HIGH PERFORMANCE ANALOG CIRCUIT DESIGN USING FLOATING-GATE TECHNIQUES

A Thesis
Presented to
The Academic Faculty

by

Guillermo J. Serrano

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Electrical and Computer Engineering

Georgia Institute of Technology
December 2007

HIGH PERFORMANCE ANALOG CIRCUIT DESIGN USING FLOATING-GATE TECHNIQUES

Approved by:

Dr. Kevin Kornegay,
Committee Chair
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Paul Hasler, Advisor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. David Anderson
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Jennifer Michaels
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Levent Degertekin
School of Mechanical Engineering
Georgia Institute of Technology

Date Approved: 26 July 2007

To my parents,

Muneca y Luis,

and to my brothers,

Luis y Pablo

ACKNOWLEDGEMENTS

I would like to thank Dr. Paul Hasler for his guidance and numerous stimulating conversations during the course of my doctoral studies. Thanks are due to the members of my committee for all the insightful comments that have helped shape this dissertation.

Many thanks to members of ICE lab for putting up with me during my six years at GaTech. Venkatesh Srinivasan and Abhishek Bandyopadyay require a special mention for being both a great friend and an excellent work partner. It has been a pleasure working with you!

Finally I want to thank my parents Muneca and Luis, and my brothers Luis and Pablo. This work would not have been possible without their love, support, and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Random device mismatch plays an important role in the design of high performance analog circuits. Parameter variations and mismatches of CMOS devices can easily limit and deteriorate the performance of analog circuits. This is mainly due to the lack of tunability of the analog circuits that necessitates the use of digital or laser trimming, special layout techniques, and calibration schemes among others. These problems can be alleviated by making use of the analog storage (programmability) and capacitive coupling capabilities of floating-gate transistors. The programmability property of floating-gate transistors is exploited in this work to compensate for mismatch and device parameter variations in various high performance analog circuits. A careful look is taken at the characteristics and behavior of floating-gate transistors; issues such as programming, precision, accuracy, and charge retention are addressed in this work.

In this work, an alternate approach to reduce the offset voltage of the amplifier is presented. The proposed approach uses floating-gate transistors as programmable current sources that provide offset compensation while being a part of the amplifier of interest during normal operation. This technique results in an offset voltage cancelation that is independent of other amplifier parameters and does not dissipate additional power. Two compact programmable architectures that implement a voltage reference based on the charge difference between two floating-gate transistors are introduced. The references exhibit a low temperature coefficient (TC) as all the transistors temperature dependencies are canceled. Programming the charge on the floating-gate transistors provides the flexibility of an arbitrary voltage reference with

a single design and allows for a high initial accuracy of the reference. Also, this work introduces a novel programmable temperature compensated current reference. The proposed circuit achieves a first order temperature compensation by canceling the negative TC of an on-chip poly resistor with the positive TC of a MOS transistor operating in the ohmic region. Flexibility and immunity to device parameters are enabled through the use of floating-gate transistors. Programmability of the ohmic resistor enables optimal temperature compensation while programmability of the reference voltage allows for an accurate current reference for a wide range of values. Finally, this work combines the already established DAC design techniques with floating-gate circuits to obtain a high precision converter. The use of floating-gate transistors allow to compensate for the inherent V_{th} mismatch of the MOS transistors. This approach enables higher accuracy along with a substantial decrease of the die size.

CHAPTER I

MISMATCH: AN ANALOG CIRCUIT PERFORMANCE LIMITATION

Integrated circuit designers generally rely on the concept of matched behavior between identically designed devices. Performance of digital and analog metal oxide semiconductor (MOS) circuits is affected by time-independent variations between identically designed transistors, better known as mismatch. In digital circuits, transistor mismatch leads to propagation delays. In analog circuits, the spread in the DC characteristics of supposedly matched transistors produces inaccurate or even anomalous circuit behavior. The impact of matching MOS transistors becomes more important as dimensions of the devices are reduced and the available signal swing decreases.

1.1 Device Mismatch

Mismatch results from either systematic or stochastic (random) effects. Manufacturing variations result in process and device parameter variations from lot-to-lot, wafer-to-wafer, die-to-die, and device-to-device. Lot-to-lot and wafer-to-wafer variations are common to all devices in the circuit; they introduce a systematic shift in the device characteristics and circuit performance. Processing and temperature gradients introduce systematic device variations which are independent of device size. Insensitivity to these systematic variations can be achieved with the use of differential circuit topologies, proper biasing techniques, and layout techniques such as symmetry and common-centroid layouts. Device-to-device variations result in random differences

between the device parameters such as doping concentration, mobility, oxide thickness, and poly-silicon granularity. These variations can not be predicted during the design phase and are dependent on the device size.

1.1.1 MOS Matching Models

Device mismatch between two geometrically identical MOS transistors has been studied extensively by several researchers, notably [46, 47, 63]. It has been observed experimentally that the treshold voltage (V_{th}) difference and the transconductace parameter (K) difference, ΔV_{th} and ΔK respectively, are the dominant sources of device mismatch. It has been shown in [63] that ΔV_{th} and ΔK variations can be modeled using a normal distribution with zero mean and variance given by

$$\sigma^2\left(\Delta V_{th}\right) = \frac{{A_{V_{th}}}^2}{W \cdot L} \tag{1}$$

$$\sigma^2\left(\frac{\Delta K}{K}\right) = \frac{{A_K}^2}{W \cdot L} \tag{2}$$

where $A_{V_{th}}$ and A_K are technology-dependent parameters, W is the gate-width, L the gate-length, and $W \cdot L$ is the device area. It can be observed from (1) and (2) that device mismatch can be countered by increasing the area of the device, resulting in a minimum area requirement for a given accuracy specification. Such an approach increases the parasitic capacitance of the device, thereby increasing the power dissipation required in order to achieve a given bandwidth.

Device mismatch in a circuit results in differences in the drain-source currents (ΔI_{ds}) between devices with identical bias voltages or in differences in the gate-source voltages (ΔV_{gs}) for transistors biased with the same current. Using a drain-source current model valid in the strong inversion region, an expression for the ΔI_{ds} variance can derived as

$$\sigma^2\left(\frac{\Delta I_{ds}}{I_{ds}}\right) = \sigma^2\left(\frac{\Delta K}{K}\right) + \left(\frac{2}{V_{gs} - V_{th}}\right)^2 \sigma^2\left(\Delta V_{th}\right). \tag{3}$$

It can be observed from (3) that the current mismatch between two identical transistors operating in strong inversion is dependent on the bias point. For typical bias points ($V_{gs} - V_{th} < 0.65V$) the relative effect of the V_{th} mismatch dominates over the K mismatch. Better matching is achieved with higher $V_{gs} - V_{th}$ values resulting in lower voltage headroom.

A similar expression, derived for a drain-source current model valid in the weak inversion region, is given by

$$\sigma^2\left(\frac{\Delta I_{ds}}{I_{ds}}\right) = \sigma^2\left(\frac{\Delta K}{K}\right) + \left(\frac{\kappa}{U_t}\right)^2 \sigma^2\left(\Delta V_{th}\right) \tag{4}$$

where $U_t = kT/q$ is the thermal voltage and κ is given by $\frac{C_{ox}}{C_{ox}+C_{dep}}$. The current mismatch between two identical transistors operating in weak inversion is bias independent as seen in (4). Maximum current mismatch between two identical transistors will occur when operating in weak inversion.

1.1.2 Bipolar Matching Models

In a complementary metal oxide semiconductor (CMOS) process parasitic bipolar transistors are typically used in voltage reference circuits. The physical causes of bipolar mismatch have not been extensively investigated. Several studies have focused on the matching of the base current (I_b) and the collector current (I_c) for a pair of identical, closely spaced transistors. Matching of two identically bipolar transistors improves with the emitter area (A_e) and can be modeled in as

$$\sigma^2\left(\frac{\Delta I_b}{I_b}\right) = \frac{{A_{I_b}}^2}{A_e} \tag{5}$$

$$\sigma^2\left(\frac{\Delta I_c}{I_c}\right) = \frac{{A_{I_c}}^2}{A_e} \tag{6}$$

where A_{I_b} and A_{I_c} are technology constants, $\frac{\Delta I_b}{I_b}$ is the relative base current mismatch, and $\frac{\Delta I_c}{I_c}$ is the relative collector current mismatch.

Variations in I_b and I_c typically manifest themselves as variations in the base-emitter voltage (V_{be}) . It is shown in [46] that the ΔV_{be} variance can be modeled

as

$$\sigma^2\left(\Delta V_{be}\right) \approx \left(U_t\right)^2 \sigma^2\left(\frac{\Delta I_C}{I_C}\right). \tag{7}$$

The collector current mismatch is the dominant error source for base-emitter voltage matching as seen in (7). Unlike MOS transistors in the strong inversion region, mismatch behavior of a bipolar transistor is independent of the bias point.

1.2 Getting Around Mismatch

Performance of analog circuits has been limited primarily due to mismatch and variations in parameter values of integrated circuit components such as transistors, resistors, and capacitors. To ensure a specific circuit performance, typically, error correction is enabled with the use of some sort of memory along with a digital to analog converter to compensate for parameter mismatch. Mismatch between design components can be corrected using a post-fabrication trim procedure such as laser annealing, laser trims, poly fuses and zener zapping. All these trimming techniques are discrete in nature, involve an area penalty, and are one-time programable. Multiple programmability is possible with the use of EEPROM memories.

The concept of a floating-gate memory was proposed in 1967 by Kahng and Sze [43]. The first commercially available floating-gate based memory, the floating-gate avalanche-injection MOS device (FAMOS) was developed in 1970 [1]. Since then floating-gate transistors have been used extensively as non-volatile digital memory elements in EPROMs, EEPROMs and Flash memories [49]. Around 1988 analog designers started to use the floating-gate memory as an analog memory [75]; an amplifier with an offset voltage of $< 5mV$ was reported in [21]. It wasn't until mid $90's$ when the number of publications related to this technology started to increase [16, 50, 82]. Apart from being memory elements, floating-gate transistors can be used as computational elements as well [7, 39].

This work uses floating-gate transistors in the design of high performance analog

circuits. Here, floating-gate transistors are used as analog memories to compensate for device mismatch and parameter variations. The proposed approach allows for a high resolution memory without the area penalty and long term retention. A careful look is taken at the characteristics and behavior of floating-gate transistors, as they form an integral part of the proposed work.

1.3 Effect of Device Mismatches on Circuit Performance

Mismatches manifest themselves as offsets voltages in amplifiers; they limit the signal resolution in circuits such as comparators and analog-to-digital converters. Degradation of the temperature behavior and low precision of voltage and current references often occur as result of variations in device parameter values. Also, matching between transistors directly impacts the achievable accuracy in digital-to-analog converters.

1.3.1 Mismatch in Amplifiers

The effect of device mismatch in an operational amplifier is measured as the input referred offset voltage (V_{off}). This offset voltage, typically dominated by V_{th} mismatch [46], is a combination of the ΔV_{th} of all transistor pairs in the circuit and is given by

$$V_{off} = \Delta V_{th_{diffpair}} + \frac{\Delta V_{th_1}}{A_1} + \frac{\Delta V_{th_2}}{A_2} + ... \tag{8}$$

where $\Delta V_{th_{diffpair}}$ is the V_{th} mismatch of the differential pair, ΔV_{th_i} is the V_{th} mismatch of the i_{th} transistor pair, and A_i is the gain from the input differential pair to the i_{th} transistor pair. For high gain stages, $A_i >> 1$, V_{off} reduces to

$$V_{off} \approx \Delta V_{th_{diffpair}}. \tag{9}$$

As seen in (9), the input referred offset voltage of an amplifier is mainly determined by the differential pair threshold voltage mismatch.

The amplifier offset voltage can be reduce using resistor trimming. Resistor trimming is usually performed using laser annealing, laser trims, poly fuses, and zener

zapping. Both laser annealing and laser trims are expensive and do not provide the flexibility of in-package trims. Trimming using poly fuses and zener zapping is discrete in nature. Thus, accuracy is limited to the smallest resistor step used. Also, using a number of zener diodes and poly fuses involve an area penalty. All of the above resistor trimming techniques are one-time programmable.

Other techniques commonly used to reduce the offset voltage include auto-zeroing, correlated double sampling, and chopper stabilization. Auto-zeroing is primarily useful for sampled data systems and is limited by issues such as charge injection, clock feed-through, 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 are the typical alternatives. However, the chopper amplifier is limited in use to low-bandwidth applications while the ping-pong approach involves the use of multiple amplifiers and multi-phase clocks that add additional overhead in terms of area and power.

In this work, an alternate approach to reduce the offset voltage of the amplifier is presented. The proposed approach uses floating-gate transistors as programmable current sources 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-gate transistors as separate trimming elements as in [21, 64] or current-mode DACs as trimming elements [45]. Unlike [3, 31], the proposed scheme is independent of other amplifier parameters and the offset cancelation by itself dissipates no additional power.

1.3.2 Mismatch in Voltage References

In CMOS technology the bandgap voltage (V_{bg}) implemented using parasitic bipolar junction transistors is the popular choice for implementing a voltage reference (V_{ref}). The bandgap reference results from the addition of a V_{be} and a scaled version of ΔV_{be}

and is given by

$$V_{ref} = V_{bg} = V_{be} + K \cdot \Delta V_{be} = V_{be} + U_t \cdot \left[\frac{R_2}{R_1} ln \left(\frac{A_{e_2}}{A_{e_1}} \right) \right] \tag{10}$$

where $\frac{A_{e_2}}{A_{e_1}}$ is the emitter area ratio of two bipolar transistors. The resulting reference voltage is restricted to 1.206, the energy bandgap of silicon. First order temperature compensation is achieved by properly choosing $\frac{R_2}{R_1}$ and $\frac{A_{e_2}}{A_{e_1}}$, as V_{be} and U_t exhibit opposite temperature behavior. This compensation results in temperature coefficients in the range of $25 - 50 ppm/^oC$ [39]. Higher order temperature effects can be reduced with more complicated schemes such as curvature correction [38].

Temperature dependence and initial accuracy of bandgap references is often degraded by device mismatch. The major source of error is caused by V_{be} mismatch, as the bandgap voltage depends on its absolute value as seen in (10). The ratios $\frac{R_2}{R_1}$ and $\frac{A_{e_2}}{A_{e_1}}$ introduce minor errors due to relative mismatches. Some kind of trimming circuit is usually required to compensate for these mismatches. Typically optimal first order temperature compensation is achieved by trimming R_2 with the use of poly-fuses, laser trimming, or digital memories at the cost of die area. Additional trimming is needed when a highly accurate reference value other than 1.206 is desired.

In this work, two compact programmable architectures that implement a voltage reference based on the charge difference between two floating-gate transistors are presented. The references exhibit a low temperature coefficient as all the transistors temperature dependencies are canceled. Programming the charge on the floating-gate transistors provides the flexibility of an arbitrary voltage reference value with a single design and also allows for a high initial accuracy of the reference.

1.3.3 Mismatch in Current References

Temperature compensation of a current reference is a difficult task as typical approaches rely on specific device parameters for proper performance. Optimal compensation and high accuracy of the reference is difficult to obtain since parameter

values can not be predicted accurately due to random variation across process, dies, and runs. Also, the current reference value is typically dictated by the compensation method; temperature compensation is only obtained for a single, non-arbitrary current value. Trimming circuits are often used to ensure proper temperature behavior of the current reference and improve its accuracy.

A common approach when designing a current reference is to apply a voltage reference (V_{ref}) to a well know resistor (R) [24, 55], resulting in

$$I_{ref} = \frac{V_{ref}}{R} = \frac{V_{bg}}{R} = \frac{V_{be} + K \cdot \Delta V_{be}}{R}. \tag{11}$$

Typically the bandgap voltage is selected to implement V_{ref} while R can be any resistive device available such as poly or MOS resistors. Temperature compensation is obtained by purposely designing a bandgap voltage with linear temperature dependence to cancel the temperature dependance of the selected resistor. The temperature dependance and accuracy of the reference will be affected by device mismatch, mainly due to ΔV_{be} and ΔR. A trimming circuit is usually required to obtain optimal temperature compensation. Also, arbitrary current reference values are not possible without additional trimming circuits.

This work introduces a novel programable temperature compensated current reference. The proposed circuit achieves a first order temperature compensation by canceling the negative temperature coefficient (TC) of an on-chip poly resistor with the positive TC of a MOS transistor operating in the ohmic region. Flexibility and immunity to device parameters are enabled trough the use of floating-gate transistors. Programmability of the ohmic resistor enables optimal temperature compensation while programmability of the reference voltage allows for an accurate current reference for a wide range of values.

1.3.4 Mismatch in Current-Steering Digital-to-Analog Converters

Current-steering DACs are based on an array of matched current sources which are unity or binary weighted. The intrinsic accuracy of a current-steering DAC is dictated by device mismatch. Architectures variants, such as the two-stage, the interpolated, and the segmented architectures, are often used. The difficulty to meet a certain intrinsic accuracy specification due to the random mismatches between the current sources, however is the same for all architectures. Large devices, randomized layouts, and laser trimmings, among others, are techniques often used to reduce mismatch and obtain higher precision. These techniques improve linearity, but at the expense of die area, power dissipation, and/or dynamic performance.

In this work, a V_{th} compensated digital-to-analog converter (DAC) is presented. The use of floating-gate transistors allow to compensate for the inherent V_{th} mismatch of the MOS transistors. This approach enables higher accuracy along with a substantial decrease of the die size.

CHAPTER II

A PROGRAMMABLE ANALOG MEMORY

2.1 Floating-Gate Transistor: An Analog Memory

The successful use of floating-gate transistors in analog circuits depends on understanding certain key aspects of these devices. A careful look is taken at the characteristics and behavior of floating-gate transistors; issues such as programming, precision, accuracy, and charge retention are addressed in this chapter.

2.1.1 Circuit Description

Figure 1 shows the circuit diagram of a pMOS floating-gate transistor along with its layout. A floating-gate transistor consists of a pMOS transistor with its gate terminal connected to a capacitor. The basic structure of a floating gate memory relies on an insulated poly-silicon layer, the floating-gate, interposed between the substrate and the control gate C_{in}. Electrons and holes in the insulated floating gate cannot escape from it, hence granting a permanent storage information. Charge on the floating node can be modify using channel hot-electron injection or Fowler-Nordheim tunneling.

External inputs to the floating gate are capacitively coupled through C_{in} and C_{tun} as seen in Figure 1. Capacitor C_{in} is typically used as the input to the transistor (analog to the gate terminal of a standard MOS transistor) while C_{tun} is only used for tunneling. The tunneling capacitor is implemented using the gate oxide between the gate poly-silicon and an n-well (commonly known as a MOS capacitor). Capacitor C_{in} can be implemented as a poly-poly capacitor as shown in Figure 1, or as a MOS capacitor. The current of the floating-gate transistor will be determined by both the capacitively coupled input voltage and the charge stored on the floating-gate.

