Correlation Learning Rule in Floating-Gate pFET Synapses

Paul Hasler, *Member, IEEE*, and Jeff Dugger, *Student Member, IEEE*

Abstract— **We study the weight dynamics of the floating-gate pFET synapse and the effects of the pFET's gate and drain voltages on these dynamics. We show that we can derive a weight update rule such that the equilibrium weight value is proportional to the correlation between the gate and drain voltages. In particular, we want a rule of the form $\tau' \Delta \dot{W} = -\Delta W + \eta E[xy]$, where x is a voltage signal on the gate terminal and y is a voltage signal on the drain terminal. We obtain this rule by making a linear approximation to the weight dynamics around a given equilibrium point. We develop this approximation by considering the basic functional form of the system dynamics and then examining the effects of the gate and drain voltages on the specifics of this form.**

We have come to view floating-gate devices not only as memory elements, but also as adaptive analog computational elements [1-8]. Considered as an analog computational element, a floating-gate device computes its output as a product of its input and an adaptive device parameter. We call the adaptive device parameter the weight. When used as an adaptive analog computational element, we call the floating-gate device a synapse, since we will use it to build electronic neural networks. A fundamental question we would like to answer is how the synapse weight adapts.

In this paper, we show how the synapse weight adapts based on correlations between voltage signals applied to the gate and the drain terminals of the floating-gate pFET synapse. We presented initial work elsewhere [9]; this paper is a more complete discussion that clearly shows the adaptation as a function of voltage signals and correlation between these signals. We focus on pFET floating-gate devices because they are available in any standard CMOS process; although we focus on pFET floating-gate devices, these results are easily extendable to nFET floating-gate devices. This work contrasts with earlier floating-gate synapse research where *learning* dynamics only resist slow timescale changes in the inputs [4], [10].

I. Overview of Floating-Gate Devices

Figure 1 shows the layout, cross-section, and circuit symbol for our floating-gate pFET device that has been described elsewhere [1-8]. A floating gate is a polysilicon gate surrounded by SiO_2. Charge on the floating gate is stored permanently, providing a long-term memory, because it is completely surrounded by a high-quality insulator. Voltage signals on the gate terminal(s), as well as the source and drain terminals, of the pFET synapse capacitively couple into the channel by way of the floating gate. Since we operate our floating-gate synapses in the subthreshold regime, the relationship between channel current, floating gate voltage (V_{fg}), and source voltage (V_s) around a bias current, I_{so}, is given by

$$I_s = I_{s0} e^{-(\kappa_p \Delta V_{fg} - \Delta V_s)/U_T}. \tag{1}$$

We do not consider the Early effect (channel-length modulation) in this discussion. For a fixed gate voltage, adding charge to the floating-gate effectively increases the gate voltage seen by the channel and therefore causes the channel current to decrease. Similarly, removing charge from the floating gate causes the channel current to increase.

The authors are associated with the Integrated Computational Electronics Laboratory at Georgia Institute of Technology, Atlanta, GA 30332-0250. E-mail: phasler@ee.gatech.edu

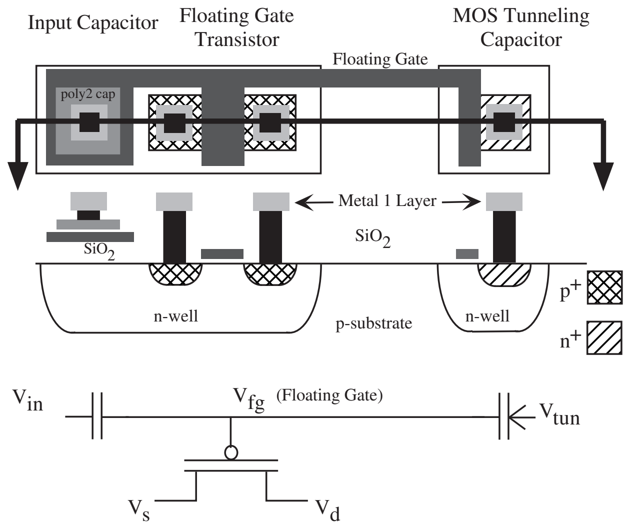


Fig. 1. Layout, cross section, and circuit diagram of the floating-gate pFET in a standard double-poly $n$well MOSIS process. The cross section corresponds to the horizonatal line slicing through the layout (or top) view. The pFET transistor is the standard pFET transistor in the nwell process. The gate input capacitively couples to the floating-gate by either a poly-poly capacitor, a diffused linear capacitor, or a MOS capacitor. We add floating-gate charge by electron tunneling, and we remove floating-gate charge by hot-electron injection. The tunneling junction used by the single-transistor synapse is a region of gate oxide between the polysilicon floating-gate and $n$well (a MOS capacitor). Between V_{tun} and the floating-gate is our symbol for a tunneling junction, a capacitor with an added arrow designating the charge flow. We use high-quality gate oxide to reduce the effects of electron trapping.

A. Electron Tunneling

We add charge to the floating gate by removing electrons. We use electron tunneling (Fowler-Nordheim tunneling [11]) to remove electrons from the floating gate. The tunneling junction is schematically represented by a capacitor coupling the tunneling voltage terminal to the floating gate

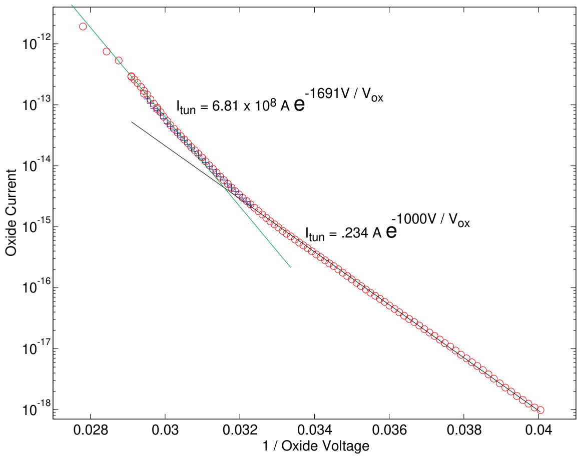


Fig. 2. Tunneling in an $n$well process. Over a wide range of oxide voltage, most of the tunneling occurs between the floating gate and n^+ diffusion region because this region is accumulated and the higher electric fields at the corner of the floating gate. Experimental measurements of electron tunneling current versus 1/oxide voltage in a 2.0μm process with 42nm gate oxide. The two straight line fits are to the classic Fowler-Nordheim expression in (2). The two different straight-line regions might be due to tunneling through intermediate traps, or due to initially tunneling through the junction edge for low oxide voltages and tunneling through the middle of the junction for high oxide voltages.

