aVLSI Array of Spiking Neurons with Dynamical Synapses

Yu M. Chi

***Abstract*— A design for a spiking neuron with dynamical synapses is presented. Both the neuron and synapse are intended for a large scale neural network connected through an AER bus. The neuron design is a standard integrate-and-fire model with the appropriate interfacing circuits. The synapse, however, models several features from biology including a conductance, between the membrane voltage and synaptic reversal potential, that dynamically changes in response to spike frequency.**

I. Introduction

At present, biological computational systems exceed the performance of digital computers by orders of magnitude in terms of power efficiency throughput and intelligence, particularly in areas like sensory systems and cognition. With this in mind and taking inspiration from nature, researchers have been focused on creating silicon based systems with the same properties as biology. One key area is properly emulating the neuron, synapse and their interactions.

In biology, synapses are an integral part of neural systems by connecting neurons to each other and enabling communications. Replicating the connectivity of synapses and their behavior, however, is challenging due to the fundamental differences between VLSI chips and the brain. In contrast to the intricate degree of interconnectivity in the brain, signaling in VLSI is limited to just a few metal layers. Creating point-to-point connections between different sites in a large array is impossible and completely precludes the ability to dynamically redefine links.

However, communications in electronic systems is fast, much more so than biological neural processes. In arrayed structures like memory or imaging systems, random access of each individual element in the array is possible by row-column addressing one cell at a time through a shared bus [1]. This can be extended to emulate synaptic connections by time-multiplexing neural events on a single high speed bus (Fig. 1). As long as bus level communications is much faster than the operations inside each cell, time domain multiplexing is not a problem and allows for the realization of biologically realistic connections between neurons by simply having the control logic route the appropriate signals from cell to cell as needed.

The task then becomes to design the actual cells in this AER array - it must be simultaneously compact and power efficient to allow for a high degree on integration while persevering all of the necessary features of real biological neurons [3] and synapses [4].

II. System Description

A. Neuron

The neuron used follows the design previously published [3]. The neuron consists of a input amplifier (M1-M2), reset logic (M3-M7), and AER interfacing circuitry (M8-M10) and is shown in Figure 2.

The neuron is first reset when the digital signals R_{ACK} and C_{ACK} are brought high from the peripheral circuits and bring the membrane voltage to ground through M4 and M3 (and also turning off M5 and M6). At this point the input amplifier acts as a standard common source amplifier with an active load. Synaptic currents raise the voltage at the gate of M1 by charging the membrane capacitance. The node at the drain of M1 and M2 begins to drop when the gate source overdrive voltage of M1 becomes large enough to generate a current greater than the bias through M2.

Subsequently, the gate voltage on M7 begins to drop, causing current to flow through M5 and M6 and onto the membrane capacitor, closing the positive feedback loop and causing the neuron to spike. The spiking threshold is set by the terminal at the source of M1 and V_{bias} which both determine the overdrive voltage needed on the gate of M1 to generate sufficient current to activate the positive feedback. When this occurs, the membrane voltage is rapidly charged through M5-M7, completely independent of the externally injected current.

After spiking, V_{mem} is held in a low state, turning on M9 and M10 which are connected to the column and row level interface circuits. The row request signal is immediately pulled high through M10 indicating a service request somewhere in that row. When the external control is ready, it asserts row scan low which then allows the periphery to localize the spike. Afterwards the neuron is again reset, completing the cycle for a single spike.

B. Synapse

The basic equation for ion flow within through a synapse can be described as,

$$I_{syn} = g_{syn} r(t)(V_{mem} - V_r), \tag{1}$$

where I_{syn} is the synaptic current, V_r is the synaptic reversal potential and $g_{syn}r(t)$ is the dynamical conductance which evolves over time. As shown in previous publications [5], the variable $r(t)$ is described by,

$$\frac{dr}{dt} = \alpha[T](1 - r) - \beta r. \tag{2}$$

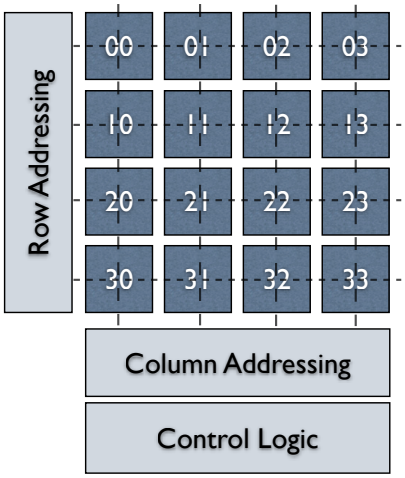

Fig. 1. Basic structure of an array of spiking neurons with an AER interface.

