On the Non-STDP Behavior and Its Remedy in a Floating-Gate Synapse

Gopalakrishnan, Roshan; Basu, Arindam

2015

Gopalakrishnan, R., & Basu, A. (2015). On the Non-STDP Behavior and Its Remedy in a Floating-Gate Synapse. IEEE Transactions on Neural Networks and Learning Systems, 26(10), 2596-2601.

<https://hdl.handle.net/10356/81587>

<https://doi.org/10.1109/TNNLS.2015.2388633>

© 2015 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. The published version is available at: [<http://dx.doi.org/10.1109/TNNLS.2015.2388633>].

Downloaded on 26 Apr 2023 21:52:58 SGT

On the “non-STDP” Behavior and its Remedy in a Floating-Gate Synapse

Roshan Gopalakrishnan, *Student Member, IEEE* and Arindam Basu, *Member, IEEE*

Abstract—This paper describes the neuromorphic VLSI implementation of a synapse utilizing a single floating-gate (FG) transistor that can be used to store a weight in a nonvolatile manner and demonstrate biological learning rules such as Spike Timing Dependent Plasticity (STDP). The experimental STDP plot (change in weight against $\Delta t = t_{post} - t_{pre}$) of a traditional FG synapse from previous studies shows a depression instead of potentiation at some range of positive values of Δt – we call this “non-STDP” behavior. In this paper, we firstly analyze theoretically the reason for this anomaly and then present a simple solution based on changing control gate waveforms of the FG device to make the weight change conform closely with biological observations over a wide range of parameters. Experimental results from a FG synapse fabricated in AMS 0.35 μ m CMOS process design are also presented to justify the claim. Finally, we present simulation results of a circuit designed to create the modified gate voltage waveform.

Index Terms—STDP, floating gate, synapse, learning, neuromorphic, VLSI, neuroscience.

I. INTRODUCTION

Over the last decade, numerous experimental studies [1] have shown that the synaptic strength is modified as a function of the spike timing difference $\Delta t = t_{post} - t_{pre}$ between the firing times t_{pre} and t_{post} of the presynaptic and postsynaptic neurons respectively. This phenomenon, called Spike Time-Dependent Plasticity (STDP), has emerged as one of several unsupervised plasticity rules that play an important role in learning and memory in the brain. The STDP based on a pair of pre- and post-synaptic spike is referred as doublet STDP (D-STDP) while the one based on triplet of synaptic spikes i.e either pre-post-pre synaptic spike or post-pre-post synaptic spike is referred as triplet STDP (T-STDP). In this paper we deal only with D-STDP and henceforth, we will imply the doublet rule when we mention STDP. The STDP model employs two phenomenon: one is long term potentiation (LTP) which is elicited by increase in synaptic weight due to the occurrence of postsynaptic spike after a presynaptic spike whereas the second one is long term depression (LTD) in which the synaptic weight is reduced when the timing relation is reversed. The STDP rule has some very good functional consequences including spike correlation based learning or finding the start of repeating spike patterns, as well as rate normalization without explicit renormalization similar to Oja’s rule [2].

This work was supported by grants RG 21/10 and ARC 8/13.

Roshan Gopalakrishnan and Arindam Basu are with VIRTUS, IC design centre of excellence, School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798 (e-mail: arindam.basu@ntu.edu.sg).

Fig. 1: Theoretical implementation of Doublet STDP (D-STDP): The plot is based on equation 1 with the parameters $A^+ = 4.6m$ and $A^- = 3m$.

Due to the popularity of STDP in computational neuroscience, neuromorphic engineers who aim to emulate brain function using VLSI have also tried to emulate this behavior in silicon. However, implementing a compact learning synapse remains one of the big challenges in the field [3]. Several recent papers have reported STDP implementations [4], [5]; however, these synapses could only hold two states in long term. The size of these synapses are also large hindering scalability of these designs. A promising solution for non-volatile analog weight storage is provided by a floating-gate (FG) device which is typically used to implement flash memory [6]. However, these earlier methods used multiple transistors in each synapse resulting in a large area of the array. A more recent work [7] used a single FG transistor for weight storage and adaptation by utilizing quantum mechanical processes of hot electron injection and tunneling based on input signal timing. Since then, it has also been used in a reconfigurable neural network as well [8], [9]. However, the experimental result of STDP using FG (Fig. 7(a) in [7]) has a difference when compared to biological findings: the FG synapse shows depression instead of potentiation for some range of positive Δt . Recently, resistive RAM based devices have also shown good promise for non-volatile weight storage and adaptation; however, this technology is still not mature and there are problems in interfacing these devices with CMOS. Hence, in this paper, we focus on improving the performance of the most compact CMOS compatible synapse reported in [7].

