CORRELATION LEARNING RULE IN FLOATING-GATE PFET SYNAPSES

Paul Hasler and Jeff Dugger

Georgia Institute of Technology
Atlanta, GA 30332-0250,
phasler@ee.gatech.edu

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 the signal on the gate terminal and y is the 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, 2, 3]. 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 we adapt the synapse weight based on the device's input and output signals. In this paper, we show how the synapse weight adapts based on correlations between signals applied to the gate and the drain of the floating-gate device.

Figure 1 shows the layout, cross-section, and circuit symbol for our floating-gate pFET device. 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. Signals on the gate terminal(s) of the pFET synapse capacitatively couple into the channel by way of the floating gate.

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. Hot-electron injection decreases the floating gate charge, and electron tunneling increases the floating gate charge. For a fixed gate voltage, decreasing the charge on the floating gate increases the effective gate voltage seen by the channel causing the channel current to increase. Similarly, increasing the charge on the floating gate lowers the effective gate voltage seen by the channel of the pFET synapse, causing the channel current to decrease. The injection current depends on the drain voltage and the channel current, whereas the tunneling current only depends upon the floating-gate voltage. We fix the tunneling voltage terminal to a constant bias voltage. The tunneling and injection currents in the device are small compared with transistor bias currents; therefore weight changes occur on a much slower timescale than do computations.

1. WEIGHT DYNAMICS OF THE FLOATING-GATE PFET SYNAPSE

We will consider floating-gate devices that use continuous electron-tunneling and hot-electron injection currents. In equilibrium, that is at the bias point, the tunneling and injection currents are equal. We define the weight resulting from the charge due to bias conditions to be equal to one.

For the basic floating-gate pFET, increasing the channel current increases the injection current, and in turn further increases the channel current. Thus, we have a positive feedback relation between channel current and injection current, This positive feedback does not give us a stable weight value that we can use in analog computations. We show experimental data of this phenomena in Fig. 2b.

To compensate for this effect, we introduce another pFET between V_{dd} and the source terminal of the floating-gate pFET. Figure 2a shows the circuit for the source-degenerated (s-d) pFET synapse. The s-d pFET synapse uses a weak exponential element—the short channel pFET, M2, in Fig. 2 —to change the injection-current dependence upon channel current [3]. The resulting source voltage is logarithmically related to the channel current. The added pFET, M2, provides source degenerative feedback to stabilize the injection current, as seen in Fig. 2b. For the s-d pFET, increasing channel current decreases the source voltage of the floating-gate pFET, decreasing its drain to source voltage, thereby decreasing its injection current.

The fundamental equation governing the weight dynamics of the floating-gate pFET synapse is given by [2, 3]:

$$\tau \frac{dw}{dt} = w^{\gamma} e^{(\Delta\hat{V}_g/V_{g0} - \Delta\hat{V}_d/V_{inj})} - w^{\beta} e^{-\Delta\hat{V}_g/V_{g1}} \qquad (1)$$

where γ, β, V_{g0}, V_{g1}, and V_{inj} are device dependent parameters and are related to quantities derived in [2, 3]. All voltages are referenced with respect to ground. The first term on the right side of (1) describes the weight dynamics due to hot-electron injection while the second term de-

