A Continuously-Adapting Analog Node Using Floating-Gate Synapses

Jeff Dugger and Paul Hasler
Georgia Institute of Technology
Atlanta, GA 30332
jeffd@ece.gatech.edu, phasler@ece.gatech.edu

***Abstract*—We demonstrate that a network of pFET single-transistor learning synapses implements a multi-input adaptive node. These floating-gate synapses are capable of four-quadrant multiplications between an input and a weight, as well as adapting to a four-quadrant correlation between the input and a *learning* signal applied as a drain voltage. Our adaptive floating-gate node structure converges to the correct solution for constant RMS input values during adaptation. Our structure accounts for device mismatch, gate variance, and drain variance effects in the learning rule. We present experimental results for our circuit.**

We have come to see floating-gate devices not just as digital memories, but as computational circuit elements with analog memory and important time-domain dynamic features all in one device. We envision large networks based on a few EEPROM type elements with simultaneous nonvolatile weight storage, matrix-vector multiplication, and continuous weight adaptation. In this approach, the memory element becomes a part of the computation, allowing for very high chip density, which will lead to the implementation of large scale adaptive arrays on a single chip.

This technology when introduced as single transistor floating-gate synapses [1], initially proposed complicated weight-update equations based upon the programming physics. We call these elements synapses because of their relationship to synapses in adaptive filters and neural networks [2], and their loose connection with biological synapses [5]. Recent research illustrates the spectrum of floating-gate applications including adaptive floating-gate dynamics [6], programmable floating-gate filters [7], and possible weight-update rules based upon correlations of voltage signals applied to circuit terminals [9], [8], [10]. In this paper, we present experimental results on an adaptive floating-gate learning node built from an array of pFET single transistor synapses. This work is the first step in building dense on-chip learning networks implementing a wide space of learning algorithms.

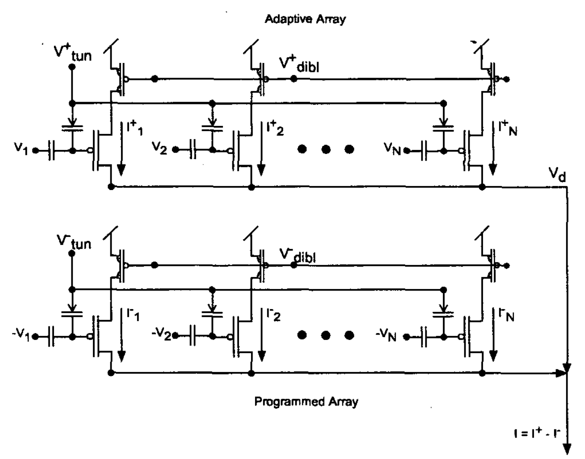


Fig. 1. The differential synapse array enables four quadrant multiplication of weights with inputs. Each floating-gate transistor behaves as a transistor amplifier with an adaptive gain. Adaption occurs through slow-timescale dynamics due to the tunneling and injection process at the floating-gates. We continuously adapt the synapses with the positive signals, and we program the synapses with negative signals to balance the steady-state current of the positive synapse. The network output is the total drain current, and the *learning* signal is the applied drain voltage. By applying the appropriate relationship between output drain current and the resulting drain voltage, we could get a variety of learning rules. Since pFET hot-electron injection is unstable when channel current is the free parameter, we stabilize each synapse element by incorporating DIBL transistors to provide source-degenerative feedback.

I. Floating-Gate Circuits Adapt to Signal Correlations

Figure 1 shows our adaptive floating-gate node, which is built from an array of differential four-quadrant synapses. The floating-gate transistor is a compact learning element which acts like an adaptive gain amplifier with built-in storage and gain adaptation. We define the adaptive gain as the weight. Gate voltages are input signals, and the drain voltage serves as a common learning signal for the node.

Because a single floating-gate device only allows positive weight values, we use a differential pair of synapses

Proc. 43rd IEEE Midwest Symp. on Circuits and Systems, Lansing MI, Aug 8-11, 2000
0-7803-6475-9/00/$10.00 ©IEEE 2000

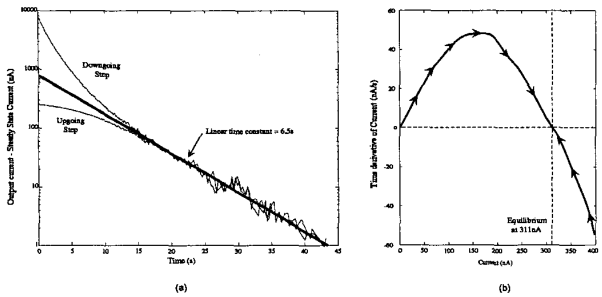


Fig. 2. Experimental measurements of floating-gate dynamics from a 0.5μ gate-length process. The gate input is a step decrease in voltage followed later by a step increase in voltage. We used a power supply voltage, $V_{dd} = 5.60V$, and tunneling voltage, $V_{tun} = 15.65V$, to set the operating point for these experiments, as well as the other measurements in this paper. (a) Convergence of the output current of a single synapse back to equilibrium after the step perturbations. The response is nonlinear with asymmetries between tunneling and injection. These nonlinear asymmetries are what allow us to compute correlations. b Time derivative of the weight ($\dot{w}$) vs. weight value (w). We graphically see that weight value converges toward its equilibrium level.