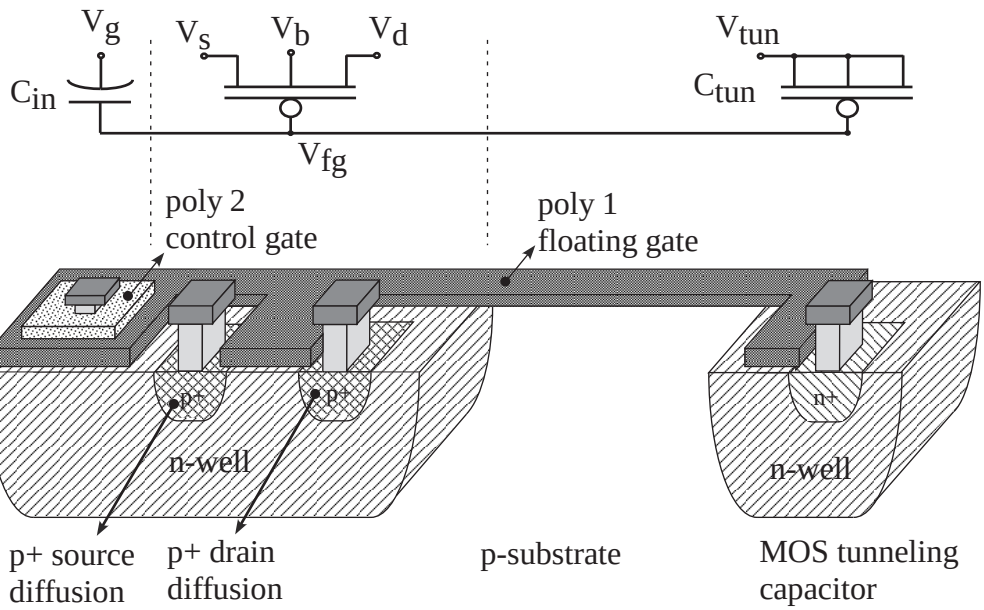


Figure 1. Floating-Gate Transistor: Circuit schematic and layout of a floating-gate pMOS transistor.

2.1.2 Analytical Description

Consider the floating-gate transistor presented in Figure 1 operating in the strong inversion region. Neglecting Early effects, assuming saturation and $V_{sb} = 0\text{V}$, the drain current I_d through the device is given by

$$I_d = \frac{K}{2}\left(V_s - \kappa V_{fg} - \kappa V_{th}\right)^2 \tag{12}$$

where K is the transconductance parameter, V_{fg} is the floating gate voltage, V_s is the source voltage, V_{th} is the threshold voltage of the device, and κ is defined as

$$\kappa = \frac{C_{ox}}{C_{ox} + C_{dep}}. \tag{13}$$

The floating gate voltage V_{fg} can be expressed as the combination of the charge Q stored on the floating node and the capacitively coupled input V_g. Assuming $V_{tun} = 0\text{V}$, V_{fg} is given by

$$V_{fg} = Vg\frac{C_{in}}{C_T} + \frac{Q}{C_T} \tag{14}$$

where $C_T = C_{in} + C_{tun} + C_{parasitic}$ is the total capacitance of the floating-node. An I_d expression in terms of V_g and Q can be obtained by substituting (14) in (12), resulting in

$$I_d = \frac{K}{2}\left(V_s - \kappa_{eff} V_{fg} - \kappa V'_{th}\right)^2 \tag{15}$$

where κ_{eff} is the effective κ of the floating-gate transistor given by

$$\kappa_{eff} = \kappa \frac{C_{in}}{C_T} = \frac{C_{ox}}{C_{ox} + C_{dep}} \frac{C_{in}}{C_T}. \tag{16}$$

and V'_{th} is the modified threshold voltage given by

$$V'_{th} = V_{th} + \frac{Q}{C_T}. \tag{17}$$

Now, consider the floating-gate transistor presented in Figure 1 operating in the weak inversion (sub-threshold) region. Again, neglecting Early effects, assuming saturation and $V_{sb} = 0\text{V}$, the drain current I_d through the device is given by

$$I_d = I'_o e^{\frac{V_s - \kappa V_{fg} - \kappa V_{th}}{U_t}} \tag{18}$$

where U_t is the thermal voltage and I'_o is a pre-exponential constant made up of fundamental device parameters [79]. Substituting (14) in (18) results in

$$I_d = I'_o e^{\frac{V_s - \kappa_{eff} V_g - \kappa V'_{th}}{U_t}}. \tag{19}$$

where all variables have been previously defined.

As seen in (15) and (19), the only difference between a floating-gate transistor and a standard MOS device relies in κ and V_{th}. The change in κ, due to the capacitive coupling as seen in (16), can be viewed as a change in the gain of the transistor. For $C_{in} >> C_{tun} + C_{parasitic}$, κ_{eff} can be approximated as κ. The change in V_{th}, due to the charge stored on the floating node as shown in (17), can be viewed as a threshold voltage shift. Thus, by removing or adding electrons from the floating node, the effective threshold voltage of a pMOS floating-gate transistor can be increased

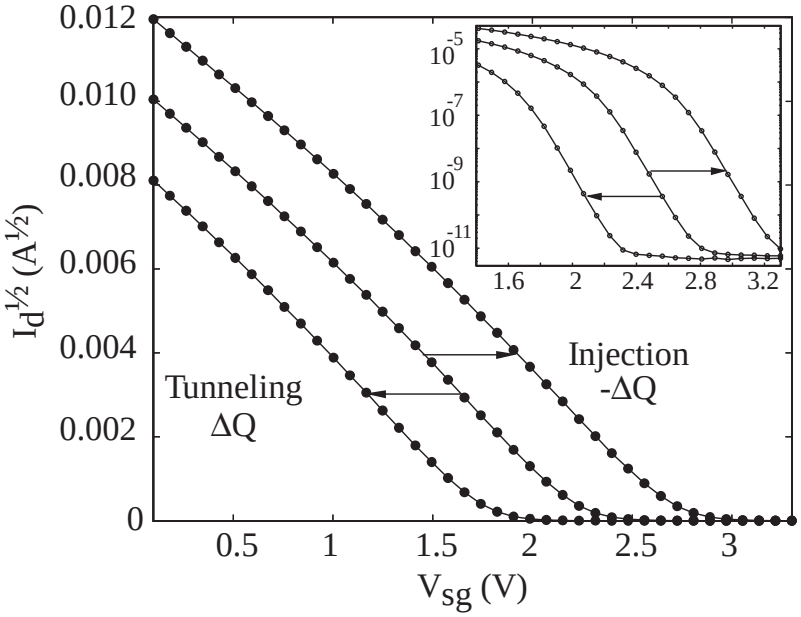


Figure 2. Threshold Voltage Modification: Drain current of a floating-gate transistor as a function of the gate voltage for different ΔQ values in the floating node. The inset figure emphasizes the sub-threshold region.

or decreased respectively, as demonstrated graphically in Figure 2. Here, the drain current of a transistor is plotted against the gate voltage for both above threshold and below threshold operation (inset).

2.2 Charge Modification Mechanisms

Programming of the floating-gate voltage is achieved by modifying its charge. In this work Fowler-Nordheim tunneling and hot-electron injection are used to add and remove electrons form the floating-gate, respectively. Tunneling is mainly used as a global erase while injection is used for accurate programming.

2.2.1 Fowler-Nordheim Tunneling

Fowler-Nordheim tunneling [34, 52] can be defined as the process of tunneling an electron across the $S_i - S_iO_2$ barrier. This phenomenon occurs when the electron has enough energy to surmount the $S_i - S_iO_2$ barrier. Tunneling is enabled by applying a high enough electric field across the tunneling junction, thus decreasing the energy

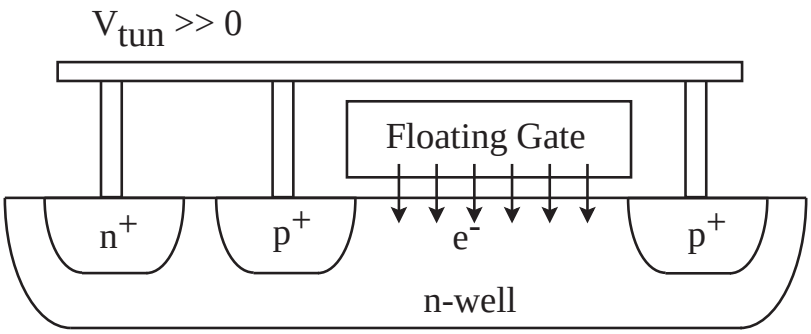


Figure 3. Electron Tunneling: Cross section of the MOS capacitor along with the biases required to enable tunneling.

barrier.

The tunneling junction is nothing but a capacitive connection to the floating-gate (see Figure 1). The floating-gate forms one terminal of the capacitor while the other terminal is called the tunneling voltage, V_{tun}. The tunneling capacitor is fabricated as a MOS capacitor formed on an n-well as shown in Figure 3. The tunneling voltage makes an ohmic contact to the n-well using an n^+ diffusion layer. It should be noted that the poly-silicon floating-gate is directly connected to the poly-silicon gate of the tunneling MOS capacitor.

In this work, tunneling is used to remove electrons from the floating node. By applying a high tunneling voltage (> 15V for a $0.5\mu m$ CMOS process), high electric fields are generated at the $S_i - S_iO_2$ interface thus enabling electrons tunnel across the oxide. The amount of charge transferred due to tunneling depends on the tunneling voltage and the amount of time the high electric field is sustained across the $S_i - S_iO_2$ barrier.

2.2.2 Hot-Electron Injection

Hot-electron injection [29,76] occurs in pMOS transistors when carriers are accelerated to a high enough energy level to surmount the $S_i - S_iO_2$ barrier. At high electric fields and in the presence of drain currents, electrons are created at the drain edge of

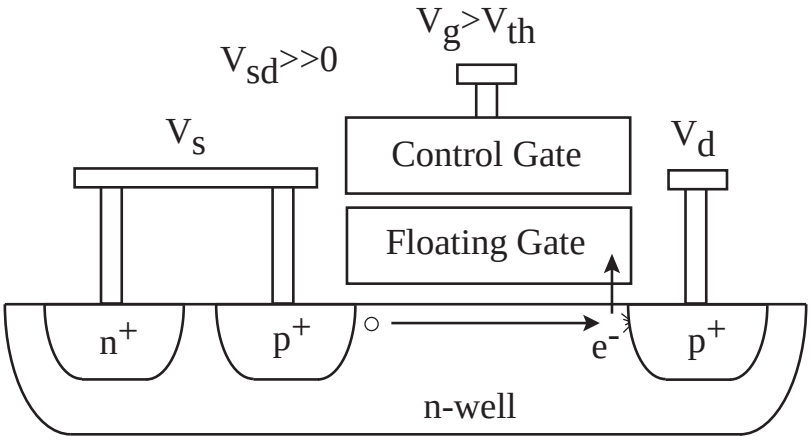


Figure 4. Electron Injection: Cross section of the floating-gate transistor along with the biases required to enable injection.

the drain-to-channel depletion region via hot-hole impact ionization. These electrons travel back into the channel region, gain sufficient kinetic energy such that they cross $S_i - S_iO_2$ barrier, and are injected onto the floating-gate. Electrons that do not cross the $S_i - S_iO_2$ barrier are swept away towards the bulk and flow as bulk currents. A pictorial diagram of the hot-electron injection phenomena is presented in Figure 4.

In this work, hot-electron injection is used to precisely add electrons to the floating gate. Injection is enabled with a high source-to-drain electric field when there is current flowing through the channel. This is physically achieved by applying a $V_{sd} >$ 5V for a certain amount of time in a $0.5\mu m$ process. The amount of charge transferred will be a function of the initial current I_{init}, the drain-source voltage V_{ds}, and the pulse time t_{pulse} [29].

2.3 Programming Precision

The accuracy to which one can program floating-gate transistors to a target current depends on the smallest drain current change that can be programmed onto a floating-gate device. In order to estimate the design choices available to improve programming

precision, a figure of merit (FOM) is defined as

$$FOM = -\log_2\left(\frac{\Delta I}{I}\right) \tag{20}$$

where ΔI is the minimum programmable change in drain current that is necessary to meet a system level accuracy specification and I is the bias current of the floating-gate transistor. It should be noted that such a definition results in the FOM being represented in the familiar binary system, as number of bits of accuracy achievable. In the discussion below, the FOM is related to floating-gate circuit parameters for operation in both the weak and strong inversion regimes such that the floating-gate transistor can be designed to achieve the required bits of precision.

2.3.1 Weak Inversion

Consider a floating-gate pMOS transistor operating in the weak inversion regime. Ignoring the device Early and body effects, the source-drain current of the device is given by

$$I = I_o e^{-\frac{\kappa}{U_t}\left(V_{fg}+V_{th}\right)} e^{\frac{V_s}{U_t}}. \tag{21}$$

A ΔV_{fg} change in the floating-gate voltage can be expressed as a ΔI change in drain current given by

$$I + \Delta I = I_o e^{-\frac{\kappa}{U_t}\left(V_{fg}+V_{th}+\Delta V_{fg}\right)} e^{\frac{V_s}{U_t}}. \tag{22}$$

Dividing (22) by (21), noting that $\Delta V_{fg} = \frac{\Delta Q}{C_T}$, and solving for $\frac{\Delta I}{I}$ the achievable change in drain current relative to the initial drain current is given by

$$\frac{\Delta I}{I} = e^{-\frac{\kappa}{U_t}\frac{\Delta Q}{C_T}} - 1 \approx -\frac{\kappa}{U_t}\frac{\Delta Q}{C_T} \tag{23}$$

where C_T is the total capacitance at the floating-gate and ΔQ is the programmed charge. It is clear from (23) that the achievable precision is inversely proportional to the charge that can be reliably transferred onto the floating-gate and directly proportional to the total floating-gate capacitance.

2.3.2 Strong Inversion

Consider a floating-gate pMOS transistor operating in the strong inversion region. Ignoring Early effects, the drain current of the device in saturation is given by

$$I = \frac{K}{2}\left[V_s - \kappa\left(V_{fg} + V_{th}\right)\right]^2. \tag{24}$$

Programming the device, such that a charge transfer of ΔQ results in a change in the floating-gate voltage of ΔV_{fg}, modifies the drain current to be

$$I + \Delta I = \frac{K}{2}\left[V_s - \kappa\left(V_{fg} + V_{th} + \Delta V_{fg}\right)\right]^2. \tag{25}$$

Dividing (25) by (24) and manipulating the algebra with the assumption that ΔV_{fg} is much smaller than the overdrive voltage, $V_{od} = V_s - \kappa\left(V_{fg} + V_{th}\right)$, results in

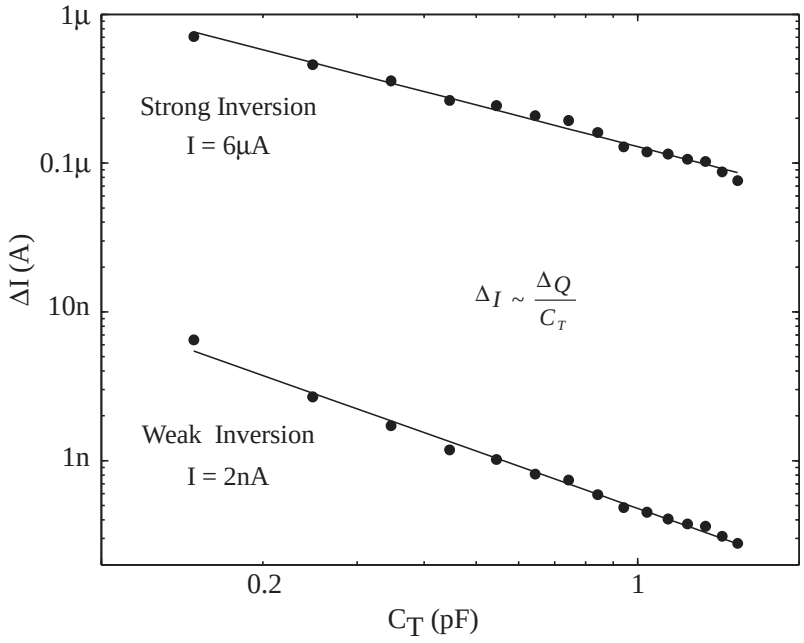
$$\frac{\Delta I}{I} = -\frac{2\kappa\Delta V_{fg}}{V_s - \kappa\left(V_{fg} - V_{th}\right)} = -\frac{2\kappa}{V_{od}}\frac{\Delta Q}{C_T} \tag{26}$$

As can be observed from (26), the achievable precision is inversely proportional to the charge that can be transferred onto the floating-gate and directly proportional to the overdrive voltage of the device and the total floating-gate capacitance. Comparing (26) with (23), it should be noted that with identical C_T and ΔQ, the achievable precision is higher in the strong inversion region than in the weak inversion region.

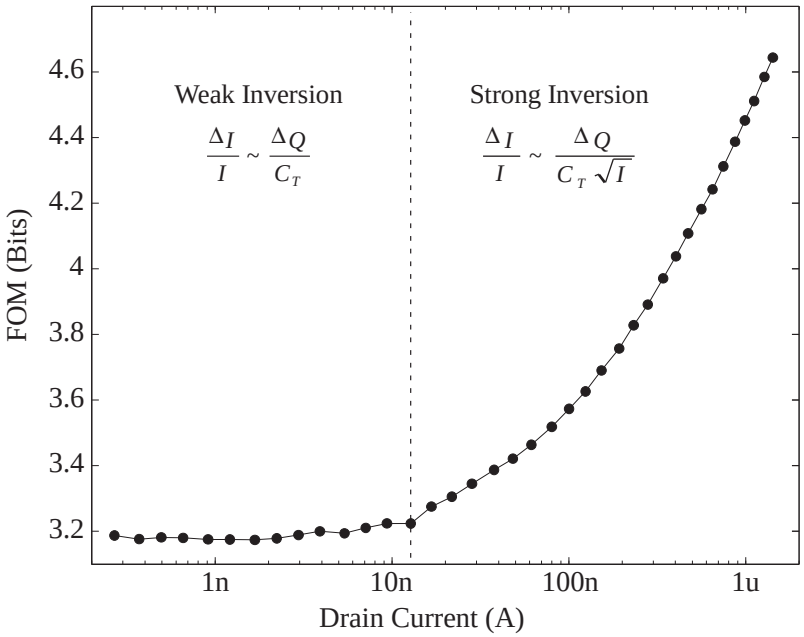
2.3.3 Experimental Results

In order to verify the theory presented above, a test chip was fabricated in a $0.5\mu m$ CMOS process. The prototype chip consisted of an array of floating-gate pMOS transistors with the same aspect ratio but with varying input capacitors such that C_T varied from one transistor to another.

Figure 5(a) shows the change in current of floating-gate transistors with different CT values and identical programmed ΔQ. As expected from (23) and (26), the change in current is inversely proportional to C_T, independent of the region of operation. Figure 5(b) shows the FOM plotted against the initial drain current with identical



(a)



(b)

Figure 5. Programming Precision: (a) Plot of ΔI as a function of the total capacitance C_T on the floating-node for a fixed ΔQ. (b) Plot of the FOM as a function of the initial current for a fixed ΔQ.

Table 1. Achievable Bits of Accuracy (FOM)

	Weak Inversion			Strong Inversion		
$\Downarrow C_T\ \Delta Q \Rightarrow$	$1e^-$	$10e^-$	$100e^-$	$1e^-$	$10e^-$	$100e^-$
10fF	11.18	7.85	4.83	13.44	10.12	6.8
100fF	14.5	11.18	7.85	16.76	13.44	10.12
1pF	17.82	14.5	11.18	20.09	16.76	13.44

programmed ΔQ. As expected from (23), the FOM is independent of the initial current in the weak inversion region. In the strong inversion region the FOM exhibits a dependance on the initial current. This can be inferred from (26) by recalling that $V_{od} = \sqrt{\frac{2\kappa I}{K}}$.

