Communications on Applied Electronics (CAE) – ISSN : 2394-4714
Foundation of Computer Science FCS, New York, USA
Volume 1 – No.7, May 2015 – www.caeaccess.org

An Improved Neural Network Design with Asynchronous Programmable Synaptic Memory

Vaishnavi.M
PG Student, ECE Department
Kalaignar Karunanidhi Institute of technology
Coimbatore, India

M.Jayasheela
HOD/PG, ECE Department
Kalaignar Karunanidhi Institute of technology
Coimbatore, India

ABSTRACT
The electrophysiological behavior of real neurons is emulated by the silicon neuron. The network of neurons helps to obtain accurate results for a complicated system which has a non-linear behavior. The network is integrated on a single VLSI device and implemented in various fields as Neural Network. Neural Network is comprises of Asynchronous circuit, Memory architecture, Neuron, and Synapse circuits. The fast access, connectivity and power hungry operation are based on the Asynchronous and memory circuits. Since the power consumption has become a major limiting factor in any VLSI design, the proposed work presents an efficient Neural Network Architecture whose power consumption is minimized by differential and symmetrical properties of the modified C-element in the controller and 10T SRAM cell in the memory Architecture respectively.

Keywords
Asynchronous, circuit, neural network, static random access memory (SRAM), very large scale integration (VLSI)

1. INTRODUCTION
Spiking neural networks represent a promising computational paradigm for solving complex pattern recognition and sensory processing tasks that are difficult to tackle using standard machine learning techniques. Many VLSI models of spiking neurons have been developed in the past and many are still being actively investigated [2]–[9]. The main goal is to integrate large numbers of these circuits on single chip, or even wafers, and create large networks of neurons, densely interconnected. The technique used to connect multiple neurons with each other is asynchronous digital circuits. It is therefore crucial and challenging to develop fast low-power circuits that implement faithful models of asynchronous circuits that manage the communication pattern.

Earlier developed circuits represent Networks of Integrate and Fire [3]-[7] neurons were not used with memories for accessing synaptic weights. The neural network exhibit a wide range of useful computational properties such as feature binding, pattern recognition, segmentation, etc.

A compact full-custom VLSI device that comprises low-power asynchronous digital circuits to implement networks of neurons with programmable synaptic weights is proposed. Earlier developed circuits [1] are low power, but it operates in "scaled-time". The main aim of using neural network is that it is fast. Therefore a new set of asynchronous circuits for interfacing the asynchronous events to conventional five-bit Static Random Access Memory (SRAM) cells, are designed to manage the storage of the network's synaptic weight [1] values. Earlier proposed methods make use of SRAM cells as digital memory storage for synaptic weights in neuromorphic chips. However, as these solutions typically require long settling times and power consuming, they are not ideal for integration in circuits that employ fast asynchronous digital communication circuits. A solution that uses fast low power SRAM cells interfaced to efficient asynchronous digital circuits is proposed. Asynchronous Controller circuits are also modified for effective and efficient operation of the network.

C-element in the Asynchronous Controller is modified on its representation to reduce the power consumption. The differential operation of the proposed C-element eliminates unwanted noises.

The leakage power in the memory block of the neural network is reduced by implementing 10T SRAM cell instead of the existing 14T SRAM memory block. The proposed architecture reduces leakage power for about 16%. Improves the Static Noise Margin as it contains separate path for read and write operation. Jeopardize nature of Static Noise Margin and Write Trip point. Bit-line leakage is reduced in Read'1' mode s additional stacking transistor M9 is used. Functions in Subthreshold Region also. Uses Bit-interleaving Word Organization which decreases the power consumption. Read SNM and Write Margin is improved.

2. THE CHIP ARCHITECTURE
The architecture of the chip is illustrated in Figure 1. The chip was fabricated using AMS 0.35µm CMOS technology. It comprises five main blocks: the Asynchronous Controller, SRAM block, the Neural-core, the Bias Generator and the AER Input/output interfaces [1].

