IMPROVED CORRELATION LEARNING RULE IN CONTINUOUSLY ADAPTING FLOATING-GATE ARRAYS USING LOGARITHMIC PRE-DISTORTION OF INPUT AND LEARNING SIGNALS

Jeff Dugger and Paul Hasler

Georgia Institute of Technology
Atlanta, GA 30332-0250,
jeffd@ece.gatech.edu

ABSTRACT

Floating-gate pFETs can be used to implement analog signal multiplication with an adaptive analog gain, creating the possibility for compact, low-power analog adaptive filters and neural networks in VLSI. We have shown that the adaptation mechanism of this device yields a weight which is proportional to correlations between terminal voltage signals. However, the weight contains dependencies proportional to the variances of these terminal voltages as well. In addition, harmonic distortion of the input signals can mask the correlation effect. In this paper, we show how to eliminate or miminize these non-idealities through logarithmic pre-distortion of our input and learning signals as well as through the use of differential circuit structures. The result is a weight update mechanism which is closer to a pure correlation learning rule.

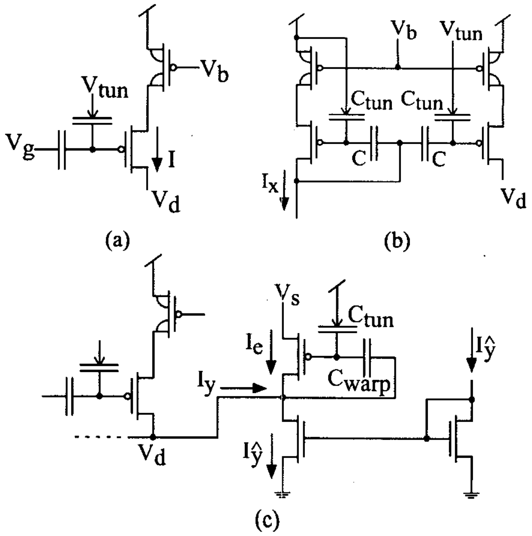


Fig. 1. (a) The source-degenerated floating-gate pFET which yields weight multiplication, weight storage, and the weight correlation learning rule. (b) The pre-distort circuit is a simple current mirror which implements $\Delta V_g = V_{mg} \ln(1 + x)$ to eliminate harmonic distortion in the output current. (c) The pre-distort circuit for the drain voltage implements $\Delta V_d = V_{md} \ln(1 + y)$ to eliminate drain variance terms in the weight value.

1. THE SOURCE-DEGENERATED PFET SYNAPSE AND CORRELATION LEARNING RULE

Floating-gate pFETs provide a device level implementation of analog signal multiplication by an adaptive analog gain [1]. These adaptive analog gain elements create the possibility of implementing compact, low-power analog adaptive filters and neural networks in VLSI. The source-degenerated floating-gate pFET device, appearing in Fig. 1a, produces a current (when operated in subthreshold) which is a product of an exponential function of its gate voltage and a weight factor [1]. This current is described as

$$I = I_b w e^{-\Delta V_g / V_{gA}}, \tag{1}$$

where I_b is the bias current, w is the weight, and $V_{gA} = \frac{1}{\kappa_x} \frac{C_T}{C_g} \frac{U_T}{\kappa}$ is a voltage scale factor. The source-degenerating feedback provides $\kappa_x \approx U_T/V_A$ (V_A is the Early voltage of the feedback transistor), while C_g is the capacitance coupling into the floating-gate, C_T is the total capacitance at the floating-gate, U_T is the thermal voltage, and κ is the capacitive coupling from the floating-gate to the channel surface potential. If we make the gate voltage amplitude small enough, then we can make a linear approximation to the exponential function and treat this device as a transconductance amplifier where the weight factor acts as an adaptive gain. Charge stored on the floating gate provides the weight, with electron tunneling and hot-electron injection providing weight update mechanisms. We previously showed that the weight update mechanism leads to a learning rule based on correlations between the gate and drain voltages of the source-degenerated floating-gate pFET [2, 3, 4].