as shown in Fig. 1. The arrow on the capacitor denotes the charge flow. Increasing the voltage across this tunneling capacitor, either by increasing the tunneling voltage (V_{tun}) or decreasing the floating-gate voltage, increases the effective electric field across the oxide, thereby increasing the probability of the electron tunneling through the barrier. Figure 2 shows measured electron tunneling current through our tunneling capacitor versus (1 / applied oxide voltage). Typical values for the oxide field range from $0.75V/nm$ to $1.0V/nm$. We start from the classic model of electron tunneling through a silicon–silicon-dioxide system [6], [11], in which the electron tunneling current is given by

$$I_{tun} = I_0 \exp\left(\frac{\mathcal{E}_o}{\mathcal{E}_{ox}}\right) = I_0 \exp\left(\frac{t_{ox}\mathcal{E}_o}{V_{tun} - V_{fg}}\right), \quad (2)$$

where $\mathcal{E}_{ox}$ is the oxide electric field, t_{ox} is the oxide thickness, and $\mathcal{E}_o$ is a device parameter that is roughly equal to $25.6V/nm$ [12]. The cause for two separate regions might be due to tunneling through intermediate traps [13], or due to initially tunneling through the junction edge for low oxide voltages and tunneling through the middle of the junction for high oxide voltages.

B. pFET Hot-Electron Injection

We use pFET hot-electron injection to add electrons (remove charge) to the floating-gate [1], [6], [7]. If we are to inject an electron onto a floating gate, the MOSFET must have a high-electric-field region ($> 10\text{V}/\mu m$) to accelerate channel electrons to energies above the silicon–silicon-dioxide barrier, and in that region the oxide electric field must transport the electrons that surmount the barrier to the floating gate. In general, the subthreshold MOSFET injection current is proportional to the source current (I_s), and is the exponential of a smooth function (f_2) of the drain-to-channel potential (Φ_{dc}). We express this relationship as follows:

$$I_{inj} = I_s e^{f_2(\Phi_{dc})}. \quad (3)$$

Because the injection current is only a weak function of the floating-gate voltage for a fixed source current (I_p) and Φ_{dc}, we neglect the gate-voltage dependence for this application [3]. Hot-electron injection by nFET devices is somewhat simpler than for pFET devices, and is described in detail elsewhere [6], [14].

In this discussion we will use pFET hot-electron injection to add electrons to the floating-gate. We use pFET hot-electron injection, because pFET hot-electron injection can not be eliminated from a CMOS process without adversely affecting basic transistor operation, and therefore will be available in all commercial CMOS processes. One might wonder how pFETs, where the current carriers are holes, inject hot electrons onto the floating gate. Figure 3a shows the band diagram of a pFET operating under bias conditions that are favorable for hot-electron injection. The hot-hole impact ionization creates electrons at the drain edge of the drain-to-channel depletion region, due to the high electric fields there. The hole impact-ionization current is proportional to the pFET source current, and is the exponential of a smooth function (f_1) of the drain-to-channel potential (Φ_{dc}). These electrons travel back into the channel region, gaining energy as they go. When their kinetic energy exceeds that of the silicon–silicon-dioxide barrier, they can be injected into the oxide and transported to the floating gate. We express this relationship as follows:

$$I_{impact} = I_p e^{f_1(\Phi_{dc})}, \quad (4)$$

where Φ_{dc} is the potential drop from channel to drain. The injection current is proportional to the hole impact-ionization current, and is the exponential of another smooth function (f_3) of the voltage drop from channel to drain. We express this relationship as follows:

$$I_{inj} = I_{impact} e^{f_3(\Phi_{dc})}. \quad (5)$$

As a result (3) models the pFET injection current.

Figure 3b shows measured injection efficiency for four source currents; injection efficiency is the ratio of the injection current (I_{inj}) to source current (I_s). The measurements for four different source current values are nearly equal, which is consistent with injection efficiency being independent of source current. Injection efficiency is approximately an exponential of a linear function in Φ_{dc} over ranges spanning 1V. The slope of the curve on this exponential scale decreases with increasing Φ_{dc}. Using this linear approximation, we can model the hot-electron injection current for a changing gate and drain-to-source (ΔV_{ds}) voltage as [3]

$$I_{inj} = I_{inj0}\left(\frac{I_s}{I_{s0}}\right)^{\alpha} e^{-\Delta V_{ds}/V_{inj}}, \quad (6)$$

