A. Nonlinear Dendrites Model with Spatiotemporal Spike Patterns

Dendritic neurons with lumped nonlinearity are able to perform a larger number of input-output mappings than is possible by a neuron with linear ones [3]. Though the performance of NLD has been tried on a pattern classification task in relation to static and spike-rate input patterns [9], no application has been made to the learning of spike timing-based decision rules. To bridge this gap, a neuron with NLD that is able to respond to temporal features of input spike patterns is presented in this paper. As presented in Fig. 1, the structure of a neuron with NLD is characterized by m identical dendritic branches and k excitatory synaptic contacts per branch. For each branch, the activation of synaptic contact is determined by one of d dimensions of an input spike pattern. At the relevant times governed by incoming spikes, the synaptic connections are initiated and the membrane voltage is calculated by weighted

978-1-4799-1484-5/14/$31.00 ©2014 IEEE

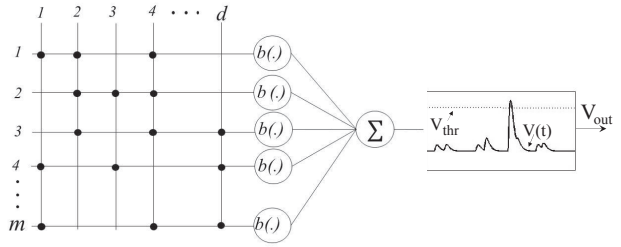

Fig. 1. The architecture of nonlinear dendritic neuron (NLD) model

sum of postsynaptic potentials (PSPs) as follows:

$$V(t) = \sum_{j=1}^{m} b\left(\sum_{i=1}^{k} w_{ij} K(t - t_{ij})\right) \tag{1}$$

where w_{ij} is the weight of the ith synapse formed on the jth branch, $b(\cdot)$ is the nonlinear activation function of the dendritic branch which is characterized by $b(z) = {z^2}/{x_{thr}}$, z is the total synaptic activation on a branch [9], K denotes the postsynaptic potential kernel and t_{ij} are times of incoming spikes. The normalized postsynaptic potentials (PSP) contributed by each incoming spike arriving at time t_i is obtained by:

$$K(t - t_i) = V_0(\exp[-(t - t_i)/\tau] - \exp[-(t - t_i)/\tau_s]) \tag{2}$$

where the parameters τ and $\tau_s = \tau/4$ denote the decay time constants of membrane integration and synaptic current respectively. The NLD neuron is designed such that the final output voltage V_{out} takes a value of either "1" or "0" depending on whether the summed membrane voltage $V(t)$ crosses a threshold voltage (V_{thr}). Similar to [1], this indicates that a neuron fires at least one spike if $V(t)$ crosses V_{thr}, otherwise it remains quiescent. After the neuron fires a spike, it is returned to the refractory state and the rest of the spikes in the incoming pattern do not affect the computation.

B. Morphological Learning Algorithm

In this section, the morphological learning rule that can modify the connections between afferent lines and nonlinear dendrites (NLD) is introduced. The inspiration of this work comes from the Tempotron learning rule [1] that learnt the temporal features of random spike patterns by updating its weights. However, for feasibility of hardware implementation, we consider the use of binary weights in this work as opposed to using high resolution weights. To start with the proposed learning rule, the cost function which measures the deviation between the maximum membrane voltage (V_{max}) generated by error patterns and the firing threshold (V_{thr}) is defined as:

$$E = \begin{cases} V_{thr} - V_{max}, & if\ positive\ pattern\ P^{+}\ is\ presented \\ V_{max} - V_{thr}, & if\ negative\ pattern\ P^{-}\ is\ presented \end{cases} \tag{3}$$

where V_{max} is the maximum value of postsynaptic potential $V(t)$ at the time t_{max}, i.e., $V_{\max} = V(t_{\max})$. With a guidance by the gradient-descent method, the change in synaptic efficacy is calculated by:

$$\Delta w_{ij} = -\frac{dE^2}{dw_{ij}} = -\left(\frac{\partial E^2}{\partial w_{ij}} + \left(\frac{\partial E^2}{\partial t_{\max}} \cdot \frac{dt_{\max}}{dw_i}\right)\right) \tag{4}$$

$$= 2\,(V_{thr} - V_{max})\,\frac{\partial\left(\sum_{j=1}^{m} b\left(\sum_{i=1}^{k} w_{ij} K(t_{max} - t_i)\right)\right)}{\partial w_{ij}} \tag{5}$$