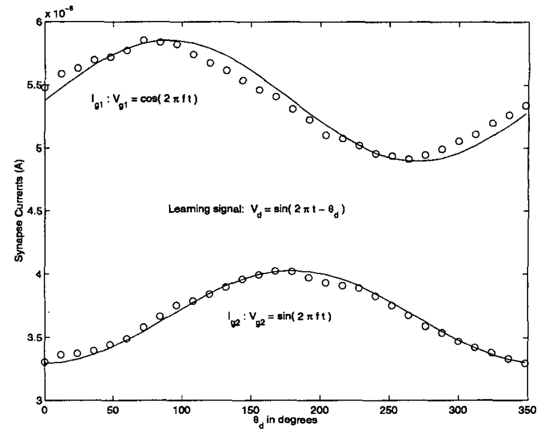


Fig. 3. Experimental measurements of positive synapse current phase correlations. We would program the negative synapses to currents of 5.5μA (w_1^-) and 3.5μA (w_2^-), and therefore both weights are either positive or negative. These results show correlations between a specific gate and global drain terminal direct the convergence of that synapse's weight.

for each input signal, to obtain signed weights [11]. In this circuit, we continuously adapt one set of weights, which we will call the positive weights, while the other set of weights, which we will call the negative weights, does not adapt. Instead, the negative weights are programmed to give zero-point references for the full four-quadrant weight values. The output current of a differential synapse is described by

$$I = I^+ - I^- = I_b \sum_i (w_i^+ + w_i^-) + g'_m \sum_i (w_i^+ - w_i^-)\Delta V_i \quad (1)$$

where I_b is the bias current for every synapse in the array, $g'_m = \kappa' I_b / U_T$ is the transconductance of the floating-gate transistors, V_i is the input into the ith synapse, and w_i^+ and w_i^- are the plus and minus weights of the ith synapse, respectively. As a result, we get a weighted sum of inputs riding on a slowly moving bias current.

The synapse weight adapts due to tunneling and injection currents at the floating node of the transistor. The dynamics of tunneling and injection processes appear in Fig. 2. We see that these processes converge to a stable equilibrium. Usually, the timescale of adaptation is much slower than the timescale of computation. These multiple timescales are essential for implementing adaptive circuits. Linearizing the dynamics around the stable equilibrium, we obtain the weight update equation [8], [10]

$$\tau' \dot{w}_i = -w_i + a\langle \Delta V_i^2 \rangle + b\langle \Delta V_d^2 \rangle + c\langle \Delta V_i \Delta V_d \rangle, \quad (2)$$

where V_d is the drain voltage, and we define $w_i = w_i^+ - w_i^-$. The constants a, b, and c are voltage-normalizing terms dependent on tunneling and injection parameters and are discussed comprehensively elsewhere [10], [9]. Solving for the steady-state, or equilibrium weight, yields

$$w_i = -\langle V_i V_d \rangle. \quad (3)$$

We program the appropriate zero-point reference weights, which cancels the constant term in (2), and constrain each input at some constant RMS value. Applying differential voltage signals for each input to both the adapting and non-adapting synapses cancels the gate and the drain variance terms in the weight update rule (2).

We want to compare this with the solution to the least mean square algorithm [3], $\vec{w} = Q^{-1}(\vec{V} d)$, where Q is the autocorrelation matrix of the input signal, $\vec{V}$ is a vector of inputs, and d is the desired learning signal. For orthogonal inputs (Q is diagonal), the weight equations are solved as

$$w_i = \frac{< (V_i)(d) >}{< V_i^2 >} \quad (4)$$

which compares well to (3) when the input variance remains constant. Thus, we can use this floating-gate node in adaptive signal processing and neural network applications.