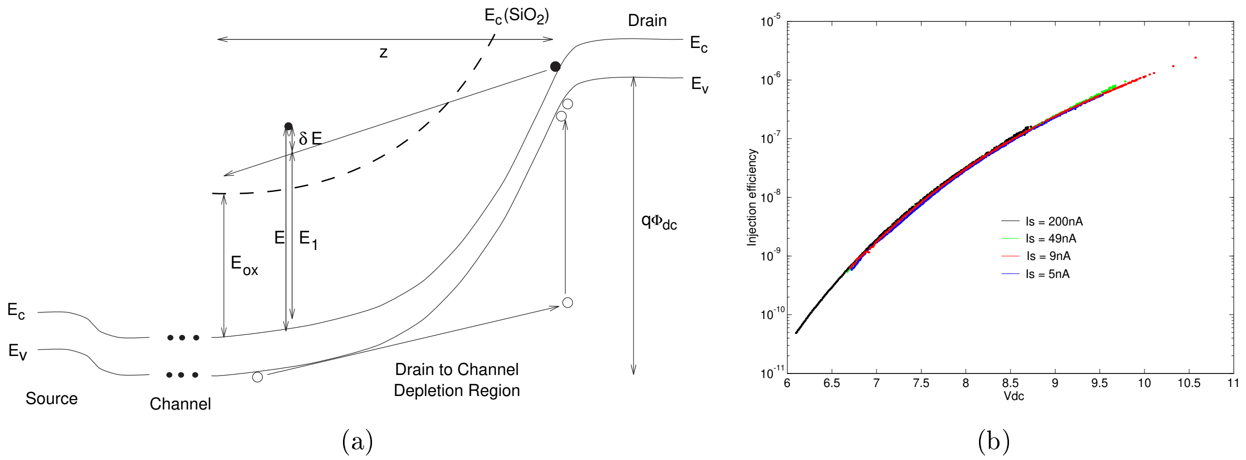


Fig. 3. pFET hot-electron injection. (a) Band diagram of a subthreshold pFET transistor under conditions favorable for hot-electron injection. E_{ox} is the Si–SiO$_2$ barrier, which is 3.1eV for no field across the oxide. The holes in the channel gain energy through the drain-to-channel depletion region with the potential drop of Φ_{dc}, resulting in hole impact ionization. The resulting electron accelerates towards the channel; if this electron gains energy larger than the silicon–silicon-dioxide barrier, then the electron has a chance to move from the channel to the floating-gate. We can model the electron at position z at energy E, as spreading a distance δE around an average trajectory, described by $E_1(z)$ defined by an average number of phonon collisions, as we previously described for nFET injection [14]. (b) Measured data of pFET injection efficiency versus the drain-to-channel voltage for four source currents. Injection efficiency is the ratio of injection current to source current. The injection efficiencies are nearly identical for the different source currents; therefore, they appear to be indistinguishable on the plot. At Φ_{dc} equal to $8.2V$, the injection efficiency increases by a factor of e for an increase in Φ_{dc} of $250mV$.

where V_{inj} is a device and bias dependent parameter, and α is $1-\frac{U_T}{V_{inj}}$. For a quiescent $\Phi_{dc}=8.2V$, a typical value for V_{inj} is $250mV$, and for α is 0.90. We have validated this model over several orders of magnitude in current and wide ranges in voltage elsewhere [6], [3].

II. Weight Update for Continuously Adapting pFET Synapses

We will consider floating-gate devices that use continuous electron-tunneling and hot-electron injection currents. To model the effect of continuous floating-gate currents, we apply Kirchoff's Current Law to the floating-gate node:

$$C_T\frac{dV_{fg}}{dt}=C_1\frac{dV_g}{dt}+C_2\frac{dV_d}{dt}+I_{tun}-I_{inj}, \tag{7}$$

where we have included the effects of electron tunneling tunneling and hot-electron injection, C_T is the total capacitance connected to the floating-gate, C_1 is the capacitance between the input and the floating-gate, and C_2 is the capacitance between the drain and the floating-gate. We have fixed the tunneling voltage terminal to a constant bias voltage.

The tunneling and injection currents in the device are small compared with the transistor bias current; therefore, changes in floating-gate charge occur on a much slower timescale than synapse computations. These two time scales allow us to define separate terminal voltages representing slow and fast timescale voltage changes. For example, we now define the change in the floating-gate voltage around its bias current as equal to $\bar{V}_{fg}+\Delta V_{fg}$ as the slow ($\bar{V}_{fg}$) and fast (ΔV_{fg}) timescale decomposition of the floating-gate voltage. We can also decompose the gate, drain, and source voltages into slow and fast timescale components. We will assume that the input voltage(s) are zero mean, that is, they only have fast-timescale components. The separation of timescales allows us to separate the floating-gate node equation (7) into two equations: one modeling the fast dynamics and one modeling the slow dynamics.

At fast timescales, we approximate the tunneling and injection currents to be negligible. We model the behavior at the fast timescales as $C_T\Delta V_{fg}=C_1\Delta V_g+C_2\Delta V_d$ where V_g has only fast components. Since we decomposed ΔV_{fg} into fast and slow components, applying this to (1), we get

$$I_s=I_{s0}We^{-\kappa_p\Delta V_{fg}/U_T}, \tag{8}$$

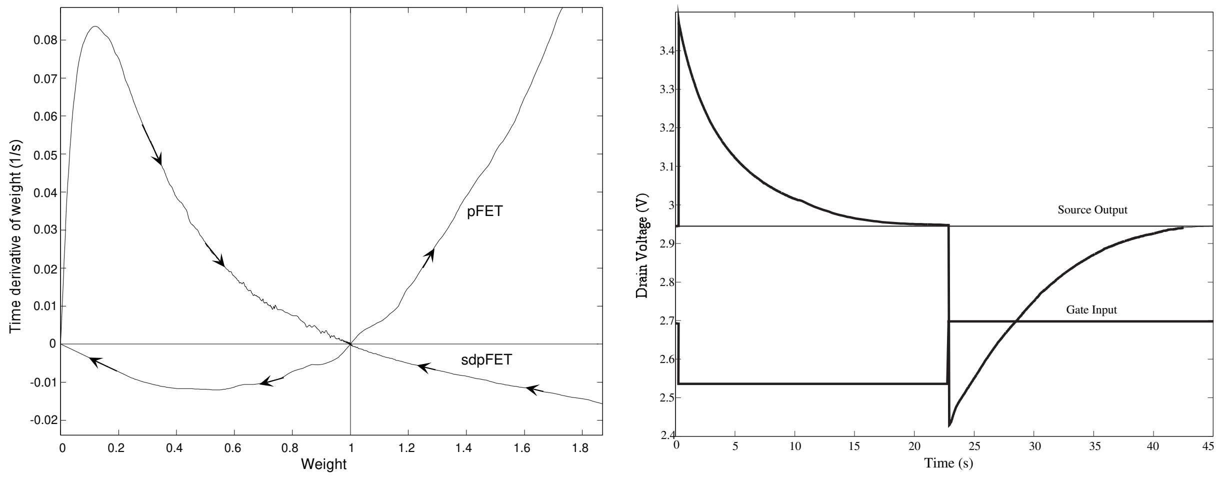
where we have defined the factor due to the slow-timescale floating-gate voltage to be the weight of the synapse

$$W=e^{-\kappa_p\bar{V}_{fg}/U_T}. \tag{9}$$

We often consider $I_{s0}W$ as an adaptive bias current. We consider fast-timescale variables as representing the computational signals of the synapse. We define the weight as $W=\frac{I_s}{I_{s0}}$ for no input signals, and therefore it is proportional to the average current at the source. At the bias point, $I_s=I_{s0}$, and therefore $W=1$. The pFET channel current depends upon a multiplication of the gate terminal voltage and an exponential function of the floating-gate charge, which we call the weight.