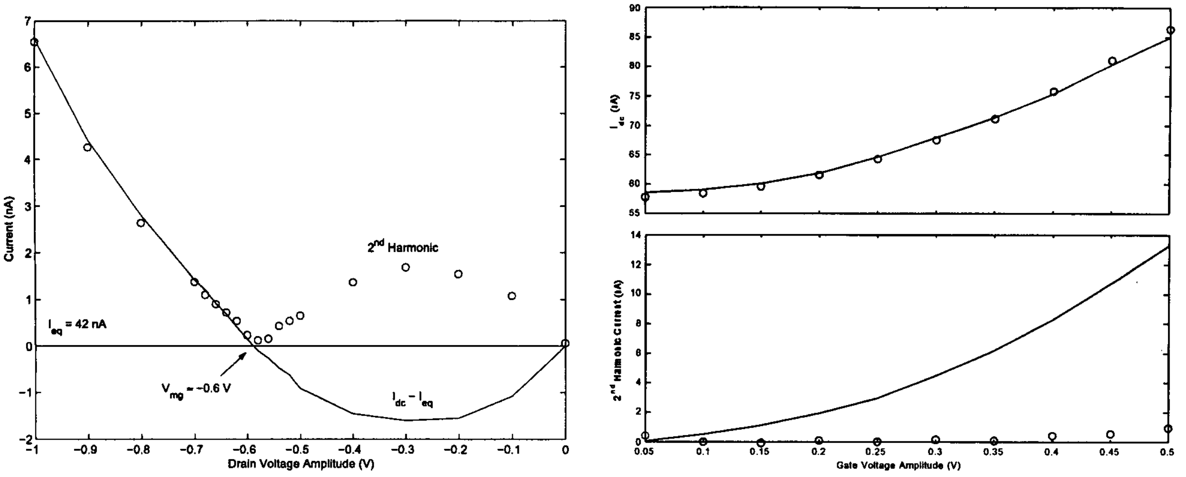
The mathematical starting point for the correlation learning rule is given by

$$\tau \frac{dw}{dt} = w^{\gamma} e^{\Delta V_g / V_{g0}} e^{-\Delta V_d / V_{inj}} - w^{\beta} e^{-\Delta V_g / V_{g1}}, \tag{2}$$

where $V_{g0} = V_{gA}/(1 - \gamma)$, $V_{g1} = V_{gA}/(\beta - 1)$, $\gamma = 2 -$

This work was partially supported by grants National Science Foundation (CISE-1068549, ECS (CAREER): 0093915, ECS-9988905) and by corporate donations to the Georgia Tech Analog Consortium by Texas Instruments and Motorola, Inc.

0-7803-7448-7/02/$17.00 ©2002 IEEE



(a) (b)

Fig. 2. (a) This plot shows the DC value (subtracting the equilibrium current) and 2^{nd} harmonic of the drain current vs. gate voltage pre-distort value. The gate pre-distortion is given by $\Delta V_g = V_{mg} \ln(1 + A \sin \omega t)$ for $\Delta V_d = 0$. We find that the appropriate pre-distort factor ($V_{mg} \approx -0.6V$)to be that which makes the 2^{nd} harmonic to be zero and coincides with the non-zero root of $I_{dc} - I_{eq}$. I_{eq} is the value the drain current assumes when there is no signal input. (b) These plots compare the dc values and second-harmonic values of the drain current vs. gate voltage amplitude both with and without pre-distortion. We see that the dc value follows a similar quadratic form in both cases, implying that there is still significant gate variance with pre-distortion for a given pre-distort value. The second-harmonic plot shows that harmonic distortion has been significantly reduced.

$\frac{1}{\kappa_x}(\frac{U_T}{V_{inj}})$, and $\beta = 1 + \frac{1}{\kappa_x}(\frac{1}{\kappa}\frac{U_T}{V_x})$. The constants V_{inj} and V_x result from models of injection and tunneling [2, 3, 4]. While we would like to obtain a learning rule purely dependent on correlations between the gate and drain voltages, the rule we actually obtain from (2) leads to weight dependencies on the variance of these two voltages as well. Also, we may not be able to make the gate voltage small enough to obtain a linear model of (1) while simultaneously keeping it large enough to see correlation behavior. The resulting harmonic distortion can mask the weight effects. In that case, we would like to eliminate harmonic distortion. Because the drain current of the source-degenerated floating-gate pFET and its equilibrium weight value are both determined by exponential functions of the terminal voltages, we explore the use of logarithmic pre-distortion to remove these undesired effects. Figure 1b shows the circuit structure which will provide gate pre-distortion, and Fig. 1c illustrates the circuit which will be used to provide pre-distortion on the drain voltage.

The experimental results reported here are from chips manufactured in the $0.5\mu m$ AMI process through MOSIS. We choose values of $V_{dd} = 8V$, $V_d = 2.15V$, $V_{tun} = 19V$, and $V_b = 7.55V$ to yield a desired range of subthreshold equilibrium currents for continuous-time adaptation, and we choose $V_d = 5V$ and $V_{tun} = V_{dd}$ to turn off the adaptation mechanism.