The asynchronous controller manages the communication between the external digital asynchronous signals and the on-chip ones. The asynchronous SRAM block is used to store synaptic weight values with the inclusion of a filter circuit that generates a dual-rail representation of the data. The neural core block consists of a column of 32x1 adaptive integrate-and-fire neurons and an array of 4x32 synapses with DAC circuits to convert the digitally encoded weight into an analog current.

From the figure, it is clear that the external input signals are encoded using a Bundled Data (BD) representation with an 18 bit wide bus for the data and two additional lines for the control signals (REQ, ACK). In the input signal 10 bits encode the X- and Y-addresses of the memory cells, five bits for synaptic weight values and three bits for synapse type to be used.

Communications on Applied Electronics (CAE) – ISSN : 2394-4714
Foundation of Computer Science FCS, New York, USA
Volume 1 – No.7, May 2015 – www.caeaccess.org

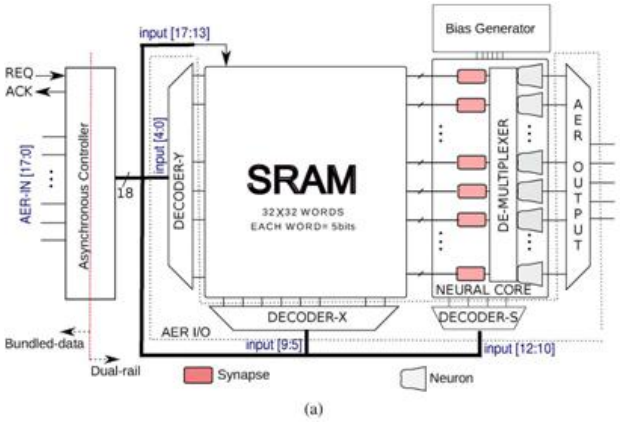


Fig 1: Chip block diagram

3. EXISTING WORK

The neural network architecture has five main blocks namely asynchronous controller, SRAM block, neural core, and Bias generator and I/O interfaces [1]. The main disadvantage of the network architecture is that it consumes more area, time and power. The main block involved in the power consuming operation are memory and asynchronous circuits. Thus to eliminate the leakage power a fast and low power asynchronous controller and SRAM block is proposed. Remaining circuits of the neural core, bias generator and the synapse are considered to be the same as past implementations [1],[11]-[15].

The conventional asynchronous controller uses the C-element and latches. The main function of this block is the translation from the BD representation to the DR. The C-element used in the asynchronous controller has a major drawback of high power consumption and delay. The Asynchronous circuit is shown in Figure 3. This circuit uses the main access circuit as C-element.

The SRAM block uses 6T SRAM cell, transmission gates and filter circuit that represents the dual rail data. The 6T SRAM cell has a disadvantage of power consumption due to looping of the inverters in write operation. Hence a 10T cell is proposed to overcome the disadvantages and it replaces the whole block of memory that contains about 14 transistors. Hence the proposed work also reduces the area.

3.1 C-Element

C-elements [16] are sequential logic devices that operate as event synchronizers.. The C-element has output switching only when all inputs are at the same logic value. If inputs A and B are same values, output Q produce a value. When inputs are different, Q keeps its previous value. The conventional CMOS C-element implements the RESET function as a pull-up network (PUN) and the SET function as a pull-down network (PDN), which is then inverted to obtain the actual output. To achieve the hysteresis state-holding functionality a weak feedback inverter is used.

Figure 2 shows the structure of the C-element. Correct sizing is advisable for this type of gate to function correctly. The PUN and PDN must be strong enough to overcome the contention current from the weak feedback inverter.