The slow-timescale floating-gate node voltage equation gives us the starting point for developing the weight dynamics of the synapse. The slow-timescale equation is

$$C_T\frac{d\bar{V}_{fg}}{dt}=I_{tun}-I_{inj}, \tag{10}$$



(a) (b)

Fig. 4. (a) Plot of the time-derivative of w versus w for the pFET, and source-degenerated pFET synapses. The arrows show the direction that the differential equation will take. This data shows that the pFET synapse will diverge from the $W = 1$ steady state, but that source-degenerated synapses will stabilize to the $W = 1$ steady state. The s-d pFET modifies the basic pFET dynamics by local negative-feedback. We use no fast-time scale movement of gate or drain voltages. (b) The behavior of the current autozeroing circuit using a source-degenerated pFET synapse. Unlike the pFET synapse, this circuit converges to its steady-state current. We use this data to measure weight changes versus time. The tunneling voltage was held at a constant value throughout this experiment; there was non-negligible tunneling and injection current throughout this experiment.

where we will neglect the slow timescale variables of gate and drain voltage, as previously mentioned. By differentiating the weight, W, given in (9) with respect to time as

$$\frac{dW}{dt} = -\frac{\kappa_p}{U_T}\frac{d\bar{V}_{fg}}{dt}W, \tag{11}$$

we get the weight update equation by substituting this result into (10):

$$\left(\frac{U_T C_T}{\kappa_p}\right)\frac{dW}{dt} = W(I_{inj} - I_{tun}) \tag{12}$$

With this formulation, we can integrate (10) over many periods at the fast-timescale, but still make only a small change in the slow-timescale output voltage. We assume that the input signals are ergodic so that we can use time averages for the expected value. Thus we shall define $E[\cdot]$ as the average of a time-varying signal, $x(t)$, over a time interval, T, that is much shorter than the slow timescale, but much longer than one period at the fast-timescale:

$$E[x(t)] = \frac{1}{T}\int_0^T x(t)dt. \tag{13}$$

The remaining issues are to substitute models of tunneling and injection currents. For a fixed tunneling voltage ($\Delta V_{tun} = 0$), the floating-gate tunneling current is given by

$$I_{tun} = I_{tun0}e^{-(\bar{V}_{fg}+\Delta V_{fg})/V_x} = I_{tun0}W^{U_T/(\kappa_p V_x)}e^{-\Delta V_{fg}/V_x}, \tag{14}$$

where V_x is dependent on the bias values of V_{tun0}, V_{fg0}, and device parameters. The floating-gate injection current around a bias injection current, I_{inj0}, is given by

$$I_{inj} = I_{inj0}W^{1+\alpha}e^{\alpha\kappa_p\Delta V_{fg}/U_T}e^{-\Delta V_d/V_{inj}}. \tag{15}$$

In equilibrium, that is at the bias point, the tunneling and injection currents are equal, so that $I_{fg0} = I_{inj0} = I_{tun0}$. Using these results with (12) gives us the fundamental equation governing the weight dynamics of the floating-gate pFET synapse as

$$\tau\frac{dW}{dt} = \begin{array}{l} W^{1+\alpha}e^{\alpha\kappa_p\Delta V_{fg}/V_{g0}} - e^{\Delta V_d/V_{inj})} \\ -W^{1+(U_T/\kappa_x V_x)}e^{-\Delta V_{fg}/V_x} \end{array} \tag{16}$$

where we define $\tau = \frac{U_T}{\kappa_p}\frac{C_T}{I_{fg0}}$, and all voltages are referenced with respect to ground.

The weight dynamics of the basic pFET synapse are unstable because of a positive feedback relationship that exists between the hot-electron injection current and the channel current. For a step increase in the gate-to-source voltage of the pFET, we get an increase in the channel current. This increases the hot-electron injection current in the pFET. Increasing the hot-electron injection current adds electrons to the floating-gate, which makes the gate-to-source voltage effectively larger, further increasing the channel current. This positive feedback relation does not give us a stable weight value that we can use in analog computations. We show experimental data of this phenomena in Fig. 4a.

To obtain a stable continuously-adapting pFET synapse, we introduce external negative feedback by using source degeneration shown in Fig. 5a. Figure 5a shows the circuit for the source-degenerated (s-d) pFET synapse [6], [5]. The circuit diagram for the s-d pFET synapse shows that we have added another pFET, M2, between V_{dd} and the source terminal of the floating-gate pFET, M1. M2 is a short-channel pFET with significant drain-induced barrier lowering (DIBL). A transistor that strongly exhibits DIBL shows an exponential change in current for a linear change

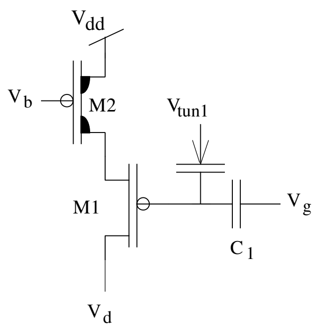


Fig. 5. The source-degenerated (s-d) pFET synapse. The s-d pFET modifies the basic pFET dynamics by local negative-feedback; the device converges to a stable equilibrium for either a constant current or voltage. The s-d pFET synapse is comprised of a floating-gate pFET synapse and a second ultra-short pFET, which provides feedback to the source terminal. We utilize the DIBL effect in short-channel MOSFETs to build a compact weakly exponential element.

in drain voltage. The resulting dependence of channel current on the floating-gate voltage for the circuit in Fig. 5a is

$$I_s = I_{s0} e^{-\kappa_x \kappa_p (\bar{V}_{fg} + \Delta V_{fg})/U_T} \tag{17}$$

where κ_x is a parameter that characterizes the sharpness of the exponential relationship of the source degenerative feedback element.