2. HARMONIC DISTORTION, GATE VARIANCE, AND GATE PRE-DISTORTION

We first remove the harmonic distortion from the pFET drain current by pre-distorting the gate voltage input. The pre-distorted voltage applied to the gate is

$$\Delta V_g = V_{mg} \ln(1 + A \sin \omega t). \tag{3}$$

In this experiment, the pre-distort factor, V_{mg}, ranges from -1 V to 0 V and the sinusoidal amplitude, A, is 0.65. This experiment is performed with the adaptation mechanism turned off such that the only effect observed is the harmonic distortion. Substituting (3) in (1) and simplifying yields

$$I = I_b(1 + A \sin \omega t)^{\alpha_g}, \tag{4}$$

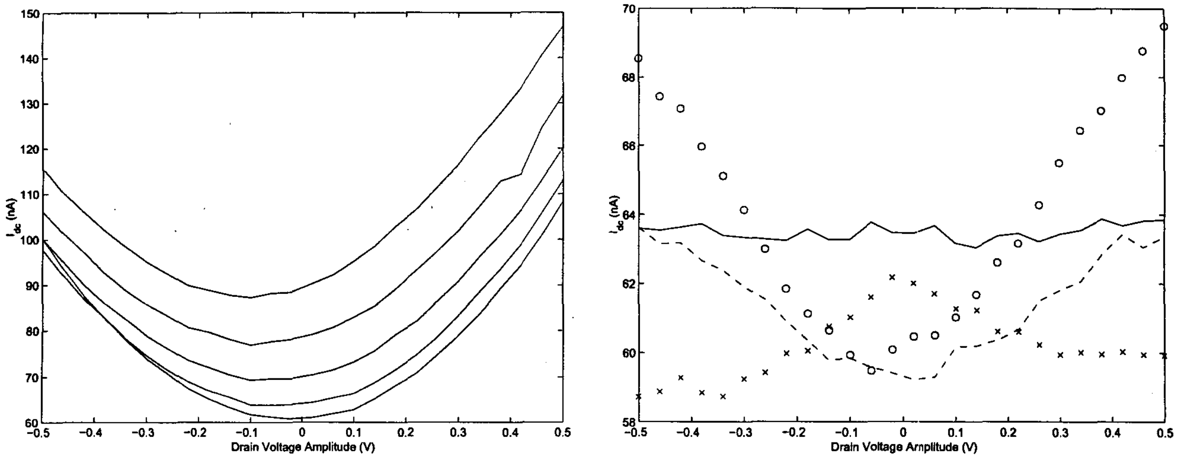
where $\alpha_g = -V_{mg}/V_{gA}$. A second-order approximation of (4) leads to

$$\begin{aligned} I \approx I_b(1 + \tfrac{1}{4}\alpha_g(\alpha_g - 1)A^2 + \alpha_g A \sin \omega t \\ - \tfrac{1}{4}\alpha_g(\alpha_g - 1)A^2 \sin 2\omega t). \end{aligned} \tag{5}$$

This second-order approximation indicates that we should look at the dc value and second harmonic of the drain current in Fig. 3 to determine the appropriate pre-distort factor that will eliminate harmonic distortion. From (5), we see that we should look for the value of V_{mg} where the second harmonic becomes zero, which gives us the appropriate pre-warp factor, V_{mg}. We also see that this coincides with the non-zero root of the dc current when the equilibrium value is subtracted. Figure 2a demonstrates that the appropriate pre-distort value for this gate is $V_{mg} \approx -0.6V$, which is the corresponding V_{gA}.

Both the dc values and the second-harmonic values of the drain current with and without pre-distortion are compared in Fig. 2b. Both sets of data are obtained with the continuous-time adaptation mechanism turned on. We obtain the non-distorted data by applying a sinusoidal input, $\Delta V_g = A_g \sin \omega t$, of various amplitudes ($0.05V < A_g < 0.5V$ in steps of 0.05 V) with no drain voltage change ($\Delta V_d = 0$). For the pre-distorted data, we use the factor, V_{mg}, found above. We choose the sinusoidal amplitude values, A, in 3 as

$$A = -\tanh(A_g/V_{mg}), \tag{6}$$



(a) (b)

Fig. 3. (a) A plot of DC value of the drain current (weight) vs. drain voltage amplitude for a pre-distorted sinusoid applied to the gate and a non-distorted sinusoid applied to the drain. The gate variance contributes to constant offsets in the data. The quadratic drain variance term masks the linear correlation term. (b) This plot shows the DC value of the drain current vs. drain voltage amplitude when $\Delta V_d = V_{md}\ln(1 + A\sin\omega t)$ with $\Delta V_g = 0$. Here we have plots for several values of V_{md}. We choose that value of $V_{md} = -0.22V$ (corresponding to the flattest curve) as the appropriate distortion factor to eliminate drain variance effects from the weight value.

