Proposal For Neuromorphic Hardware Using Spin Devices

[1]Mrigank Sharad, [2]Charles Augustine, [1]Georgios Panagopoulos, [1]Kaushik Roy
[1]Department of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA
[2]Circuit Research Lab, Intel labs, Intel Corporation, Hillsboro, OR, US
msharad@.purdue.edu

Abstract: We present a design-scheme for ultra-low power neuromorphic hardware using emerging spin-devices. We propose device models for 'neuron', based on lateral spin valves that constitute of *nano-magnets* connected through metal-channels. Such magneto-metallic neurons can operate at ultra-low terminal voltage of ~20 mV, resulting in small computation energy. Use of domain wall magnets as programmable 'synapse' and as 'integrating neuron' is proposed. Magnetic tunnel junctions are employed for interfacing the spin-neurons with charge-based devices like CMOS, for large-scale networks. Device-circuit co-simulation-framework is used for simulating such hybrid designs, in order to evaluate system-level performance. We present the design of different classes of neuromorphic architectures using the proposed scheme that can be suitable for different applications like, analog-data-sensing, data-conversion, cognitive-computing, associative memory, programmable-logic and analog and digital signal processing. We show that the spin-based neuromorphic designs can achieve 15X-300X lower computation energy for these applications, as compared to state of art CMOS designs.

I. INTRODUCTION

Neural-networks (NN) constitute a powerful computation paradigm that can algorithmically outperform Von-Neumann schemes in numerous data-processing applications [1]-[7]. However, CMOS based hardware implementations of neuromorphic architectures prove inefficient in terms of power consumption and area-complexity. On one hand, digital designs consume large amount of area, whereas, on the other hand, analog designs, although compact, lead to power hungry solutions. This has limited the scope of neural networks to algorithms and software.

In order to tap the potential of neuromorphic computation at the hardware level, the device-circuit models for the neuron and the synapse, apart from being compact, should also achieve low power consumption. In this work we propose the application of spin-devices in NN hardware design that can help achieve these goals.

A multi-input lateral spin valve (LSV) can perform non-Boolean, analog-mode computation like majority-evaluation that can be used to model neuron functionality [1]-[5]. Programmable spin injection strength of domain wall magnet (DWM) can be used to implement a compact synapse [1], [6]. Current driven-motion of domain wall along a magnetic nano-strip can be used to realize an ultra low voltage 'integrating neuron'. Such compact, low-resistance, magneto-metallic devices can perform analog-mode-computation, while operating at ultra-low magnitude, pulsed voltage-supply, thereby simultaneously achieving low power consumption as well as small area. We use magnetic-tunnel-junctions (MTJ) to interface the proposed device models for neuron with CMOS, in order to realize different classes of neuromorphic architectures, dedicated to different applications.

In brief, we propose an entirely novel hardware-design scheme which exploits specific spin-device characteristics to perform ultra low energy neuromorphic computation. The presented work involves innovation in device-modeling as well as in the associated circuit-design. It also addresses the architecture level issues related to such a heterogeneous integration, in order to arrive at a comprehensive design solution.

Rest of the paper is organized as follows. Section 2 introduces the spin-based device models for neurons. Circuit integration scheme for the proposed devices is described in section 3. Section 4 presents some examples of different neuromorphic architectures based on the proposed scheme. The performance and prospects of the proposed design scheme is discussed in section 5. Finally section 6 concludes the paper.

II. SPIN BASED DEVICE MODELS FOR NEURON

In this section we present different models for neurons based on spin-devices. First, two different device-models for 'summing – neurons' are discussed. In such a neuron, all input signals are clock synchronized and concurrent. Hence the 'integration' operation in the 'integrate and fire' functionality of a neuron can be simply replaced by summation. Both 'local' as well as 'nonlocal' spin torque can be used to implement the proposed device models [4]. Following the 'summing-neurons' we briefly describe DWM based 'integrating-neuron' model.

A. Bipolar Spin Neuron

Fig. 1 shows the device structure for biopolar spin neuron [2], [4], [5], [7]. It constitutes of an output magnet m_1 with MTJ based read-port (using a reference magnet m_5), and two anti-parallel input magnets m_2 and m_3, with their 'easy-axis' parallel to that of m_1. A preset-magnet m_4, with an orthogonal easy-axis, is used to implement current-mode Bennett-clocking (BC) [1, 4].