A. dynamics of s-d pFET synapse

The stabilizing effects of adding M2 can be seen in Fig. 4a. The DIBL current-voltage characteristic of M2 guarantees that the decrease in the source voltage of M1 due to an increase in channel current will be sufficient to provide the negative feedback desired. Increasing the channel current in M2 causes a decrease in the source voltage of M1, this in turn decreases the drain-to-source voltage of M1. Decreasing the drain-to-source voltage of M1 decreases the channel current, so that hot-electron-injection current also decreases. Fig. 4b shows dynamics (measured) to get plots for the s-d pFET in Fig. 4a.

Using our previous derivation for the pFET synapse, we can derive the weight update rule for our s-d pFET synapse. We have defined the weight of this s-d pFET synapse as

$$W = e^{-\kappa_x \kappa_p \bar{V}_{fg}/U_T}, \tag{18}$$

and we get the resulting weight update equation as

$$\frac{U_T C_T}{\kappa_x \kappa_p} \frac{dW}{dt} = W(I_{inj} - I_{tun}). \tag{19}$$

We modify the model of tunneling and injection currents for this s-d pFET synapse as

$$I_{tun} : I_{tun} = I_{fg0} W^{(\beta-1)} e^{-\Delta V_g / V_{g1}}$$

$$I_{inj} : I_{inj} = I_{fg0} W^{(\gamma-1)} e^{(\Delta V_g/V_{g0} - \Delta V_d/V_{inj})} \tag{20}$$

where $\beta = 1 + \frac{1}{\kappa_x}\left(\frac{1}{\kappa_p}\frac{U_T}{V_x}\right)$, $V_{g1} = \frac{1}{\kappa_x}\frac{C_T}{C_1}\frac{U_T}{\kappa_p}\frac{1}{(\beta-1)}$. $\gamma = 2 - \frac{1}{\kappa_x}\left(\frac{U_T}{V_{inj}}\right)$, $V_{g0} = \frac{1}{\kappa_x}\frac{C_T}{C_1}\frac{U_T}{\kappa_p}\frac{1}{(1-\gamma)}$. Recall that in equilibrium, that is at the bias point, the tunneling and injection currents are equal, so that $I_{fg0} = I_{inj0} = I_{tun0}$. Using these results with (12) gives us the fundamental equation governing the weight dynamics of the s-d floating-gate pFET synapse as

$$\tau \frac{dW}{dt} = W^{\gamma} e^{(\Delta V_g/V_{g0} - \Delta V_d/V_{inj})} - W^{\beta} e^{-\Delta V_g/V_{g1}} \tag{21}$$

where we define $\tau = \frac{1}{\kappa_x}\frac{C_T U_T}{I_{fg0}\kappa_p}$, and all voltages are referenced with respect to ground.

III. Weight Dynamics of the Floating-Gate pFET Synapse

To understand the basic form of the weight dynamics, we look at the simplest case of (21) when we have no fast timescale signals in gate or drain voltages. This allows us to see the dependence of $\frac{dW}{dt}$ on w. The resulting equation is

$$\tau \frac{dW}{dt} = W^{\gamma} - W^{\beta}. \tag{22}$$

For the source degenerated pFET, we experimentally observe the weight dynamics by noting that the weight of the floating-gate synapse is directly proportional to the channel current when the signal voltages (fast timescale) are zero [1], [2], [5]. We measure the channel current indirectly by measuring the source voltage, since the source voltage is logarithmically related to the channel current due to the DIBL FET operating with subthreshold currents. To see the weight dynamics, we apply a slow timescale step to the gate voltage, ΔV_g, and observe the source voltage, ΔV_s. The gate voltage input and the source voltage response are shown in Fig. 4b. We obtain the channel current, and thus W by taking the exponential of this value. A finite difference approximation using successive values of W gives $\frac{dW}{dt}$. The s-d pFET dynamics given by these measurements appears in Fig. 5b. Figure 4a also illustrates the dynamics of this equation for the basic floating-gate pFET synapse. We observe there are two equilibrium points for both systems, $W_{eq} = 0$ and $W_{eq} = 1$. We are interested in the weight dynamics around $W_{eq} = 1$, since the weight cannot go below zero, and this steady state will not change with different inputs. Figure 4a verifies that the basic pFET synapse is unstable, but the source degenerated pFET is stable.

Since we are only interested in the linear dynamics around $W_{eq} = 1$, we substitute $W = 1 + \Delta W$ into (22), use the binomial expansion, and keep only the linear terms of the result to obtain

$$\tau \Delta \dot{W} = (1 + \gamma \Delta W) - (1 + \beta \Delta W) = -(\beta - \gamma)\Delta W. \tag{23}$$

From (23) it is seen that $\beta > \gamma$ yields the stable dynamical equation. This relation between β and γ is determined by the source degenerated pFET.

The functional form of the dynamics, illustrated in Fig. 4a, also hold when ΔV_g and ΔV_d are non-zero. Non-

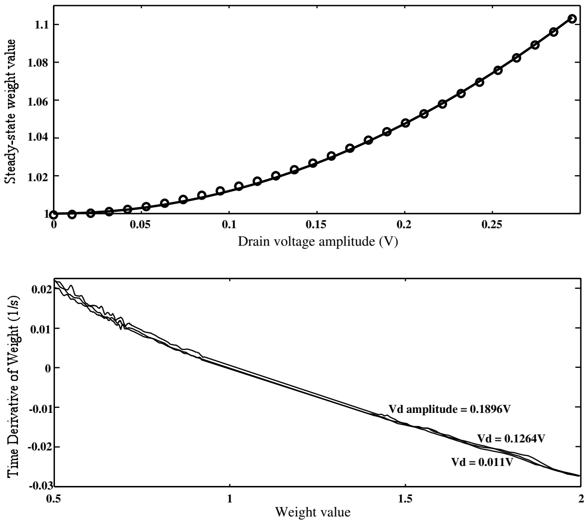


Fig. 6. Change in the dynamical weight equation due to sinusoidal signal amplitudes at only the drain terminal. The gate terminal is remaining at a fixed voltage through this experiment. We plot the $\frac{dW}{dt}$ versus w for three drain input amplitudes, as well as we plot the equilibrium w (W_{eq}) versus drain input amplitude.