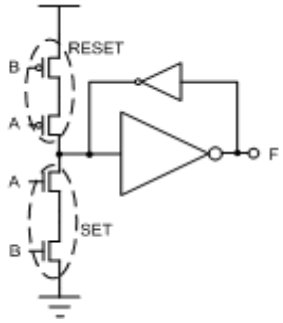
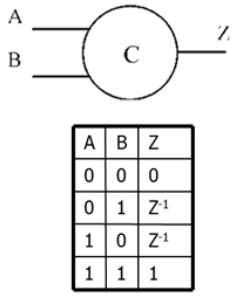


Fig 2: Conventional C-element

Another undesirable characteristic of the conventional gate is that as supply voltage is reduced, it will cease to operate correctly sooner than the static design due to the relative strengths of the PUN, PDN, and weak inverter being affected by the reduced supply voltage.



A	B	Z
0	0	0
0	1	Z^{-1}
1	0	Z^{-1}
1	1	1

Fig 3: Symbol and Truth Table for C-element

Figure 3 shows the conventional C-element and the truth table that explains the operation of the architecture. The C-element is implemented in the Asynchronous Controller Circuit in the Neural Network.

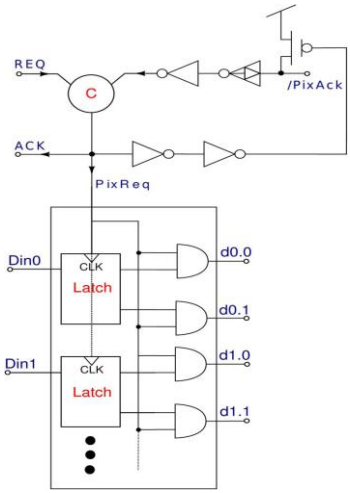


Fig 4: Asynchronous Controller Circuits

Figure 4 shows the Asynchronous Controller Circuits that has the REQ and ACk signals that are asynchronous in nature. The work of the controller is to synchronize the signals and produce an ACK signal to the synapse circuit.

3.2 Conventional SRAM Cell

Figure 5 shows the Memory Architecture and the 6T cell, Transmission gates and output circuits. The SRAM architecture has 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 is used to implement the memory cells. The memory array has 32 X32 words, each word comprising five bits. The circuit operation is explained [1],[16]. The main disadvantage in this architecture is that in a non partitioned scheme, every column will have an active memory cell, uselessly discharging the bit-lines of the un-accessed columns and resulting in wasted power needed to pre-charge these bit-lines back to their original value. During write operation unwanted switching between cross coupled inverters causes power dissipation.

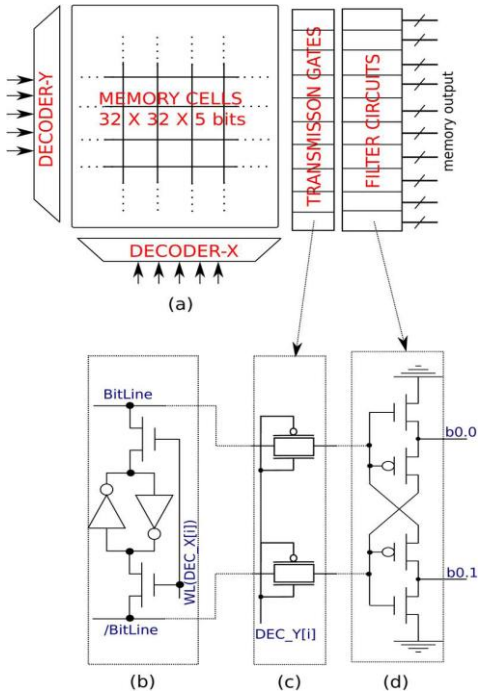


Fig 5: (a) Memory Architecture. (b) 6T Cell (c) Transmission Gate (d) Memory Output filter Circuit to produce DR

4. PROPOSED WORK

The neural network architecture has five main blocks namely asynchronous controller, SRAM block, neural core, and Bias generator and I/O interfaces [1].

4.1 Proposed C-Element

To avoid these flaws the C-element is designed to operate in low supply voltage, low power and high speed. The proposed circuit is shown in Figure 5. In this circuit, the output state is maintained through a feed-back conducting path of three transistors in the pull-up tree or the pull-down tree. This circuit is completely ratio less.