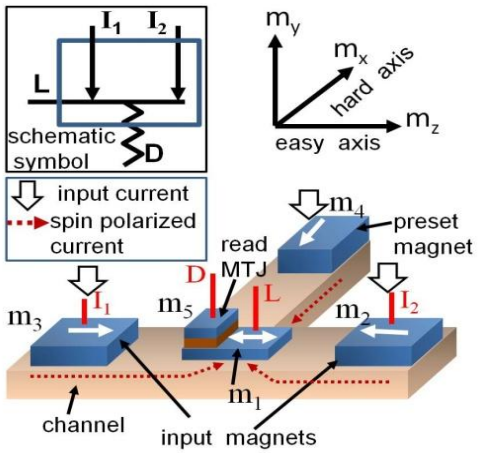

Fig. 1. Bipolar Spin Neuron with local spin injection and decoupled read-write [4].

A current pulse input through m_4, presets the output magnet, m_1, along its hard-axis. The preset pulse is overlapped with the synchronous input current pulses received through the magnets m_2 and m_3. After removal of the preset pulse, m_1 switches back to its easy-axis. The final spin-polarity of m_1 depends upon the sign of the difference ΔI, between the current inputs through m_2 and m_3. The lower limit on the magnitude of ΔI (hence, on current per-input for the neuron), for deterministic switching, is imposed by the thermal-noise in the output magnet, and, imprecision in Bennett-Clocking (BC). The effects of these non-idealities have been included in device simulation (fig. 2).

Transfer-function of an artificial neuron can be expressed as the sign-function of weighted sum of inputs, where the individual weights can be either positive or negative. In the proposed device, the neuron functionality is realized by connecting all the positive-

weight inputs (excitatory inputs) to its right-spin input-magnet and vice-versa. The output magnet, in effect, evaluates the sign function with the help of Bennett-clocking, where the right-spin state can be regarded as the 'firing state'

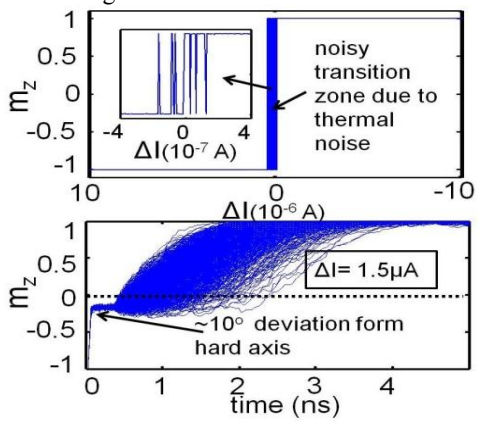

Fig. 2 Due to noise in the neuron-magnet and imprecise BC (leading to $m_z \neq 0$ during preset), larger ΔI (hence, current for inter-neuron signaling) is required for correct switching, than the ideal case. Minimum inter-neuron signaling current can be determined on the basis of bit-error rate (BER) resulting from these effects. .

B. Multi-input spin neuron with DWM synapse

The device operation explained above can be extended to a multi-input lateral spin valve (LSV) with programmable inputs in the form of DWM (fig. 3a), to realize a compact neuron-synapse unit [1], [6] (fig, 4).

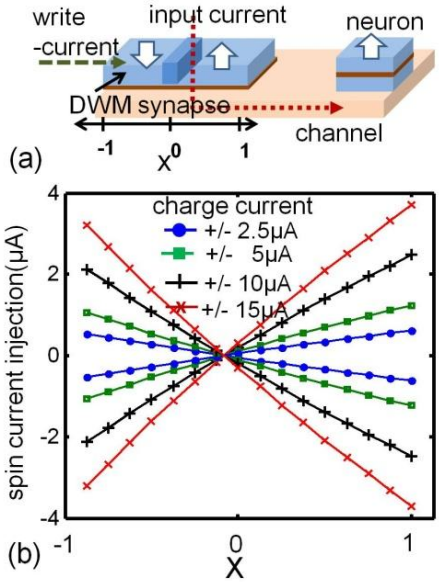

Fig. 3 (a) Domain wall synapse with channel interface (b) Spin polarization strength current injected through DWM as a function of DW location

A DWM constitutes of opposite spin-polarity domains separated by a non-magnetic transition region, termed as the domain wall (DW). The DW can be moved along the nano-magnetic strip by current injection. Hence, a DWM interfaced with the metal channel of an LSV acts as a programmable spin-injector or a spin-mode synapse [1]. The spin-potential in the central region of the channel (around the ground lead below the output magnet) depends upon the sum of spin currents injected by all the DWM synapses and in turn determines the firing or non-firing state of the neuron, post-Bennett clocking. Fig. 5 depicts the plot for spin-potential in the central region of the channel, surrounding the output magnet of a 16-input neuron, under input conditions corresponding to firing and non-firing conditions. It shows that, in case of a firing event, the entire channel is dominantly at a positive spin potential and vice-versa.