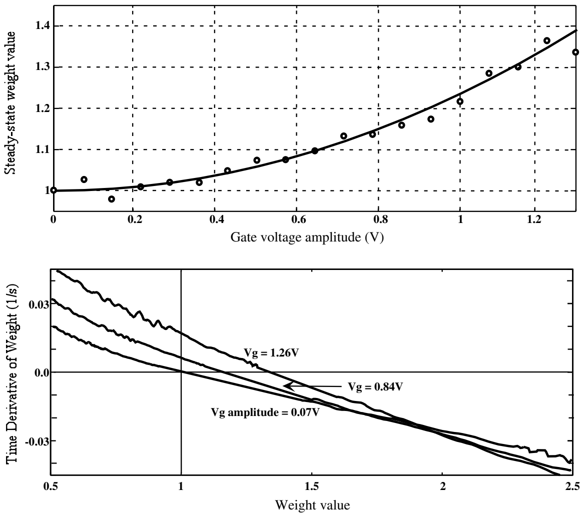


Fig. 7. Change in the dynamical weight equation due to sinusoidal signal amplitudes at only the gate terminal. The drain voltage is remaining at a fixed voltage through this experiment. We plot $\frac{dW}{dt}$ versus W for three gate input amplitudes, as well as we plot Equilibrium W (W_{eq}) versus gate input amplitude.

zero values of these variables will affect the non-zero equilibrium weight, the slope of the dynamics around this equilibrium weight, and the position of the separatrix between the two equilibrium weights. In this discussion we are mainly interested in the effects of the voltages on the non-zero equilibrium weight.

IV. Effects of Drain Voltage on the Equilibrium Weight

Our approximation of the weight dynamics begins by considering the effects on the equilibrium weight due to drain voltage alone. We start with the weight equation when $\Delta V_g = 0$. The weight dynamics then become

$$\tau \frac{dW}{dt} = W^{\gamma} e^{-\Delta V_d / V_{inj}} - W^{\beta}. \tag{24}$$

Figure 6 shows experimental measurements of $\frac{dW}{dt}$ vs. W for fast timescale drain voltage signals of various amplitudes and no gate voltage signal. We apply a slow timescale step input to the gate such that we get a sweep of $\frac{dW}{dt}$ vs. W. We apply a fast timescale input, $\Delta V_d = V_1 \sin(\omega t)$, at various amplitudes to view the effects of fast timescale signals on the weight dynamics. As before, we measure the source voltage to obtain the weight dynamics.

To observe the effects of the fast timescale signals on the weight equilibrium, we take the expected value and solving for W_{eq} we get

$$W_{eq} = E[e^{-\Delta V_d / V_{inj}}]^{1/(\beta-\gamma)} \tag{25}$$

Next we use a quadratic approximation of $e^{-\Delta V_d / V_{inj}}$ by truncating the Taylor series. We use a quadratic approximation because it is the minimum order expansion needed to show correlations in our circuit. Taking the expected value of each term of the approximation yields

$$W_{eq} \approx (1 + \frac{1}{2} E[(\Delta V_d / V_{inj})^2])^{1/(\beta-\gamma)} \tag{26}$$

The linear term for ΔV_d disappears since its expected value is zero. We use the binomial expansion approximation to eliminate the exponent $1/(\beta - \gamma)$. Substituting $W_{eq} = 1 + \Delta W_{eq}$ and solving for ΔW_{eq} gives

$$\Delta W_{eq} \approx \frac{1}{2} \frac{1}{(\beta - \gamma)} E[(\Delta V_d / V_{inj})^2]. \tag{27}$$

The data for this experiment uses a sinusoidal input, $\Delta V_d = V_1 \sin(\omega t)$. Substituting this into 27, we get

$$\Delta W_{eq} \approx \frac{1}{4} \frac{1}{(\beta - \gamma)} (V_1 / V_{inj})^2. \tag{28}$$

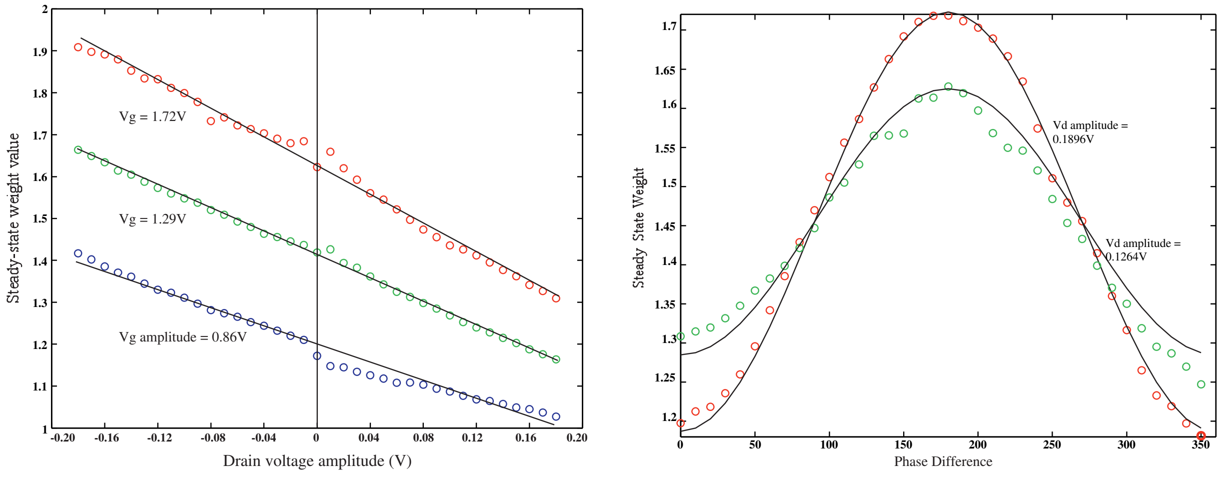
Figure 6 shows W_{eq} vs. V_1 with $V_{inj} \approx 500$ mV. The dependence of w_{eq} on V_1 is quadratic, but the V_{inj} term is large compared to the range of operation so that V_d has no real effect on the weight. Thus the effect of V_d on W_{eq} can be neglected.