An advantage of this implementation is that it is symmetrical with respect to the inputs. For the circuit to have the same pull-up and pull-down resistance the normal N-tree and P-tree transistors, except those of the output inverter, must be made half the size.

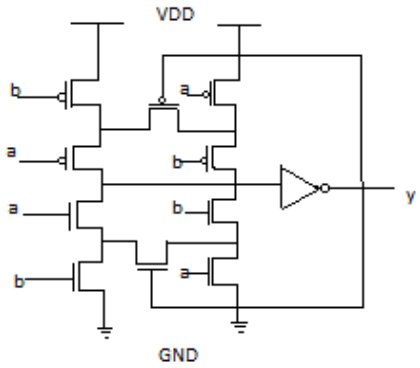


Fig 6: Proposed C-element

Proposed C-element is definitely the right choice for low-power and high-performance applications because it has the advantages of both: minimum overhead for latching with no resistance against output switching. Furthermore, implementation has a nice symmetrical topology with respect to the inputs. The results obtained through this study are not limited to the C-element and may be extended to similar circuits in which the state of their outputs must be latched. For example, this comparison demonstrated that minimizing the number of transistors dedicated to latching and avoiding topologies that resist output switching may result in significant energy savings.

4.2 Proposed SRAM Cell

Figure 7 shows the Proposed Memory Architecture with 10T SRAM cell representation. The proposed SRAM contains 10 transistors. It contains 10 transistors.T1-T3 and T2-T4 are cross coupled inverters. T5 & T6 are transistors for accessing; T7 & T8 constitute a separately used for read operation. T9 & T10 act as switches that connect the read path to any one of the write bit lines depending upon the last written data. M5 & M6 are controlled by word line (WL) signal which is asserted by the DECODER X circuit during write cycle only. T7 & T9 are controlled by read word line (RWL) signal which is asserted by the DECODER Y circuit. Hence when we use 10t SRAM cell instead of the earlier used memory architecture, we have a great area minimization that eliminates the use of transmission gates and the filter circuits.

Write operation of the proposed 10T SRAM cell begins with WL going high which turns on T5 M6 transistors and write operation is performed similar to that of conventional 6T SRAM cell. Read operation begins first pre-charging RBL line to full swing voltage & after that DEC-Y signal is asserted.

This turns on T7and T9 transistors. When Q="1", so M8 will be off & hence no discharge current flows through read path, but when Q="0" then T8 will be on & hence RBL will discharge through T7, M8, M9/M10 to BL/BLB. This decrease in voltage of RBL is detected by sense amplifier.

In order to reduce the read power, the swing of read bit line (RBL) is reduced. During read 0 operation charge is shared between RBL and BL or RBL and BLB depending on the fact that the last written data was logic 0 or 1 respectively. Due to charge sharing RBL does not discharge completely & stays at mid level voltage & so in the next cycle pre-charge circuitry consumes less power to drive the bit line from mid level value to the full swing voltage.

