

## SYNAPTIC MEMORY IN SEQUENTIAL CIRCUIT USING 10T SRAM IN NEURAL NETWORK

R.Sangeetha,  
PG Scholar,  
P.A. College of Engineering and Technology,  
Pollachi, Coimbatore, India.  
sangeethar0191@gmail.com

Mr. M. Yuvaraja,  
Head of ECE Department,  
P.A. College of Engineering and Technology,  
Pollachi, Coimbatore, India.  
yuvarajmuthusamy@gmail.com

**Abstract-**A hybrid analog/digital VLSI is implemented in a spiking neural network with programmable synaptic weights. Static Random Access Memory module is interfaced to a fast current-mode event-driven DAC for producing synaptic currents with the appropriate amplitude. These currents produce realistic temporal dynamics. SRAM cells acts as a transceiver and receiving asynchronous events in input then performing neural computation with hybrid analog / digital circuits on the input spikes. Eventually producing digital asynchronous events in an output. Input, output, and synaptic weight values are transmitted to/ from the chip using a common communication protocol based on the Address Event Representation (AER). Tanner tool is used to implement the neural architecture. The neural core representation is possible to interface with the device to a workstation or a micro-controller. It explore the effect of different types of Spike-Timing Dependent Plasticity learning algorithms. It is for updating the synaptic weights values in the SRAM module. To improve the efficiency of the output signal and to avoid the spikes a new set of neural core structure is used, it have a more number of filters are used . It is feed forward network based system. While using 10T SRAM in proposed work to reduced a power consumption and delay. Increase the efficiency of the system and more low pass filter in the neural core is used to reduce spike in a input signal.

**Keyword:** silicon neuron, silicon synapse, spike-timing dependent plasticity (STDP),

**spiking, static random access memory (SRAM), synaptic dynamics.**

### I. INTRODUCTION

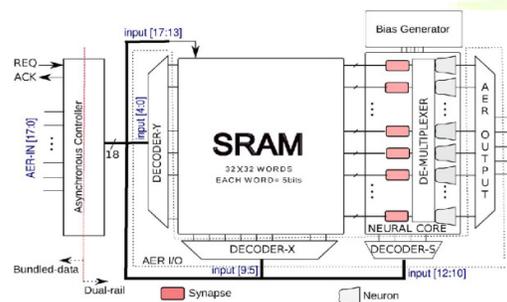
Spiking neural networks[5] represent a promising computational paradigm for solving complex pattern recognition and sensory processing tasks that are difficult to tackle using standard machine vision and machine learning techniques. The chip using a common communication protocol based on the Address Event Representation (AER) and Dependent Plasticity (STDP) learning algorithms for updating the synaptic weights[13] values in the SRAM module. In memory design using STDP design power dissipation is reduced Efficiency of memory store is increased. In this specific scenario, this suggests the design of full custom analog/digital Very Large Scale Integration (VLSI) neuromorphic systems [11]. However, to meet the requirement of real-time interaction with the environment, some of the recently proposed VLSI design solutions that operate only on “accelerated time” scales are not suitable. Similarly, neural VLSI solutions that focus on large-scale systems simulations are not ideal, as they compromise the low-power or compactness requirements. compact full-custom VLSI device that comprises low-power sub-threshold analog circuits and asynchronous digital circuits to implement networks of spiking neurons with programmable synaptic weights .

Implementation of neural computation is performed in the analog domain while the communication of spikes between neurons is carried out asynchronously in the digital domain. Specifically, the analog circuits

implement neural and synaptic dynamics in a very compact and power efficient way, while digital asynchronous circuits implement a real-time event (spike) based communication protocol[3]. We designed a new set of asynchronous circuits for interfacing the asynchronous events to conventional five-bit Static Random Access Memory (SRAM) cells, to manage the storage of the network's synaptic weight values the programmable SRAM cells can update the network's synaptic weights using the same asynchronous communication protocol used to transmit spiking events across the network. Also the idea of programming different parameters in spiking neural networks[9], such as synaptic weights[13], or even dendritic tree and synaptic routing structures, is not new. However, as these solutions typically require long settling times, they are not ideal for integration in circuits that employ fast asynchronous digital event-based communication circuits. Here we propose a solution that uses both SRAM cells and fast Digital to Analog Converters (DACs) interfaced to asynchronous digital circuits, to either set the synaptic weights.

## II. EXISTING TECHNIQUES

Neural computation is performed in the analog domain while the communication of spikes between neurons is carried out asynchronously in the digital domain.