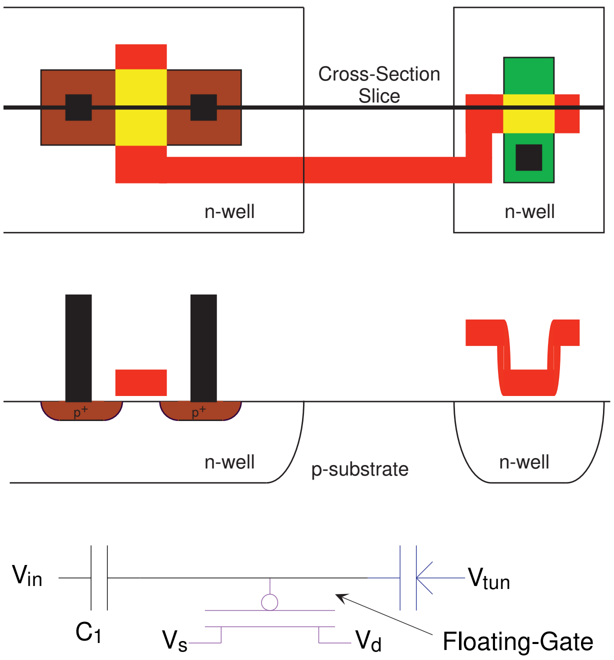
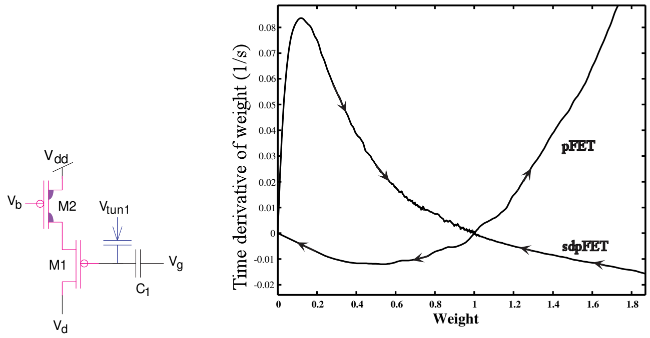


Figure 1: Layout, cross section, and circuit diagram of the floating-gate pFET in a standard double-poly $n$well MOSIS process. From the layout, we see that the floating-gate is a polysilicon layer with no contacts to other layers. The cross section corresponds to the horizontal line slicing through the layout (or top) view. The pFET transistor is the standard pFET transistor in the $n$well process. The gate input capacitively couples to the floating-gate by either a poly-poly capacitor, a diffused linear capacitor, or a MOS capacitor, as seen in the circuit diagram (not explicitly shown in the other two figures). 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.



(a) (b)

Figure 2: (a) Circuit diagram of the s-d pFET synapse. 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. 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. We utilize the DIBL effect in the upper pFET, M2, using a very short channel length to build a compact weakly exponential element. (b) Plot of the time-derivative of w versus w for the pFET, and s-d 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 s-d synapses will stabilize to the $w = 1$ steady state. The s-d pFET modifies the basic pFET dynamics by local negative-feedback. We use zero $\Delta\hat{V}_d$ and zero $\Delta\hat{V}_g$

scribes the weight dynamics due to electron tunneling. The inputs, ΔV_d or ΔV_g, are the sum of a slow rate signal, $\Delta\bar{V}_d$ or $\Delta\bar{V}_g$, and a fast rate signal, $\Delta\hat{V}_d$ or $\Delta\hat{V}_g$.

To understand the basic form of the weight dynamics, we look at the simplest case of (1) when $\Delta\hat{V}_g = \Delta\hat{V}_d = 0$. This allows us to see the dependence of $\frac{dw}{dt}$ on w. The resulting equation is

$$\tau\frac{dw}{dt} = w^{\gamma} - w^{\beta}. \qquad (2)$$

Figure 2b illustrates the dynamics of this equation for the basic floating-gate pFET and for a source degenerated pFET. From Figure 2b, we observe that 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 can not go below zero, and this steady state will not change with different inputs. Figure 2b also shows that the basic pFET synapse is unstable, but the source degenerated pFET is stable. The functional form of the dynamics illustrated in Fig. 2b also holds when $\Delta\hat{V}_g$ and $\Delta\hat{V}_d$ are non-zero. Non-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 equilibrium weights.

2. EFFECTS OF DRAIN AND GATE SIGNALS ON THE EQUILIBRIUM WEIGHT

Our approximation of the weight dynamics begins by considering the effects on the equilibrium weight due to only drain voltage or only gate voltage. We start with the weight equation when either $\Delta\hat{V}_g = 0$ or $\Delta\hat{V}_d = 0$, respectively. The weight dynamics then become