Table 1 presents quantitative numbers for the FOM for both the weak inversion and strong inversion regions based on the theory developed above. The FOM has been calculated for different values of ΔQ and C_T using a κ of 0.7, a U_t of $26m$V, and a V_{od} of $250m$V (for strong inversion). Using the above developed theory and depending on the region of operation of the floating-gate transistor, one can design a floating-gate transistor such that a target accuracy specification is met.

2.4 Automated Rapid Programming of Floating-Gates

With the recent increase in the number of floating-gate devices used as analog elements in a single IC [10, 23, 80], accurate and fast programming has become essential for the viability of this approach. A first-principle model for hot-electron injection has been presented in [30]. Adaptation of this model for programming purposes is not trivial because it is computationally complex and intensive. This work presents novel method that can be used to program large floating-gate arrays at fast rates with 0.2% of accuracy over a wide range of currents (3.5 decades). The proposed programming algorithm minimizes programming time by choosing an optimal injection rate for a given target current. A characterization of the injection rate of the transistors is done

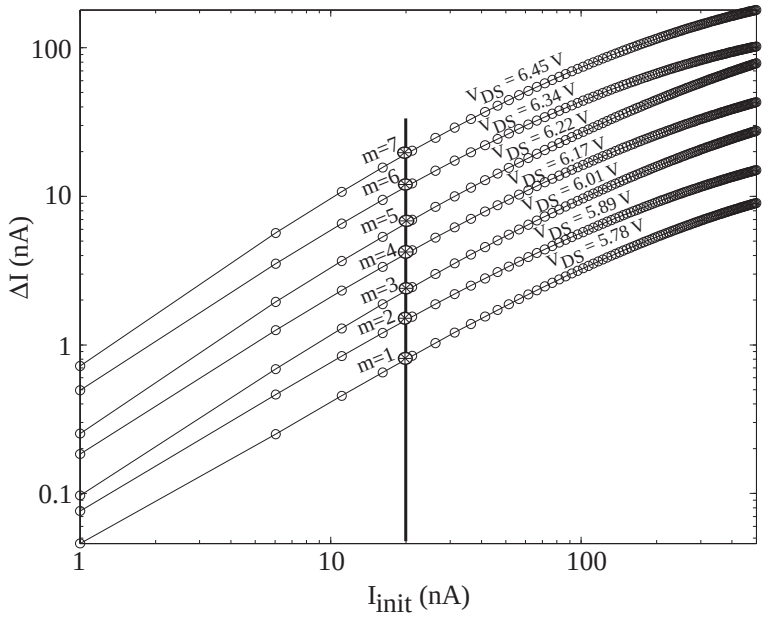


Figure 6. Injection Rate Characterization: A plot of ΔI as a function of I_{init} for several V_{ds} values.

once as part of a calibration procedure. The mathematical model developed condenses this characterization data into a few parameters, thus simplifying its complexity.

2.4.1 Array Characterization

As discussed in Section 2.2.2, the injection rate of a floating-gate transistor is a function of the initial current (I_{init}), the drain-source voltage (V_{ds}), and the pulse width (t_{pulse}). Characterization of the array is performed by measuring the injection rates of the elements in the array for different I_{init} and V_{ds} values. This is shown graphically in Figure 6, where ΔI is plotted against I_{init} for several V_{ds} values. A first order fit would suffice for sub-threshold currents but for above threshold currents a second order gives a better estimate. Figure 7 shows ΔI against V_{ds} for an specific I_{init}. This plot was obtained from the family of curves in Figure 6 as demonstrated by the vertical line. A linear regressed fit can be used to model this variation.

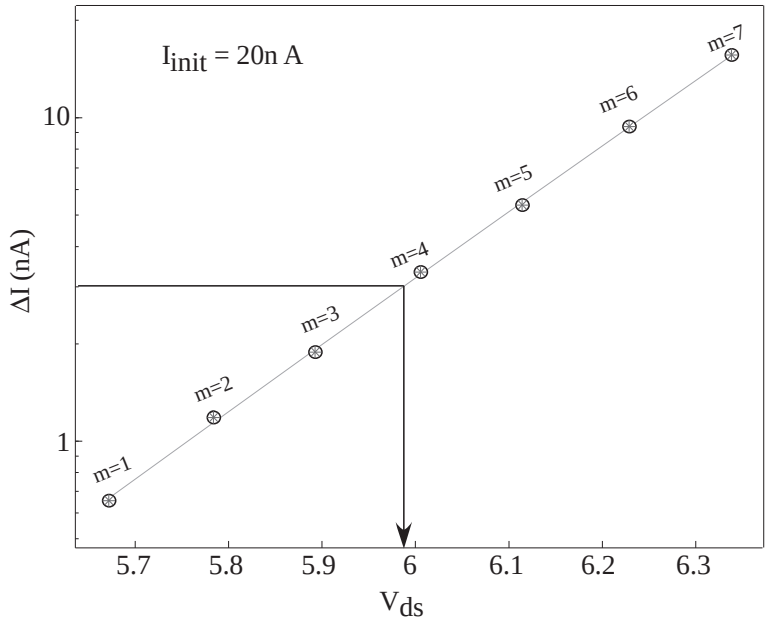


Figure 7. Injection Rate Characterization: A plot of ΔI as a function of V_{ds} for a given I_{init}.

2.4.2 Mathematical Model

The variation of ΔI with $I_{initial}$ (see Figure 6) can be modeled by a second order equation as

$$\log\left(\frac{\Delta I}{I_{S_0}}\right) = k_2 \cdot \left[\log\left(\frac{I_{initial}}{I_{S_0}}\right)\right]^2 + k_1 \cdot \log\left(\frac{I_{initial}}{I_{S_0}}\right) + k_0 \tag{27}$$

where I_{S_0} is a bias current with units of amperes and k_0, k_1 and k_2 are unitless functions of V_{ds}. Figure 8 shows the variation of the k parameters of (27) with V_{ds}. It can be seen that k_1 and k_2 are constants while k_0 is a linear function of V_{ds}. Therefore the change in relative current is linearly dependent on V_{ds}, which corroborates the results shown Figure 7.

The variation of ΔI with V_{ds} (see Figure 7) can be modeled by a linear equation as

$$\log\left(\frac{\Delta I}{I_{S_0}}\right) = m \cdot V_{ds} + f \tag{28}$$

where m, with units of 1/Volts, and f, unitless, are functions of $I_{initial}$. Figure 9(a) shows the variation of m with $I_{initial}$ while Figure 9(b) shows the variation of f with

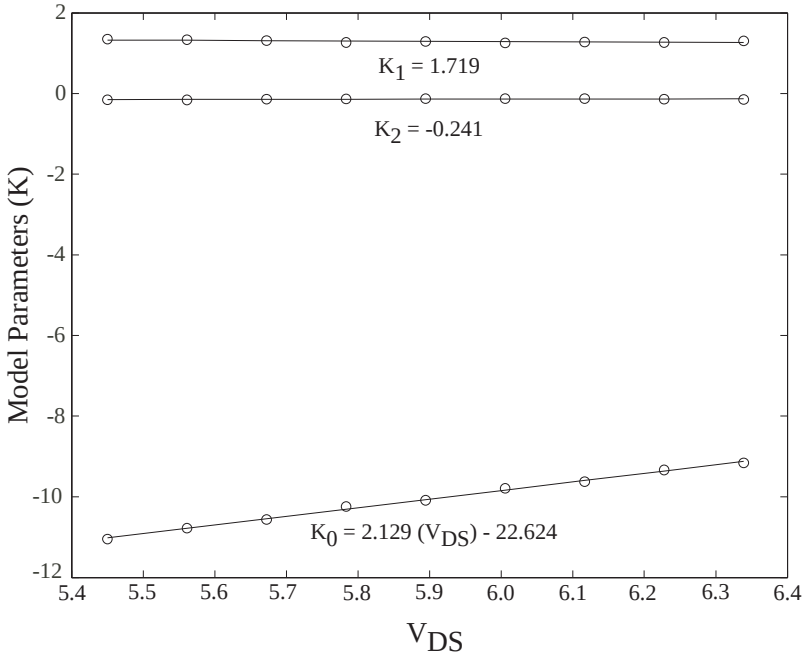


Figure 8. Mathematical Modeling of Injection: Variation of model parameters k_1, k_2, and k_3 with V_{ds}.

$I_{initial}$. Both variations can be approximated by using second order relationships. The solid lines show the fit for each data set. Hence,

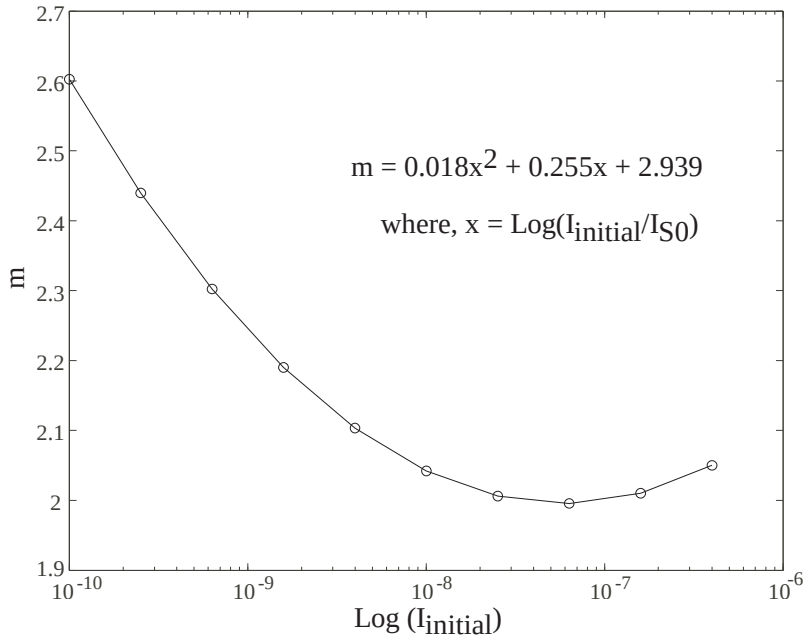
$$m = a_2 \left[\log\left(\frac{I_{initial}}{I_{S_0}}\right)\right]^2 + a_1 \log\left(\frac{I_{initial}}{I_{S_0}}\right) + a_0 \tag{29}$$

$$f = b_2 \left[\log\left(\frac{I_{initial}}{I_{S_0}}\right)\right]^2 + b_1 \log\left(\frac{I_{initial}}{I_{S_0}}\right) + b_0 \tag{30}$$

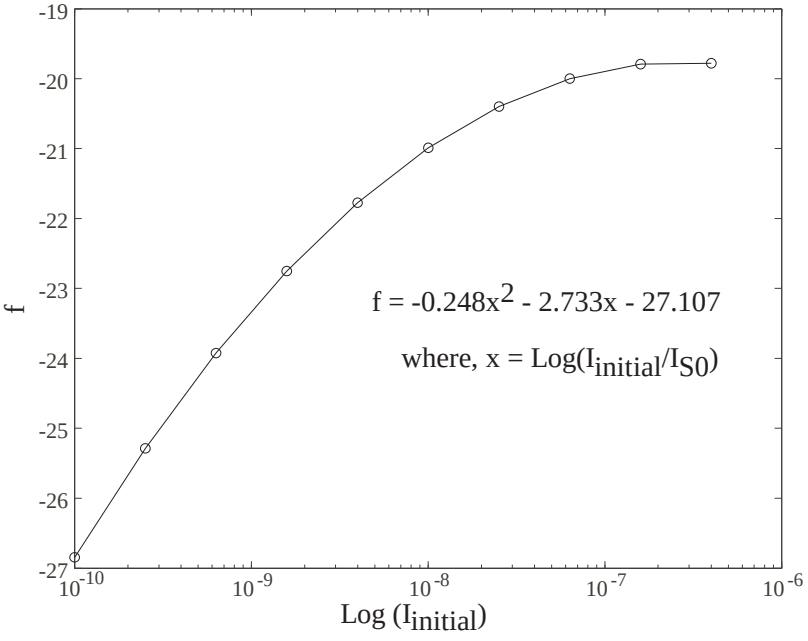
where a_0, a_1 and a_2 are regressed parameters with units of 1/Volts, and b_0, b_1 and b_2 are unitless regressed parameters. This quadratic behavior of m and f with $I_{initial}$ corroborates (27).

By substituting (29) and (30) in (28) an expression for V_{ds} in terms of $I_{initial}$ and ΔI can be obtained as

$$V_{ds} = \frac{\log\left(\frac{\Delta I}{I_{S_0}}\right) - b_2\left[\log\left(\frac{I_{initial}}{I_{S_0}}\right)\right]^2}{a_2\left[\log\left(\frac{I_{initial}}{I_{S_0}}\right)\right]^2 + a_1 \log\left(\frac{I_{initial}}{I_{S_0}}\right) + a_0} + \frac{-b_1 \log\left(\frac{I_{initial}}{I_{S_0}}\right) - b_0}{a_2\left[\log\left(\frac{I_{initial}}{I_{S_0}}\right)\right]^2 + a_1 \log\left(\frac{I_{initial}}{I_{S_0}}\right) + a_0}. \tag{31}$$



(a)



(b)

Figure 9. Mathematical Modeling of Injection: (a) Variation of model parameter m with $I_{initial}$. (b) Variation of f with $I_{initial}$.

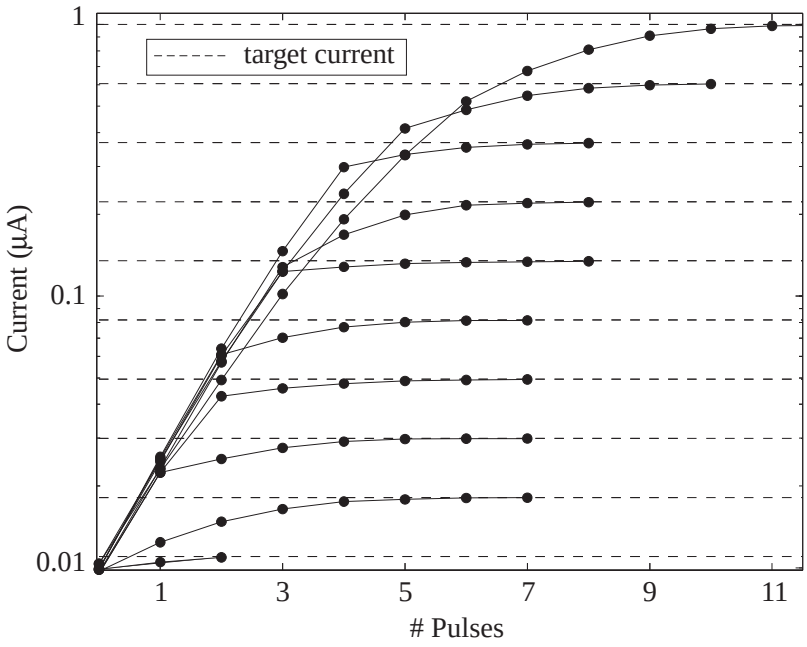


Figure 10. Target Programming: Asymptotic approach toward programming various target current.

Since this equation is a direct calculation of V_{ds}, the computational complexity has been reduced, only 6 parameters (a_0, a_1, a_2, b_0, b_1, and b_2) need to be stored and no regression has to be done before each injection. For a given I_{init} and a desired ΔI an optimal V_{ds} value can be obtained for injection.

2.4.3 Experimental Results

The algorithm was tested using large floating-gate arrays in 0.25μm and 0.5μm CMOS processes. Figure 10 shows how ten different elements were programmed asymptotically using the proposed algorithm. Each element had an initial current of 10nA (an arbitrary choice). The dashed lines show the target currents. The programming procedure stops when an element has been programmed to within 0.2% of the target current. The number of pulses increase with increasing target currents. The average number of pulses required are 7-8 for programming currents within 2 decades.

This algorithm can be used to program both deep sub-threshold currents and above threshold currents. The average percentage error (for 50 samples for each

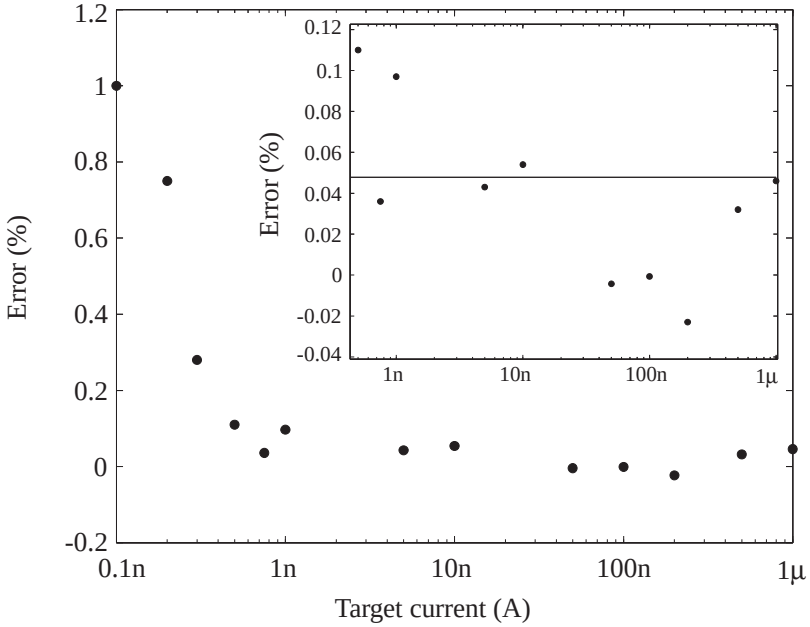
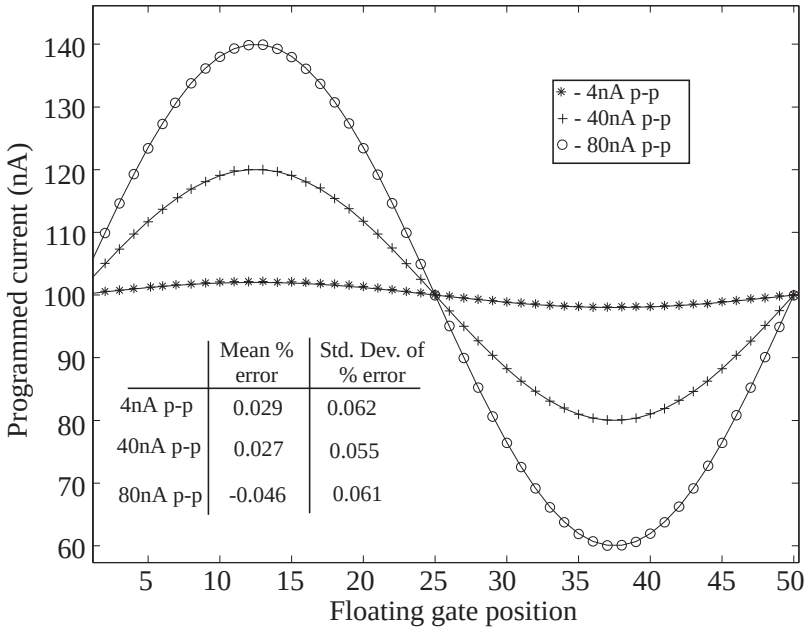


Figure 11. Programming Accuracy: Average percentage error of a 50 sample data for a wide range of programmed currents.