To ameliorate the STDP graph of a FG synapse obtained in [7], a minimum hardware overhead solution is proposed in this paper. The solution contains a modified gate voltage waveform in red dashed line as shown in Fig. 2 while the gate voltage waveform proposed in previous work [7] is shown in black line in the same figure. The paper is organized as follows: Section II provides a brief explanation of STDP model. Section III-A introduces the operation of a floating gate synapse. To vindicate the novelty proposed in the paper, a mathematical model of the weight update is presented in III-B. The experimental results and its discussion are included

in section IV followed by conclusions in the last section.

II. STDP SYNAPTIC MODIFICATION RULE

An essential requirement for learning in real and artificial neural networks is synaptic modification. In biology, chemical synapses (simply called synapses in the rest of the paper) are specialized structures to permit the communication between two neurons with an associated synaptic strength or weight. Learning typically implies the modification of synaptic weight due to the activities of the pre- and post-synaptic neurons. There are two types of synaptic weight change: one is potentiation or increase of weight while the other is depression or reduction in weight. In case of STDP, these two modifications are based on the timing of pre- and post-synaptic spikes as described next.

A. Doublet STDP

In doublet STDP, potentiation occurs when a postsynaptic spike succeeds a presynaptic spike; otherwise depression happens. The weight changes can be governed by a temporal learning window. The temporal learning window for STDP can be expressed as [10]

$$\Delta w = \begin{cases} \Delta w^+ = A^+ e^{(-\frac{\Delta t}{\tau_+})} & \text{if } \Delta t \geq 0 \\ \Delta w^- = -A^- e^{(\frac{\Delta t}{\tau_-})} & \text{if } \Delta t < 0 \end{cases} \quad (1)$$

where $\Delta t = t_{post} - t_{pre}$ is the time difference between a post-synaptic and pre-synaptic spike, τ_+ and τ_- are the time constants of the learning window, and A^+ and A^- are the maximal weight changes for potentiation and depression, respectively. The theoretical graph for the above equation is simulated using MATLAB and is shown in Fig. 1 with the parameters being obtained by data fitting as explained in [10].

III. FLOATING GATE SYNAPSE

A. Basic Operation

The equation for drain current of a subthreshold saturated pFET whose well is tied to V_{dd} is given by [7]

$$I_d = I_{s0} e^{\kappa(V_{dd} - V_{fg})/U_T} \quad (2)$$

where U_T is the thermal voltage and κ is the gate coupling coefficient. In order to generate excitatory post synaptic currents (EPSCs) similar to the decaying exponential waveform in biological observations, we apply a triangular waveform with unequal slopes at the input [7]. The fast decreasing part of the input triangular waveform with slope S_1 results in a quick rise of the synaptic current while the slow linearly increasing part with slope S_2 determines the exponential decay in the output current due to the exponential relationship between the gate voltage and drain current of the MOS transistor. The synaptic current is negligible for the maximum gate voltage. In this paper we modify the triangular waveform present at the input without harming the generation of EPSCs similar to biological observation.

The synaptic weight adaptation for obtaining different plasticity rule depends on the signals on the FG MOSFET. The terminal voltages of the floating gate transistor shown in

the Fig. 2(a) are the gate voltage (V_g), drain voltage (V_d) and tunneling voltage (V_{tun}). The specifications of terminal voltage waveforms are given in Fig. 2(b).