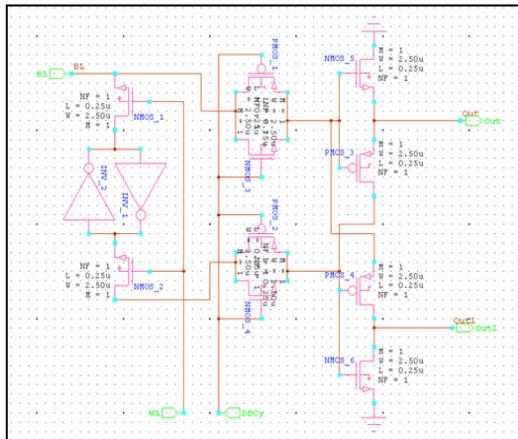


**Fig 1 Main Block Diagram Of Neural Network**

### A. SYNAPTIC MEMORY

The SRAM architecture is illustrated in Fig. 2. Two row and column decoders

receive five bits each, encoded in dual-rail, and generate a one-hot code at the output. A standard six-transistor circuit (6T SRAM design) is used to implement the memory cells. The memory array has  $32 \times 32$  words, each word comprising five bits. An output filter produces a dual-rail representation of the data. During idle mode, when there is no input, the Bitline and /Bitline signals are pulled up to VDD and the output of the filter circuits [b0.0 and b0.1 are both set to Gnd. During a "Read" operation, the X-decoder of the memory block selects a column (via the WL word-line); the Bitline and /Bitline signals of the five memory cells in the selected column are then set to values that correspond to the content of the five-bit memory word; and the Y-decoder enables the transmission gates of the corresponding row, thus allowing all the driven Bitline and /Bitline signals of the selected word to reach the input of the filter circuit. Finally, the filter circuit generates dual-rail data from the Bitline and /Bitline signals, setting either of the b0.0 or b0.1 lines to VDD, according to the content of the memory cell. The content of the memory block is programmed by setting the Write\_enable signal to VDD and transmitting the five bits that represent the content of the memory cells together with the standard address-event data. In this "write" mode the memory bits can drive the set of Bitline and /Bitline signals belonging to the row selected by the Y-decoder input data. As the X-decoder input data enables only one of the SRAM column WL wordlines, only the memory cell with the corresponding X- and Y-address will change its content. In addition to being stored in the 6T SRAM cell, the content of the memory word is also passed through to the neural-core, for producing synaptic currents[3] with the desired amplitude.

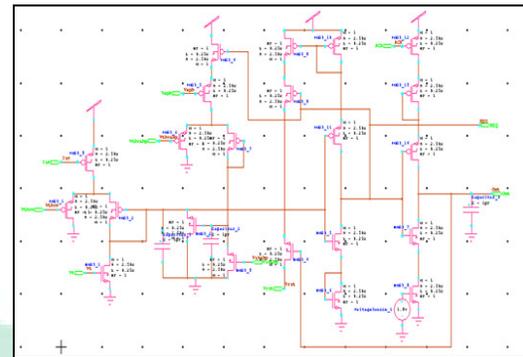


**Fig. 2. Memory architecture.**

## B. NEURAL CORE

The neural-core block comprises 32 Integrate-and-Fire (I&F) neurons, four synapse circuits (three excitatory and one inhibitory) per neuron, and a synapse address-demultiplexer circuit.

**1) Neuron Circuit:** The neuron circuit is the Adaptive exponential I&F neuron described in but with an extra free parameter corresponding to the neuron's reset potential. The circuit diagram of this new design is shown in Fig. 3. The neuron's input DPI integrates the input current until it reaches the neuron's threshold voltage. At this point there is an exponential rise due to the positive feedback in the silicon neuron's circuit that causes the neuron to generate an action potential. The membrane potential is then reset to the neuron's [10] tunable reset potential. extremely low power, consuming about 7 pJ per spike. In addition, the circuit is extremely compact compared to alternative designs, while still being able to reproduce interesting dynamics, such as spike-frequency adaptation[5]



**Fig 3. Neuron circuit**

**2) Synapse Circuit:** The synapse circuit includes three main functional blocks (see Fig. 4): a DPI to implement the synaptic dynamics; a DAC circuit to generate the appropriate weighted current fed in input to the DPI; and a validity-check circuit to activate the DAC when there is valid data at its input, and to produce an acknowledge signal fed back to the asynchronous controller[2]. As the output of the memory block generates valid DR representation data, the synapse validity-check block raises its PiXAck signal and feeds the memory content data to the DAC. The PiXAck signals of all synapses are wire-OR'ed together. The result is used by the asynchronous controller to complete the handshaking cycle. The type of synapse circuit selected depends on the address-event data sent to the neural-core S-decoder. The asynchronous data and control paths of SRAM and neural-core blocks are independent. For correct operation, the S-decoder output should be ready before the weight bits are sent to the synapse DAC. timing assumption that the Decoder-S data path is faster than the memory access-time. The memory access time includes both the decoding time and the time required for the Bitline signals to be driven by the memory control circuits.