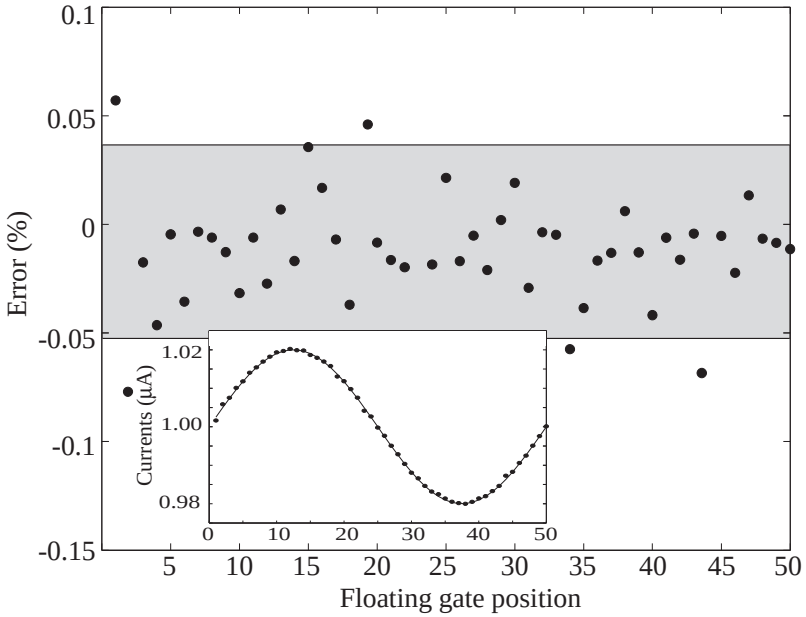
target current) for a target current range of $100nA - 1\mu A$ is plotted in Figure 11. The algorithm can be used to program target currents within 0.2% of accuracy over 3.5 decades (as shown in inset) in both 0.25μm and 0.5μm CMOS processes.

Floating gate architectures have been used as analog computational elements in a number of applications [10,23]. Arbitrary waveforms can be stored on-chip accurately using the proposed algorithm. Figure 12(a) shows sine waves of $4nA_{pp}$, $40nA_{pp}$, and $80nA_{pp}$ that were programmed onto 50 floating gates. The DC of these sine waves were 100nA. The bubbles indicate the programmed values while the solid lines represent the ideal targeted sine waves. The pulse width used for this experiment was 20μs. The average number of pulses required to approach the target currents asymptotically were 4-6 pulses. It can be observed that the maximum error is less than 0.2%. Figure 12(b) shows the percentage error of a $40nA_{pp}$ sine wave programmed over a DC current of 1μA (see inset). A maximum error of 0.07% was obtained.

As an application specific example [10] an 8x8 DCT kernel was programmed on



(a)



(b)

Figure 12. Analog Storage: (a) Current values for a row of 50 floating-gate transistors programmed as sine wave of $4nA_{pp}$, $40nA_{pp}$, and $80nA_{pp}$. (b) Percentage error for a $40nA_{pp}$ sine wave programmed at a DC current of $1\mu A$.

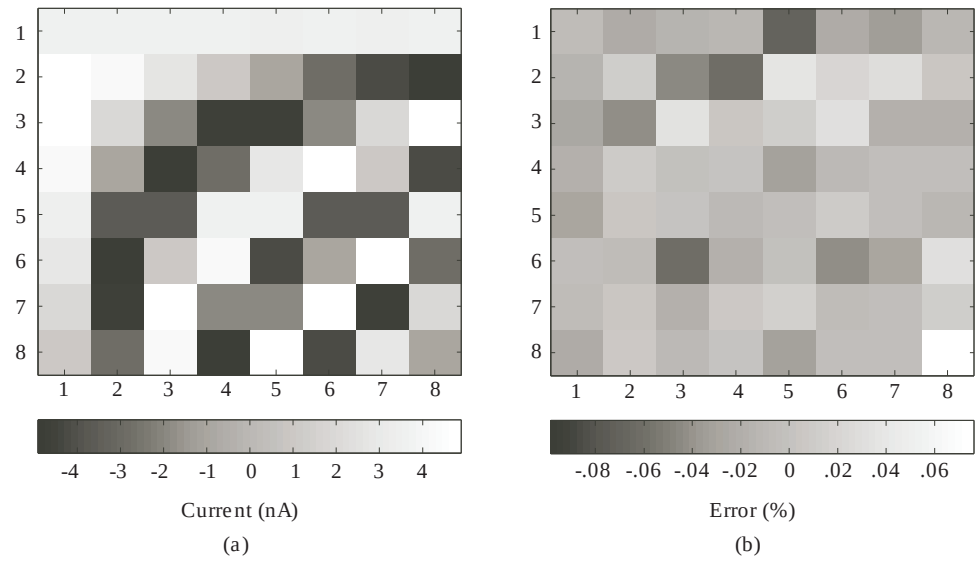


Figure 13. Analog Computational Elements: (a) A 8x8 DCT kernel programmed onto a floating-gate transistor array around a DC current value of $10nA$. (b) Percentage error of the 8x8 DCT kernel.

a floating-gate array around a DC current value of $10nA$. A pictorial diagram of the programmed DCT is shown in Figure 13(a). The maximum average deviation of the programmed values was 0.08% as seen in Figure 13(b).

2.5 Charge Retention

Floating-gate transistors inherently have good charge retention capabilities on account of the gate being surrounded by a high quality insulator. Initial investigations of floating-gate retention were carried out by observing the drain current of a floating-gate device for long periods of time. Figure 14(a) shows the drain current of a floating-gate transistor measured over a period of 380 hours. The drain current was programmed from a current of $< 1\mu A$ to an initial value of $30\mu A$ and displayed a mean value of $29.93\mu A$ with a standard deviation of $28nA$ (see Figure 14(b)). The current exhibits a short-term drift in the beginning beyond which no significant drift can be observed. This short-term drift is on account of the interface trap sites settling to a new equilibrium after programming [64]. Similar results have been observed in a

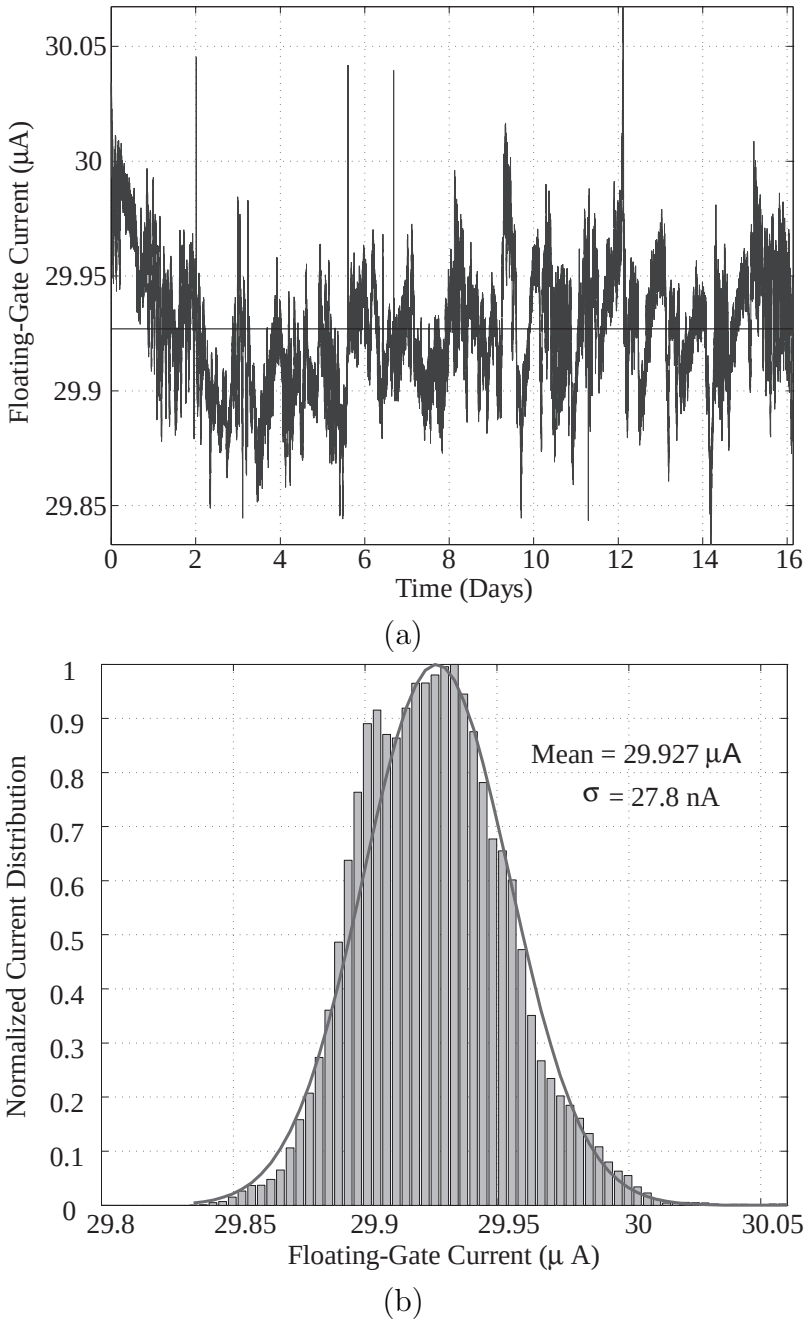


Figure 14. Long Term Retention:(a) Drain current of a floating-gate transistor in a $0.5\mu m$ CMOS process measured over 16 days. (b) Drain current distribution of a floating-gate transistor current measured over 16 days.

$1.5\mu m$ CMOS process [4]. Although this is a good indicator of the charge retention capabilities of floating-gate transistors, accurate estimates of the long-term charge retention can be made through accelerated life time tests.

2.5.1 Long Term Charge Loss

Long-term charge loss in floating-gate transistors occur due to the phenomenon of thermionic emission [18, 21, 60, 64]. The amount of charge lost is a function of both temperature and time and is given by

$$\frac{Q(t)}{Q(0)} = e^{-t\upsilon \cdot e^{\frac{-\phi_B}{kT}}} \tag{32}$$

where $Q(0)$ is the initial charge on the floating-gate, $Q(t)$ is the floating-gate charge at time t, υ is the relaxation frequency of electrons in poly-silicon, ϕ_B is the $S_i - S_iO_2$ barrier potential in electron-volts, k is the Boltzmann's constant, and T is the temperature. As expected from (32), charge loss in floating-gates is a slow process that is accelerated at high temperatures.

Floating-gate charge loss is measured indirectly by measuring the change in the transistor's threshold voltage. Programming floating-gate transistors by adding/removing charge modifies the threshold voltage of the device, V_{th}, as given by,

$$V_{th} = V_{th}^{'} + \frac{Q}{C_T} \tag{33}$$

where, Q is the floating-gate charge, $V_{th}^{'}$ is the threshold voltage of a non floating-gate device, and C_T is the total capacitance at the floating node. Using (33) the charge loss in a floating-gate can be derived as

$$\frac{Q(t)}{Q(0)} = \frac{V_{th}(t) - V_{th}^{'}}{V_{th}(0) - V_{th}^{'}} \tag{34}$$

where $V_{th}(t)$ indicates the threshold voltage of the device after time t and $V_{th}(0)$ represents the initial programmed threshold voltage.

In order to estimate the amount of charge loss in floating-gate transistors the parameters υ and ϕ_B from (32) need to be extracted for the desired process as they

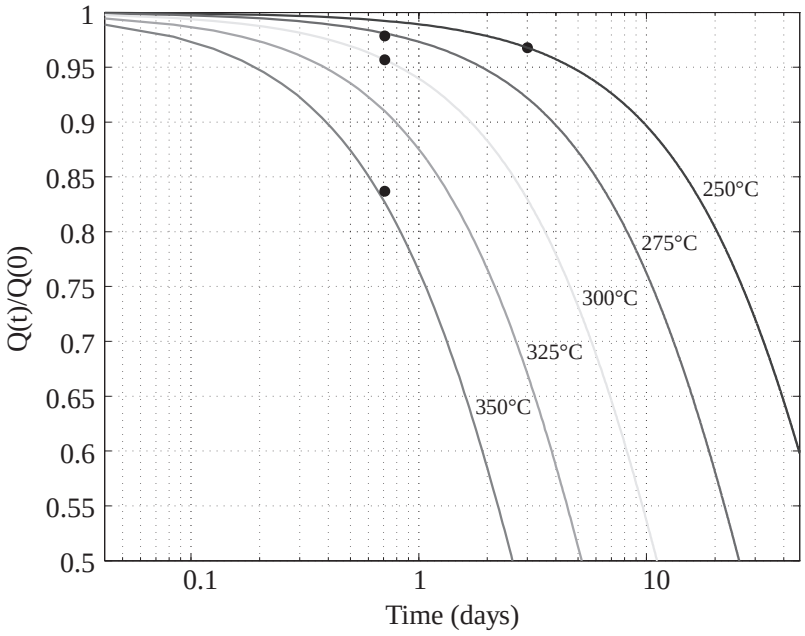


Figure 15. Charge Retention: Charge loss measured at different temperatures and time periods as estimated from threshold voltage changes is plotted using ∘'s along with the extrapolated theoretical fits in solid lines.

exhibit a wide spread in their values. Extraction of these parameters was done for a $0.5\mu m$ process by programming the floating-gate transistors to a threshold voltage of $-0.5V$ and storing them at high temperatures for a predefined time period. The change in threshold voltage was measured and using (34) the relative charge loss $\frac{Q(t)}{Q(0)}$ was estimated. Expressions for υ and ϕ_B can be derived from (32) as

$$\phi_B = \frac{kT_1T_2}{T_1 - T_2} ln\left[\frac{t_2}{t_1}.\frac{ln(x_1)}{ln(x_2)}\right] \tag{35}$$

and

$$\upsilon = \frac{-ln(x_1)}{t_1 \cdot e^{(\frac{\phi_B}{kT_1})}} \tag{36}$$

where $x_i = \frac{Q_i(t)}{Q_i(0)}$ and the subscript i denotes different data points. Using (35) and (36), the values for the barrier potential and the relaxation frequency were extracted to be $0.9eV$ and $60s^{-1}$ for a $0.5\mu m$ CMOS process. Values of $0.0618eV$ and $55ms^{-1}$ were extracted for a $0.35\mu m$ CMOS process. Figure 15 shows the measured floating-gate charge loss along with a theoretical extrapolated fit using the estimated model

Table 2. Floating-Gate Retention in 10 Years

	10% Programmed Change			**50% Programmed Change**		
Temperature	$\Delta Q/Q$	ΔV_{fg}	$\Delta I/I$	$\Delta Q/Q$	ΔV_{fg}	$\Delta I/I$
$25°C$	$10^{-3}\%$	$36.7nV$	$2\times10^{-4}\%$	$10^{-3}\%$	$156nV$	$9\times10^{-4}\%$
$90°C$	0.62%	$16.4\mu V$	0.06%	0.62%	$65\mu V$	0.47%
$140°C$	18.2%	$0.45mV$	1.8%	18.2%	$1.92mV$	10.7%

parameters. The measured data agrees well with the theoretical prediction; this result has been observed across many floating-gate devices.

Using the extrapolated values for the parameters υ and ϕ_B, retention numbers were estimated using (32) for transistors operating in the weak inversion. Table 2 summarizes the data retention numbers for the $0.5\mu m$ process for two different cases of programmed difference currents, namely, a 10% change and a 50% change for a time period of 10 years for different temperatures. A total floating-gate capacitance of $100fF$ and a κ of 0.7 has been assumed for these calculations. As can be observed from Table 2, the percentage change in charge over time at different temperatures is the same, irrespective of the programmed current difference between devices. However, the change in the floating-gate voltage with time for different temperatures depends on the programmed current difference. This is on account of the fact that larger the current difference, larger is the difference in charge and so larger is the absolute change in the charge with time for a given temperature. This results in the change in the floating-gate voltage being larger for a larger current difference. For the same reasons, the percentage change in programmed currents is larger for the case of the 50% programmed current difference as against the 10% programmed current difference. As can be observed, floating-gate transistors display excellent charge retention capabilities.

Direct tunneling through the gate oxide (gate leakage) is a limitation for charge retention in floating-gate transistors for oxide thicknesses less than $5nm$ which are

typical for finer line processes ($< 0.25\mu m$). However, in smaller dimension processes, floating-gate transistors can be implemented using the available thick oxide transistors thereby preserving their charge retention capabilities.

2.5.2 Radiation Effects

It is a well known fact that radiation affects the charge retention capabilities of floating-gate transistor memories. For example, exposure of floating-gate transistors to ultra-violet (UV) radiation is a commonly used technique in EPROM devices to erase the information stored on the floating-gate [1]. Exposing the floating-gate transistor to high energy UV rays imparts sufficient kinetic energy to the electrons stored on the floating-gate to surmount the $Si - SiO_2$ barrier, thereby erasing the floating-gate.

The use of floating-gate memories for space applications has motivated the study of the radiation effects on their retention capabilities. In particular, flash memories have been studied after radiation exposure by several authors [48, 59, 66], regarding both total ionizing dose effects[1] and single event upsets[2]. In both cases, the most radiation sensitive part of commercial flash memories is the complex circuitry external to the floating-gate memory cell array [59]. The loss of the charge stored in the floating gate of a programmed cell and the consequent threshold voltage shift ΔV_{th} have been less frequently investigated in literature [22, 51]. It has been demonstrated in [51] that floating gate charge loss upon heavy ion irradiation is not negligible and is more evident when decreasing the floating gate area. At this point, no studies have been reported regarding the radiation effects in floating-gate memories when used as analog memories instead of digital memories.

[1] Cumulative long term ionizing damage due to protons and electrons.

[2] A change of state caused by a high-energy particle strike to a sensitive node in a micro-electronic device, such as in a microprocessor, semiconductor memory, or power transistors.

Table 3. Summary of Floating-Gate Transistor Performance

Parameter	Value
Technology	0.25μm and 0.5μm N-well CMOS
Floating-Gate Dim.(W/L)	6λ / 4λ
Array size	96×16
Maximum % Error	$< \pm 0.2\%$
Pulse Width	20μs
Global Erase	Fowler-Nordheim tunnel
Programming mechanism	Hot–electron injection
Current range	150pA to 1.5μA
Avg. no. of pulses for programming	7-8
Charge Loss 10 years @ 25^oC	0.001%

2.6 Summary

The key design issues when using floating-gate transistors in analog circuits has been discussed. Equations that can be used in designing the W/L and C_T of the floating-gate transistors for a given precision have been derived. It has also been demonstrated that the programming accuracy will depend of the region of operation. A predictive algorithm that can be used to program large arrays of floatinggate elements at fast rates has been presented. Experimental measurements for different applications showed an accuracy of $< 0.2\%$ over a wide range of target currents (over 3.5 decades). Charge loss in floating-gate transistors has been studied using the framework of a thermionic emission model. A negligible loss in charge has been extrapolated for 10 years at 25^oC for a $0.5\mu m$ CMOS process. A compilation of the experimental results presented in this chapter can be seen in Table 3. In summary, the non-volatile charge retention when combined with programmability, makes floating-gate transistors well suited for use in high performance analog circuits.

CHAPTER III

A PRECISION CMOS AMPLIFIER

This chapter presents a floating-gate based offset cancelation scheme for a folded cascode amplifier. The proposed approach results in continuous-time operation of the amplifier with long-term offset cancelation that obviates the need for any refresh circuitry. This scheme is no limited to low-bandwidth applications and does not use sampling techniques, hence avoids such issues as charge injection, clock feed-through, and under-sampled wide-band noise.

3.1 Previous Work

Mismatches between MOS transistors pose a serious challenge to analog circuit designers and most commonly manifest themselves as an offset voltage in operational amplifiers. Techniques commonly used to reduce the offset voltage in amplifiers include auto-zeroing, correlated double sampling, and chopper stabilization [31]. Auto-zeroing is primarily useful for sampled data systems and is limited by issues such as charge injection, clock feed-through, and wide-band noise folding into the baseband on account of under-sampling. For a continuous-time operation, chopper stabilization or ping-pong amplifiers are typical alternatives. The chopper amplifier is limited in use to low-bandwidth applications while the ping-pong approach involves the use of multiple amplifiers and multi-phase clocks that add additional overhead in terms of area and power.