where the second term in (4) becomes 0 because $\partial(V_{thr} - V_{\max})/\partial t_{\max} = 0$ by definition of t_{max} [1]. By substituting (3) and (1) in (4), (5) is obtained. From (3) to (5), the weight updating rule is obtained as:

$$\Delta w_{ij} = (V_{thr} - V_{max})\, b'_j K(t_{\max} - t_i) \tag{6}$$

where b' denotes the derivative of b. As mentioned earlier, the weights of the NLD only takes the binary values, i.e., $w_{ij} = 1$ if a connection exists and 0 otherwise (Note that we allow multiple connections between an input and a dendritic branch; therefore, effective weight of a connection can be an integer larger than 1). Hence, we cannot directly modify the weights by adding the Δw_{ij} term derived here. Instead, the term Δw_{ij} in (6) is redefined as a correlation term, c_{ij} i.e., $c_{ij} = \Delta w_{ij} = (V_{thr} - V_{max})\, b'_j K(t_{\max} - t_i)$ and is used to guide the process of swapping connections. At every iteration of the learning process, the synapse with the lowest c_{ij} averaged over an entire batch of patterns will be replaced with the highest c_{ij} synapse in a candidate replacement set. The mechanism of the learning process is presented as follows:

1) The learning process starts with random initialization of total synaptic connections ($s = m \times k$) from the uniform distribution of $1 : N$ afferents.
2) In each iteration of the training process, a random input spike pattern is presented to the NLD neuron. The activation of synapses on each dendritic branch is determined and the cell membrane voltage ($V(t)$) in equation (1) is calculated for all the input patterns.
3) From the calculated $V(t)$, the maximum membrane voltage (V_{max}) is observed and classification output is determined, i.e. the patterns are correctly classified if $V_{max} > V_{thr}$ is satisfied for P^{+} and $V_{max} < V_{thr}$ for P^{-}.
4) For all misclassified patterns, the correlation term, c_{ij} is computed by equation (6) for each synapse. Based on the obtained c_{ij} averaged over the entire pattern set, a random set T of n_T synapses was targeted for possible replacement.
5) The poorest-performing synapse (minimum c_{ij}) in T is replaced with the best-performing synapse (maximum c_{ij}) from another randomly chosen replacement set R containing n_R of the N afferent lines.
6) Synaptic connections are modified if the number of learnt patterns increases. If there is no increment, a new replacement set R is chosen. If the number of learnt patterns does not change for 40 iterations, it is assumed that

the learning algorithm encountered a local minimum. In this case, we modify synaptic connections even though there is no change in the number of learnt patterns. By doing so, the learning algorithm may attempt to escape from local minimum.

7) The learning steps (2) to (6) are repeated until all the patterns are correctly learnt, or 100 local minima are encountered, or the maximum number of iterations is reached.

C. Determination of Optimal Threshold

Since we do not have an arbitrary multiplicative weight in our neural model, the range of maximum voltages generated in response to a fixed temporal spike pattern is limited. This is similar to the problem faced in [10]. Hence, improper selection of threshold may largely degrade the classification performance since a very large V_{thr} $(= m \times k \times K_{max})$ may never be crossed by $V(t)$. To define a proper threshold (V_{thr})

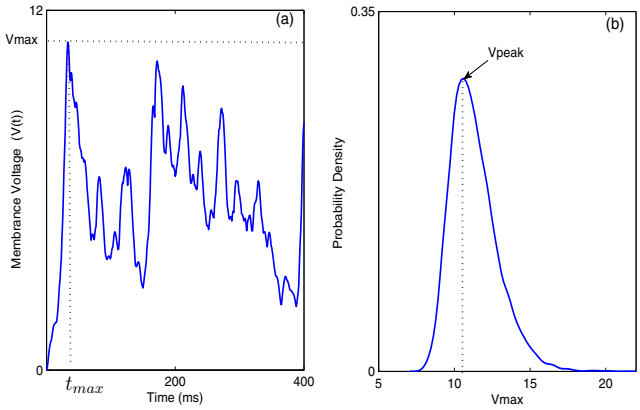

Fig. 2. (a) Membrane Potential $V(t)$ for a spike pattern (b) Probability distribution of V_{max}

level, the maximum value of $V(t)$, i.e., V_{max} at time t_{max} for random input spike patterns is firstly determined (Fig. 2 (a)). This process is repeated for a large number of randomly generated patterns. From the V_{max} obtained over this entire set, the probability distribution of V_{max} is generated and is used to determine the optimal threshold V_{thr}. As shown in Fig. 2 (b), this is equal to the voltage corresponding to the peak of probability distribution function, V_{peak}.