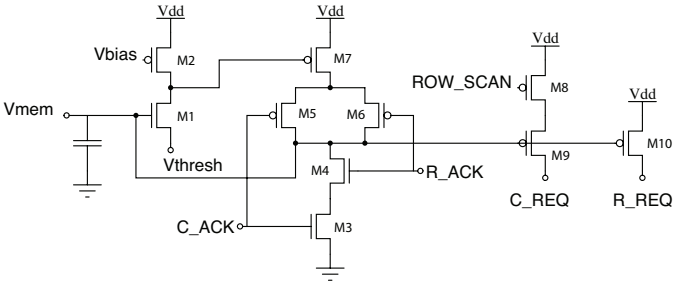

Fig. 2. Integrate and Fire Neuron [3] Schematic

The terms α and β are constants and $[T]$ is represents the neurotransmitter concentration.

For a pulsed input $[T]$ at time t_0, $r(t)$ has two closed form solutions [5]:

$$r(t-t_0) = r_\infty + (r(t_0) - r_\infty)e^{\frac{-(t-t_0)}{\tau_r}} \tag{3}$$

during the $[T]$ pulse and

$$r(t-t_1) = r(t_1)e^{-\beta(t-t_1)} \tag{4}$$

starting at time t_1 when the pulse ends. The terms β and τ_r represent time constants and r_∞ is a constant. In addition, the terms $r(t_0)$ and $r(t_1)$ are the initial conditions right before and after the pulse, respectively.

In the interest of designing circuits that emulate this behavior, it is necessary to obtain an intuitive understanding of the mechanics of the $r(t)$ equation and begin translating the appropriate variables and operations to voltages, currents and transistors. To this end, the first conductance equation can be seen as a sum of RC step responses to incoming spikes in $[T]$ followed by an RC decay during the interspike interval.

Therefore, to a first order approximation, the synapse circuit must sum an incoming spike, generate an exponential decay and use that to modulate a transconductance between the membrane voltage and synaptic reversal potential. A complete schematic of the circuit achieving these aims is shown in Figure 3. The design follows the general layout of the DPI

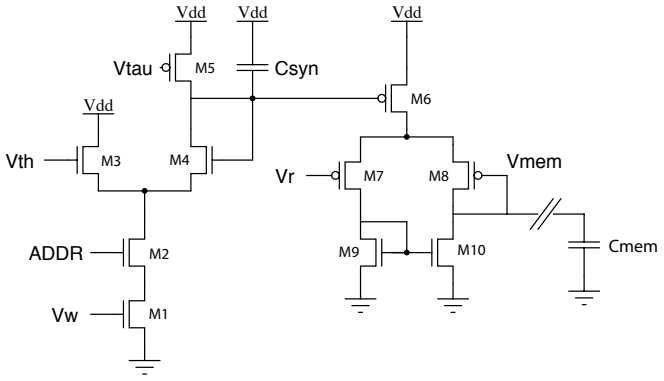

Fig. 3. Conductance Based Synapse Schematic. V_{mem} and C_{mem} are shared with the neuron.

synapse previously presented [2], but with simplifications and the addition of a full OTA conductance stage.

The synapse is activated through M1 and M2, whose gates are connected to the column and row level control logic. One axis sets an analog value, V_w, on the gate of M1. The other axis then turns on the synapse through the digital pulse signal ADDR which causes current to flow through M4 and discharge the bottom plate of C_{syn} (assuming that the gate of M4 is still much higher than V_{th}).

For a V_w below the threshold voltage the effect of a pulse with a duty cycle of, t_p, is,

$$V_{syn} = V_{syn}(t_0) - \frac{I_s e^{\frac{\kappa_n V_w}{V_T}}}{C_{syn}} t_p. \tag{5}$$

The term $V_{syn}(t_0)$ is the pre-event voltage and the entire operation can be modeled as simply a step from the initial voltage, for spike duty cycle much shorter than the spike period.

The PMOS, M5, subsequently recharges C_{syn} with a constant current set by the bias value V_τ. Assuming that $I_{M1} >> I_{M5}$, V_{syn} now ramps back towards the supply rail,

$$V_{syn}(t) = V_{syn}(t_1) + \frac{I_s e^{\frac{\kappa_p (V_{dd} - V_\tau)}{V_T}}}{C_{syn}} t. \tag{6}$$

At this point, the sum and decay of incoming action events to the synapse has been represented by a linear step and linear ramp. To examine the full effect of a series of input spikes and the dynamical behavior of the synapse, it is now necessary to consider the effect of M5 and the resulting current generated through M6 by the voltage on V_{syn}. The transistors M4 and M5 form a differential pair with the gate of M4 tied to C_{syn} and the gate of M5 set to a bias value, V_{th}. Assuming an initial condition where the synapse is at rest and C_{syn} is fully charged to V_{dd}, pulses through ADDR cause the voltage on C_{syn} to drop, by an amount determined only by V_w, the pulse duty cycle and pulse frequency. Without M5, the voltage on C_{syn}, for a sufficiently high frequency spike input, will cause the gate of M6 to drop enough such that M6 enters strong inversion and eventually saturate the synapse.

