A Precision CMOS Amplifier Using Floating-Gates For Offset Cancellation

Venkatesh Srinivasan*, Guillermo J. Serrano†, Jordan Gray‡ and Paul Hasler§

School of Electrical and Computer Engineering
Georgia Institute of Technology, Atlanta, Georgia 30332–0250
Email: *vsriniva@ece.gatech.edu; †gserrano@ece.gatech.edu; ‡jgray@ece.gatech.edu; §phasler@ece.gatech.edu

***Abstract*— A long-term offset cancellation scheme that enables continuous-time amplifier operation is described. Offset cancellation is achieved by programming floating-gate transistors that form an integral part of the amplifier's architecture. The offset voltage of a single-stage folded cascode amplifier is reduced to $\pm 25\mu V$ in a $0.5\mu m$ digital CMOS process. The offset voltage drift is $0.5\mu V$ over a period of 10 years at $25°C$ and varies by a maximum of $130\mu V$ over a temperature range of $170°C$.**

I. Offset Removal

A floating-gate based offset cancellation scheme is presented that results in a continuous-time amplifier with long-term offset cancellation. A prototype amplifier has been fabricated with its offset voltage reduced to $\pm 25\mu V$. The use of floating-gates for correcting mismatches in analog circuitry is particularly advantageous as it offers programmability, long-term retention and can be fabricated in a standard digital CMOS process. This approach involves no sampling and hence avoids such issues as charge injection, clock feedthrough and undersampling wideband noise that are serious limitations to autozeroing and correlated double sampling [4]. Also, unlike chopper stabilization [4], the proposed scheme is not limited to low-bandwidth applications while offering continuous-time operation with comparable offset reduction.

Fig. 1 shows a conceptual representation of the offset cancellation scheme. Floating-gate transistors are used as programmable current sources ($I_{os'}$) that provide offset compensation while being a part of the amplifier of interest during normal operation. Such an approach results in a compact architecture with a simple design strategy that avoids the overhead of using floating-gates as separate trimming elements as in [1], [2] and the offset cancellation by itself dissipates no additional power.

Section II of the paper introduces floating-gates, its programming and results from long-term charge retention experiments. Section III describes the prototype folded-cascode amplifier that uses floating-gate programmable current sources for offset cancellation. Experimental results from a $0.5\mu m$ digital CMOS process are presented in section IV. Finally, section V compares the proposed scheme with alternate approaches and concludes by summarizing the results and implications of the proposed approach.

II. Floating-Gate MOS Transistor

A floating-gate MOS transistor is a transistor whose polysilicon gate is completely surrounded by SiO_2, a high quality insulator. This creates a potential barrier that prevents charge stored on the floating-gate from leaking off of the floating node. Fig. 2(a) shows the circuit schematic and layout of a single-poly floating-gate pMOS transistor. In order to maintain the non-volatile charge storage of the floating-gate, external inputs are capacitively coupled through an input capacitor C_{in}. It should be noted that the second polysilicon layer shown in Fig. 2(a) is used primarily to implement the input capacitor.

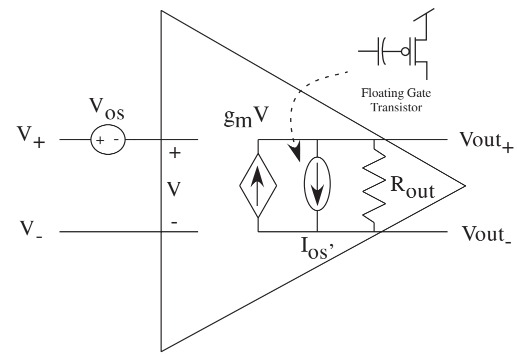


Fig. 1. **Offset Cancellation Macromodel**: The offset voltage of the amplifier V_{os} is cancelled by programming an offset current $I_{os'}$ in the opposite direction on floating-gate transistors.