III. MORPHOLOGICAL LEARNING FOR CLASSIFICATION

A. Input Pattern Generation

In order to test the morphological learning's ability, its performance was assessed on classifying a population of single spike latency patterns [1]. To perform the task, P spike patterns are generated, each of which was randomly assigned to one of the two classes: positive class (P^+, Class 1) and negative class (P^-, Class 2) . For each spike pattern, $\mathrm{X} = (x_1, x_2, \ldots, x_N)$ consists of spikes arriving at N afferents, each of which spiked only once at a time drawn independently from a uniform distribution between 1 and $T = 400$ ms.

For instance, consider random latency patterns with spike times of 12 afferent inputs. A positive pattern P^+ and negative pattern P^- are shown in Fig. 3 (a). The morphological learning rule is required to correctly classify such random patterns.

B. Learning to Classify Latency Spike Patterns

The NLD neuron was trained to correctly classify random input spike patterns belonging to the two classes 1 and 2. Patterns are said to be correctly classified if their associated membrane voltage, $V(t)$ reached its maximum value, V_{max} at time $t^+_{\max}$ and $t^-_{\max}$ and satisfied the condition, $V^+_{\max} > V_{thr} > V^-_{\max}$. As shown in Fig. 3 (b), when the conditions are satisfied, the neuron fires at least one spike in response to P^+ patterns while it remains below threshold for P^- patterns.

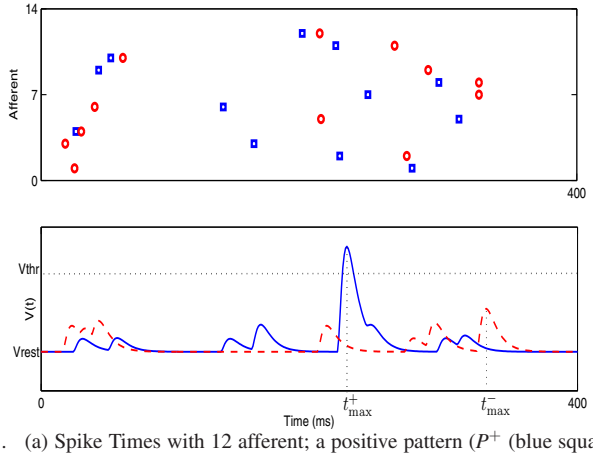

Fig. 3. (a) Spike Times with 12 afferent; a positive pattern (P^+ (blue square) and a negative pattern P^- (red circle) (b) Generation of V(t) for positive and negative patterns

In the process of learning, the grouping of synchronous spikes on a dendritic branch will be encouraged for the patterns in Class 1 while it is prevented for those in Class 2. As presented in Fig. 4, with the proposed learning rule, the connections of NLD are modified such that for a learnt pattern in Class 1, afferents receiving synchronous spikes are grouped together on a dendrite which allows neuron membrane voltage $V(t)$ to exceed V_{thr} while a trained pattern in Class 2 is characterized by asynchronous spikes arriving on a dendrite, resulting in $V(t)$ falling below V_{thr}.

IV. RESULT AND DISCUSSION

Throughout the experiment, the design parameters m, k, x_{thr}, τ, n_T, n_R and T, were chosen as 50, 10, 6, 15 ms, 25, 25 and 400 ms respectively. The details about the selection of parameters x_{thr} are discussed in [9]. There is no default rule for the selection of NLD parameters, which were selected by the trial-and-error method. As discussed in Section II-C, the firing threshold V_{thr} is chosen at the peak location of V_{max} distribution over 10000 random input spike patterns. Hence, V_{thr} is set at 11.

To start with the classification task, the NLD neuron in Fig. 1 was trained for random latency patterns as generated in Section III-A. To see the effectiveness of the proposed

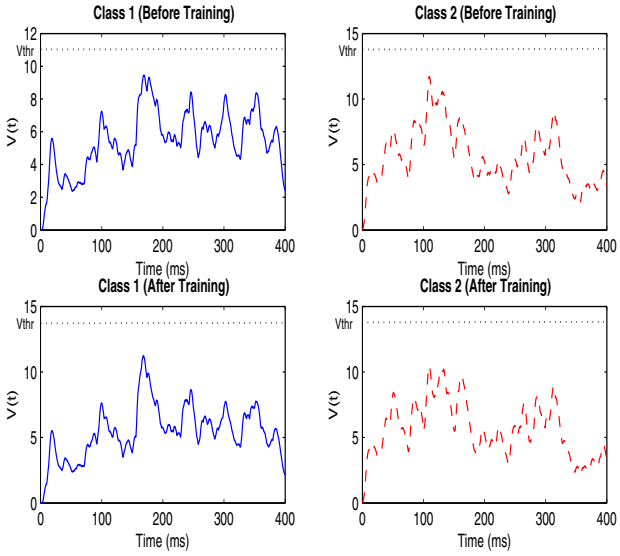

Fig. 4. Membrane potential generated by a pattern in Class 1 and 2 respectively.