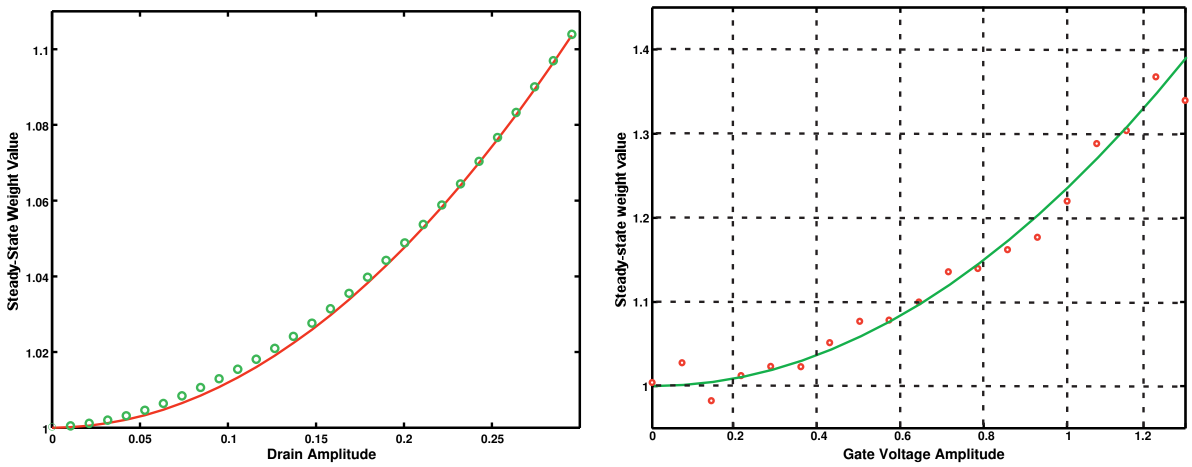
$$\Delta\hat{V}_g = 0: \;\; \tau\frac{dw}{dt} = w^{\gamma}e^{-\Delta\hat{V}_d/V_{inj}} - w^{\beta}.$$

$$\Delta\hat{V}_d = 0: \;\; \tau\frac{dw}{dt} = w^{\gamma}e^{\Delta\hat{V}_g/V_{g0}} - w^{\beta}e^{-\Delta\hat{V}_g/V_{g1}}. \qquad (3)$$

By taking the expected value of both sides of (3), we solve for the equilibrium weight as

$$\Delta\hat{V}_g = 0: \;\; w_{eq} = E[e^{-\Delta\hat{V}_d/V_{inj}}]^{1/(\beta-\gamma)}$$

$$\Delta\hat{V}_d = 0: \;\; w_{eq} = \left(\frac{E[e^{-\Delta\hat{V}_g/V_{g0}}]}{E[e^{-\Delta\hat{V}_g/V_{g1}}]}\right)^{1/(\beta-\gamma)}. \qquad (4)$$



(a) (b)

Figure 3: Change in the equilibrium weight due to sinusoidal signals; these signals sweep over the operating ranges of interest. (a) Change in the equilibrium weight due to sinusoidal signal amplitudes at only the drain terminal. The gate terminal is remaining at a fixed voltage through this experiment. (b) Change in the equilibrium weight (W_{eq}) versus gate input amplitude due to sinusoidal signal amplitudes at only the gate terminal. The drain voltage is remaining at a fixed voltage through this experiment.

Next we use a quadratic approximation of $e^{-\Delta\hat{V}_d/V_{inj}}$ and $e^{-\Delta\hat{V}_g/V_{g0}}$ by truncating their Taylor series expansions. We use quadratic approximation because it is the minimum order expansion needed to show correlations in our circuit. The linear term for $\Delta\hat{V}_d$ and $\Delta\hat{V}_g$ disappear since their expected values are zero. Substituting $w_{eq} = 1 + \Delta w_{eq}$ and solving for Δw_{eq} gives

$$\Delta\hat{V}_g = 0: \;\; \Delta w_{eq} \approx \frac{1}{2}\frac{1}{(\beta-\gamma)}E[(\Delta\hat{V}_d/V_{inj})^2].$$

$$\Delta\hat{V}_d = 0: \;\; \Delta w_{eq} \approx \frac{1}{2}\frac{1}{(\beta-\gamma)}E[(\Delta\hat{V}_g/V_z)^2], \qquad (5)$$

where $1/V_z^2 = 1/V_{g0}^2$ - $1/V_{g1}^2$. We used the binomial expansion approximation to eliminate the exponent $1/(\beta-\gamma)$.