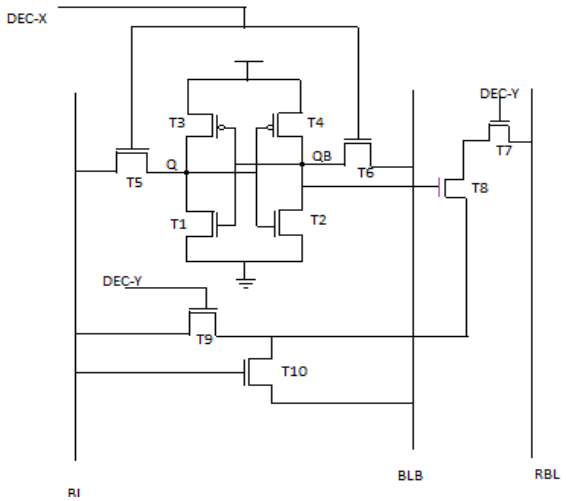


Fig 7: Proposed Memory Architecture With 10T SRAM Cell Representation

For write power reduction the proposed technique does not use pre-charge circuitry for the write bit lines instead a write driver is used that drives the bit lines high or low depending upon the data value. Write power reduction is possible only after read 0 operation, this can be understood as consider a write 0 operation followed by a write 1 operation. In this case initially BL=0 and BLB=1. During read 0 the read discharge flows through read path & T9 to BL & this charges BL partially. In next cycle of write 1 the write driver has to drive BL to logic 1 from an intermediate level voltage & this reduces the write power.

The proposed C-element and memory architecture is implemented in the Neural Network Architecture and the results are shown in next section.

5. SIMULATION RESULTS

The simulation work of both the existing and proposed circuits of the asynchronous controller and SRAM cell is done in the tanner EDA tool and the results are tabulated.

Figure 7 shows the simulated output of the proposed Asynchronous Controller

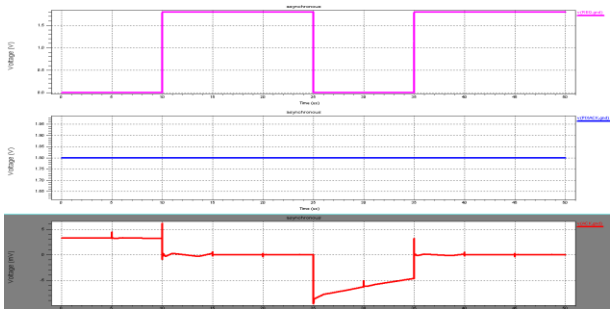


Fig 8: Simulated Output of the Proposed Asynchronous Controller

Figure 8 shows the simulated output of the proposed Asynchronous Controller. From the figure, it is clear that as the REQ and ACK arrives, the controller processes the input and provides trigger pulse for the SRAM architecture for storing the synaptic weight.

The proposed Circuit of SRAM cell is represented as Symbol and then instantiated for 32×32 layout to be organized for the Memory Architecture. This chip architecture can be implemented in any Complex Circuits that leads the user to obtain fast and accurate result of the system. The main application of the Neural Network chip are Artificial Intelligence and Robotics.

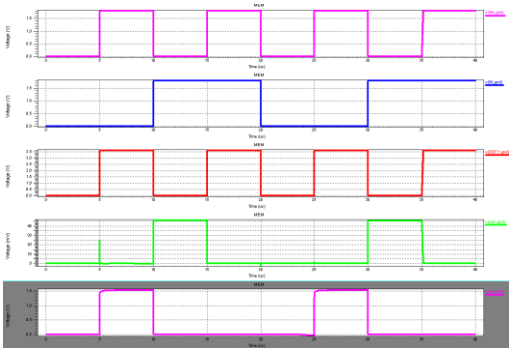


Fig 9: Simulated Output of the Proposed SRAM cell

Figure 9 shows the simulated output of the proposed SRAM cell. From the figure it is observed that the operation of the SRAM cell depends on the output of the decoder circuits and the Asynchronous Controller. The inputs to the SRAM cell are DECODER-X, DECODER-Y, Bit-line and Word-line and the output of the SRAM is sent to the synapse circuit to select the synapse as per the weights stored.