Synaptic weight modification in a FG MOSFET uses a combination of hot-electron injection (HEI) and Fowler-Nordheim tunneling [7]. HEI adds electrons on to the floating gate node, which reduces the voltage on the floating gate thus resulting in more current through the transistor hence increasing the weight of the synapse. On the other hand tunneling takes away electrons from the FG node, which increases the voltage of floating gate node thus reducing the drain current of the transistor and the synaptic weight. At every pre-synaptic spike, a triangular gate voltage waveform is generated while at every post-synaptic spike, a triangular tunneling voltage and a drain voltage pulse is generated as illustrated in the Fig. 2(c) and (d). Hence, injection and tunneling currents (and the resultant weight change) also depend on the time difference between pre- and post-synaptic spike times. The governing equations for injection and tunneling are given as [11], [7]

$$\begin{aligned} I_{inj} &= I_{inj0} (I_d/I_{s0})^\alpha e^{-\Delta V_{ds}/V_{inj}} \\ I_{tun} &= I_{tun0} e^{(V_{tun} - V_{fg})/V_{ox}} \end{aligned} \quad (3)$$

where I_d is the drain current, $\alpha = 1 - U_T/V_{inj}$, V_{ox} and V_{inj} are process dependent parameters.

B. Theoretical Model of Learning

If we compare the result obtained from the experimental measurement using original gate voltage waveform in Fig. 7(a) of [7] with the theoretical graph of STDP in Fig. 1, we notice that there is depression happening instead of potentiation on the positive x-axis of the experimental curve of STDP. We refer to this artifact as “non-STDP” behavior. To understand this effect, we first develop a simplified mathematical model of the change in weight for a FG synapse. The weight of the synaptic device is defined as [7]:

$$w = e^{\frac{-\kappa V_{fg}}{U_T}} \quad (4)$$

Hence, equations to predict the change in FG voltage effectively predict the change in weight. The following assumptions are made in deriving the theoretical equations:

- 1) To ensure small change in weight at very large negative and positive values of Δt , V_{g_init} has to be high enough so that $V_{tun_max} - V_{g_init}$ is small enough for negligible tunneling. Similarly $\Delta V_g = V_{g_init} - V_{g_min}$, should be small enough so that $V_{tun_init} - V_{g_min}$ is small enough for negligible tunneling.
- 2) Strong coupling from gate to floating gate node where as a weak coupling from tunneling node to floating gate node. This is justified since typically gate capacitance, $C_g \gg$ tunneling capacitance, C_{tun} .
- 3) The gate voltage waveform falls to its minimum value instantaneously. In other words, $S_1 \gg S_2, S_3$ as shown in Fig. 2(b). Also, we use T_g to denote the temporal duration of gate voltage with the understanding that $T_g = T_{g_ori}$ for the original waveform proposed in [7] while $T_g = T_{g_mod}$ for our proposed modification.

Fig. 2: (a) Floating gate MOSFET : The different terminal voltages V_{tun} , V_g , V_d as shown above are used to obtain the desired plasticity rule. Both the original gate waveform used in [7] and the novel technique proposed here are shown explicitly. (b) Nominal parameter values used in experiments and simulations for different terminal voltage waveforms of a FG synapse in this work. (c) The plot shows the timing diagram for different events in the case of $\Delta t > 0$. (d) Architecture of a neural network with M inputs and N neurons need $M \times N$ synapses and only M gate voltage generators.

We can now derive the slow time scale equation [7] for change in FG voltage due to tunneling and injection as:

$$C_T \frac{d\bar{V}_{fg}}{dt} = I_{tun} - I_{inj} = C_T \frac{d\bar{V}_{fg2}}{dt} - C_T \frac{d\bar{V}_{fg1}}{dt} \quad (5)$$

where C_T is the total capacitance on the FG node and $\bar{V}_{fg}$ denotes change on a slow time scale.

1) Case 1: $\Delta t > 0$: First, we consider the case of $\Delta t > 0$ i.e the positive axis of STDP curve and combine equations (3) and (2) to get:

$$C_T \frac{d\bar{V}_{fg1}}{dt} = -I_{inj0} (e^{\kappa(V_{dd}-V_{fg})/U_T})^\alpha e^{-\Delta V_{ds}/V_{inj}} \quad (6)$$

where $\bar{V}_{fg1}$ is the slow time scale change in V_{fg} due to injection only. Since change in V_{fg} on the RHS happens due to coupling from the gate voltage, we can write:

$$V_{fg} = V_{fg_min} + \frac{C_g}{C_T} S_2 t \quad (7)$$

where S_2 is positive slope of V_g and C_g is the capacitance connected between the gate terminal and the floating gate terminal. Here V_{fg_min} is not same as V_{g_min} due to initial charge stored on the FG. Substituting equations (7) in equation (6), we get