We use sinusoidal inputs to obtain the experimental data shown in Fig 3. Applying either a sinusoidal input to the drain terminal, $\Delta\hat{V}_d = V_1\sin(\omega t)$, or to the gate terminal, $\Delta\hat{V}_g = V_2\sin(\omega t)$, and taking the time average, we get the following analytical model from (5):

$$\Delta\hat{V}_g = 0: \;\; \Delta w_{eq} \approx \frac{1}{4}\frac{1}{(\beta-\gamma)}(V_1/V_{inj})^2,$$

$$\Delta\hat{V}_d = 0: \;\; \Delta w_{eq} \approx \frac{1}{4}\frac{1}{(\beta-\gamma)}(V_2/V_z)^2. \qquad (6)$$

Figure 3a 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 affect on the weight. Thus the effect of V_d on w_{eq} can be neglected. Figure 3b shows w_{eq} vs. V_2 It is seen that the dependence of w_{eq} on V_2 is quadratic. We assume that $E[(\hat{V}_g/V_z)^2]$ 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.

3. CORRELATION LEARNING RULE

When both $\Delta\hat{V}_g$ and $\Delta\hat{V}_d$ are not zero, we obtain the full equilibrium weight equation through a derivation similar to that used in the preceding sections.

$$\Delta w_{eq} \approx \frac{\frac{E[\Delta\hat{V}_g^2]}{2V_z^2} - \frac{E[\Delta\hat{V}_g\Delta\hat{V}_d]}{V_{g0}V_{inj}} + \frac{E[\Delta\hat{V}_d^2]}{2V_{inj}^2}}{\beta-\gamma} \qquad (7)$$

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

$$\Delta w_{eq} \approx -\frac{1}{\beta-\gamma}\frac{E[\Delta\hat{V}_g\Delta\hat{V}_d]}{V_{g0}V_{inj}} \qquad (8)$$

This result shows that we can get a floating-gate device that computes the correlation of the two inputs.

We can see this correlation effect by applying sinusoidal gate and drain voltages in two different experiments. Assume the input for $\Delta\hat{V}_d = V_1\sin(\omega t)$ and $\Delta\hat{V}_g = V_2\sin(\omega t + \theta)$ Substituting these two inputs into (8) and computing the expected value, we get

$$\Delta w_{eq} \approx -\frac{1}{2}\frac{1}{\beta-\gamma}\frac{V_1V_2\cos(\theta)}{V_{g0}V_{inj}}. \qquad (9)$$

For the first experiment, consider the two sinusoidal inputs with no phase difference ($\theta = 0$). Figure 4a shows w_{eq} vs. $\Delta\hat{V}_d$ for various values of $\Delta\hat{V}_g$. We see that for fixed values of V_g, that there is a negative linear dependence of Δw on V_d as equation (8) suggests. In the second experiment, we examine weight equilibrium dependence on input phase