V. Effects of Gate Voltage on the Equilibrium Weight

Substituting $\Delta V_d = 0$ in the weight equation allows us to observe the effects of gate voltage alone. This results in

$$\tau \frac{dW}{dt} = W^{\gamma} e^{\Delta V_g / V_{g0}} - W^{\beta} e^{-\Delta V_g / V_{g1}}. \tag{29}$$

Figure 7 shows experimental measurements of $\frac{dW}{dt}$ vs. W for fast timescale gate voltage signals of various amplitudes and no drain voltage signal. We get these dynamics using an experimental procedure similar to that used to measure drain voltage effects.



(a) (b)

Fig. 8. Change in the dynamical weight equation due to correlated sinusoidal signal amplitudes between the gate and drain terminals. (a) Change in the dynamical weight equation due to different size drain amplitudes for three different gate amplitudes. (b) Change in the dynamical weight equation due to various degrees of correlation (phase difference) between sinusoidal input signals at the gate and drain terminals.

As in the drain voltage case, we take the expected value, find the equilibrium weight, and make the quadratic approximation to get

$$W_{eq} \approx \left(\frac{1+\frac{1}{2}E[(\Delta V_g/V_{g0})^2]}{1+\frac{1}{2}E[(\Delta V_g/V_{g1})^2]}\right)^{\frac{1}{(\beta-\gamma)}}. \tag{30}$$

Applying the binomial expansion twice, and keeping only those terms relevant to the quadratic approximation we simplify the fraction and the exponent to get

$$W_{eq} \approx (1+\frac{E[(\Delta V_g/V_{g0})^2]}{2(\beta-\gamma)})(1-\frac{E[(\Delta V_g/V_{g1})^2]}{2(\beta-\gamma)}). \tag{31}$$

We multiply this out and truncate to the terms appropriate for our quadratic approximation in ΔV_g. Finally, substituting $W_{eq} = 1 + \Delta W_{eq}$ we get

$$\Delta W_{eq} \approx \frac{1}{2}\frac{1}{(\beta-\gamma)}E[(\Delta V_g/V_z)^2] \tag{32}$$

where $1/V_z^2 = 1/V_{g0}^2 - 1/V_{g1}^2$.

Applying a sinusoidal input $\Delta V_g = V_2\sin(\omega t)$, and taking the time average, we get

$$\Delta W_{eq} \approx \frac{1}{4}\frac{1}{(\beta-\gamma)}(V_2/V_z)^2. \tag{33}$$

Figure 7 shows W_{eq} vs. V_2 It is seen that the dependence of W_{eq} on V_2 is quadratic. We assume that $E[(V_g/V_z)^2]$ is constant and can be subtracted out of the weight dynamics equation, along with W = 1. This assumption is realistic if we apply automatic gain control to the input voltage.

VI. Equilibrium Weight is Determined by Correlations Between Gate and Drain Voltages

When we have fast timescale voltage signals on both the gate and the drain, we obtain the full equilibrium weight equation through a derivation similar to that used in the preceding sections:

$$\Delta W_{eq} \approx \frac{\frac{E[\Delta V_g^2]}{2V_z^2} - \frac{E[\Delta V_g \Delta V_d]}{V_{g0}V_{inj}} + \frac{E[\Delta V_d^2]}{2V_{inj}^2}}{\beta-\gamma}. \tag{34}$$

As we have previously stated, we can neglect the effects of the drain voltage. We can also assume that the signal energy in the gate voltage is constant. This allows us to subtract off the term due to gate voltage alone, counting it as part of the DC operating point of the circuit. Taking these two points into consideration, we obtain the result

$$\Delta W_{eq} \approx -\frac{1}{\beta-\gamma}\frac{E[\Delta V_g \Delta V_d]}{V_{g0}V_{inj}}. \tag{35}$$

This result shows that we can adjust the synapse weight according to correlations between the gate and drain terminal voltage signals.

We can see this correlation effect by applying fast timescale gate and drain voltages in two different experiments. The experimental procedure follows that in the earlier cases; we apply a slow timescale step to the gate and fast timescale sinusoidal voltages to the gate and drain terminals. Assume the input for $\Delta V_d = V_1\sin(\omega t)$ and $\Delta V_g = V_2\sin(\omega t + \theta)$. Substituting these two inputs into (35) and computing the expected value, we get

$$\Delta W_{eq} \approx -\frac{1}{2}\frac{1}{\beta-\gamma}\frac{V_1V_2\cos(\theta)}{V_{g0}V_{inj}}. \tag{36}$$

For the first experiment, consider the two sinusoidal inputs with no phase difference ($\theta = 0$). Figure 8a shows W_{eq} vs. ΔV_d for various values of ΔV_g. We see that for fixed values of gate voltage, that there is a negative linear dependence

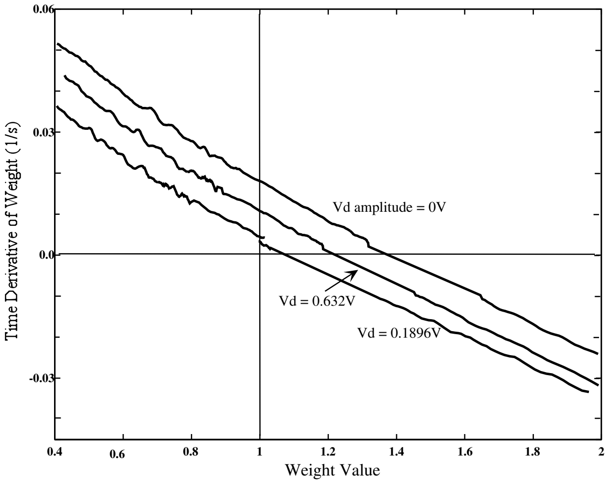


Fig. 9. dW/dt vs. W for varying ΔV_d and fixed ΔV_g. This data gives strong evidence for our learning rule, $\tau' \Delta\dot{W} = -\Delta W + \eta E[xy]$.

of ΔW on V_d as equation (35) suggests. In the second experiment, we sweep θ from 0 to 2π. For $\theta \neq 0$ we have Figure 8b shows W_{eq} vs. θ. We see definite correlations due to phase differences where $\Delta W_{eq} \propto -\cos\theta$.