$$C_T \frac{d\bar{V}_{fg1}}{dt} = -A e^{-Xt} \quad (8)$$

where

$$A = I_{inj0} e^{\alpha \kappa (V_{dd}-V_{fg_min})/U_T} e^{-\Delta V_{ds}/V_{inj}} \quad (9)$$

$$X = \frac{\alpha \kappa C_g S_2}{C_T U_T}$$

Referring to Fig. 2(c), significant amount of injection happens in the time from Δt to $\Delta t + T_d$, where ΔV_{ds} is constant and significant. Also, since T_d is very small compared to T_g , we can assume that V_g is constant during the drain pulse. Hence, we finally get:

$$C_T \Delta \bar{V}_{fg1} = -A T_d e^{-X \Delta t} \quad (10)$$

The above equation for $\Delta \bar{V}_{fg1}$ is a function of Δt and is shown in Fig. 3(a) (plotted in red dashed line and marked as “Original V_g ”).

Now let us analyze the contribution of tunneling to $\Delta \bar{V}_{fg}$ denoted as $\Delta \bar{V}_{fg2}$. With reference to Fig. 2(c) and from assumption 1, the effect of tunneling is significant only from Δt to T_g . Hence, we have:

$$C_T \int_0^{\Delta \bar{V}_{fg2}} d\bar{V}_{fg} = \int_{\Delta t}^{T_g} I_{tun} dt \quad (11)$$

Substituting I_{tun} from equation (3), we further get:

$$C_T \Delta \bar{V}_{fg2} = \int_{\Delta t}^{T_g} I_{tun0} e^{(V_{tun}-V_{fg})/V_{ox}} dt \quad (12)$$

where

$$V_{tun} = \begin{cases} V_{tun_init} & \text{if } 0 < t < \Delta t \\ V_{tun_max} + S_3(t - \Delta t) & \text{if } \Delta t < t < T_g \end{cases} \quad (13)$$

where S_3 is the negative slope of V_{tun} . Thus substituting equations (13) and (7) into (12), we finally get:

$$\begin{aligned} C_T \Delta \overline{V_{fg2}} &= B e^{\frac{-S_2 \Delta t}{V_{ox}}} \int_{\Delta t}^{T_g} e^{Yt} dt \\ &= B' (e^{YT_g} - e^{Y\Delta t}) e^{\frac{-S_2 \Delta t}{V_{ox}}} \end{aligned} \quad (14)$$

where

$$\begin{aligned} Y &= \frac{S_3 - \frac{C_g S_2}{C_T}}{V_{ox}} \\ B &= I_{tun0} e^{\frac{V_{tun,max} - V_{fg,min}}{V_{ox}}} \\ B' &= B/Y \end{aligned} \quad (15)$$

The above equation for $\Delta \overline{V_{fg2}}$ is also a function of Δt and is shown in Fig. 3(a). It can be seen that the effect of tunneling persists longer than injection.

2) *Case 2: $\Delta t < 0$:* Now we consider the case of $\Delta t = (t_{post} - t_{pre}) < 0$ i.e the negative axis of STDP curve. Similar to what we have done above, let us see the effect of tunneling and injection separately.

For the contribution of injection to $\overline{V_{fg}}$, we could see that injection happens only during the initial small period, T_d of time axis where the drain current of the MOS is almost zero. So we can completely neglect the effect of injection on $\overline{V_{fg}}$ in this case.

Now let us consider the contribution of tunneling to $\overline{V_{fg}}$. Similar to the analysis above, using assumption 1, we have:

$$C_T \int_0^{\Delta \overline{V_{fg2}}} dV_{fg} = \int_{-\Delta t}^{T_{tun}} I_{tun} dt \quad (16)$$

While performing the integration, there arises two cases for the upper limit of integration based on whether $-\Delta t + T_g$ is greater or lesser than T_{tun} . Also, similar to the case of positive Δt , here we have:

$$\begin{aligned} V_{fg} &= V_{fg,min} + \frac{C_g}{C_T} S_2 (t - (-\Delta t)) \\ V_{tun} &= V_{tun,max} + S_3 t \end{aligned} \quad (17)$$