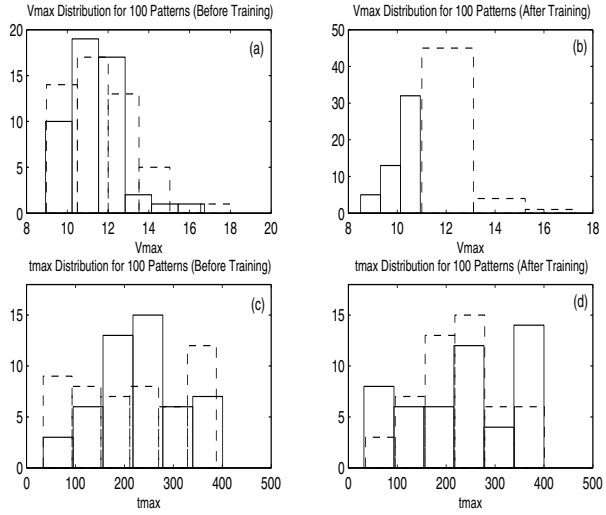

Fig. 5. The distribution of V_{max} and t_{max} before and after training. Dashed lines indicate Class 1 while solid lines indicate Class 2.

method, the classification performance was measured for a small number ($= 100$) of input patterns. The results in Fig. 5 (a) and (b) show that the proposed method can efficiently perform the classification task. A clear separation between the V_{max} distributions for Class 1 and Class 2 shows that the proposed method is able to respond to the random single spike latency patterns by shifting V_{max} away from the V_{thr} in opposite directions for the patterns belonging to the two classes. Next, we consider the distribution of t_{max} distribution before (Fig. 5 (c)) and after training (Fig. 5 (d)). Both the distributions are widely spread over time implying that the connections were modified such that the spikes in an input pattern that arrive closer to the initial t_{max} for that pattern, are grouped together on a dendrite, thereby not changing the t_{max} values significantly. Further, the NLD neuron was also trained for larger number of input patterns (500 and 1000

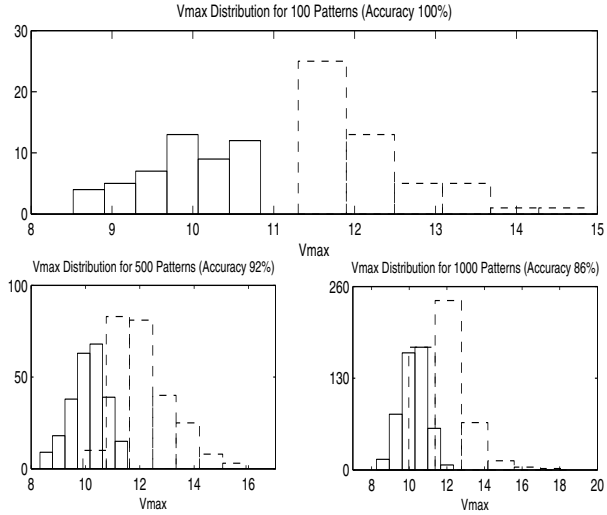

Fig. 6. The distribution of V_{max} for a set of 100, 500 and 1000 random latency spike patterns

patterns) as presented in Fig. 6. It shows that the proposed method can perform the classification task quite well by achieving accuracies of 92 and 86 % for 500 and 1000 patterns respectively.

Next, the performance of the proposed method is compared with the Tempotron learning [1] that learnt single spike latency patterns by using weight updating rule. Since we are interested in the performance of these algorithms in their hardware implementations, which are plagued by mismatch, we consider the performance of the Tempotron when its weight is quantized at different resolutions. Further, we do the quantization in two ways: either after the completion of training or as a step within the training procedure. The comparison results in Fig. 7 show that the Tempotron using floating-point numbers achieves better performance compared to the proposed method. However, when the high resolution weights are quantized at 2-bit level, its performance is worse than the proposed method. Also, it can be seen that the performance is better when the quantization is performed within the learning loop. This is to be expected since the learning algorithm can now try to correct this quantization error as well.