where A_g is the set of amplitudes for the non-distorted case and V_{mg} is the pre-distort factor. This equation gives us comparable voltage amplitudes in both the non-distorted and pre-distorted cases. In Fig. 2b we see that dc current still displays a quadratic dependence, thus gate variance still has a significant effect even with pre-distortion. This happens because the pre-distort value for harmonic distortion cancellation and that for optimal gate-variance cancellation are not the same. This can be seen by comparing the exponential voltage scale factors in (1) and (2). Figure 2a compares the second-harmonic values of the drain current from the non-distorted and pre-distorted cases and shows that the harmonic distortion has been nearly eliminated in the pre-distorted case.

3. DRAIN VARIANCE EFFECTS, DRAIN PRE-DISTORTION, AND SIGNAL CORRELATIONS

Now that we have eliminated the effects of harmonic distortion, we observe the weight that results for non-zero gate and drain voltages. Figure 3a illustrates how the effects of gate and drain variance can mask the desired correlation dependence of the weight. We wish to examine how pre-distortion of the drain voltage affects the resulting weight value.

An analysis of the drain pre-distort circuit illustrated in Fig. 1c leads to the following

$$V_d = V_s - V_{md}\ln[(I_{\hat{y}} - I_y)/I_b], \tag{7}$$

where $V_{md} = \frac{C_T}{C_{warp}}\frac{U_T}{\kappa}$. In this paper, we set the target-signal current in (7) to a constant bias value ($I_{\hat{y}} = I_b$), and we define the learning signal to be $y = I_y/I_b$. This leads to the drain voltage function

$$V_d = V_s - V_{md}\ln(1 - y) \tag{8}$$

which is used in our learning rule. All of the experiments described here have the drain signal of $y = A\sin(\omega\, t + \phi)$.

We obtain the appropriate pre-distort factor for the drain by observing the dc current value vs. drain voltage amplitude for several pre-distort factors. Figure 3b exhibits the results of this experiment. The appropriate pre-distort factor gives the flattest curve. As seen from the figure, this value is $V_{md} \approx -0.22V$.

With this pre-distort value for the drain voltage, we observe the resulting weight values when we apply pre-distorted sinusoidal signals to both the gate and drain voltages. We hope to see the correlation effects more clearly. In Fig. 4a, we have plotted the dc current value vs. the drain voltage amplitude. We see a set of curves described by a tanh function, which is due to the relation given in (6) as applied to the drain. Plotting these dc current values vs. the actual drain signal amplitude, A, in Fig. 4b, we see a set of straight lines of various slopes. Steeper slopes correspond to larger gate voltage amplitudes. These lines correspond to $w \propto -E[xy]$ where x is the pre-distorted gate signal and y is the pre-distorted drain signal. This plot demonstrates that we indeed obtain clear linear correlation behavior with pre-distortion. In both plots we have subtracted offsets due to the gate variance.

4. CANCELLING CONSTANT OFFSET AND GATE VARIANCE IN THE WEIGHT

To obtain a pure correlation learning rule, we must remove the constant offset and gate-variance terms appearing in the equilibrium weight. In the ideal case of perfectly matched devices, a continuously-adapting differential pair of floating-gate transistors (Fig. 5a) provides the simplest way to remove these effects. However, Fig. 5b illustrates how mismatch between floating-gate devices results in significant differences in gate variance, rendering the differential scheme unworkable. Differences in weight