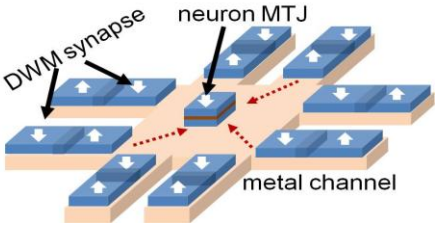

Fig. 4 Spin-based neuron model with three inputs (DWMsynapses). The free layer of the neuron MTJ is in contact with the channel and its polarity, after preset, is determined by spin polarity of combined input current in the channel region (ground terminal) just below it.

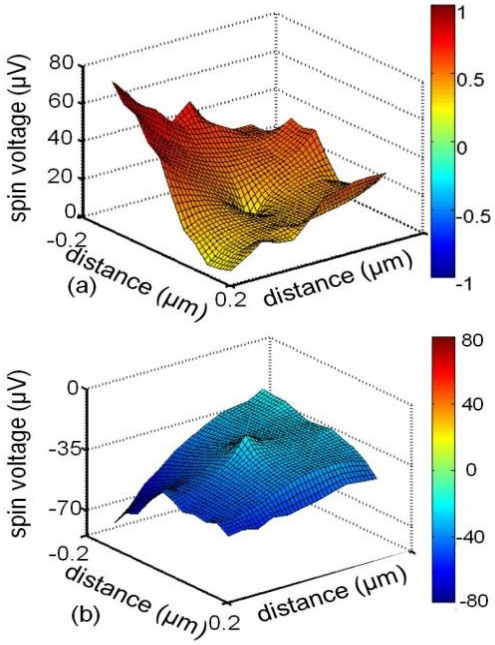

Fig. 5 (a) Channel spin potential of a 16 input neuron under firing condition (b) Channel spin potential under non-firing conidtion.

C. Integrating Neuron Based on Domain Wall Magnet

Spiking neural network is the most recent and evolving topology of neural networks [9]. Among different NN classes, it is regarded as the closest analogue to the biological neural network.

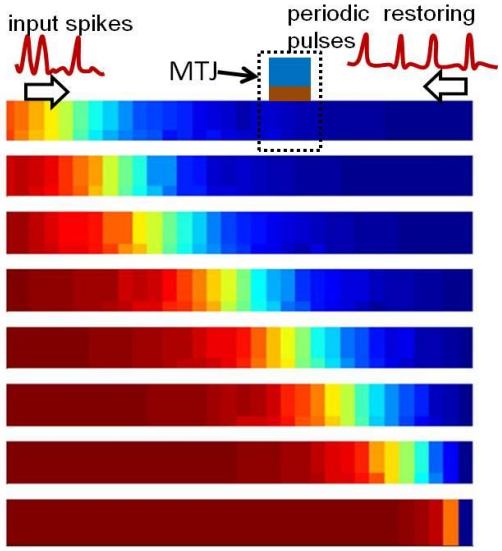

Fig. 6 Integrating neuron using DWM stripe: periodic restoration spikes are used to model 'leaky integration' in the neuron.

It employs asynchronous communication between neurons using spikes. This necessitates time-domain integration of input-signals. Conventionally, dedicated capacitors have been employed for low

speed SNN, while analog integrators have been used for getting higher performance. This once again presents the similar bottle neck of area and power consumption as described in the introduction. We propose the use of DWM stripe to realize time domain integration of input spikes. Firing state of the neuron can be detected using an MTJ (fig. 6). A DWM based integrating neuron allows spike transmission across ultra low terminal voltage and also mitigates the area overhead of capacitor. Hence it can lead to low power and compact SNN design.