The synapse DAC circuit is activated by both the Decoder-S output and the validity-check block. The five bits that encode the weight value control switches on a corresponding number of branches, each connected to a current source, programmable via the bias-generator block. In principle, for perfect binary encoding the current in each

branch should be twice as large as the current in the previous branch. But we chose to have five independent current sources in order to fine tune them and compensate for mismatch effects across the synapse population. The sum of the currents from the five branches of each synapse DAC produces the final current, used by the corresponding DPI synapse circuit[8]. We bias the DPI circuit in its linear range to implement a linear first-order low-pass filter.

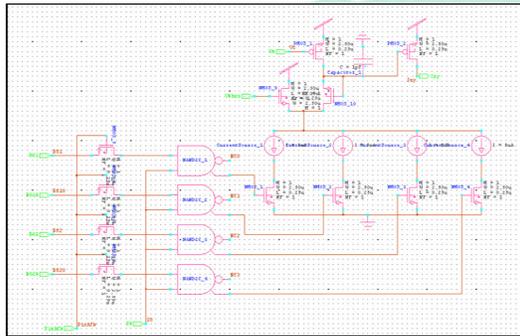


Fig 4 . Synapses circuit

time-division multiplexing it in time to integrate their independent contributions. The DPI output current will therefore be the integral of the weighted current pulses produced by the address-events sent to the memory-cells of the corresponding row.

### III MODIFIED 10T SRAM AND NEURAL CORE

10T SRAM cell performs better than 6T SRAM cell in terms of reliability and stability. 6T SRAM cell has less reliability at low supply voltage due to degradation in noise margins. Simple changes are made in neural core to reduce a power consumption. Increase the transistor level in low pass filter to reduce the spiking noise[4] in given input signal.

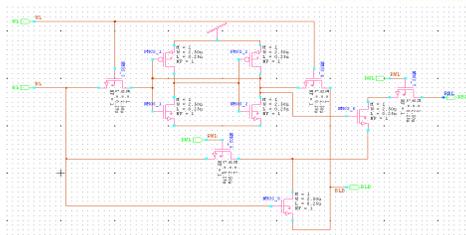


Fig 5. 10T SRAM

In read mode, WL is enabled and VGND is forced to 0 V while W\_WL remains disabled. The disabled W\_WL makes data nodes ('Q' and 'QB') decoupled from bitline during the read access. Due to this isolation, the read SNM of our 10T cell is almost same as the hold SNM of conventional 6T cell. Since hold SNM is much larger than read SNM in the 6T cell, read stability is remarkably improved in our 10T cell (Fig. 5). Depending on the cell data value, one of the bitlines starts discharging after WL is enabled. In our 10T cell, the read value is developed as an inverted signal of cell data. During write mode, both WL and W\_WL are enabled to transfer the write data to cell node from bitlines. As discussed in the introduction, weak writability is another major challenge for subthreshold SRAMs. Since our 10T cell has series access transistors, writability is a critical issue. In some previous subthreshold SRAMs  $V_{dd}$  is collapsed to enhance writability. However, it also degrades hold stability of the SRAM cells in other row sharing the line. To operate this technique successfully, each row should have individual line, resulting in large area penalty (more than 50% in thin-cell layout assuming poly pitch)

The neuron circuit is the Adaptive exponential I&F neuron but with an extra free parameter corresponding to the neuron's reset potential.

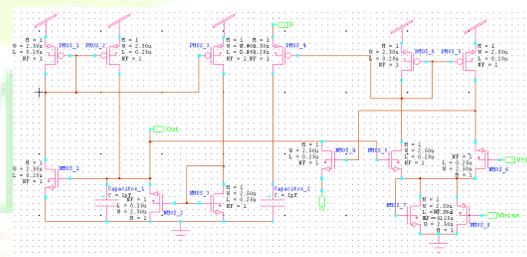
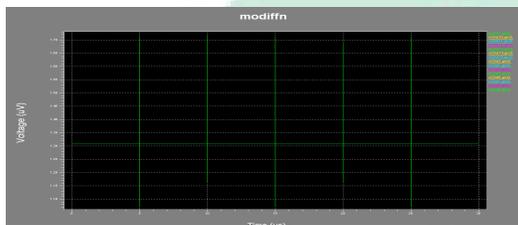