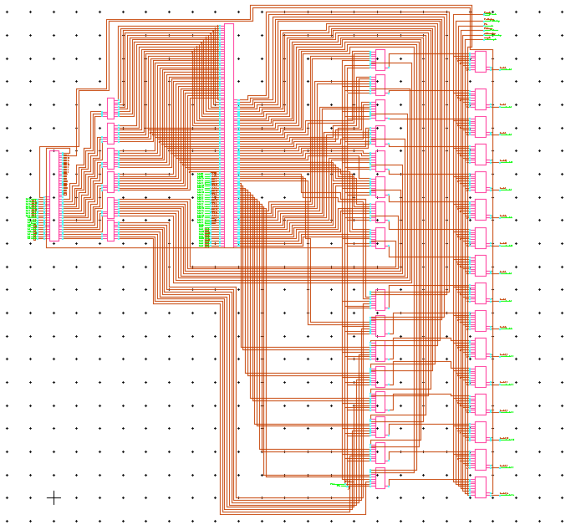


Fig 10: Proposed Neural Network Architecture

Figure 10 shows the Proposed Neural Network Architecture .The proposed Circuit of SRAM cell is represented as symbol and then it is instantiated for 32×32 layout to be organized for the Memory Architecture. Similarly the symbols of Synapse circuit, Neuron Circuit are obtained and are interconnected using wires to obtain the whole Neural Network Chip Architecture.

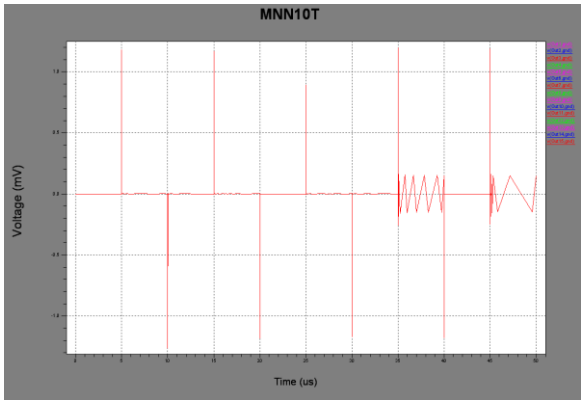


Fig 11: Simulated Output Waveform of the Proposed Neural Network

From the figure it is clear that the response of the neuron circuit to the input current from the synapse. the plot is done between the V_{mem} and the time in seconds spikes are obtained at frequent intervals of time . The neuron circuit integrates the input current until the threshold voltage is reached and resets the membrane potential voltage to reset potential applied to the circuit. So the spikes appear at equal intervals of time.

Table 1 Comparison results of Existing and Proposed Asynchronous controller

ASYNCH CONTOLLER	AVG POWER	MIN POWER	MAX POWER	DELAY
EXISTING	1.79x 10^{-2} w	2.05x10^{-2} w	1.78x10^{-2} w	3x10^{-5} sec
PROPOSED	3.15x10^{-4} w	4.08x10^{-4} w	2.39x10^{-9} w	2.5x10^{-5} sec

Table 1 shows the results of comparison between the conventional and the proposed circuit of the asynchronous controller and it infers that the proposed asynchronous controller reduces the power about 98% and the delay is reduced about 17%.

Table 2 Comparison Results of Existing and Proposed SRAM cell

SRAM CELL	AVG POWER	MIN POWER	MAX POWER	DELAY
EXISTING 14T	1.903x10^{-4} w	2.22x10^{-9} w	1.59x10^{-3} w	5x10^{-6} sec
PROPOSED 10T	1.39x10^{-9} w	1.03x10^{-10} w	1.25x10^{-5} w	3x10^{-5} sec

Table 2 shows the power results of the existing and proposed SRAM cell. the table infers that proposed SRAM cell reduces the power about 98% and the delay is reduced about 94%.

Table 3 Comparison results of existing and proposed neural network architecture

NEURAL NETWORK	AVG POWER	MIN POWER	MAX POWER	DELAY
CONVENTIONAL	4.12x10^{-2} w	3.5x10^{-1} w	3.25 x10^{-4} w	5x10^{-6} sec
MODIFIED	2.27x10^{-3} w	2x10^{-2} w	2x10^{-7} w	4x10^{-5} sec

Table 3 infers that the power consumption of the proposed Neural Network is reduced to about 95% and the delay is reduced to an extent of 20%.