III. CIRCUIT INTEGRATION SCHEME

Since spin-signals cannot be transmitted over a distance longer than a few spin diffusion length of the metal channel [1], we employ charge mode signaling to interconnect the spin neurons. A dynamic CMOS latch (fig. 7) senses the state of the neuron MTJ and drives the current source transistor, which transmits synapse current to all the fan-out neurons (shown schematically using the symbol given in fig. 1). Note that, the synapse input currents, involved in computation, flow across a small terminal voltage ΔV (fig. 7), thereby reducing the static power consumption resulting from large number of synaptic communications. A detailed description of the circuits can be found in [1]-[8].

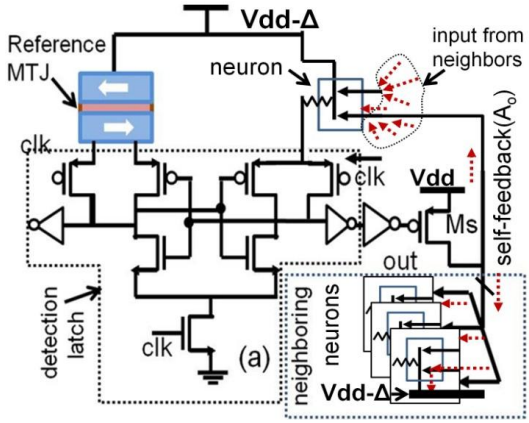

Fig. 7 Differential MTJ latch for inter-neuron current-mode signaling using deep triode current source (DTCS) transistor M_s.

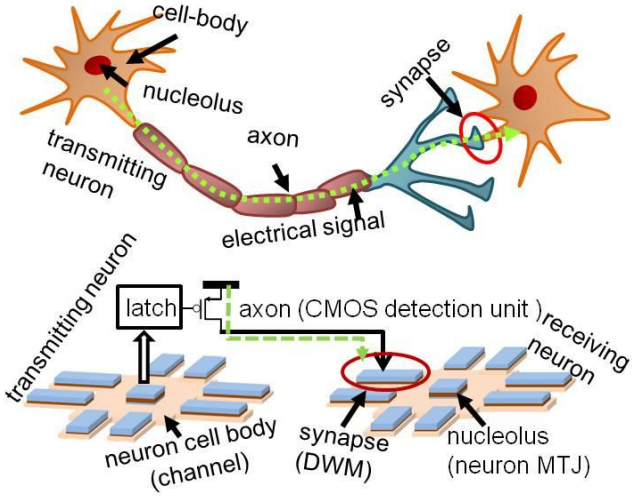

Fig. 8 Correspondence of the spin-CMOS Hybrid NN to biological neural network

The aforemensioned detection scheme can be applied to both the neuron strcutures described in the previous section. The use of differential supply voltage facilitates mitigation of common-mode noise across the device terminals resulting form clocking and other noise sources. The DWM-based neuron for SNN requires small amount of additional control logic that can be implemented in fully digital mode.

The resulting spin-CMOS hybrid design scheme can be extended to different classes of neuromorphic architectures. Fig. 8 depicts an interesting analogy between the biological neural network and the spin-CMOS hybrid neural network using the multi-input neuron with DWM synapses. The neuron magnet acts as the firing site, i.e., the nucleolus, the metal channel can be compared to the cell body of the neuron , spin potential in the central region of the channel is analogous to electrochemical potential in the neuron cell body which determines the firing/non-firing state of the neuron, the CMOS detection and transmission unit can be compared to axon of the biological neuron that transmits electrical signal to the receiving neuron, and finally the DWM acts as the synapse.

IV.NEUROMORPHIC ARCHITECTURES USING SPIN NEURONS

In this section we briefly discuss the design of few different topologies of neuromorphic architectures using the proposed spin devices.