Fig 6. Modified neuron core

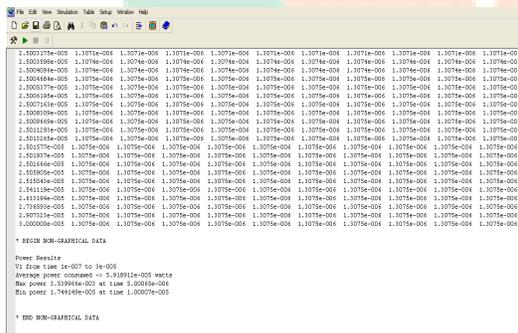
$P_1$  and  $P_2$  act as a digital low pass filter. Then  $P_3$  &  $N_3$  &  $N_2$  act as an amplifier using current mirror. current mirror is nothing but a transmission drain and gate is connected together.  $P_4$  &  $N_4$  act as reset circuit as well as  $P_6$  &  $P_5$ ,  $N_5$  &  $N_6$ ,  $N_7$  &  $N_8$  act as a low pass filter.

## V. RESULT AND DISCUSION

The proposed circuit is simulated using tanner in the high-performance 130-nm technology. The supply voltage used in the simulation is 0.5 V. S-Edit is a schematic capture tool that supports integrated analog simulation with automatic conversion from cadence and view draw schematics. User can run simulations and cross-probe from S-Edit, making the design process real-time and more efficient.



**Fig 7. Simulation result of neural architecture**



**Fig 8. Power Report of Neural Architecture**

| Structure         | Existing System    | 10T SRAM            | Modified neural Architecture |
|-------------------|--------------------|---------------------|------------------------------|
| Technology        | 130nm              | 130nm               | 130nm                        |
| CMOS              |                    |                     |                              |
| Supply voltage    | 1.8v               | 1.8v                | 1.8v                         |
| Power consumption | 4.12674e-002 watts | 2.2718 3e-003 watts | 5.91022e-002 watts           |

**Table 1. performance comparison**

From the table 1 it is to be known that the proposed neuron architecture consumes less power compared with other circuit blocks. In all the techniques the technology and the supply voltage are same. Due to the low power filter power consumed is reduced.

## VI. CONCLUSION

A novel neuromorphic VLSI device comprising both a spiking neural-core with biophysically realistic analog synapse and neuron circuits, as well as a fully asynchronous digital memory block. it is possible to integrate fast digital circuits next to very slow analog ones, using time constants that span over seven orders of magnitude, and to obtain remarkable performance figures with low mismatch. The proposed work is to improve the efficiency of the output signal and to avoid the spike a new set of neural core structure is used it have a more no of filter to eliminate the spikes in a given input signal. it is feed forward network based system. While using 10T SRAM reduced a power consumption and delay. Increase the efficiency of the system and more low pass filter in the neural core is used to reduce spike in a input signal.

## REFERENCES

1. Saber Moradi and Giacomo Indiveri 'An event based neural network architecture with an asynchronous synaptic memory', vol 8, No 1
2. A. Martin and M. Nystrom, "Asynchronous techniques for system-on chip design," Proc. IEEE, Vol. 94, pp. 1089–1120, 2006 asynchronous static random access memory," in Proc. IEEE Systems Conf., 2011, pp. 277–280.
3. C. Bartolozzi and G. Indiveri, "Synaptic dynamics in analog VLSI," Neural Comput. Vol. 19, no. 10, pp. 2581–2603, Oct. 2007.

4. E. Izhikevich, "Simple model of spiking neurons," IEEE Trans. Neural Netw., Vol. 14, no. 6, pp. 1569–1572, 2003.
5. J. Brader, W. Senn, and S. Fusi, "Learning real world stimuli in a neural network with spike-driven synaptic dynamics," Neural Comput., Vol. 19, pp. 2881–2912, 2007
6. M. Valle, "Analog VLSI implementation of artificial neural networks with supervised on-chip learning," Anal. Integr. Circuits Signal Process., Vol. 33, No. 3, pp. 263–287, 2002. pp. 130–133.
7. S. Moradi and G. Indiveri, "A VLSI network of spiking neurons with asynchronous static random access memory," in Proc. IEEE Biomedical
8. S. Ramakrishnan, R. Wunderlich, and P. Hasler, "Neuron array with plastic synapses and programmable dendrites," in Proc. IEEE Biomedical Circuits and Systems Conf., Nov. 2012, pp. 400–403.
9. S. - C. Liu and T. Delbruck, "Neuromorphic sensory systems," Current Opin. Neurobiol., Vol. 20, no. 3, pp. 288–295, 2010.
10. Y. Wang and S.-C. Liu, "Multilayer processing of spatiotemporal spike patterns in a neuron with active dendrites," Neural Comput., Vol. 8, pp. 2086–2112, 2010