VII. Hebbian Learning Rule from Approximated Weight Dynamics

We now present the approximated weight update rule for the linearized region using the results of the previous sections. The final result is

$$\begin{aligned} \tau' \Delta\dot{W} \quad \approx \quad & -\Delta W - \tfrac{1}{\beta-\gamma}\tfrac{1}{V_{g0}V_{inj}} E[\Delta V_g \Delta V_d] \\ & + \tfrac{1}{2} E[(\Delta V_g / V_z)^2] + \tfrac{1}{2} E[(\Delta V_d / V_{inj})^2] \end{aligned} \tag{37}$$

We derive the effects of the terminal voltages on the rate of the weight dynamics using methods similar to the equilibrium weight derivations. A new time constant, τ', is defined by the following

$$\begin{aligned} \tfrac{\tau}{\tau'} \quad \approx \quad & \beta - \gamma + \gamma \tfrac{E[\Delta V_g \Delta V_d]}{V_{g0}V_{inj}} \\ & -\tfrac{1}{2}(\gamma / V_{g0}^2 - \beta / V_{g1}^2) E[(\Delta V_g)^2] - \tfrac{\gamma}{2} E[(\tfrac{\Delta V_d}{V_{inj}})^2] \end{aligned} \tag{38}$$

Neglecting the drain voltage effects as before, and assuming the effects due to the gate voltage are constant and can be absorbed into other circuit constants, we obtain the final approximation to (21) as

$$\tau' \Delta\dot{W} \approx -\Delta W - \frac{1}{\beta - \gamma}\frac{1}{V_{g0}V_{inj}} E[\Delta V_g \Delta V_d] \tag{39}$$

where τ' is approximated as

$$\frac{\tau}{\tau'} \approx \beta - \gamma + \gamma \frac{E\{\Delta V_g \Delta V_d\}}{V_{g0}V_{inj}} \tag{40}$$

Figure 9 shows $\frac{dW}{dt}$ vs. W for varying ΔV_d and fixed ΔV_g. If we let $\eta = \frac{1}{(\beta-\gamma)V_{g0}Vinj}$, $x = \Delta V_{fg}$, and $y = -\Delta V_d$, then we can rewrite (39) as

$$\tau' \Delta\dot{W} = -\Delta W + \eta E[xy]. \tag{41}$$

This is a Hebbian learning rule, based on the correlations between the signals x and y. We would like to use this learning rule as the basis for several neural network and adaptive system algorithms such as principal components analysis, adaptive filtering, self-organizing neural maps, and supervised learning neural networks [15]. Our future research will explore these potential applications of the floating-gate pFET synapse.

VIII. Acknowledgements

The authors wish to thank Girish Patel for his useful comments and edits on this manuscript.

References

[1] P. Hasler, C. Diorio, B. A. Minch, and C. Mead, "Single Transistor Learning Synapses," *Advances in Neural Information Processing Systems 7*, MIT Press, 1995, pp. 817-824.
[2] P. Hasler, C. Diorio, B. A. Minch, and C. Mead, "Single transistor learning synapses with long term storage," *Proceedings of the International Symposium on Circuits and Systems*, Seattle, 1995, vol. III, pp. 1660-1663.
[3] P. Hasler, B.A. Minch, C. Diorio, and C.A. Mead, "An autozeroing floating-gate amplifier", *IEEE Transactions on Circuits and Systems II*, **Special Issue**, 2000.
[4] Paul Hasler, "Continuous-Time Feedback in Floating-Gate MOS Circuits," *IEEE Transactions on Circuits and Systems*, **Special Issue**, 2000.
[5] Paul Hasler, Chris Diorio, and Bradley A. Minch, "A four-quadrant floating-gate synapse," *Proceedings of IEEE International Symposium on Circuits and Systems*, Monterey, 1998.
[6] P. Hasler, Foundations of Learning in Analog VLSI, California Institute of Technology, February 1997.
[7] Paul Hasler, B. A. Minch, and C. Diorio, "Adaptive Circuits Using pFET Floating-Gate Devices," *20th Anniversary Conference on Advanced Research in VLSI*, Atlanta, March 1999, pp. 215-229.
[8] Paul Hasler, Bradley A. Minch, Jeff Dugger, and Chris Diorio, "Adaptive Circuits and Synapses Using pFET Floating-Gate Devices," in Gert Cauwenbergs, editor, *Learning in Silicon*, Kluwer Academic, 1999, pp. 33-65.
[9] Paul Hasler and Jeff Dugger, "Correlation Learning Rule in Floating-Gate pFET Synapses," *IEEE International Symposium on Circuits and Systems*, Orlando, vol. V, 1999, pp. 387–400.
[10] Chris Diorio, Paul Hasler, Bradley A. Minch, and Carver A. Mead, "A Floating-Gate MOS Learning Array with Locally Computed Weight Updates," *IEEE Transactions on Electron Devices*, vol. 44, no. 12, 1997, pp. 2281-2289.
[11] M. Lenzlinger and E.H. Snow, "Fowler–Nordheim tunneling into thermally grown SiO_2," *Journal of Applied Physics*, vol. 40, no. 1, 1969, pp.278–283.
[12] C. Mead, "Scaling of MOS technology to submicrometer feature sizes," *Journal of VLSI Signal Processing*, vol. 8, 1994, pp. 9-25.
[13] Y. Xu, *Electron Transport through Thin Film Amorphous Silicon — A Tunneling Study*, Ph.D. dissertation, Stanford University, 1992.
[14] Paul Hasler, Andreas Andreou, Chris Diorio, Bradley A. Minch, and Carver Mead, "Impact ionization and hot-electron injection derived consistently from Boltzmann transport," *VLSI Design*, in Press.
[15] John Hertz, Anders Krogh, and Richard G. Palmer. *Introduction to the Theory of Neural Computation*. Addison-Wesley, 1991.