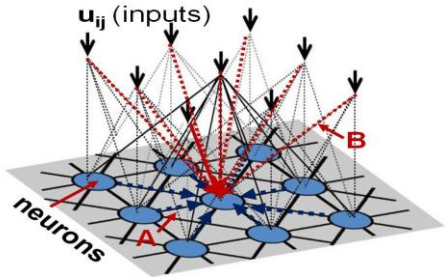

Fig. 9. CNN architecture with 3x3 neighbourhood connectvity

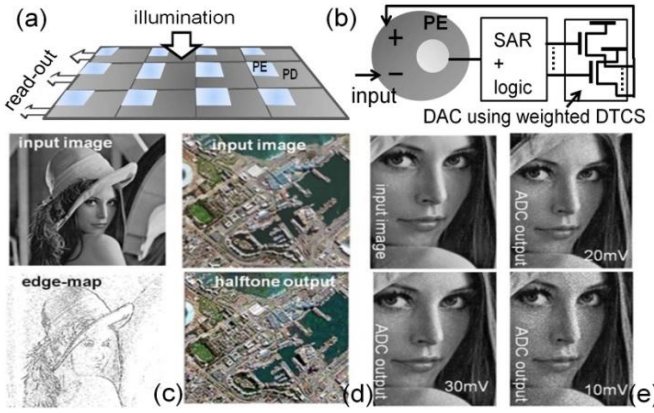

Fig. 10 (a) On sensor image processing architecture (b) SAR-ADC using spintronic neuron, and simulation results for (c) edge-detection, (d) half-toning, and (e) digitization (using spin-CMOS hybrid SAR-ADC : lowering ΔV increases % noise and hence degrades accuracy)

Cellular Neural Network (CNN) employs neighborhood connectivity and recursive operation [4], [5], [7]. This class of computation has been found to be highly suitable for several image processing applications, which essentially involve processing of pixel neighborhoods in a parallel fashion. Both the summing neurons described in section 2 can be used to implement CNN design with the help of CMOS transistors [4], [7]. The bipolar spin neuron leads to an application specific design, whereas, the multi-input spin-neuron can be employed for a programmable image processing engine. The multi-input spin neuron is specifically suited for CNN design, as the number of inputs in such a device is limited by the channel spin-diffusion length (and hence is more suitable for neighborhood connectivity) [2]. In [5], [7] we showed that the proposed design scheme can lead to 2 to 3 order of magnitude lower computation energy for analog image sensing, digitization and processing (fig. 10)., as compared to CMOS based mixed signal designs. As mentioned before, the main advantage

comes from ultra low voltage, pulsed operation of spin neurons that are applied to analog computation.

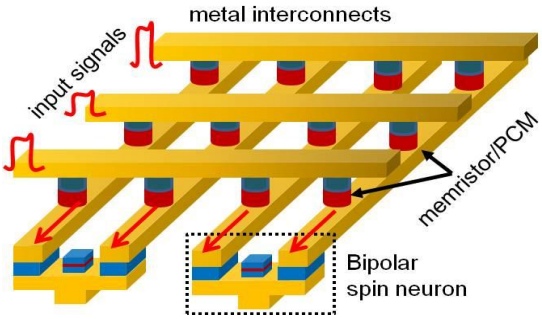

Fig. 11 Cross-bar NN architecture using memristors as synapses and bipolar spin neurons.

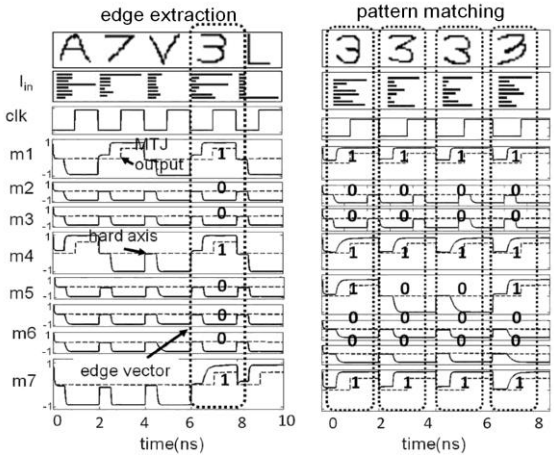

Fig. 12. Charcter recognition simulated using feed forward NN with multi-input neurons [1].

Emergence of programmable conductive elements, like TiO_2 memristor and phase change memory (PCM), have provided a compact model for neuron. However current-mode signal processing with CMOS-neurons in such circuits leads to power hungry designs. Application of the proposed bipolar spin-neuron in such a design, however, can lead to ultra low power NN hardware. Fig. 11 shows a cross-bar neural network architecture using memristor (/PCM) synapses and spin neurons [2]. Depending upon the polarity of the connectivity between an input line and a neuron, one of the two memristive junctions between them is driven to off state, while the other is programmed to match the required weight magnitude. The spin-neurons facilitate ultra-low voltage, pulsed synaptic communication across the cross-bar metal interconnects, thereby reducing the static-power consumption resulting from large number of inter-neuron signals per-cycle in a large-scale array. Such a design can provide ultra low power solution to several interesting applications, like, logic in memory, associative memory, programmable logic and pattern matching.