Correcting analog circuit mismatch using resistor trimming is an alternate technique. Resistor trimming is usually performed using laser annealing, laser trims, poly fuses, and zener zapping. Both laser annealing and laser trims are expensive and do not provide the flexibility of in-package trims. Trimming using poly fuses and zener

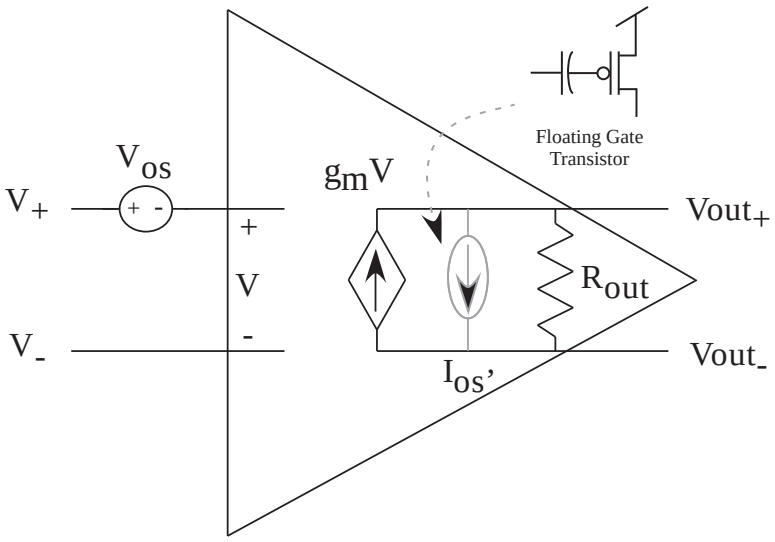


Figure 16. Amplifier Macromodel: The inherent input referred offset voltage V_{os} of the amplifier is canceled by programming an offset current I_{os} using floating-gate transistors.

zapping is discrete in nature and therefore accuracy is limited to the smallest resistor step used. Also, using a number of zener diodes and poly fuses involves an area penalty. All of the above resistor trimming techniques are one-time programmable.

The use of floating-gate transistors to correct for mismatch in analog circuitry has been investigated by other authors as well [21, 64]. The approach in [21] results in an uni-directional offset cancelation. This requires an intentional offset creation of the correct polarity during the design phase of the amplifier for proper operation. This intentional offset creation has been cited as the reason for the degradation of the offset voltage temperature sensitivity [21]. The work in [64] introduces a trimming circuitry based on floating-gate transistors to produce a difference current which is then used as a building block to compensate for mismatch induced errors. The proposed approach in this paper is conceptually similar to that in [64] in that it uses a differential current to trim offsets. However, the difference current is created using just two floating-gate transistors which then form an integral part of the amplifier of interest. This results in an advantage in terms of both area and design overhead.

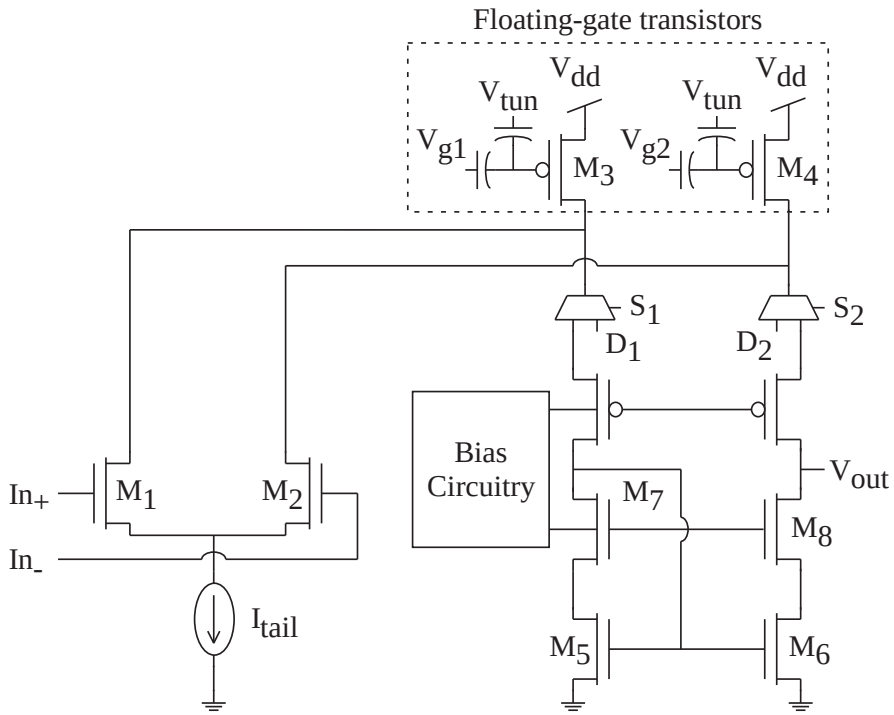


Figure 17. Floating-Gate Based Offset Compensation: A circuit implementation of the proposed offset removal scheme in a single stage folded-cascode amplifier.

3.2 Proposed Offset Voltage Compensation

The amplifier macro-model shown in Figure 16 illustrates the concept of using floating-gate transistors to cancel the input offset voltage of an amplifier. Here the inherent offset voltage V_{os} of the amplifier is nullified by an offset current I_{os} in the opposite direction. A programmable I_{os} is enabled through use of floating-gate transistors, thus allowing for offset correction after fabrication. Such an approach results in a compact architecture with a simple design strategy that avoids the overhead of using floating-gate transistors as separate trimming elements as in [21, 64] or current-mode DACs as trimming elements [45]. Also, the proposed offset cancelation scheme is independent of other amplifier parameters unlike other approaches [3, 31] and the offset cancelation by itself dissipates no additional power.

3.2.1 Amplifier Architecture

A practical implementation of the proposed approach is shown in Figure 17 through the use of a single stage folded cascode amplifier. The only distinction from the typical folded cascode topology [6] is the use of floating-gate transistors for M_3 and M_4. Here the offset current I_{os} is programmed as the current difference I_3-I_4 between M_3 and M_4 such that inherent offset voltage of the amplifier V_{off} can be canceled. Multiplexors S_1 and S_2 allow the isolation of transistors M_3 and M_4 for programming purposes. During normal operation, M_3 and M_4 behave as typical current sources.

A key advantage of this architecture is that the programming transistors are an integral part of the amplifier thereby simplifying the design process. Initially, design of a typical folded cascode is performed; all transistors are sized according the specifications required. Offset compensation is then added by making M_3 and M_4 floating gate transistors. Appropriate switches are then added to isolate the floating-gate transistors during programming. Finally C_{in} is sized to meet the C_T requirement given by the FOM (refer to Section 2.3).

3.2.2 Input Referred Offset Voltage Cancelation

The analytical expression for the relationship between the input referred offset voltage and the programmed currents is given by

$$V_{off} = V_{off}{}' + \frac{I_3 - I_4}{g_{mn_1}} = V_{off}{}' + \Delta V_{fg_{3-4}} \frac{g_{mp_3}}{g_{mn_1}} \tag{37}$$

where $V_{off}{}'$ is the true uncompensated offset voltage of the amplifier, g_{mn_1} is the transconductance of M_1, g_{mp_3} is the transconductance of M_3, and $\Delta V_{fg_{3-4}}$ is the difference between the floating-gate voltages of M_3 and M_4. As seen in (37), the offset voltage can be canceled by programming a $\Delta V_{fg_{3-4}}$ value opposite to $V_{off}{}'$.

The minimum offset achievable will be determined by the minimum $\Delta V_{fg_{3-4}}$ possible. Using (37), the offset voltage change can be expressed as

$$\Delta V_{off} = \Delta V_{fg} \frac{g_{mp_3}}{g_{mn_1}} = \frac{\Delta Q}{C_T} \frac{g_{mp_3}}{g_{mn_1}} \tag{38}$$

where $\Delta V_{fg} = \frac{\Delta Q}{C_T}$. Offset voltage requirements can be guaranteed by properly sizing C_T and $\frac{g_{mp_3}}{g_{mn_1}}$ as seen in (38).

3.2.3 Offset Voltage Temperature Sensitivity

Assuming the device mismatch is dominated by the ΔV_{th} [46], the temperature sensitivity of the offset voltage can be analyzed by rewriting (37) as follows

$$V_{off} = \Delta V_{th_{1-2}} + \Delta V_{th_{5-6}} \frac{g_{mn_5}}{g_{mn_1}} + \Delta V_{fg_{3-4}} \frac{g_{mp_3}}{g_{mn_1}} \tag{39}$$

where $\Delta V_{th_{1-2}}$ is the threshold voltage mismatch between M_1 and M_2, and $\Delta V_{th_{5-6}}$ is the threshold voltage mismatch between M_5 and M_6. The only temperature dependance arises from the $\frac{g_{mp_3}}{g_{mn_1}}$ ratio in second term, due to the difference in the mobility coefficient of electrons and holes.

Assuming all transistors are operating in strong inversion, the offset voltage expression as a function of the transistors bias currents can be obtain from (39) as

$$V_{off} = \Delta V_{th_{1-2}} + \Delta V_{th_{5-6}} \sqrt{\frac{K_{n5}}{K_{n1}} \frac{I_3 - I_1}{I_1}} + \Delta V_{fg_{3-4}} \sqrt{\frac{K_{p3} I_3}{K_{n1} I_1}}. \tag{40}$$

The offset voltage temperature dependance can be approximated to a first order by differentiating (40) with respect to temperature resulting in

$$\begin{aligned} \frac{\delta V_{off}}{\delta T} = & \frac{\Delta V_{th_{1-2}}}{2} \frac{g_{mn_5}}{g_{mn_1}} \left[\frac{I_3}{I_3 - I_1} \left(\frac{1}{I_3} \frac{\delta I_3}{\delta T} - \frac{1}{I_1} \frac{\delta I_1}{\delta T} \right) \right] \\ & + \frac{\Delta V_{fg_{3-4}}}{2} \frac{g_{mp_3}}{g_{mn_1}} \left[\frac{n_n - n_p}{T} + \frac{1}{I_3} \frac{\delta I_3}{\delta T} - \frac{1}{I_1} \frac{\delta I_1}{\delta T} \right] \end{aligned} \tag{41}$$

where n_p and n_n are the mobility temperature coefficients of a pMOS and an nMOS transistor, respectively , and T is the temperature in Kelvins.

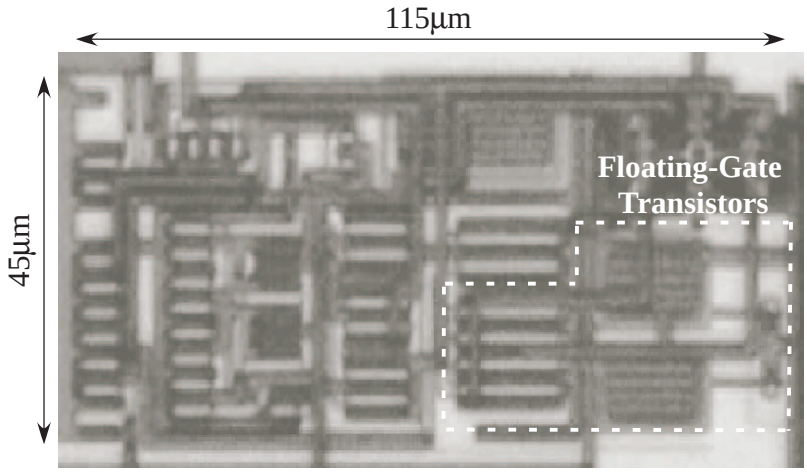


Figure 18. Amplifier Die Micrograph: The prototype circuit occupies an area of $115\mu m \times 45\mu m$ in a $0.5\mu m$ CMOS process.

If I_1 and I_3 are biased from the same source, such that $\frac{1}{I_1}\frac{\delta I_1}{\delta T} = \frac{1}{I_3}\frac{\delta I_3}{\delta T}$, (41) can be reduced to

$$\frac{\delta V_{off}}{\delta T} = \frac{\Delta V_{fg_{3-4}}}{2}\frac{g_{mp_3}}{g_{mn_1}}\left[\frac{n_n - n_p}{T}\right] = \frac{V_{off}{}'}{2}\left[\frac{n_n - n_p}{T}\right]. \tag{42}$$

The temperature dependence of the offset voltage is solely dictated by the difference between device parameters n_p and n_n, as shown in (42).

3.3 Experimental Results

A prototype amplifier was fabricated in a $0.5\mu m$ CMOS process. The chip micrograph of the proposed amplifier is shown in Figure 18. The total area of the amplifier excluding the buffer is $115\mu m \times 45\mu m$. The additional area occupied by the input capacitors and the switches on account of using floating-gate transistors is $45\mu m \times 45\mu m$. The amplifier was designed to operate in the strong inversion region and was tested with a 3.3V power supply.

Applying (37) and programming the drain currents of transistors M_3 and M_4, the prototype amplifier was programmed to five different offset voltages in steps of $10mV$ ranging from $-20mV$ to $+20mV$. Figure 19 shows the DC transfer characteristics of the amplifier configured as a comparator with the non-inverting terminal held

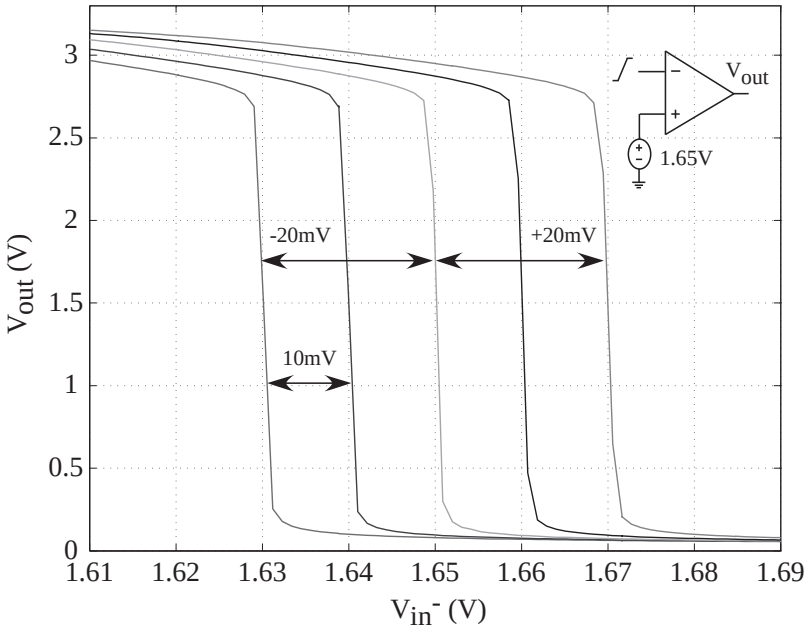
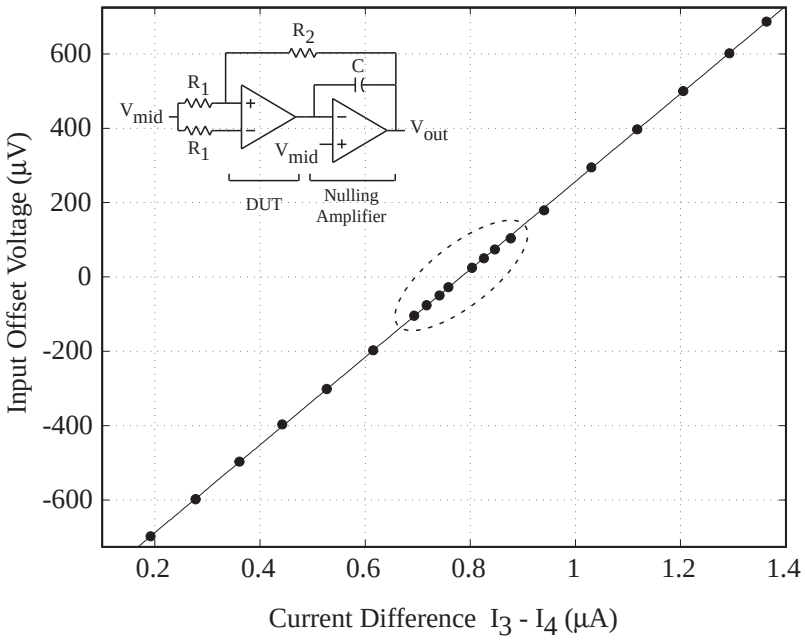


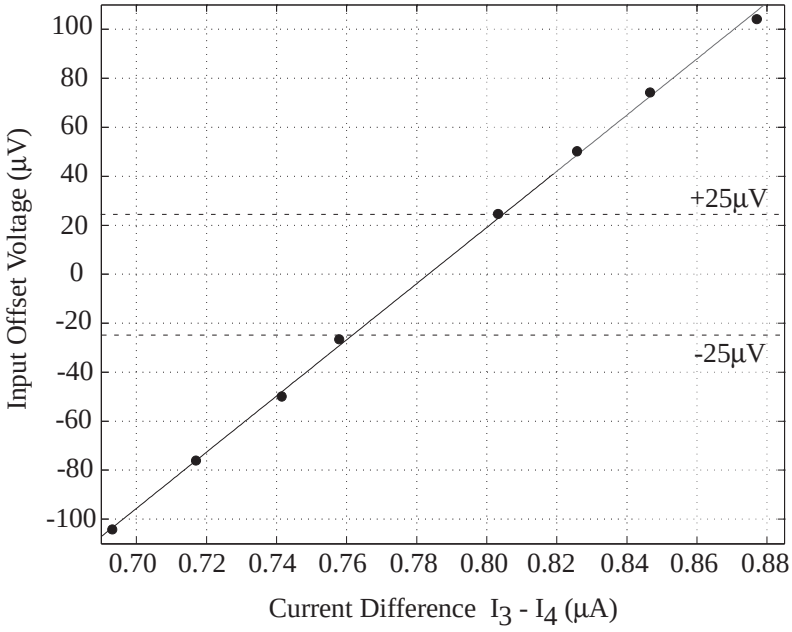
Figure 19. Offset Programming: Transfer curves of the prototype amplifier configured as a comparator with offset voltages programmed in $10mV$ steps.

at $1.65V$. As can be observed the comparator trip points are evenly spaced $10mV$ apart as programmed. Also, it can be seen that the amplifier can be programmed to display different offset voltages with both positive and negative polarities. This clearly demonstrates the programmable nature of the approach, a feature that could be exploited when designing, for instance, comparators.

Accurate measurements of the offset voltage were made by using the amplifier under test along with a second amplifier configured as a nulling amplifier forming a servo loop [20] as seen in the inset of Figure 20(a). Figure 20(a) shows the measured input referred offset voltage of the amplifier plotted against the various programmed floating-gate transistor difference currents in steps of $100\mu V$. The measured data shows a linear dependence of the offset voltage with the programmed difference currents as expected from (37). As can be observed in Figure 20(b), that zooms into the region encircled in Figure 20(a), the offset voltage of the prototype amplifier can be reduced to $25\mu V$. Experimentally, it is possible to program current increments as low



(a)



(b)

Figure 20. Offset Cancelation: (a) Plot of the input offset voltage against $I_3 - I_4$, programmed in $100\mu V$ steps. (b) Optimal voltage cancelation; a minimum offset voltage of 25μV was obtained.