With the differential pair in place and a spike frequency high enough to cause C_{syn} to discharge the gate of M4 close to V_{th}, the effect of M5 becomes significant and the current onto the capacitor is,

$$I_{M4} = I_{M1} \frac{e^{\frac{\kappa V_{syn}}{V_T}}}{e^{\frac{\kappa V_{syn}}{V_T}} + e^{\frac{\kappa V_{th}}{V_T}}}. \quad (7)$$

The differential pair implements a means to ensure that the synapse always operates in weak inversion by gradually limiting the value V_{syn} to V_{th} and also to lowpass filter the temporal response of the synapse to high frequency input spike trains.

The voltage generated on to C_{syn} is now converted to a current on M6. For the discharge phase,

$$I_{M6} = I_s e^{\frac{\kappa_p (V_{dd} - V_{syn}(t_1))}{V_T}} e^{\frac{\kappa_p (V_{dd} - \frac{I_{M5}}{C_{syn}} t)}{V_T}}, \quad (8)$$

where the first exponential corresponds to the initial $r(t_1)$ constant and the second is the exponential decay with a time constant governed by the current from M5 (set via V_τ).

At this point, we have modeled the response of $g_{syn} r(t)$ with the current through M6 - all that is left is to model the actual conductance between the membrane voltage and synaptic reversal potential. The easiest way is to simply use a transconductance amplifier with the output connected back to the input. Conveniently, the G_m of a standard CMOS opamp/OTA is set by it's bias current (supplied through M6 in this case).

Analyzing the operation of M6-M10 is easy and follows standard analog circuit design. Ignoring the feedback between the output and M8 for the moment and assuming all transistors are operating in sub-threshold, the current through M8 is

$$I_{M8} = \frac{I_{M6}}{1 + e^{\kappa \frac{V_{mem} - V_r}{V_T}}}. \quad (9)$$

Similarly, the current through the other branch is

$$I_{M7} = \frac{I_{M6}}{1 + e^{\kappa \frac{V_r - V_{mem}}{V_T}}}. \quad (10)$$

The current through M7 flows through the diode connected M9 which mirrors the current to M10. Hence the current flowing out of the drains of M8 and M10 is the difference between the preceding equations. When V_{mem} is approximately $4V_T$ less than the synaptic reversal potential, V_r, all of the current flows directly through M8. When the membrane voltage nears the reversal potential, the current output can be linearized and written as,

$$I_{syn} = g_{syn}(V_r - V_{mem}), \quad (11)$$

where the differential conductance, g_{syn}, is determined by

$$g_{syn} = \frac{I_{M6} \kappa}{2V_T}, \quad (12)$$

the same exponential current pulse from M6.

Connecting the output V_{mem} to the gate of M8 simply converts the transconductance amplifier into a conductor with a conductance modulated by the exponentially decaying pulse generated through M6 and completes the synapse. Current stops flowing the membrane capacitance when the membrane voltage is equal to the synaptic reversal potential (although a spike should have already occurred at this point).

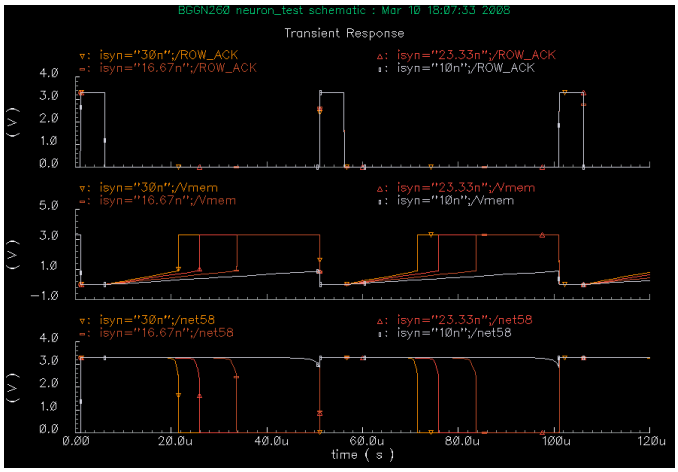

Fig. 4. Simulation results for the neuron. The top trace is the external reset signal, the middle trace the membrane voltage, and the bottom trace is the output of the common source amplifier.

III. Results

The aforementioned circuits for the neuron and synapse were implemented in a $0.5\mu M$ CMOS process and simulated.

The neuron was tested by setting the source of M1 to ground and injecting a constant current ranging from $10nA$ to $30nA$ onto the membrane capacitance. In each case a ramp was observed on V_{mem} from the current until approximately $900mV$ where the positive feedback loop activated and caused a fast spike, as expected (Fig. 4).