II. From Learning Synapses to Learning Nodes

The synapse is the basic element in any learning system. Nodes are the next computational element to consider in adaptive networks. While a synapse is usually a single adaptive weight element, a node combines many synapses

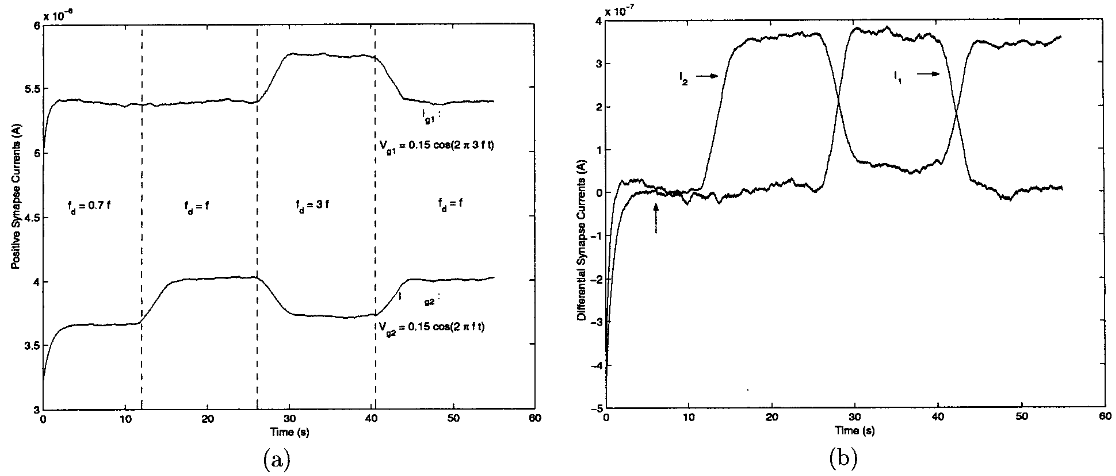


Fig. 4. Experimental measurements of frequency correlations for sinusoidal inputs to the two-input node given by $V_{g1} = sin(2\pi 3ft)$ and $V_{g2} = sin(2\pi ft)$. The learning signal, $V_d = sin(2\pi f_d t)$, takes on three different frequencies $f_d = 0.7f, f, 3f$. (a) Adaptation of synapse weights for various frequencies: $f_d = 0.7$, neither side adapts; $f_d = f$, only V_{g2} adapts; $f_d = 3f$, only V_{g1} adapts. (b) Current outputs from our differential synapses. We programmed our negative weights to offset the positive synapse current measured with input signals with the same signal variance. We see that the synapse that has an identical synapse input frequency as the drain signal has a non-zero weight.

by summation. In some cases, the node output is simply the weighted sum of inputs. In other cases, the node output is a nonlinear function of the weighted sum. In this paper we consider only the linear case. Combining several single transistor learning synapses results in a single node element for an adaptive circuit. An adaptive floating-gate node circuit provides a foundation to investigate many adaptive networks and learning algorithms.

In the first experiment, we show adaptation to phase correlations of fixed-phase sinusoids applied to each synapse input with a sinusoid of variable phase applied as the learning signal. Figure 3 shows the results of our first experiment. The inputs to the node are two sinusoidal signals, with the same frequency (f): $V_{g1} = 0.15cos(2\pi ft)$ and $V_{g2} = 0.15sin(2\pi ft)$, where the drain *learning* signal is $V_d = 0.3sin(2\pi ft - \theta_d)$ From (3) we expect steady-state synapse weights as

$$w_1 = sin(2\pi\theta_d) \text{ and } w_2 = -cos(2\pi\theta_d), \qquad (5)$$

The experimental data in Fig. 3 shows the resulting positive weight values when we swept the phase. Experimental data shows close agreement with analytic results.

As (3) predicts and Fig. 3 verifies, we get an anti-correlation instead of a correlation learning rule, where in the case of V_{g2} and V_d, the minimum value occurs when θ_d is 0° and the maximum value occurs when θ_d is 180°. Simply negating the learning signal applied to the drain terminal will yield the desired correlation rule.

In the second experiment, we show adaptation to frequency correlations when we apply a different frequency to each of the synapse inputs, and we apply another sinusoid as a learning signal. Figure 4 shows experimental time-course measurements from our adaptive circuit where the inputs are sinusoids at a fundamental and related third harmonic frequency, and the drain voltage is also another sinusoid. We show the signals from the positive synapses and again from the differential synapses. We programmed the negative synapses to eliminate the steady-state current of the positive synapses for similar input sizes. This approach also compensates for the mismatch between synapse equilibrium points. We observe that the circuit identifies a correlation between its input and the drain *learning* signal.