A. Programming a Floating-Gate Transistor

Programming involves adding or removing charge from the floating-gate thereby modulating the threshold voltage of the device. This is achieved through the physical phenomena of hot-electron injection that adds electrons to the floating-gate and Fowler-Nordheim tunneling that removes electrons. In this approach, tunneling is used primarily as a global erase (achieved by capacitive coupling through C_{tun}) while precision programming is achieved using hot-electron injection [6]. Such a scheme has a number of advantages over a tunneling based programming scheme as in [1], [2]. These include faster programming (as the logarithmic dependence of tunneling makes precision programming highly time-consuming [5]), avoiding special processing steps such as ultra-thin tunneling oxide and the use of high voltages of both positive and negative polarities.

In this work, programming is achieved by first isolating the floating-gate transistor from the rest of the circuitry and applying a sufficient source-drain voltage for a specific period of time that is based on the desired target floating-gate drain current [6]. Also, using such a scheme, programming accuracies within 0.2% of the target currents have been achieved and currents as low as $1nA$ have been programmed [6]. Fig.

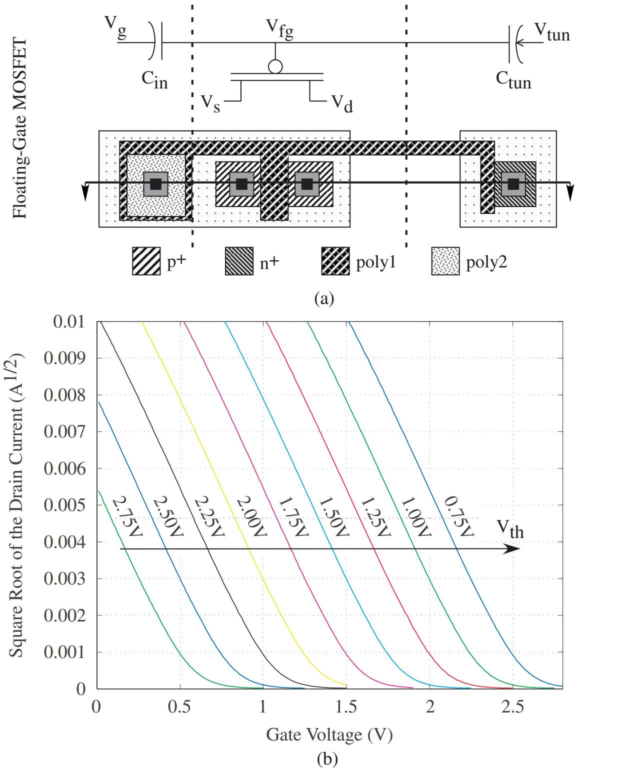


Fig. 2. **pFET floating-gate transistor: Schematic, Layout and Programming**: (a) Circuit schematic and layout of a pFET floating-gate transistor. (b) DC sweeps showing a floating-gate pFET programmed to different threshold voltages ranging from $0.75V$ to $2.75V$.

2(b) demonstrates programmability in floating-gate transistors by programming the device to different threshold voltages.

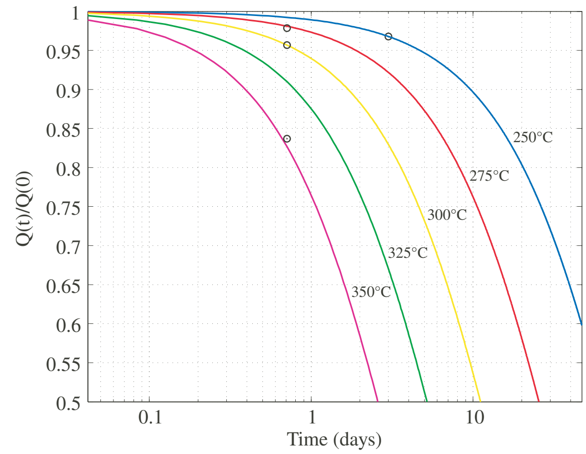
B. Charge Retention in Floating-Gate MOSFETs

Floating-gate transistors inherently have good charge retention capabilities on account of the gate being surrounded by a high quality insulator. Long-term charge loss in floating-gates occur due to the phenomenon of thermionic emission [8], [7]. The amount of charge lost is a function of both temperature and time and is given by,

$$\frac{Q(t)}{Q(0)} = exp\left[-t\upsilon \cdot exp\left(\frac{-\phi_B}{kT}\right)\right] \tag{1}$$

where, $Q(0)$ is the initial charge on the floating-gate, $Q(t)$ is the floating-gate charge at time t, υ is relaxation frequency of electrons in poly-silicon, ϕ_B is the $Si-SiO_2$ barrier potential, k is the Boltzmann's constant and T is the temperature. As expected from (1), charge loss in floating-gates is a slow process that is accelerated at high temperatures.