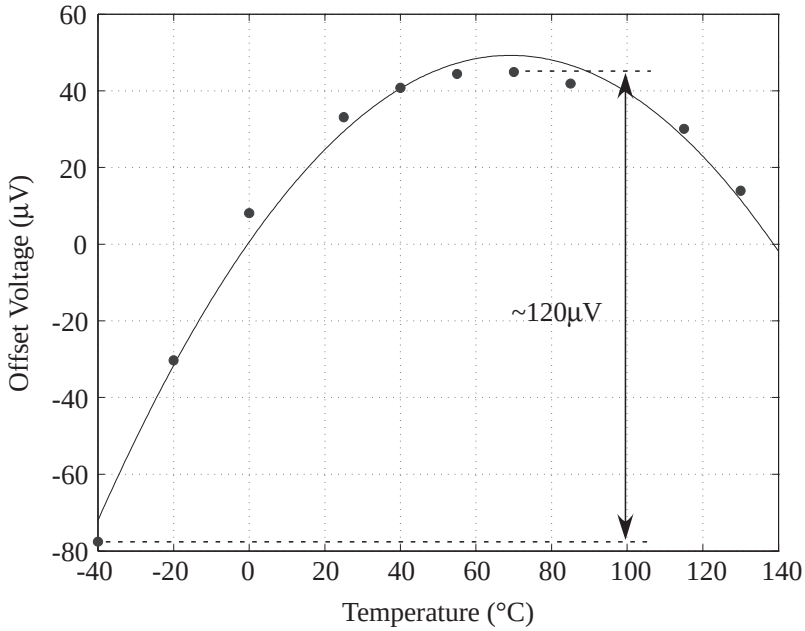


Figure 21. Offset Voltage Temperature Sensitivity: Temperature variation of the input referred offset voltage across for a temperature range of 170^oC.

as $0.1nA$. Theoretically, this indicates that offset voltages in the $100's$ of nano-volts range are possible to achieve. At present however, the primary limitation has been the internal noise of the amplifier itself.

Figure 21 shows the temperature sensitivity of the input offset voltage. The offset voltage was measured for temperatures ranging from $-40\ ^\circ C$ to $130\ ^\circ C$ after been programmed to $37\mu V$ at $25^\circ C$. A maximum change of $130\mu V$ was observed over the full temperature range of $170\ ^\circ C$.

In order to experimentally observe the offset drift with time, the amplifier was programmed to an initial offset voltage of around $-50\mu V$ from an initial offset voltage of $1mV$ and then it was measured continuously over a period of 110 hours. Figure 22 shows the measurement of the offset voltage as a function of time. As can be observed, the offset voltage exhibits an initial short term drift of about $-10\mu V$ on account of the interface trap sites settling to a new equilibrium. Beyond the initial short-term drift, the offset voltage drift is negligible as expected from earlier measurements on

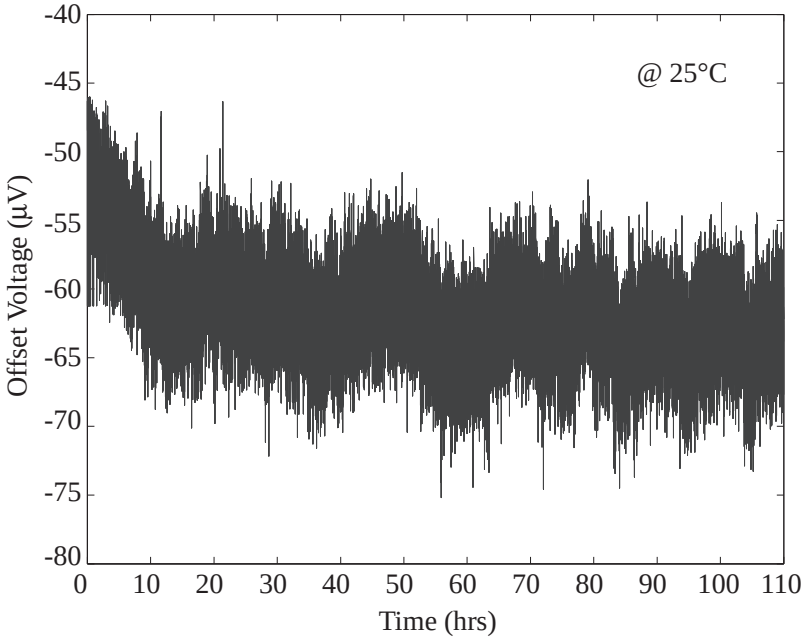


Figure 22. Offset Voltage Drift: Offset voltage variation as a function of time for a period of 110 **hours.**

floating-gate charge retention.

3.4 Summary

An amplifier topology that uses floating-gate transistors to compensate for the inherent offset voltage has been presented. The approach places minimal overhead on the amplifier design with non-volatile storage of the 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 25μV . The offset voltage exhibits a temperature sensitivity of 130μV over a temperature range of 170^oC. A summary of the experimental results is shown in Table 4. Table 5 presents a qualitative comparison of the proposed scheme with the other techniques commonly used to reduce the offset voltage, previously discussed in Section 3.1. The main advantages of the proposed approach over these techniques are: 1) it involves no sampling and hence avoids such issues as charge injection and clock feed-through that are serious limitations to

Table 4. Summary of Amplifier 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)	$115\mu m \times 45\mu m$

auto-zeroing and ping-pong schemes, 2) unlike chopper stabilization, is not limited to low-bandwidth applications, 3) it can provide a continuous range of offset voltages rather than discrete values offered by the resistor trimming and the DAC based scheme.

Table 5. Qualitative Comparison of Different Amplifier Offset Voltage Cancelation Schemes

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

CHAPTER IV

A PROGRAMMABLE LOW TC CMOS VOLTAGE REFERENCE

This chapter presents two floating-gate based low TC voltage references. The proposed circuits generate a voltage reference as the difference between to differently programmed floating-gate transistors. This results in a programable, low TC , and area efficient reference that does not requires trimming circuitry.

4.1 Previous Work

Voltage references are critical components in both analog and digital systems. The accuracy, temperature sensitivity, and drift of references impact the performance of many circuit blocks such as analog-to-digital converters, digital-to-analog converters, and power management circuitry. Typically, the first-order compensated bandgap voltage reference [19, 36] is the preferred architecture; it provides temperature coefficients in the order of $50 ppm/^oC$. Mismatch between design components are corrected using a post-fabrication trim procedure while higher order temperature effects are reduced for by using schemes such as curvature correction [71].

Several techniques have been proposed [57] for modifying the bandgap reference voltage to provide voltages less than the bandgap voltage of silicon. The structure in [9] uses native nMOS transistors while those in [42, 54] are architectures that avoid low-threshold voltage devices. These architectures require matched resistors with mismatch being addressed at the expense of area and costly post-fabrication schemes such as laser trimming. Also, all of the above schemes restrict the output voltage to a single value that is set during the design phase thereby limiting the range of reference

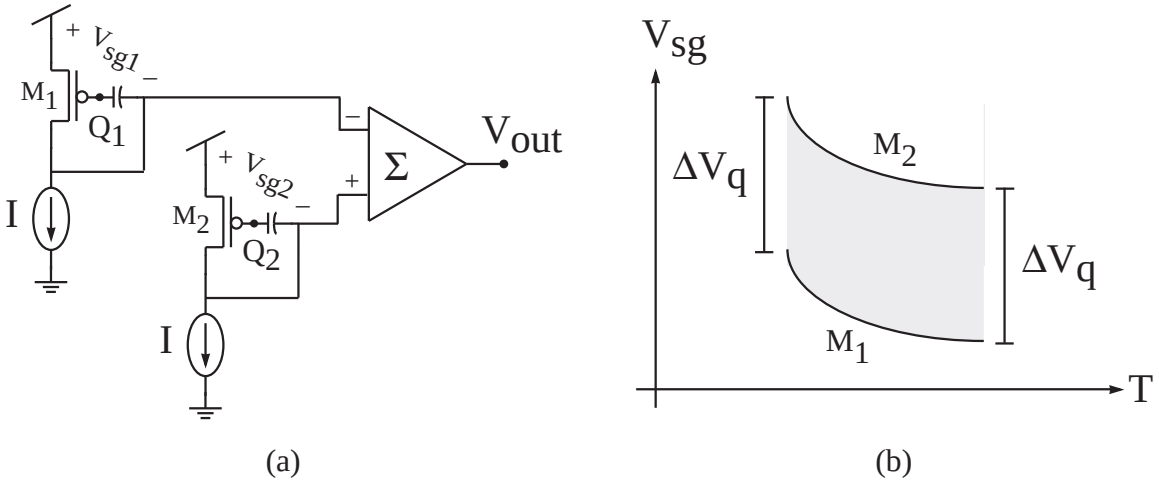


Figure 23. Conceptual representation of the proposed reference: (a) Schematic diagram of the proposed approach. (b) Graphical representation of the proposed approach.

voltages.

A number of alternate techniques have been proposed [17, 53, 61] to design voltage references wherein, the reference voltage is independent of the energy bandgap of silicon. The approach in [17] uses transistors fabricated with different threshold voltages to generate a voltage reference while [53] exploits the difference in temperature behavior between nMOS and pMOS transistors. A voltage reference has been demonstrated based on poly-silicon gate work function difference in [61]. The use of floating-gate in building voltage references has been demonstrated in [8] with temperature coefficients of $< 1ppm/^oC$. Excellent performance is obtained with a high complexity circuit that in not scalable and does no support low voltage operation.

4.2 Proposed Voltage Reference

The conceptual representation of the proposed approach in designing a voltage reference is shown in Figure 23(a). To produce a temperature-insensitive output, the reference is obtained as the difference between the effective source-gate voltages of two floating-gate transistors. Biasing the transistors with identical currents, results in an output voltage that is directly proportional to the differential charge stored in

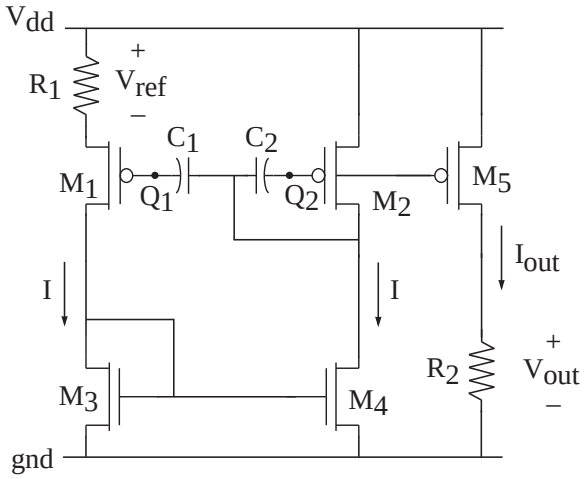


Figure 24. Proposed Voltage Reference #1: Simplified circuit schematic of the proposed programable voltage reference #1.

the transistors. The temperature dependencies of the threshold voltage and overdrive voltage are canceled due to the differential nature of the structure as shown graphically in Figure 23(b). In Figure 23(a), the operation of finding the difference between the charge on the floating-gates is shown explicitly using an operational amplifier for ease of understanding. In the practical implementation of the concept, the circuit architecture is such that the subtraction occurs without the need of an amplifier.

4.3 Reference Architecture #1

A practical implementation of the proposed concept is shown in Figure 24. The proposed circuit is similar to the popular K-multiplier circuit [5] with the difference being that transistors M_1 and M_2 are designed to be floating-gate transistors. Assuming M_1 and M_2 are identical and their currents match, $C_1 = C_2 = C$, and the source and bulk terminals of M_1 are connected together, the reference voltage V_{ref} can be expressed as

$$V_{ref} = V_{sg_2} - V_{sg_1} = \frac{Q_2 - Q_1}{C} = \frac{\Delta Q_{2-1}}{C} \tag{43}$$

where $\Delta Q_{2-1} = Q_2 - Q_1$ is the charge difference between the floating-gate transistors M_1 and M_2. A more realistic expression for V_{ref} can be written by taking in account

the threshold voltage mismatch between M_3 and M_4. Assuming strong inversion operation, (43) can be rewritten as

$$V_{ref} = \frac{\Delta Q_{2-1}}{C_T} - \Delta V_{th_n}\sqrt{\frac{K_n}{K_p}} \tag{44}$$

where ΔV_{th_n} is the threshold voltage mismatch between M_3 and M_4. Although ΔV_{th_n} will not affect V_{ref}, due to the reference programmability, it will introduce some temperature behavior. This analysis ignores the Early effect and assumes that the input capacitance and the total floating-gate capacitance of the two transistors are matched.

The proposed architecture enables a programmable voltage reference as well as a programmable current reference. Using (43), the output current I_{out} and the output voltage V_{out} can be expressed as

$$I_{out} = \frac{\Delta Q_{2-1}}{C}\frac{1}{R_1} \tag{45}$$

and

$$V_{out} = \frac{\Delta Q_{2-1}}{C}\frac{R_2}{R_1} \tag{46}$$

respectively. Arbitrary I_{out} or V_{out} values are possible by modifying Q_1 and Q_2 as seen in (45) and (46). When used as a voltage reference, R_1 size can be used as a design parameter for a predetermined power consumption at a given V_{ref}.

A key design issue is the sizing of the input capacitance C_2 of the floating-gate transistor M_2. The capacitive division caused by C_2 needs to be large enough to keep M_2 in saturation. The bias current and the capacitive ratio should be designed such that the gate voltage of M_2 obeys the following condition

$$V_{gs_2} < \frac{V_{th_p}}{1 - \frac{C_2}{C_T}}. \tag{47}$$

Ideally a capacitive ratio of 1 ensures that M_2 is in saturation for all values of V_g. However, designing according to (47) will minimize circuit area.

4.3.1 Reference Temperature Sensitivity

Assuming the source and bulk terminals of M_1 are connected together, the temperature sensitivity of V_{ref}, obtained by differentiating (44) with temperature, can be written as

$$\frac{\delta V_{ref}}{\delta T} = \frac{\Delta Q_{2-1}}{C}\frac{1}{C}\frac{\delta C}{\delta T} - \frac{\alpha \Delta V_{th_n}}{2T}\sqrt{\frac{K_n}{K_p}}. \tag{48}$$

where α is the difference in temperature coefficients of the electron and hole mobilities. Here, it is also assumed that $C = C_1 = C_2 = C_T$ and ΔQ_{2-1} displays zero temperature dependence. The temperature sensitivity of V_{ref} arises form the temperature dependance of the input capacitor of the floating-gate transistor and temperature coefficient difference between the electron an hole mobility as seen in (48). Temperature coefficient for poly-poly capacitors range from $20ppm/^oC$ to $50ppm/^oC$. Typically, the electron and hole mobility temperature coefficient is modeled as -1.5. However, for doping concentrations greater than $10^{17}/cm^3$, the temperature coefficient of electron mobility is given by -1.2 while that of the hole mobility is given by -1.9 [32]. This error can be mitigated by ensuring that the nMOS transistor pair $M_3 - M_4$ match very well.

If the bulk terminal of M_1 is connected to V_{dd}, (44) can be rewritten as

$$V_{ref} = \frac{\Delta Q_{2-1}}{C_T} + \gamma\sqrt{2\phi_F + V_{ref}} - \gamma\sqrt{2\phi_F} - \Delta V_{thn}\sqrt{\frac{K_n}{K_p}} \tag{49}$$

where ϕ_F is the Fermi potential of the bulk and γ is the body effect coefficient which is constant and independent of temperature. Note that ϕ_F, K_p, and K_n are temperature dependent. The first order temperature dependance of V_{ref} is now given by

$$\frac{\delta V_{ref}}{\delta T} \approx \frac{-\gamma\phi_F}{T}\left[\frac{1}{\sqrt{2\phi_F}} - \frac{1}{\sqrt{2\phi_F + V_{ref}}}\right] - \frac{\alpha \Delta V_{thn}}{2T}\sqrt{\frac{K_n}{K_p}}. \tag{50}$$

A V_{th} mismatch between M_1 and M_2 occurs due the connection of the bulk terminal of M_1 to V_{dd}. This V_{th} mismatch results in a degradation of the temperature behavior as seen in (50).

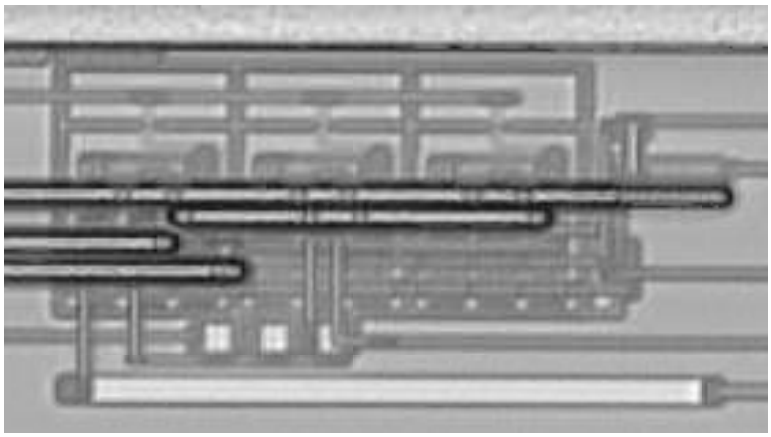


Figure 25. Voltage Reference #1 Die Micrograph: Chip micrograph of the prototype voltage reference in a $0.35\mu m$ CMOS process.

4.3.2 Reference Minimum Power Supply

The proposed architecture is advantageous in that it is well suited for low power supply operation. Notice that since the reference is essentially a circuit that operates at DC, long channel devices can be used and therefore cascoding can be avoided. For the circuit in Figure 24, the expression for the minimum power supply requirement can be written as

$$V_{dd_{min}} = V_{ref} + V_{th_n} + V_{dsat_3} + V_{dsat_1} \tag{51}$$

where V_{dsat_1} and V_{dsat_3} are the minimum drain-source voltages required to keep transistor M_1 and M_3, respectively, in the saturation region. Typical numbers for a $0.35\mu m$ CMOS process include, V_{th_n}=0.5V and V_{dsat_1}=V_{dsat_3}=0.3V. Using these numbers, a $V_{dd_{min}}$=1.8V can be used to obtain a maximum reference voltage of 0.7V . Modifications to the reference circuit, such as using a DC level-shifting current mirror [54], can result in lower supply voltage operation.

4.3.3 Experimental Results

Figure 25 shows the chip micrograph of the prototype reference fabricated in a $0.35\mu m$ CMOS process. The reference just occupies $0.0022mm^2$ of area, excluding buffers.

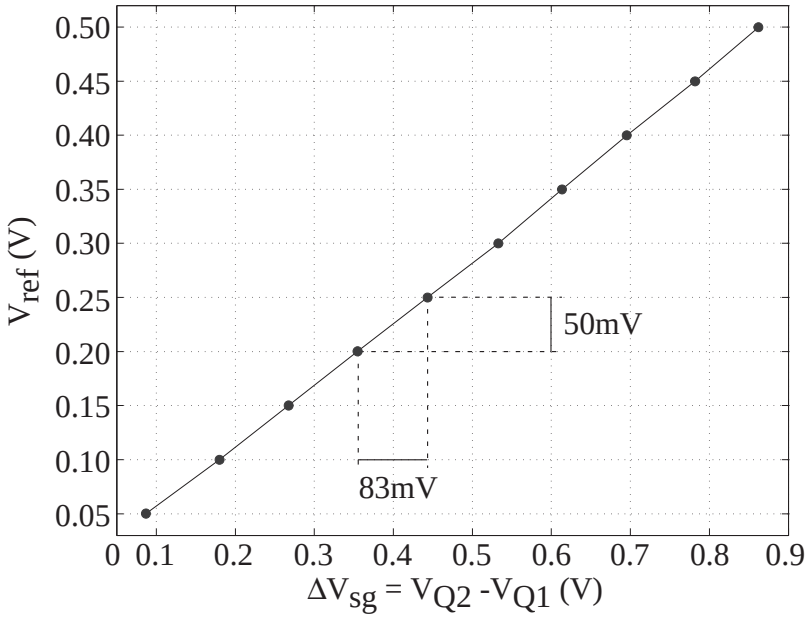


Figure 26. Voltage Reference #1 Programmability: Plot of V_{ref} against the programmed voltage difference between M_2 and M_1.