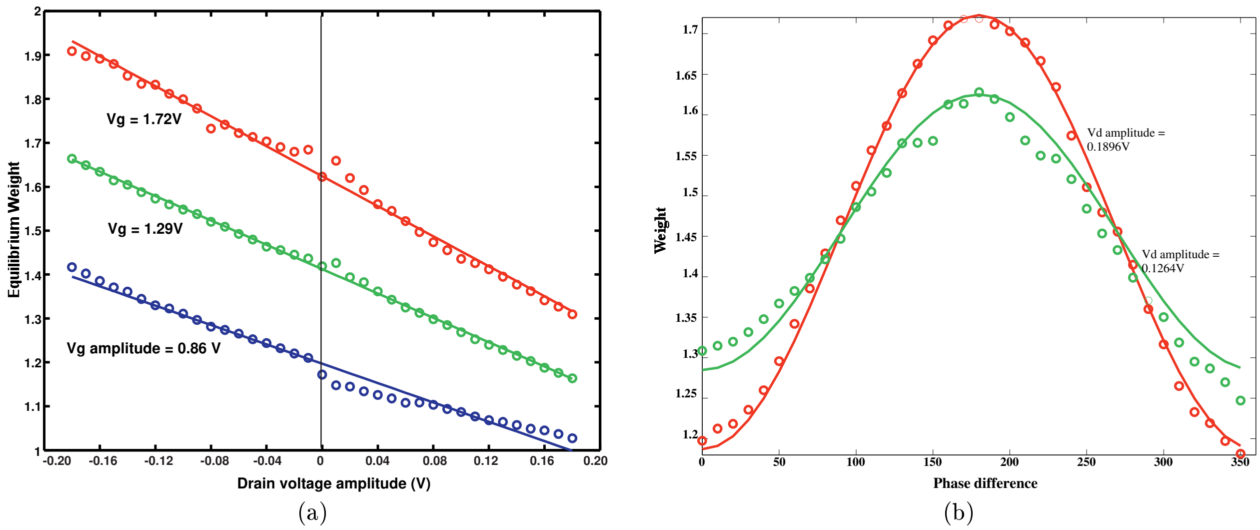


Figure 4: Change in the dynamical weight equation due to correlated sinusoidal signal amplitudes between the gate and drain terminals. (a) Change in the dynamical weight equation due to different size drain amplitudes for three different gate amplitudes. (b) Change in the dynamical weight equation due to various degrees of correlation (phase difference) between sinusoidal input signals at the gate and drain terminals. The amplitude of the input sine wave was 1.29V.

differences. Figure 4b shows w_{eq} vs. θ over the range 0 to 2π. We see definite correlations due to phase differences where $\Delta w_{eq} \propto -cos\theta$.

We now present the approximated weight update rule for the linearized region using the results of the previous sections. A linear expansion around $w = 1$ is derived by substituting $w = 1 + \Delta w$ into (2) or (1) and using the binomial expansion. Keeping only the linear terms of the expansion yields

$$\tau \Delta\dot{w} = (1 + \gamma\Delta w) - (1 + \beta\Delta w) = -(\beta - \gamma)\Delta w. \quad (10)$$

The full weight update equation after expanding (1) to second order and taking the expected value of sides is:

$$\begin{aligned} \tau' \Delta\dot{w} \quad \approx \quad & -\Delta w - \frac{1}{\beta-\gamma}\frac{1}{V_{g0}V_{inj}}E[\Delta\hat{V}_g\Delta\hat{V}_d] \\ & +\frac{1}{2}\frac{1}{\beta-\gamma}E[(\Delta\hat{V}_g/V_z)^2] + \frac{1}{2}\frac{1}{\beta-\gamma}E[(\Delta\hat{V}_d/V_{inj})^2] \end{aligned} \quad (11)$$

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 as approximately equal to $\tau/(\beta-\gamma)$, but a detailed analysis of (1) shows a weak dependence on the signal correlations which corresponds with experimental data. 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 (1) as

$$\tau' \Delta\dot{w} \approx -\Delta w - \frac{1}{\beta - \gamma}\frac{1}{V_{g0}V_{inj}}E[\Delta\hat{V}_g\Delta\hat{V}_d] \quad (12)$$

If we let $\eta = \frac{1}{(\beta-\gamma)V_{g0}Vinj}$, $x = \Delta\hat{V}_{fg}$, and $y = -\Delta\hat{V}_d$, then we can rewrite (12) as

$$\tau' \Delta\dot{w} = -\Delta w + \eta E[xy]. \quad (13)$$

This is a Hebbian learning rule, based on the correlations between the signals x and y [4]. This learning rule could serve 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.

4. REFERENCES

[1] 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.

[2] 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. Also at http://users.ece.gatech.edu/ phasler.

[3] P. Hasler, B. A. Minch, and C. Diorio, "Adaptive Circuits Using pFET Floating-Gate Devices," ARVLSI'99, Atlanta, March 1999.

[4] John Hertz, Anders Krogh, and Richard G. Palmer. Introduction to the Theory of Neural Computation. Addison-Wesley, 1991.