Predicting charge loss in floating-gates requires the estimation of the parameters υ and ϕ_B for the $0.5\mu m$ process used in the design. The values of parameters υ and ϕ_B were estimated to be $60s^{-1}$ and $0.9eV$ using experimentally measured values of charge loss for different temperatures and time periods. The charge loss was measured indirectly by measuring the change in threshold voltage of the device when exposed to high temperatures ($> 250°C$) for a prolonged period of time.



	Programmed 10% change in current			Programmed 50% change in current		
Temperature	**ΔQ/Q**	**ΔV_{fg}**	**ΔI/I**	**ΔQ/Q**	**ΔV_{fg}**	**ΔI/I**
25°C	1e-3%	36.7nV	2e-4%	1e-3%	156nV	9e-4%
90°C	0.62%	16.3µV	0.06%	0.62%	65µV	0.57%
140°C	18.2%	1.8mV	1.8%	18.2%	1.92mV	10.7%

Fig. 3. **Floating-Gate Retention**: The plot shows the measured charge loss (○'s) plotted with an extrapolated theoretical fit (solid) for different temperatures and time. The table summarizes the percentage change in the floating-gate charge, voltage and current for two different cases: (a) 10% programming change from initial (b) 50% programming change from initial.

Fig. 3 shows the measured floating-gate charge loss along with a theoretical extrapolated fit using the estimated model parameters. Also in Fig. 3, a summary of the percentage change in floating-gate charge, voltage and currents for two different cases of (a) 10% programming change from initial and (b) 50% programming change from initial are given. The measured data agrees well with the theoretical prediction and the trends observed in Fig. 3 have been observed across many floating-gate devices. The values in the table inset in Fig. 3 have been evaluated using (1) and assuming a subthreshold operation. No significant change can be extrapolated for programmed currents for a period of 10 years at room temperature, indicating good charge retention in floating-gate devices.

III. Amplifier Architecture

A single stage folded cascode amplifier shown in Fig. 4 demonstrates a practical implementation of the proposed approach. The currents through the floating-gate transistors $M3$ and $M4$ are programmed such that they cancel the offset arising from mismatches in the input differential pair ($M1$, $M2$), the cascoded current mirrors ($M5-M8$) and transistors $M3-M4$ themselves. During normal operation, the switches $S1$ and $S2$ are set such that the floating-gate transistors are a part of the operational amplifier. During programming, the floating-gates are isolated from the amplifier and a difference current $\Delta I(I3-I4)$ is programmed such that the offset voltage is nullified.

The input referred offset voltage of the amplifier is related to the difference in the floating-gate currents (ΔI_{fg}) by,

$$V_{off} = V_{off}{}' + \frac{\Delta I_{fg}}{g_{m1}} \tag{2}$$

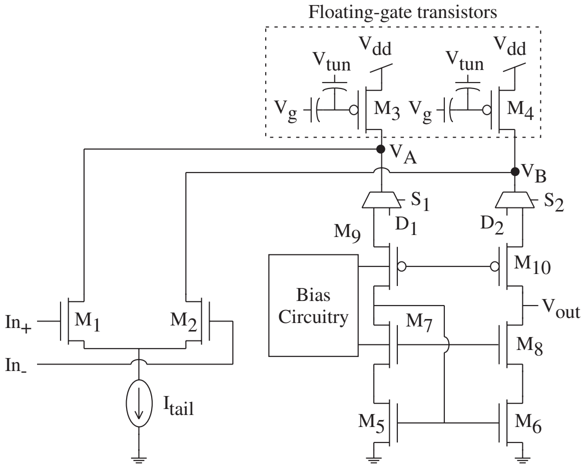


Fig. 4. **Operational Amplifier Circuit Schematic**: A single stage folded cascode amplifier that uses floating-gate transistors both as part of the amplifier and as trimming elements.