The prototype circuit was design to operate in the strong inversion region and used 50KΩ resistor to limit the bias currents in the μA range.

The programming capability of the proposed voltage reference is clearly demonstrated in Figure 26. Here, a plot of the programmed reference voltage as a function of the threshold voltage difference between transistors M_2 and M_1 is presented. The plot is linear as is implicitly conveyed in the theoretical equation (43). In order to estimate the accuracy achievable with the proposed scheme, the reference voltage was programmed in steps of 1mV from a value of 0.25V to 0.26V . Figure 27 shows the measured curve along the deviation of the programmed reference from the target value. As can be observed, the average error due to programming is within 40μV. This clearly demonstrates the high accuracy that is possible on account of the programmable nature of the reference voltage. The above accuracy of 40μV has been achieved at the package level which is a significant advantage over other schemes such as laser trimming that are techniques applied at the wafer level. A wafer level technique suffers from the drawback of the trimmed value changing drastically on account

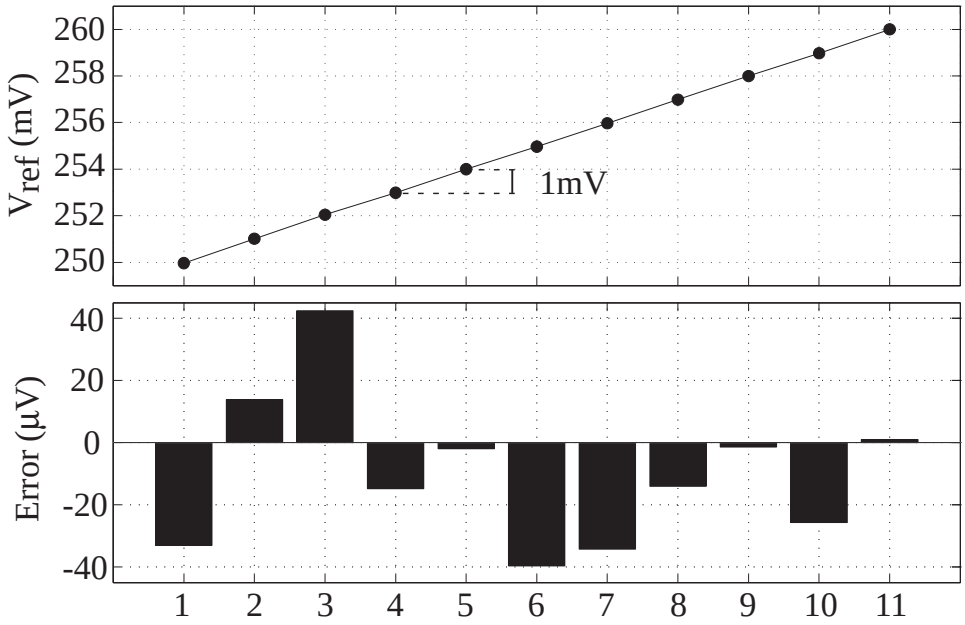
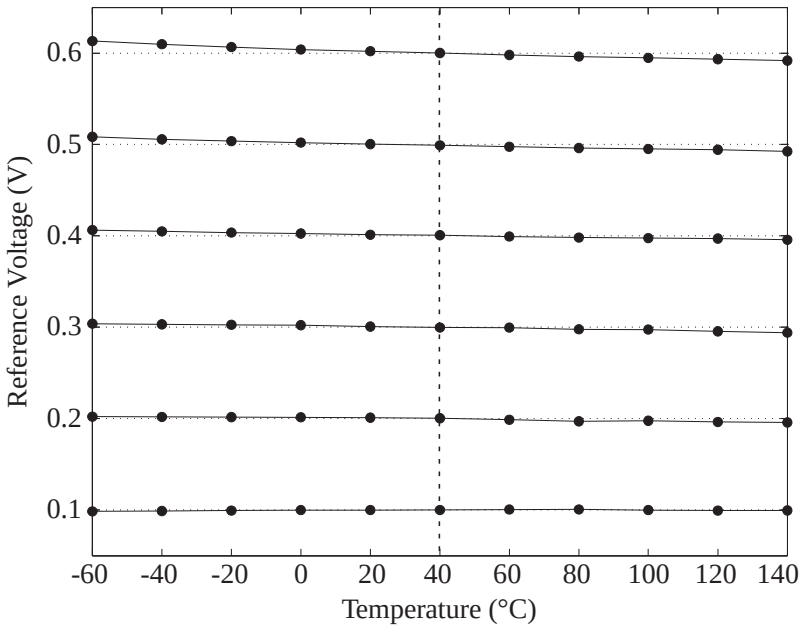


Figure 27. Voltage Reference #1 Accuracy: Data point and error plot for accurate programming of V_{ref}.

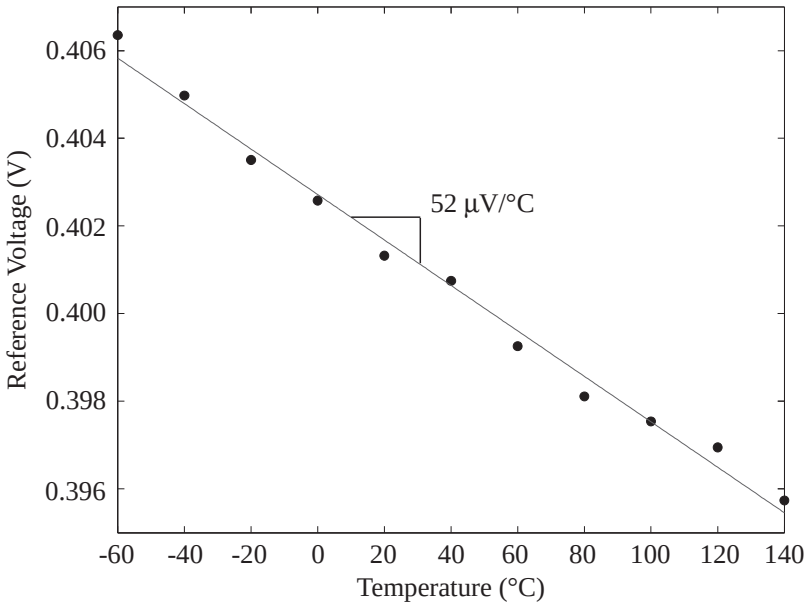
of the stresses induced by the packaging process and the package itself.

Figure 28(a) shows experimental results for the temperature dependence of the prototype chip. The reference voltage was programmed to five different values ranging from 100mV to 500mV at room temperature and measured across temperature for a range of -60^oC to 140^oC. Figure 28(b) provides a more detailed view of the temperature dependence of the reference voltage for $V_{ref} = 0.4V$. The reference voltage displays a linear temperature dependence of $52\mu V/^oC$ or $130ppm/^oC$. The strong linear dependance with temperature is mainly due to the body-effect of M_1 as it's bulk terminal was connected to V_{dd} in the prototype chip (see Section 4.3.1).

The temperature sensitivity of the reference voltage as a function of the reference voltage is shown in Figure 29. As expected from (50), the temperature sensitivity increases as a function of V_{ref} . A maximum TC of $183ppm/^oC$ was obtained for $V_{ref} = 0.6$V, while a minimum TC of $100ppm/^oC$ was obtained for $V_{ref} = 0.1$V. These results were corroborated by solving (50) via numerical analysis. As can be



(a)



(b)

Figure 28. Voltage Reference #1 Temperature Variation: (a) Voltage reference variation with temperature for different programmed V_{ref} values. (b) Voltage reference variation with temperature for $V_{ref} = 0.4V$.

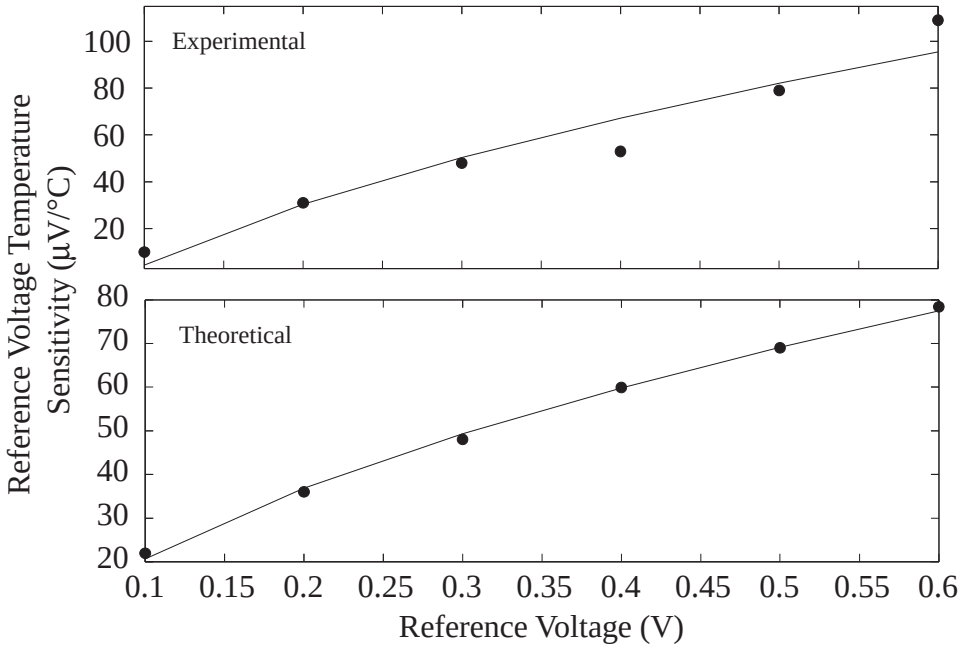
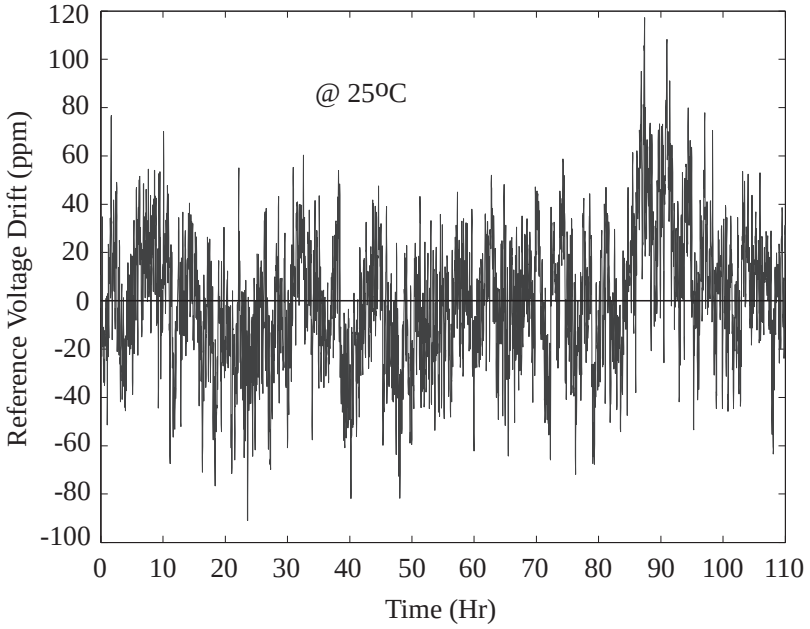


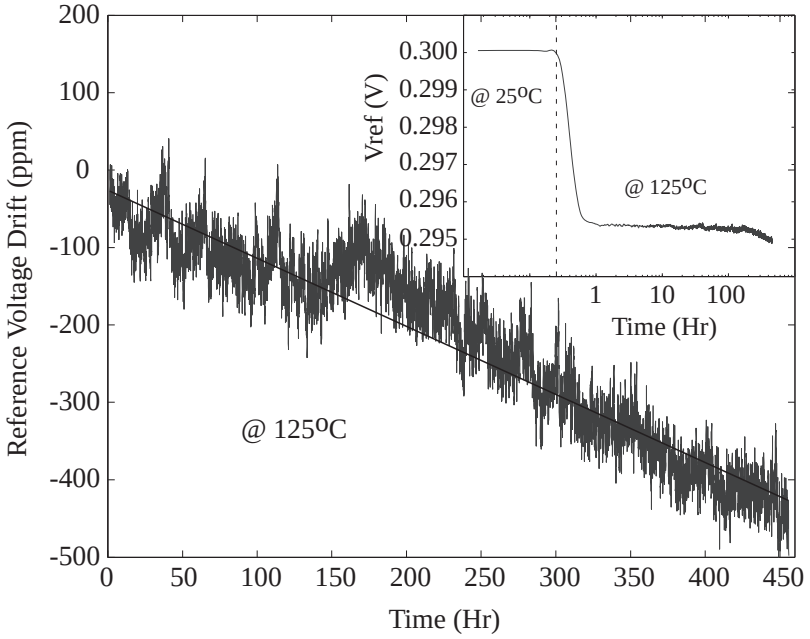
Figure 29. Voltage Reference #1 Temperature Sensitivity: Experimental and theoretical plots of the reference voltage sensitivity against their respective programmed V_{ref} values.

observed also in Figure 29, the measured data and theoretical predictions match closely. Both the temperature coefficient and its dependence on the reference voltage can be reduced by eliminating the body-effect in transistor M_1.

Figure 30(a) shows the reference voltage drift at room temperature for a period of approximately 100 hours. A negligible change in the reference voltage was obtained. Figure 30(b) shows V_{ref} at 125^oC for a period of approximately 450 hours; a net change of $400\mu V$ was recorded. The inset shows the same data on a log scale. A small jump of approximately $5mV$ occurs, as expected from Figure 29, due to the increase in temperature from 25^oC to 125^oC. It has been observed that the charge drift greatly reduces after a burn in period of around 1 day at temperatures above 300^oC.



(a)



(b)

Figure 30. Voltage Reference #1 Drift: (a) Measured reference voltage drift against time at 25^oC. (b) Measured reference voltage drift against time at 125^oC.

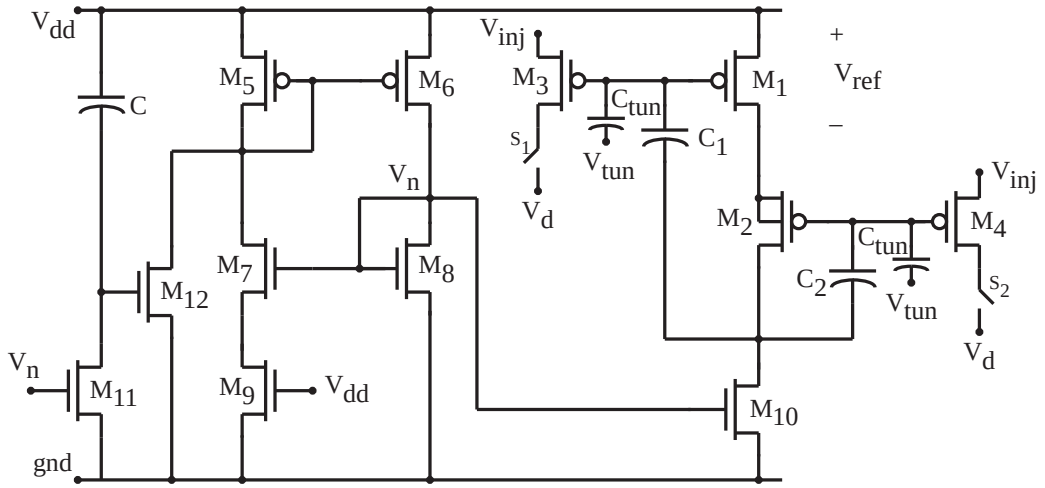


Figure 31. Proposed Voltage Reference #2: Simplified circuit schematic of the proposed programable voltage reference #2.

4.4 Reference Architecture #2

Figure 31 shows the schematic diagram of the proposed reference. Transistors M_1 and M_2 are chosen to be floating-gate transistors, with their input voltage coupled through capacitors C_1 and C_2, respectively. The charge stored on these transistors is represented by Q_1 and Q_2, whit a differential value of ΔQ_{2-1}. The reference voltage is taken as the difference between the effective source-gate voltages of M_1 and M_2. With M_1 and M_2 having identical aspect ratios, and $C_1 = C_2 = C$, the reference voltage reduces to

$$V_{out} = \frac{\Delta Q_{2-1}}{C_T}. \tag{52}$$

An arbitrary reference value can be obtained by proper programming of Q_1 and Q_2 as seen in (52).

Transistors $M_5 - M_9$ form a self-biased current reference circuit used to bias M_{10}. A startup circuit composed of $M_{11} - M_{12}$ and C is included. Transistors $M_3 - M_4$ and capacitors $C_{tun_1} - C_{tun_2}$ are used to modify the charge of their respective floating-nodes. The reference chip is set to the desired voltage at the factory through the test pins V_{inj}, V_d, and V_{tun}. Switches S_1 and S_2 allow for the selection of M_3 and

M_4 respectively. After the reference is programmed, M_3 and M_4 are turned off, thus avoiding additional power consumption.

4.4.1 Reference Temperature Sensitivity

Assuming the source and bulk terminals of M_1 are connected together, the temperature sensitivity of V_{ref} , obtained by differentiating (52) with temperature, can be written as

$$\frac{\delta V_{ref}}{\delta T} = \frac{\Delta Q_{2-1}}{C}\frac{1}{C}\frac{\delta C}{\delta T}. \tag{53}$$

Here, it is also assumed that $C = C_1 = C_2 = C_T$ and ΔQ_{2-1} displays zero temperature dependence. The temperature sensitivity of V_{ref} arises form the temperature dependance of C_1 and C_2. Typically, C_1 and C_2 are built as poly-poly capacitors and exhibit a TC of around $20ppm/^oC$ to $50ppm/^oC$.

4.4.2 Reference Minimum Power Supply

The principle of operation of the proposed voltage reference depends on M_1 and M_2 being operated in the saturation region. Low-voltage operation of the reference is achieved by lowering the effective threshold voltage of M_1 and M_2 through the programming of the common mode charge Q. The minimum supply voltage is given by

$$V_{dd_{min}} = V_{ref} + V_{dsat_2} + V_{dsat_{10}} \tag{54}$$

while the minimum output voltage is limited to V_{dsat_1}. Very low power operation is possible since M_1 and M_2 are biased with the same current, thus avoiding the high current values that would otherwise have been necessary for matching.

4.4.3 Experimental Results

Figure 32 shows the chip micrograph of the prototype reference fabricated in a $0.5\mu m$ CMOS process. The reference just occupies $0.0098mm^2$ of area, excluding buffers. The prototype circuit operates with a bias current of $1\mu A$ and consumes just $3.3\mu W$.