Substituting equations (17) and (3) into equation (16), we get case 1: $-\Delta t + T_g < T_{tun}$

$$\begin{aligned} C_T \Delta \overline{V_{fg2}} &= B e^{\frac{-C_g S_2 \Delta t}{C_T V_{ox}}} \int_{-\Delta t}^{-\Delta t + T_g} e^{Yt} dt \\ &= B' (e^{Y(T_g - \Delta t)} - e^{-Y\Delta t}) e^{\frac{-C_g S_2 \Delta t}{C_T V_{ox}}} \end{aligned} \quad (18)$$

case 2: $-\Delta t + T_g > T_{tun}$

$$\begin{aligned} C_T \Delta \overline{V_{fg2}} &= B e^{\frac{-C_g S_2 \Delta t}{C_T V_{ox}}} \int_{-\Delta t}^{T_{tun}} e^{Yt} dt \\ &= B' (e^{YT_{tun}} - e^{-Y\Delta t}) e^{\frac{-C_g S_2 \Delta t}{C_T V_{ox}}} \end{aligned} \quad (19)$$

where Y , B and B' are as given above.

The above equations for $\Delta \overline{V_{fg}}$ can be converted to change in weight [7] and is plotted as a function of Δt in Fig. 3(b). Both original and modified cases of gate voltage waveform is plotted in the figure. It is clearly seen that the theoretical model does predict a depression of weight for some range of positive values of Δt . While such a characteristic (recently

Fig. 3: Weight variation in FG synapse obtained from theory for the two different gate voltage waveforms. (a) shows the effect of injection and tunneling on ΔV_{fg} for positive Δt . (b) shows the effect of both injection and tunneling on $\Delta w/w$.

termed as triphasic STDP) might be useful in some cases [12], a large majority of the computational neuroscience studies use the traditional STDP rule and it is important for hardware emulators of software algorithms to be able to faithfully reproduce the characteristics of the nominal STDP rule. Next, we shall analyze the reason behind this “non-STDP” part and suggest a method to rectify this.

C. Proposed STDP protocol for FG synapse

To create a learning rule characteristic similar to Fig. 1, we have to design the FG synapse so that $\Delta \overline{V_{fg}}$ due to injection dominates over $\Delta \overline{V_{fg}}$ for tunneling in the entire range of positive Δt . On the other hand, we want the opposite effect for all negative values of Δt . From Fig. 3(b) we see however that for a range of positive values of Δt , tunneling dominates over injection. The reason for this can be realized by studying the injection and tunneling characteristics separately in Fig. 3(a). We can see that the effect of injection reduces to zero at $\Delta t \approx 10$ ms (Fig. 3(a)) while the effect of tunneling persists till ≈ 30 ms. This results in tunneling dominating for $\Delta t > \approx 10$ ms creating depression.

The reason for the sharp decline of the potentiation curve in Fig. 3(a) is that injection depends directly on drain current which reduces exponentially fast with Δt (equation 10). Thus there is a large part of the gate waveform (the part for which $V_g > \approx V_{g,min} + 0.2$ V) which does not create significant post-synaptic current or potentiation. Intuitively, we can eliminate this part and get a modified gate control voltage waveform as shown in Fig. 2. The effect of this modification on the weight change curve is shown in Fig. 3(b) (blue curve marked “modified V_g ”). As predicted, the magnitude of depression for positive values of Δt is much reduced now. For certain sets of bias parameters (e.g. if $V_{tun,max}$ is much lower or $V_{g,min}$ is higher), the original gate waveform also does not produce significant depression for $\Delta t > 0$; however, even in those cases statistical variations between different FG devices on a chip may cause “non-STDP” behavior. The modified waveform provides a more bio-realistic STDP curve for a

Fig. 4: Experimental results for the STDP curve obtained from a FG device for (a) $V_{tun,max} = 12V$ and $T_{tun} = 300ms$ (b) $V_{tun,max} = 11.9V$ and $T_{tun} = 300ms$ (c) $V_{tun,max} = 11.9V$ and $T_{tun} = 400ms$. Along with the experimental result, simulation results are also included for comparison. These curves are plotted for both original and modified cases of gate voltage waveform, V_g . The deviations in simulation results are most likely due to parasitic coupling of floating-gate to substrate.

much wider set of parameters and hence is more robust. We shall present results to prove this fact in the next section.

IV. RESULTS

We present experimental measurements from a floating gate synapse designed in AMS $0.35\mu m$ CMOS process and match it with theoretical predictions. For higher accuracy, the simulation results are obtained from solving the original differential equations for ΔV_{fg} and not the simplified theoretical model presented earlier.