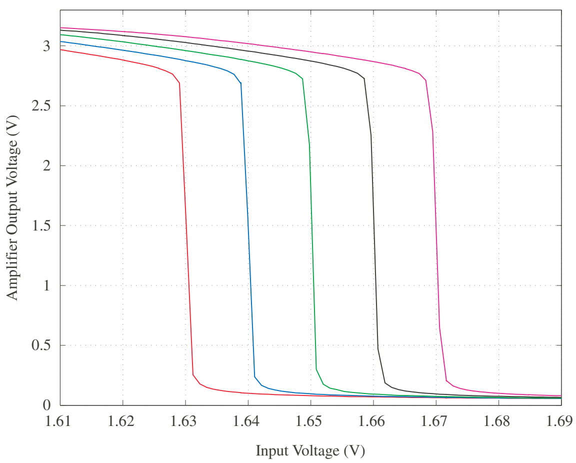


Fig. 5. **Open Loop DC Transfer Characteristics**: The input offset voltage of the amplifier was programmed to five different values in steps of $10mV$ with the non-inverting terminal of the amplifier set at $1.65V$.

where, $V_{off}{}'$ is to a first approximation the true uncompensated offset voltage of the amplifier and g_{m1} is the transconductance of the input differential pair.

A key advantage of this architecture is that the programming transistors are an integral part of the amplifier thereby simplifying the design process. Initially, all transistors including $M3$ and $M4$ are assumed to be non floating-gate transistors and are designed to meet the amplifier's specifications. Next, transistors $M3$ and $M4$ are made floating-gates with the input capacitance C_{in} chosen large enough such that the capacitive division from the external gate input V_g to the floating-node is close to one. Appropriate switches ($S1$ and $S2$) are added to isolate the floating-gates during programming and C_{tun} is added to provide a global erase capability using tunneling. With such a scheme, transistors already sized appropriately to meet the amplifier's specification are used for offset cancellation thereby requiring no additional design time and minimal additional area.

Offset cancellation involves one-time programming on account of the long-term charge retention of floating-gates. Therefore, on-chip complexity has been reduced through the use of an off-chip programming infrastructure. A custom board was designed with circuit blocks for current measurement, logic blocks for programming control, and interface circuitry to a computer in the loop for automatic control of programming. The scheme is described in detail in [6] along with programming speeds and accuracy. Such a programming infrastructure automates floating-gate programming and hence offset cancellation, thereby, making the proposed scheme practical from a commercial production standpoint.