6. CONCLUSION AND FUTURE WORK

An improved Neural Network based Asynchronous Programmable Synaptic Memory is proposed. The proposed block of the Asynchronous Controller and SRAM cells are implemented in Neural Network architecture. The proposed Network operates in low-power, delay- minimized, and accurate operation but while considering the existing and proposed Asynchronous Controller block the proposed block has increase in area. But this drawback is hidden when implemented in the whole architecture as the proposed SRAM cell has reduction in area than the existing SRAM cell. Due to less power dissipation, the proposed chip architecture has significant cost saving. The Performance comparison results shows that the leakage power is rated down about 95% and reduces the delay of the operation about 86% while at the same time satisfying the compactness and compatibility with asynchronous logic constraints.

The future work is concentrated on comparing various previously used neuron circuits and to present a complete low power, compact, high speed neuron circuit for implementing in Neural Network.

7. REFERENCES

[1] An Event-Based Neural Network Architecture With an Asynchronous Programmable Synaptic Memory Saber Moradi, Student Member, IEEE, and Giacomo Indiveri, Senior Member, IEEE 1932-4545/$31.00 © 2013 IEEE

[2] C. Mead, Analog VLSI and Neural systems. Reading, MA: Addison- Wesley, 1989.

[3] M. Mahowald and R. Douglas, "A silicon neuron," Nature, vol. 354, pp. 515–518, 1991.

[4] A. van Schaik, "Building blocks for electronic spiking neural networks," Neural Networks, vol. 14, no. 6–7, pp. 617–628, Jul–Sep 2001.

[5] K. Hynna and K. Boahen, "Space–rate coding in an adaptive silicon neuron," Neural Networks, vol. 14, pp. 645–656, 2001.

[6] G. Indiveri, "A low-power adaptive integrate-and-fire neuron circuit," in Proc. IEEE International Symposium on Circuits and Systems. IEEE, May 2003, pp. IV–820–IV–823.

[7] L. Alvado, J. Tomas, S. Saighi, S. Renaud-Le Masson, T. Bal, A. Destexhe, and G. Le Masson, "Hardware computation of conductance-based neuron models," Neurocomputing, vol. 58–60, pp. 109–115, 2004.

[8] M. Simoni, G. Cymbalyuk, M. Sorensen, and S. Calabrese, R.L. De- Weerth, "A multiconductance silicon neuron with biologically matched dynamics," IEEE Transactions on Biomedical Engineering, vol. 51, no. 2, pp. 342–354, February 2004.

[9] J. Schemmel, K. Meier, and E. Mueller, "A new VLSI model of neural microcircuits including spike time dependent plasticity," in Proceedings of the IEEE International Joint Conference on Neural Networks, vol. 3. IEEE, July 2004, pp. 1711–1716.

[10] J. Arthur and K. Boahen, "Recurrently connected silicon neurons with active dendrites for one-shot learning," in IEEE International Joint Conference on Neural Networks, vol. 3, July 2004, pp. 1699–1704.

[11] E. Farquhar and P. Hasler, "A bio-physically inspired silicon neuron," IEEE Transactions on Circuits and Systems???I, vol. 52, no. 3, pp. 477– 488, March 2005.

[12] K. M. Hynna and K. Boahen, "Neuronal ion-channel dynamics in silicon," in 2006 IEEE International Symposium on Circuits and Systems, May 2006, pp. 3614–3617.

[13] J. Arthur and K. Boahen, "Synchrony in silicon: The gamma rhythm," IEEE Transactions on Neural Networks, vol. 18, pp. 1815–1825, 2007.

[14] J. Wijekoon and P. Dudek, "Compact silicon neuron circuit with spiking and bursting behaviour," Neural Networks, vol. 21, no. 2–3, pp. 524–534, March–April 2008.

[15] M. Singh and S. Nowick, "High-throughput asynchronous pipelines for fine-grain dynamic datapaths," in Proc. IEEE 6th Int. Symp. Advanced Research in Asynchronous Circuits and Systems, 2000, pp. 198–209.

[16] C-element - Wikipedia, the free encyclopedia en.wikipedia.org/wiki/C-element.