A. Experimental Results: Robustness of “non-STDP” Behavior

In order to justify the robustness of proposed STDP protocol in suppressing the “non-STDP” phenomenon, two parameters, $V_{tun,max}$ and T_{tun} have been varied and the STDP curve has been measured. All other parameters are fixed to the values mentioned in Fig. 2(b). The experimental result for the parameters $V_{tun,max} = 12V$ and $T_{tun} = 300ms$ is taken first and are plotted in Fig. 4(a). It can be seen that the modified V_g case is indeed able to suppress the “non-STDP” behavior compared to the original case. Next the parameter $V_{tun,max} = 12V$ is changed to $V_{tun,max} = 11.9V$ keeping other parameter, $T_{tun} = 300ms$ unchanged and the corresponding STDP curve is plotted in Fig. 4(b). In this case, we can see that even the original gate voltage waveform, V_g does not exhibit “non-STDP” behavior. However, when we change $T_{tun} = 300ms$ to $T_{tun} = 400ms$ keeping $V_{tun,max} = 11.9V$, the “non-STDP” behavior appears again for the original V_g but not for the proposed V_g as shown in Fig. 4(c). These results indicate that the “non-STDP” behavior does exist for a range of parameters in the original FG synapse and our proposed method is able to robustly reduce that effect.

B. Simulation Results: Effect of Variations

Since the measurement of a STDP curve is very time consuming, we next performed simulations of the differential equations 5 to explore the effect of parametric variations in more details. The effect of statistical variations between FG devices on a chip is first considered by varying V_{inj} and V_{ox}

from their nominal values by a maximum of 5% (conservative estimate [13]). For each parameter set, we find B^+ which is defined as the least magnitude of $\Delta w/w$ for $\Delta t > 0$. For a bio-realistic STDP curve, $B^+ \approx 0$ while for “non-STDP” behavior, $B^+ < 0$. From the results shown in Fig. 5(a), we can conclude that B^+ for the modified V_g case is always closer to the bio-realistic result. Furthermore, to show that these device parameter variations are significant, we also plot the same results against A^+ (maximum potentiation) and A^- (maximum depression) in Fig. 5(b). It can be seen that the variations of A^+ and A^- are quite large proving that this robustness for modified V_g holds for significant variations. Figures 5(c) and (d) show similar plots of B^+ when bias parameters $V_{tun,max}$ and T_{tun} are varied. Again, we see that B^+ for modified V_g is always closer to 0 than that for the original gate waveform. Further, we can see from Fig. 5(d) that the original gate waveform will have $B^+ \approx 0$ for any bias condition for which $|A^+| \gg |A^-|$. This corresponds to the measurement results shown in Fig. 4(b).

V. GATE VOLTAGE GENERATOR: VLSI IMPLEMENTATION

Fig. 6(a) shows a possible circuit to implement the modified gate voltage waveform. It has very little overhead compared to the gate waveform generator circuit in [7] (PreSynaptic Computation block in Fig.3). The pre-synaptic pulse is input to the current starved inverter composed of M1, M3 and M4. When the input pulse arrives, the output of the inverter, V_{out} , falls quickly to $V_{g,min}$ (around 2.3V as shown in the Fig. 6(b)). Next, when the input becomes zero the output rises slowly to V_{dd} with a time constant set by the V_{biasp} . This part is very similar to the original circuit in [7]. To create the modified waveform, V_{out} has to be pulled up quickly to V_{dd} whenever V_{out} crosses a trip point (TP) voltage higher than $V_{g,min}$. For example, in the earlier measurement results this critical voltage is approximately $V_{g,min} + 0.2V$. We add the transistors M5, M6 and M2 to implement this function. M5 and M6 form a common source amplifier that quickly pulls up V_{out} through the positive feedback action of M2 when V_{out} reaches TP. This TP value can be set by the voltage V_{thr} or the bias voltage V_{bias} that sets the bias current I_{bias} in the common source amplifier. Simulation results for this circuit are shown in Fig.