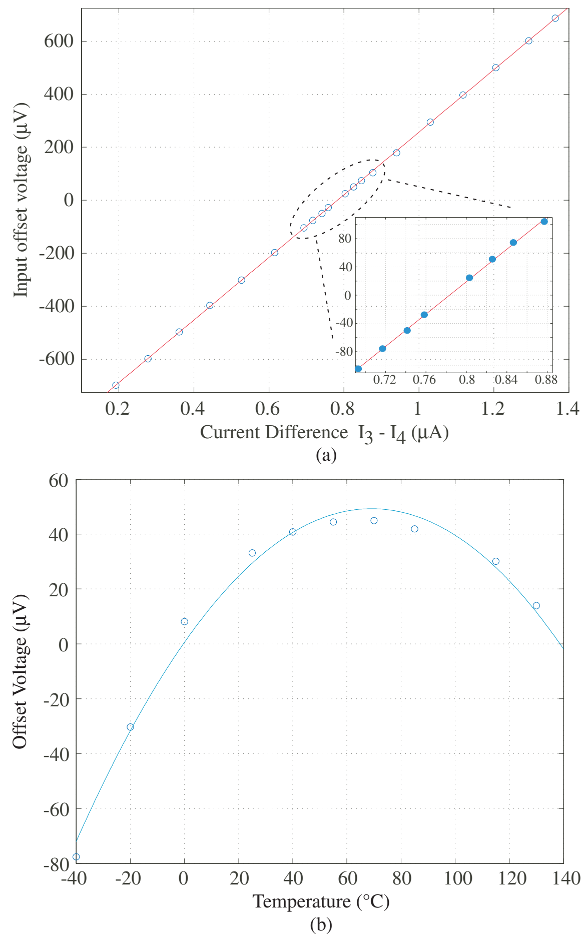


Fig. 6. **Input Offset Voltage**: (a) The input offset voltage vs. floating-gate difference current is linear with the lowest offset being $\pm 25\mu V$ as seen from the inset. (b)The input offset voltage is plotted against temperature and a maximum change of $130\mu V$ is observed over a temperature range of $-40\ °C$ to $130\ °C$.

IV. Experimental Results

The proposed amplifier was designed and fabricated in a $0.5\mu m$ digital CMOS process. Initial results on the programmability of the offset voltage is shown in Fig. 5 where the amplifier is configured as a comparator with its offset programmed in steps of $10mV$ ranging from $-20mV$ to $+20mV$. Accurate measurements of the offset voltage is made by using the amplifier under test along with a second amplifier configured as a nulling amplifier forming a servo loop [9].

TABLE I

OPERATIONAL AMPLIFIER SUMMARY OF PERFORMANCE

Parameter	Value
Supply Voltage	$3.3V$
Technology	$0.5\mu m$ CMOS
Input Common Mode Range	$1.2V - 3.1V$
Output Voltage Swing	$0.2V - 3.1V$
Input Offset Voltage	$\pm 25\mu V$
Offset Voltage Drift with Temperature	$130\mu V/170°C$
Offset Voltage Drift @ $25°C$ for 10 yrs	$0.5\mu V$
Open Loop Gain	$63dB$
Unity Gain Bandwidth @ $C_L = 20pF$	$10MHz$
Phase Margin	$60°$
Common Mode Rejection Ratio	$73dB$ (Simulation)
Power Supply Rejection Ratio	$77dB$ (Simulation)
Input Referred Noise (rms)	$8.9\mu V$ (Simulation)
Slew Rate	$5V/\mu s$
Settling Time (10 Bit) for $100mV$ Step	$105ns$
Power Dissipation (Incl. Buffer)	$8.25mW$
Area (Excl. Buffer)	$115\mu m\times 45\mu m$

Using such an approach, the offset voltage of the prototype amplifier has been reduced as low as $\pm 25\mu V$. Fig. 6(a) shows the measured input referred offset voltage of the amplifier plotted against the programmed floating-gate difference currents. The measured data shows a linear dependence of the offset voltage with the programmed difference currents as expected from (2).

Fig. 6(b) shows the sensitivity of the input offset voltage with temperature. The offset voltage was measured for temperatures ranging from $-40\ °C$ to $130\ °C$. A maximum change of $130\mu V$ was observed over the full temperature range of $170\ °C$. Based on data from the long-term charge retention in floating-gates, the offset voltage drift with time can be estimated and this has been found to be $0.5\mu V$ at $25°C$ over a period of 10 years. Table I summarizes the performance of the amplifier and the chip micrograph is shown in Fig. 7 that highlights the compactness of the proposed approach.

V. COMPARISONS AND CONCLUSIONS

Table II presents a comparison of the proposed scheme with other techniques commonly used to reduce offset voltage such as auto-zeroing, chopper stabilization [4] and a continous-time autozeroing using a ping-pong [10] scheme. Auto-zeroing is primarily useful for sampled data systems and is limited by issues such as charge injection, clock feedthrough and wide-band noise folding into the baseband on account of undersampling. For a continuous-time operation, chopper stabilization or continuous-time auto-zeroing such as a ping-pong amplifier [10] are the typical alternatives. The chopper amplifier is, however, limited in use to low-bandwidth applications [4]. The ping-pong approach involves the use of multiple amplifiers and multi-phase clocks that add additional overhead in terms of area and power. The proposed floating-gate approach involves none of the above tradeoffs and the offset cancellation by itself dissipates no additional power. The approach places minimal overhead on the amplifier design with non-volatile storage of offset reduction information.