Only at 4-bit quantization level, the classification performance of Tempotron (93 and 89 % accuracy for 500 and 1000 patterns) is comparable to our proposed method (92 and 86 % accuracy for 500 and 1000 patterns respectively) that uses only 1-bit or binary weights. This underlines the importance of our proposed method in robustly implementing spike timing based classifiers with low-resolution analog synapses using nano-scale CMOS or non-CMOS devices.

V. CONCLUSIONS

A morphological learning rule that can be used to find the optimal connection matrix of neurons with nonlinear dendrites (NLD) and binary synapses is presented. To see the effectiveness of the proposed method, the NLD neuron is trained for a classification task with spatiotemporal spike inputs.

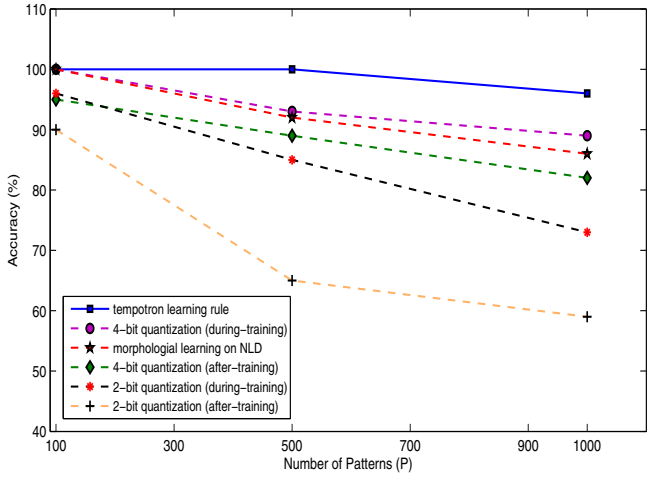

Fig. 7. Comparison studies on classification performance, for Tempotron learning [1] and morphological learning rule

The results (classification accuracy at 100, 92 and 86 % for 100, 500 and 1000 random input spike patterns respectively) depict that the NLD neuron with 1-bit weights trained by the morphological learning rule achieves comparable performance to the Tempotron learning rule using 4-bit weights. This makes our proposed system amenable for hardware realization in the face of statistical variations which plague deep sub-micron CMOS processes as well as nanoscale non-CMOS devices like memristors.

REFERENCES

[1] R. Gutig and H. Sompolinsky, "The tempotron: a neuron that learns spike timing-based decisions", *Nature Neuroscience*, vol.9, no.3, pp. 420-428, 2006.

[2] R. Johansson and I. Birznieks, "First spikes in ensembles of human tactile afferents code complex spatial fingertip events", *Nature Neuroscience*, vol.7, pp. 170-177, 2004.

[3] P. Poirazi and B. W. Mel, "Impact of active dendrites and structural plasticity on the memory capacity of neural tissue", *Neuron*, vol.29, no.3, pp. 779-796, 2001.

[4] A. Basu, C. Petre, and P. Hasler, "Bifurcations in a Silicon Neuron", In *Proceedings of the International Symposium on Circuits and Systems*, pp. 428-31, May 2008.

[5] J. Arthur and K. Boahen, "Synchrony in silicon: The gamma rhythm", *IEEE Transactions on Neural Networks*, vol. 6, pp. 1815-1825, Nov 2007.

[6] K. Boahen, "Point-to-point connectivity between neuromorphic chips using address events", *IEEE Transactions on Circuits and Systems II*, vol. 47, no. 5, pp. 416-34, May 2000.

[7] S. Brink, S. Nease, P. Hasler, S. Ramakrishnan, R. Wunderlich, A. Basu and B. Degnan, "A learning-enabled neuron array IC based upon transistor channel models of biological phenomenon", *IEEE Transactions on Biomedical Circuits and Systems*, vol. 7, no. 1, pp. 71-81, Feb. 2013.

[8] Y. Wang and S. C. Liu, "Multilayer processing of spatiotemporal spike patterns in a neuron with active dendrites", *Neural Computation*, vol.22, no.8, pp. 2086-2112, 2010.

[9] S. Hussain, R. Gopalakrishnan, A. Basu and S. C. Liu, "A Morphological Learning: Increased Memory Capacity of Neuromorphic Systems with Binary Synapses exploiting AER based Reconfiguration", *Proceeding of the International Joint Conference on Neural Networks*, pp. 1-7, Aug. 2013.

[10] S. Hussain, A. Basu, M. Wang and T. Hamilton, "DELTRON: Neuromorphic Architectures for Delay based Learning", *Proceeding of the Asia Pacific Conference on Circuits and Systems*, pp. 304-307, Dec. 2012.