Fig. 5: Parametric variations of STDP curve from FG synapse. Plot of simulation results obtained for B^+ due to variations in the device parameters V_{inj} and V_{ox} . The data is plotted as a function of (a) V_{inj} and V_{ox} and (b) A^+ and A^- . Plot of simulation results obtained for B^+ due to variations in the bias parameters $V_{tun,max}$ and T_{tun} . The data is plotted as a function of (c) $V_{tun,max}$ and T_{tun} and (d) A^+ and A^- .

Fig. 6: (a) Circuit to generate modified gate voltage waveform. The trip point (TP) voltage for sudden rise of the waveform to V_{dd} is controlled using V_{thr} . (b) The SPICE simulation result of the circuit for different V_{thr} voltage exhibit different thresholds. The circled transistors (M1 and M6) can be replaced by FG devices to correct for mismatch in slope and TP.

6(b) and operates as expected. Further, transistors M1 and M6 may be replaced by FG devices to eliminate mismatch in an array of such circuits [8].

VI. CONCLUSION

We have presented a robust STDP implementation using single FG transistor acting as a learning synapse for VLSI spiking neural networks. Compared to earlier work in [7], our proposed method allows the STDP curve of the FG device to be similar to the biological one over a much wider range of parameters. This is achieved by simply modifying the gate voltage waveform—hence, the hardware overhead compared to [7] is minimal (e.g. only .01% area increase for 100 input 100 neuron system). A simplified mathematical model for learning is presented to intuitively illustrate the benefit of our proposed method. We also present measurement results from a $0.35\mu\text{m}$

chip to justify the claim. Finally, a circuit implementation of the modified gate voltage generator is also shown.

REFERENCES

- [1] L. F. Abbott and S. B. Nelson, “Synaptic plasticity: taming the beast,” *Nature neuroscience*, vol. 3, pp. 1178–1183, Nov. 2000.
- [2] J. Sjostrom and W. Gerstner, “Spike-timing dependent plasticity,” *Scholarpedia*, vol. 5, no. 2, pp. 1362, 2010.
- [3] M. Milev and M. Hristov, “Analog implementation of ANN with inherent quadratic nonlinearity of the synapses,” *IEEE Transactions on Neural Networks*, vol. 14, no. 5, pp. 1187–1200, Sep. 2003.
- [4] J. Arthur and K. Boahen, “Synchrony in silicon: The gamma rhythm,” *IEEE Transactions on Neural Networks*, vol. 18, no. 6, pp. 1815–1825, Nov. 2007.
- [5] G. Indiveri, E. Chicca, and R. Douglas, “A VLSI Array of Low-Power Spiking Neurons and Bistable Synapses With Spike-Timing Dependent Plasticity,” *IEEE Transactions on Neural Networks*, vol. 17, no. 1, pp. 211–221, Jan. 2006.
- [6] S. C. Liu and R. Mockel, “Temporally learning floating-gate VLSI synapses,” in *IEEE ISCAS*, May. 2008, pp. 2154–57.
- [7] S. Ramakrishnan, P. Hasler, and C. Gordon, “Floating Gate Synapses with Spike-Time-Dependent Plasticity,” *IEEE Transactions on Biomedical Circuits and Systems*, vol. 5, Jun. 2011.
- [8] 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.
- [9] A. Basu, S. Ramakrishnan, C. Petre, S. Koziol, S. Brink, and P. Hasler, “Neural dynamics in reconfigurable silicon,” *IEEE Transactions on Biomedical Circuits and Systems*, vol. 4, no. 5, pp. 311–319, Oct. 2010.
- [10] J.-P. Pfister and W. Gerstner, “Triplets of spikes in a model of spike timing-dependent plasticity,” *The Journal of Neuroscience*, pp. 9673–9682, Sep. 2006.
- [11] P. Hasler, A. Basu, and S. Koziol, “Above threshold pfet injection modeling intended for programming floating-gate systems,” in *IEEE ISCAS*, May. 2007, pp. 1557–1560.
- [12] A. Waddington, P. Appleby, M. De Kamps, and N. Cohen, “Triphasic spike-timing-dependent plasticity organizes networks to produce robust sequences of neural activity,” *Frontiers in Computational Neuroscience*, vol. 6, no. 88, Nov. 2012.
- [13] S. Brink, *Learning in Silicon: A Floating-gate Based Biophysically Inspired, Neuromorphic Hardware System with Synaptic Plasticity*, Ph.D. thesis, Georgia Institute of Technology, Atlanta, USA, 2012.