Spiking neural networks based on memristive cross-bar arrays can realize self-learning networks for cognitive computing. Such a design employs some additional control circuits in each neuron to implement synaptic weight modification according to specific learning rules. But, most of the power consumption in all such networks results from synaptic communication, which can be greatly reduced using DWM based integrating-neurons.

V. DESIGN PERFORMANCE

Fig. 13 pictorially depicts the device-circuit co-simulation framework employed in this work to assess the system level performance for different neuromorphic architectures. The device models for neurons have been benchmarked with experimental data on LSVs' and DWM [1]-[7]. The corresponding behavioural models are used for circuit and system level simulation.

Fig. 14 shows the estimated energy benefits of the proposed design scheme over state of art CMOS for different applications. The large benefits for analog applications (2-3 orders of magnitude) comes from the fact that ultra low voltage pulsed operation of spin-based neurons greatly reduce the static power consumption resulting from conventional analog circuits. For applications involving binary signal processing more than 15X-30X lower computation energy has been estimated.

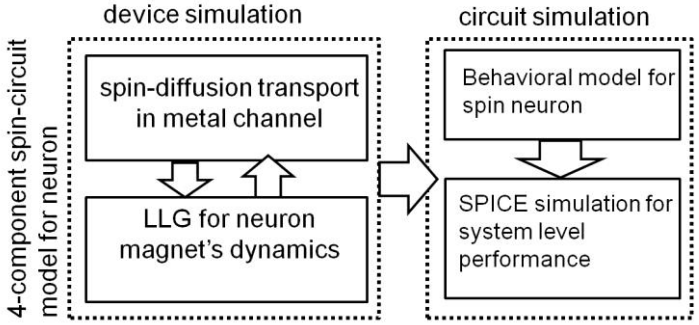

Fig. 13. Device circuit co-simulation framework employed in this work

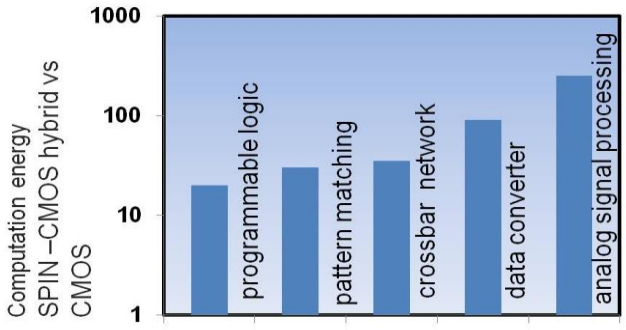

Fig. 14. Energy benefits of the proposed design scheme over CMOS for different applications.

VI. CONCLUSION

We proposed spin-based device models for neuron that can facilitate the deign of ultra-low power neuromorphic-computation hardware. We developed device-circuit co-simulation framework to assess the performance of heterogeneous neuromorphic designs that employ the proposed neurons. We obtained highly promising estimates for common data processing applications that show 20X-300X improvement in computation energy as compared to state of art CMOS design. The research presented in this work involves device-circuit-architecture co-design and can lead to a comprehensive design solution for neuromorphic hardware.

ACKNOWLEDGEMENT: This research was funded in part by Nano Research Initiative and by the INDEX center

REFERENCE

[1] M. Sharad et. al, "Spin-based Neuron Model with Domain Wall Magnets as Synapse", Ultra Low Power Neural Networks Using Spin Devices", IEEE Transaction on Nanotechnology, 2012.

[2] M. Sharad et. al., "Cognitive Computing with Spin Based Neural Networks", DAC 2012

[3] M. Sharad, "Spin-based Neuron-Synapse Units for Ultra Low Power Neural Networks", International Joint Conference on Neural Networks, 2012.

[4] M. Sharad et. al.,"Spin Neurons for Ultra Low Power Computational Hardware", DRC 2012.

[5] M . Sharad et. al., " Ultra Low Energy Analog Image Processing Using Spin Based Neurons", Nanoarch 2012.

[6] M. Sharad et. al, arXiv:1205.6022

[7] M . Sharad et. al, arXiv:1206.2466

[8] M. Sharad et. al, SRC, 2012, Publication IDs' : P062825, P063566, P063568, P063569.

[9] Wolfgang Maass, "Networks of spiking neurons: The third generation of neuralnetwork models" Elsevior Neural Networks, 1987.