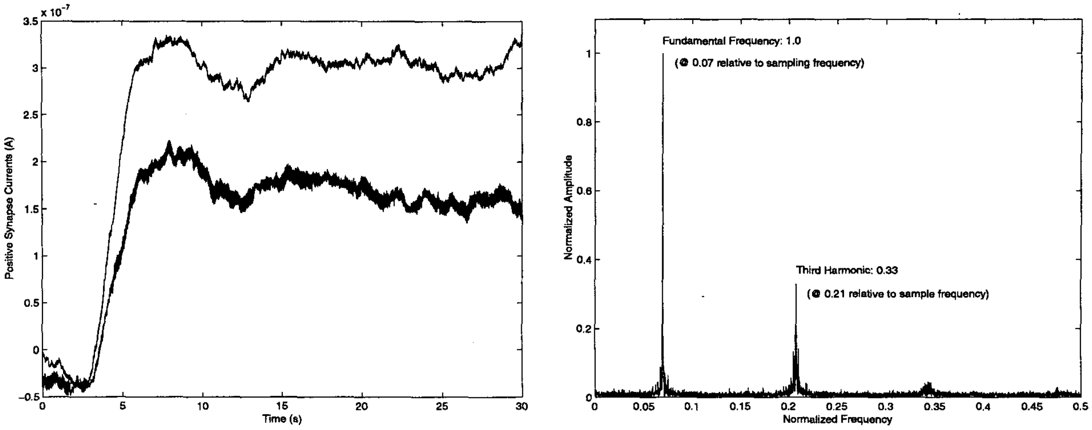
III. Learning a Square Wave from Sinusoidal Inputs

We present another example where we train the network to learn the appropriate Fourier coefficients for the components of a square wave. Our experimental results agree with theoretical expectations. From the definition of Fourier series a periodic signal can be expressed as:

$$s(t) = \sum_n a_n cos(2\pi nft) + \sum_n b_n sin(2\pi nft), \qquad (6)$$

where the coefficents result from the correlation of harmonically related sinusoids with the input signal given as:

$$a_n = \langle s(t), cos(2\pi nft)\rangle, \; b_n = \langle s(t), cos(2\pi nft)\rangle. \qquad (7)$$



(a) (b)

Fig. 5. Experimental measurements of a square wave learning signal applied to V_d (a) Time course of steady-state currents showing convergence of weights. (b) Spectrum of output current shows amount of each node input frequency matched to frequency components in learning signal. The frequency axis is normalized; the frequency does not affect these results until it approaches the adaptation rate. We obtain 1 and 1/3 for the fundamental and third harmonics as expected. The fifth harmonic appears due to the drain voltage coupling into the floating gate through overlap capacitance.

As a result, we expect our adaptive circuit to converge to these coefficients.

Figure 5a shows experimental time-course measurements from our adaptive circuit where the inputs are sinusoids at a fundamental and related third harmonic frequency, and the drain voltage learning signal is a square wave. We show the convergence of the signals in Fig. 5a for the positive synapses. Figure 5b shows the normalized amplitude and frequency of the Fast Fourier Transform of the resulting signal to make it easy to compare relative amplitudes at relative frequencies. From this experimental data, we get the expected square wave Fourier coefficients for the fundamental and third harmonics. This experiment demonstrates this circuit's behavior in extracting Fourier coefficients.

IV. Conclusions

We have demonstrated that the single transistor learning synapse can be utilitzed to implement an adaptive node with many inputs. Using the differential four-quadrant synapse structure, we get convergence to the correct solution, when we assume constant RMS values for the input signals. Our structure accounts for device mismatch, gate variance, and drain variance effects in the learning rule. We have presented experimental results for nodes with two inputs. This work is the foundation for building large, adaptive multinode networks.

References

[1] Paul Hasler, Chris Diorio, Bradley A. Minch, and Carver A. Mead, "Single transistor learning synapses," in Gerald Tesauro, David S. Touretzky, and Todd K. Leen (eds.), *Advances in Neural Information Processing Systems 7*, MIT Press, Cambridge, MA, 1995, pp. 817-824.

[2] S. Haykin, "Neural Networks: A Comprehensive Foundation," *Prentice-Hall*, 1999.

[3] Clarkson, Peter M. *Optimal and Adaptive Signal Processing*, CRC Press, Boca Raton, 1993

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

[5] P.S. Churchland and T.J. Sejnowski, *The Computational Brain*, MIT Press, Cambridge, MA, 1992.

[6] Paul Hasler, "Continuous-Time Feedback in Floating-Gate MOS Circuits," *IEEE Transactions on Circuits and Systems*, in Press.

[7] Matt Kucic, AiChen Low, Paul Hasler, and Joe Neff, "A Programmable Continuous-time Floating-Gate Fourier Processor," *IEEE Transactions on Circuits and Systems II*, in Press.

[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 on 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] Paul Hasler and Jeff Dugger, "Correlation Learning Rule in Floating-Gate pFET Synapses," *IEEE Transactions on Circuits and Systems II*, in Press.

[11] Paul Hasler, Chris Diorio, and Bradley A. Minch, "A four-quadrant floating-gate synapse," *IEEE International Symposium on Circuits and Systems*, Monterey, 1998, pp. 29-32.