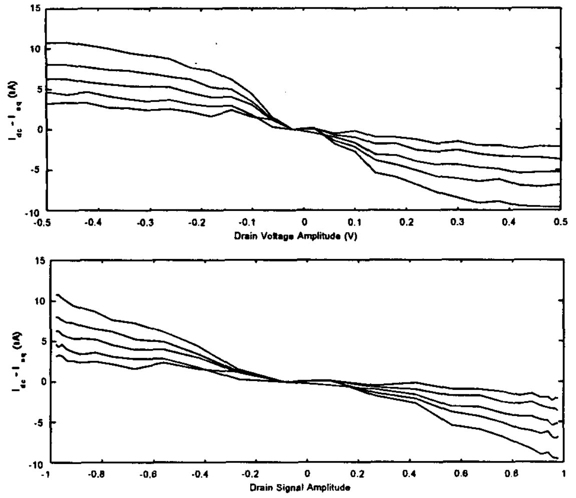
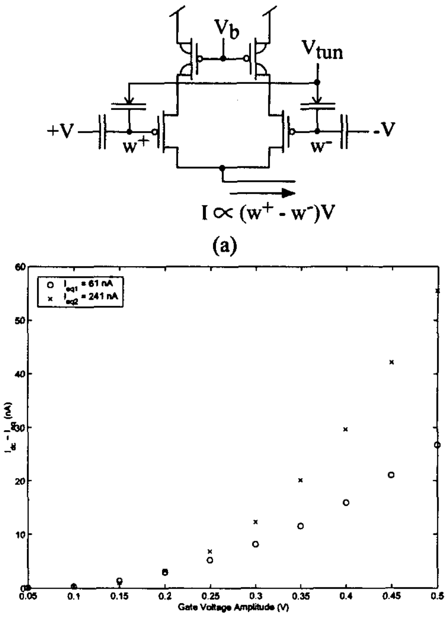


Fig. 4. The top figure is a plot of the dc current (weight) minus the equilibrium current value and those current offsets due to gate variance vs. the amplitude of the drain voltage. We see that the quadratic effect due to drain variance has been eliminated. The curves presented have a $tanh$ form due to the relationship between the signal amplitude and drain voltage amplitude because of pre-distortion. The bottom figure shows the same data plotted vs. signal amplitude revealing linear correlations. Steeper slopes correspond to larger gate voltages. Thus, we see that $w \propto -E[xy]$, where x and y are sinusoidal signals to be pre-distorted, is verified.



(a)

(b)

Fig. 5. (a) In the ideal situation of no mismatch, an adapting differential pair of floating-gate devices would cancel both the gate-variance and constant terms out of the correlation learning rule. (b) Plot of dc current values (weights) minus equilibrium values for two different source-degenerated floating-gate pFETs vs. gate voltage amplitude. Tunneling junction mismatch leads to significant differences in the gate-variance terms in the correlation learning rule. The legend demonstrates the large variation of equilibrium current values also due to mismatch.

offsets (due to differences in equilibrium current values) can be compensated using a differential floating-gate pair where one side continuously-adapts and the other side of the pair has its floating-gate charge programmed to cancel the offset. To eliminate the gate-variance problem without eliminating mismatch, we could force every input to a fixed amplitude using an automatic gain control circuit. The gate-variance term would then be a known constant to be programmed out along with the weight offset.

To remove gate variance by eliminating mismatch, we recognize that mismatch is due to the tunneling junction of the floating-gate device. In our current devices, most tunneling occurs at the edge of the tunneling junction; increasing the length of this edge for each device would tend to improve matching. Alternatively, we can attempt to tunnel directly through the well of the floating-gate device. Currently, we are exploring these methods to obtain an ideal correlation learning rule.

5. CONCLUSIONS

We have eliminated two nonlinear effects in our adaptive floating-gate devices which tend to mask the correlation term in our weight. We achieved this by pre-distorting the gate inputs to eliminate harmonic distortion and by pre-distorting the drain voltage to eliminate drain-variance terms in the weight equation. We presented simple circuits which can provide the appropriate pre-distortion functions in each case.

To cancel gate variance and dc offsets, we discussed a differential structure. We presented data that showed significant differences in weight offsets and gate variance between devices due to tunneling process mismatch. We conclude that miminimizing gate-variance effects will lead to significant improvements in the correlation learning rule.

6. REFERENCES

[1] M. Kucic, P. Hasler, J. Dugger, and D. Anderson, "Programmable and Adaptive Analog Filters using Arrays of Floating-Gate Circuits," in *2001 Conference on Advanced Research in VLSI* (E. Brunvard and C. Myers, eds.), pp. 148–161, Los Alamitos, CA: IEEE Computer Society, 2001.

[2] P. Hasler and J. Dugger, "Correlation learning rule in floating-gate pfet synapses," in *1999 IEEE International Symposium on Circuits and Systems*, vol. 5, (Orlando, FL.), pp. 387–390, May 30 - June 2 1999.

[3] P. Hasler, B. Minch, J. Dugger, and C. Diorio, "Adaptive circuits and synapses using pfet floating-gate devices," in *Learning on Silicon* (G. Cauwenberghs and M. Bayoumi, eds.), pp. 33–65, Norwell MA: Kluwer Academic Press, 1999.

[4] P. Hasler and J. Dugger, "Correlation learning rule in floating-gate pfet synapses," *IEEE Transactions on Circuits and Systems II*, vol. 48, pp. 65–73, January 2001.