For the synapse, a constant spike train of $1.25KHz$ was used to stimulate the synapse. This in turn generated a steady state saw-tooth wave on C_{syn}. The exponential current pulse through M6 can be seen on Figure 5. As expected, the actual current into the membrane capacitance is a function of the difference between V_{mem} and V_r.

The entire synapse-neuron pair was simulated by testing its response to different spike frequencies (Fig. 6). As expected, higher frequency spike trains cause the neuron to spike earlier than lower frequencies. The difference is accentuated by the summing action on C_{syn} (causing an exponentially increasing current to be generated through M6).

IV. Conclusion

A neuron with dynamical synapses is presented. The synapse models several key features of biological synapse including the dynamical conductance between the synaptic reversal potential and the membrane voltage. The next step is design the AER peripheral interface circuits and test an array-wide of spiking neurons.

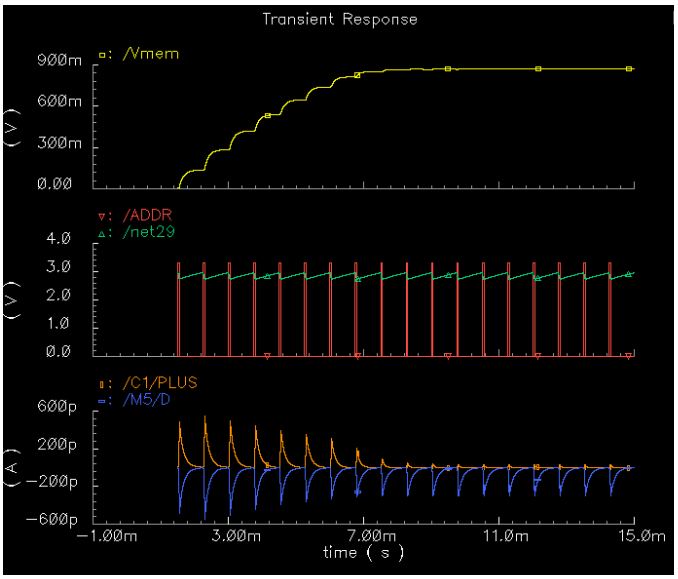

Fig. 5. Simulation results for the synapse. The top trace is the membrane potential. The second trace overlays incoming action event pulses with the voltage on C_{syn}. The bottom trace shows the current through $M6$ (positive going) and the current onto the membrane capacitance (bottom going).

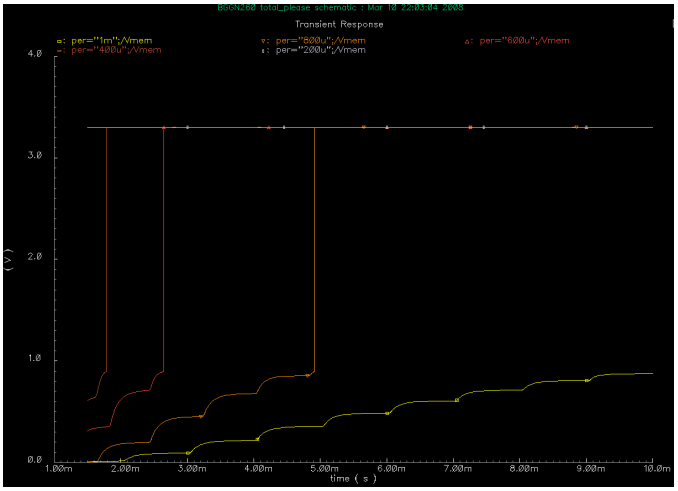

Fig. 6. Membrane voltage for different spike input frequencies.

REFERENCES

[1] K. A. Boahen, "Point-to-point connectivity between neuromorphic chips using address events, " *IEEE Transactions on Circuits and Systems II,* vol. 47, pp. 416-434 2000.

[2] C. Bartolozzi and G. Indiveri, "Synaptic Dynamics in Analog VLSI," *Neural Computation,* 19:(10) 2581-2603, Oct, 2007.

[3] D.H. Goldberg, G. Cauwenberghs and A.G. Andreou, "Probabilistic Synaptic Weighting in a Reconfigurable Network of VLSI Integrate-and-Fire Neurons" *Proc. IEEE Int. Symp. Circuits and Systems,* May. 2001.

[4] R.J. Volgelstein, U. Mallik, J.T. Vogelstein and G. Cauwenberghs, "Dynamically Reconfigurable Silicon Array of Spiking Neurons With Conductance-Based Synapses," *IEEE Trans. Neural Networks,* vol. 18 (1) pp. 253-265, 2007.

[5] A. Destexhe, Z.F. Mainen and T.J. Sejnowski, "An efficient method for computing synaptic conductances based on a kinetic model of receptor binding," *Neural Computation,* 6: 14-18, 1994.