A prototype amplifier has been fabricated in a $0.5\mu m$ standard digital CMOS process and trimmed to an offset voltage of $\pm 25\mu V$. The offset voltage exhibits a temperature sensitivity of $130\mu V$ over a temperature range of $170\ °C$. Accelerated life-time testing indicate an offset voltage drift of $0.5\mu V$ at $25\ °C$ for 10 years. In summary, programmability to a low input offset voltage coupled with a negligible long-term drift makes this approach attractive for offset reduction in operation amplifiers.

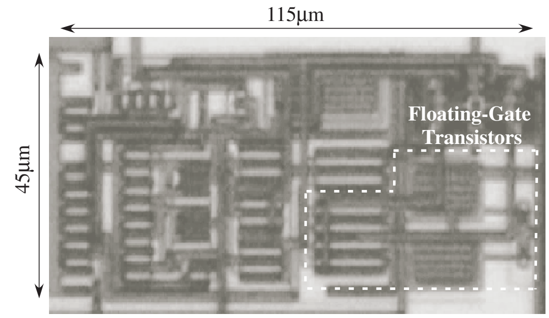


Fig. 7. **Amplifier die micrograph**: The chip micrograph of the prototype operational amplifier excluding the output buffer is shown to occupy an area of $115\mu m\times 45\mu m$.

TABLE II

COMPARISON OF OFFSET CANCELLATION SCHEMES

	FGate	Autozero	Chopper	Ping-Pong
Mode	Continuous	Sampled	Continuous	Continuous
Offset (V_{os})	Low	Moderate	Low	Moderate
Bandwidth	High	High	Low	High
Complexity	Low	Moderate	High	Moderate
$1/f$ Noise	No effect	Reduced	Reduced	Reduced
Extra Power	Low	Moderate	Moderate	Moderate
Extra Area	Low	Moderate	Moderate	Moderate
V_{os} Removal	Long-Term	Periodic	Continuous	Periodic

REFERENCES

[1] L. R. Carley, "Trimming analog circuits using floating-gate analog MOS memory," *IEEE Journal of Solid-State Circuits*, vol. 24, pp. 1569–1574, Dec. 1989.

[2] E. Sackinger and W. Guggenbuhl, "An analog trimming circuit based on a floating-gate device," *IEEE Journal of Solid-State Circuits*, vol. 23, pp. 1437–1440, Dec. 1988.

[3] F. Adil, G. Serrano, and P. Hasler, "Offset removal using floating gate circuits for mixed-signal systems," in *Southwest Symposium on Mixed-Signal Design*, Feb. 2003, pp. 190–195.

[4] C. C. Enz and G. C. Temes, "Circuit techniques for reducing the effects of op-amp imperfections: autozeroing, correlated double sampling and chopper stabilization," *Proceedings of the IEEE*, vol. 84, pp. 1584–1614, Nov. 1996.

[5] T. M. S. Kinoshita, M. Nagata, and A. Iwata, "A PWM analog memory programming circuit for floating-gate MOSFETs with $75\mu s$ programming time and 11-Bit updating resolution," *IEEE Journal of Solid-State Circuits*, vol. 36, pp. 1286–1290, May 2003.

[6] G. Serrano, P. Smith, H. Lo, R. Chawla, T. Hall, C. Twigg, and P. Hasler, "Automated Rapid Programming of Large Arrays of Floating-gate Elements," in *Proceedings of the International Symposium on Circuits and Systems*, vol. I, May 2004, pp. 373–376.

[7] C. Bleiker and H. Melchior, "A four-state EEPROM using floating-gate memory cell," *IEEE Journal of Solid-State Circuits*, vol. 22, pp. 460–463, June 1987.

[8] H. Nozama and S. Kokyama, "A thermionic electron emission model for charge retention in SAMOS structures," *Japanese Journal of Applied Physics*, vol. 21, pp. L111–L112, Feb. 1992.

[9] M. Burns and G. Roberts, Eds., *An Introduction to mixed-signal IC test and measurement*. Oxford University Press, 2001.

[10] I. E. Opris and G. A. Kovacs, "A rail-to-rail ping-pong op-amp," *IEEE Journal of Solid-State Circuits*, vol. 31, pp. 1320–1324, Sept. 1996.