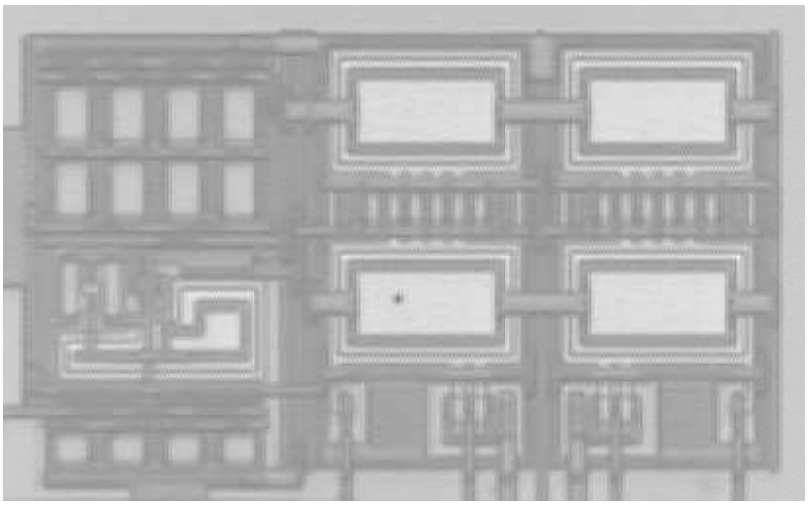


Figure 32. Voltage Reference #2 Die Micrograph: Chip micrograph of the prototype voltage reference in a $0.5\mu m$ CMOS process.

Figure 33(a) shows the temperature sensitivity of the proposed reference for a $V_{ref} = 1.5V$. The reference voltage displays a linear temperature dependence of $32\mu V/^oC$ or $22ppm/^oC$. The linear dependance with temperature is mainly due to the poly-poly capacitors C_1 and C_2 as discussed in Section 4.4.1. The temperature sensitivity of the reference voltage as a function of the reference voltage is shown in Figure 33(b). As expected from (53), the temperature sensitivity is fairly constant across the whole range. The proposed reference exhibits a maximum TC of $30ppm/^oC$.

4.5 Summary

Two simple compact programable voltage reference circuits have been presented. The voltage reference is obtained, in both cases, as the difference in charge between two floating-gate transistors. This technique allows for a programmable reference along with a low TC. The proposed reference architecture #1 exhibited temperature coefficients of $< 180ppm/^oC$ for a voltage range of $0.05V - 0.6V$. For the reference architecture #2, temperature coefficients of $< 30ppm/^oC$ were obtained for a voltage range of $0.9V - 2.7V$. An initial accuracy of $40\mu V$ was achieved for both architectures. A summary of the experimental results is shown in Table 6. Table 7 presents

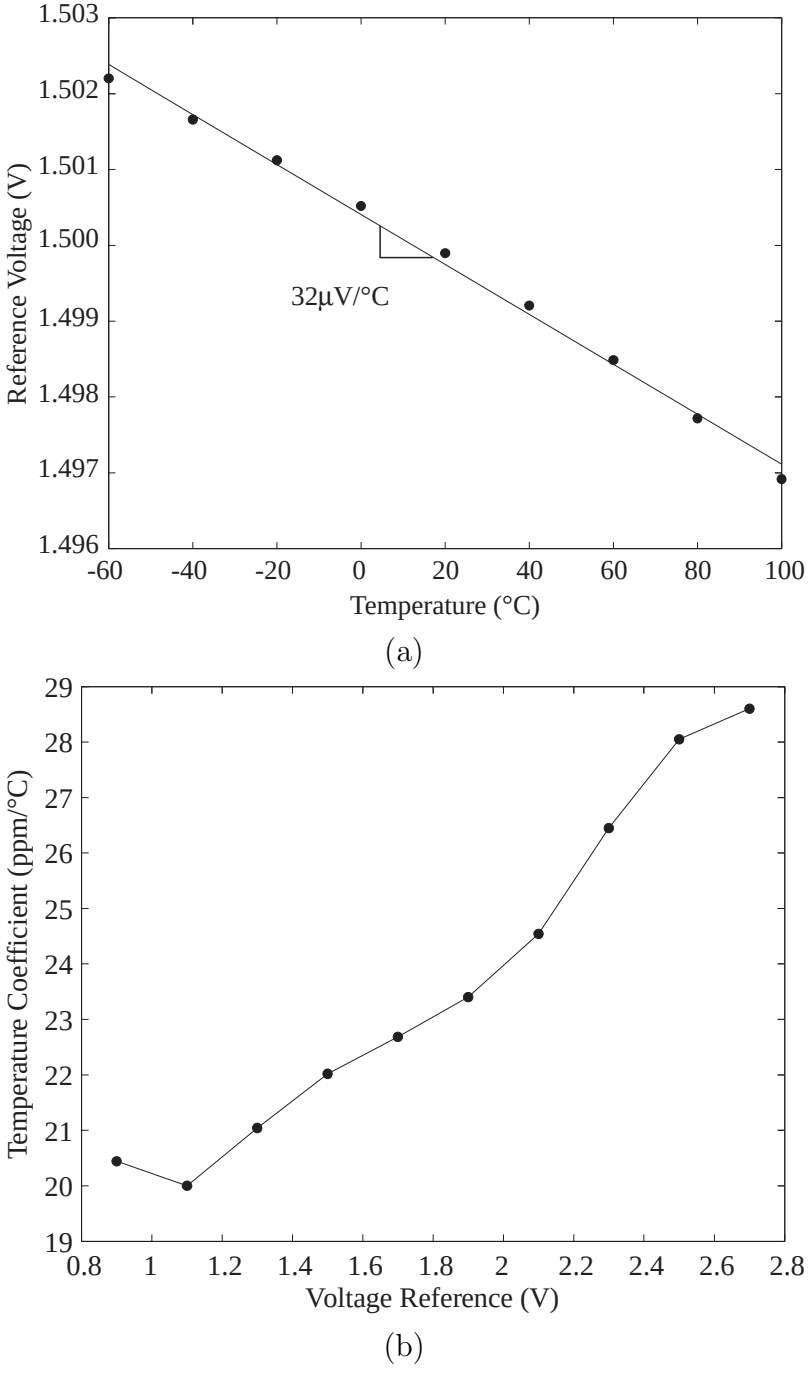


Figure 33. Voltage Reference #2 Temperature Sensitivity: (a) Voltage reference variation with temperature for $V_{ref} = 1.5V$. (b) Experimental plot of the reference voltage sensitivity against their respective programmed V_{ref} values.

Table 6. Summary of Voltage Reference Performance

Parameter	**Value - Ref #1**	**Value - Ref #2**
Supply Voltage	$2.5V$	$3.3V$
Technology	$0.35\mu m$ CMOS	$0.5\mu m$ CMOS
Voltage Range	$50mV - 600mV$	$0.9V - 2.7V$
Temperature Coefficient	$< 180ppm/^{\circ}C$	$< 30ppm/^{\circ}C$
Temperature Range	$-60\circ C - 140\circ C$	$-60\circ C - 100\circ C$
Initial Accuracy	$\pm 40\mu V$	$\pm 40\mu V$
Power Dissipation	$50\mu W$ ($V_{ref} = 0.5V$)	$3.3\mu W$
Area	$52\mu m \times 42\mu m$	$134\mu m \times 73\mu m$

the performance comparison of the proposed current reference with some of the proposed architectures in the literature. The main advantages of the proposed approach over these techniques are: 1) it allows for a very accurate reference without the use of additional trimming circuitry, 2) unlike any other schemes, the reference value is no dictated by device parameters, it can be programmed to any arbitrary value, 3) it exhibits a relative low TC for a wide range of values.

Table 7. Performance Comparison for Different Voltage References

Parameter	Unit	This Work	Leung [53]	Leung [54]	Buck [2]	Ahuja [8]
Technique		Floating-Gate	Weighted ΔV_{GS}	Bandgap	Bandgap	Floating-Gate
CMOS Technology	μm	0.5	0.6	0.6	0.6	1.5
Supply Voltage	V	3.3	1.4	0.98	3.7	$2.V$
V_{ref} **Range**	V	$0.9 - 2.7^a$	0.309	0.603	1.12	$0.5 - 5.0^a$
Temperature Coefficient	$ppm/^\circ C$	< 30	36.9^b	15^b	134^b	< 1
Temperature Range	$^\circ C$	-60 to 100	0 to 100	0 to 100	0 to 70	-40 to 85
Voltage Drift @ 10 years	ppm	400	–	–	–	24
Initial Accuracy	mV	± 0.04	± 19.26	–	± 0.50	± 0.20
Power Dissipation	μW	3.3	13.6	17.6	1400	1.3
Area	mm^2	0.0098	0.055	0.24	0.40	1.6

[a]The voltage reference can be programmed to any arbitrary value within the specified range.
[b]The temperature compensation specified is obtained with the help of trimming circuitry.

CHAPTER V

A PROGRAMMABLE LOW TC CMOS CURRENT REFERENCE

This chapter presents a programable temperature compensated CMOS current reference. The proposed circuit achieves a first order temperature compensation by canceling the negative TC of an on-chip poly resistor with the positive TC of a MOS transistor operating in the ohmic region. Programmability of the current reference is enabled with the use of floating-gate transistors, thus allowing arbitrary current values to be set accurately. The temperature compensation is independent of the reference value; a low TC reference is possible for a wide range of currents.

5.1 Previous Work

A current reference is an essential circuit on any analog and mixed signal system, as is used to establish the quiescent condition for many different circuits such as oscillators, amplifiers, and PLL's among others. Many circuit topologies have been proposed to reduce the temperature sensitivity [24, 28], improve the line regulation [55], and increase the precision [25, 65] of current references. Most of the published work has focused on minimizing their temperature dependence.

Some of the proposed architectures [24, 41, 55] use a variation of the bandgap voltage reference circuit to obtain a low TC current reference. These approaches take advantage of the opposite TC and the linear temperature dependence of ΔV_{be} and V_{be}. Others [25, 33, 38] exploit the temperature dependence of the MOS transistor parameters V_{th} and μ. A temperature coefficient of $4ppm/^oC$ was obtained in [44] with the use of a bipolar process. All CMOS current references [24, 65] have

reported experimental results in the range of $50ppm/^oC - 400ppm/^oC$ for first-order temperature compensation. With use of the second-order compensation techniques, [33, 38, 41] have shown that temperature coefficients in the $10's\ ppm/^oC$ are possible; no experimental data have been reported.

The use of programable transistors, when building a current source, has been shown only in [73,78]. In [78] temperature compensation is achieved by programming currents with opposite TC ; experimental results showed a 2% variation over a limited range of 45^oC to 75^oC. In [73] a programmable current source is introduced briefly without any temperature compensation.

5.2 A Programmable Current Reference

Figure 34(a) shows the circuit diagram of the proposed programmable current reference. The current reference consists of a programmable voltage reference V_{ref}, a resistor R, and an amplifier. Assuming the amplifier has infinite gain, the voltage across the resistor R will be forced to V_{ref}, resulting in

$$I_{ref} = \frac{V_{ref}}{R}. \tag{55}$$

The voltage reference circuit, encircled in Figure 34(a), is composed by M_1, M_2, C_{tun1}, and C_1. Assuming M_1 is off (all terminals grounded) and $C_1 >> C_{tun1}, C_{parasitic}$, the voltage reference will be given by

$$V_{ref} = V_{Q_1} = \frac{Q_1}{C_1} \tag{56}$$

were Q_1 is the charge stored on C_1, a poly-poly capacitor. Figure 34(b) shows the layout diagram of the voltage reference. The reference voltage is connected to the input transistor of the amplifier with a poly line; transistors M_1 and M_2 share the gate terminal. Inputs to this terminal are capacitively coupled through C_1 and C_{tun1}, thus creating a floating node (see Figure 34(a)). The voltage V_{ref} can be set arbitrarily by modifying Q_1 with M_1 [35] as seen in (56).

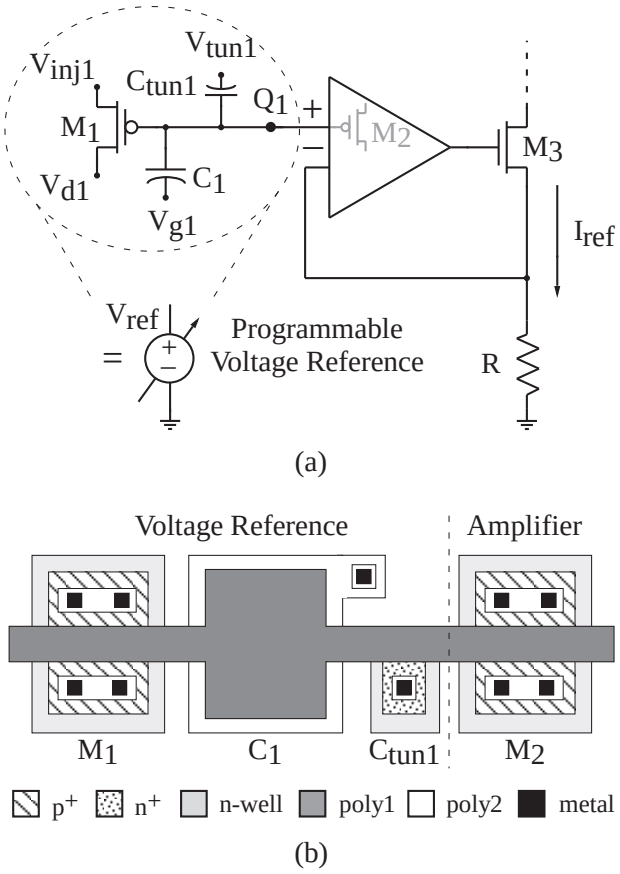


Figure 34. Programmable Current Reference: (a) Schematic diagram of the proposed programmable current reference. (b) Layout diagram of the programmable voltage reference composed by M_1, M_2, C_1, and C_{tun1}.

By substituting (56) in (55), the expression for I_{ref} can be rewritten as

$$I_{ref} = \frac{Q_1}{C_1}\frac{1}{R}. \tag{57}$$

An arbitrary I_{ref} value can be obtained after fabrication by modifying Q_1. Programmability of the current reference allows for compensation of parameter variability as the absolute value of R could vary as much as 30%.

The temperature coefficient of I_{ref} can be obtained as

$$TC_{I_{ref}} = \frac{1}{I_{ref}}\frac{\delta I_{ref}}{\delta T} = -\frac{1}{R}\frac{\delta R}{\delta T} - \frac{1}{C_1}\frac{\delta C_1}{\delta T} \approx -\frac{1}{R}\frac{\delta R}{\delta T} \tag{58}$$

were $\frac{1}{R}\frac{\delta R}{\delta T}$ and $\frac{1}{C_1}\frac{\delta C_1}{\delta T}$ are the temperature coefficients of R and C_1 respectively. The

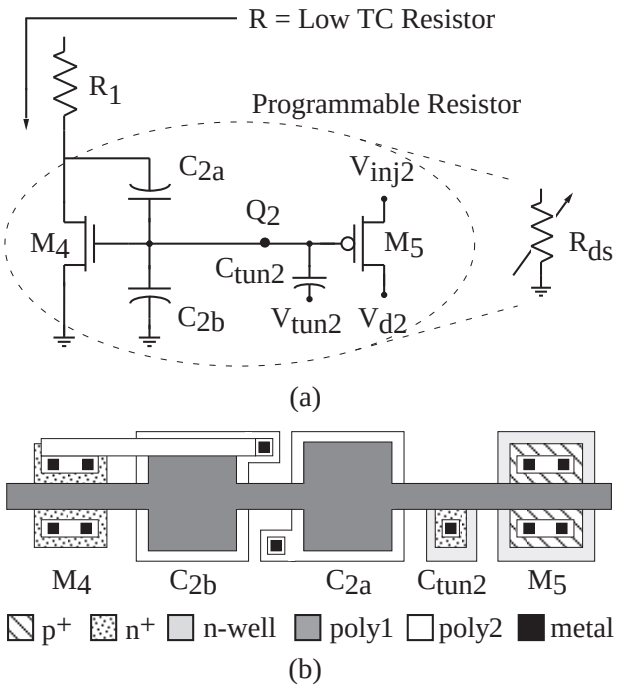


Figure 35. Temperature Compensated Resistor: (a) Circuit schematic of the proposed temperature compensated resistor. (b) Layout diagram of the ohmic resistor composed by M_4, M_5, C_{tun2}, C_{2_a}, and C_{2_b}.

temperature dependence of I_{ref} will be dictated by R; temperature coefficient for poly-poly capacitors range from $20ppm/^oC - 50ppm/^oC$, thus is assumed to be negligible. The floating-gate charge Q_1 does not exhibit any temperature variations. A low TC current reference can be obtained with a low TC resistor.

5.3 A Temperature Compensated Resistor

Figure 35(a) shows the schematic diagram of the proposed resistor R. The resistor is a series combination of R_1, a high poly resistor, and R_{ds}, a MOS transistor (M_4) operating in the ohmic region. Resistance characteristics and temperature behavior of the ohmic resistor are examined next, followed by a detailed discussion of the proposed low TC resistor.

5.3.1 A Programmable Resistor

The ohmic resistor circuit is composed of M_4, M_5, C_{tun2}, C_{2_a}, and C_{2_b}, as shown encircled in Figure 35(a). Transistor M_4, along with capacitors C_{2_a} and C_{2_b}, form a linearized resistor [62]. Figure 35(b) shows the layout diagram of the ohmic resistor. The gate terminals of M_4 and M_5 share a poly1 connection; inputs to this terminal are capacitively coupled through C_{2a}, C_{2b}, and C_{tun2}, thus creating a floating node (see Figure 35(a)). Charge on this floating node can be set arbitrarily by modifying Q_2 via M_5 [35].

Assuming there is a charge Q_2 stored in the floating node, M_4 operates in the ohmic region[1], and M_5 is off (all terminals grounded), the ohmic resistance R_{ds} can be approximated as

$$R_{ds} \approx \frac{1}{\mu_n C_{ox} \frac{W}{L} \left(V_{Q_2} - V_{th_n}\right)} \approx \frac{1}{K_n \left(V_{Q_2} - V_{th_n}\right)} \tag{59}$$

where μ_n is the mobility of charge carriers, C_{ox} is the oxide capacitance, W and L are M_4 dimensions, $V_{Q_2} = \frac{Q_2}{C_2}$ is the voltage due to Q_2, V_{th_n} is the threshold voltage, and $K_n = \frac{W}{L}\mu_n C_{ox}$. It can be seen from (59) that R_{ds} can be modified with V_{Q_2} to any arbitrary value, after fabrication.

The temperature sensitivity $\frac{\delta R_{ds}}{\delta T}$ and the first-order temperature coefficient $TC_{R_{ds}}$ of R_{ds} can be shown to be

$$\frac{\delta R_{ds}}{\delta T} = \left[-\frac{1}{\mu_n}\frac{\delta \mu_n}{\delta T} + \frac{1}{V_{Q_2} - V_{th_n}}\frac{\delta V_{th_n}}{\delta T}\right] \cdot \left[\frac{1}{K_n \left(V_{Q_2} - V_{th_n}\right)}\right] \tag{60}$$

and

$$TC_{R_{ds}} = \frac{1}{R_{ds}}\frac{\delta R_{ds}}{\delta T} = -\frac{1}{\mu_n}\frac{\delta \mu_n}{\delta T} + \frac{1}{V_{Q_2} - V_{th_n}}\frac{\delta V_{th_n}}{\delta T} = \frac{n}{T} - \frac{\alpha}{V_{Q_2} - V_{th_n}} \tag{61}$$

respectively, were T is the temperature, n is the mobility temperature coefficient, and α is the threshold voltage temperature sensitivity. The temperature behavior of

[1]The equations derived in this Section assume that M_4 operates in the strong inversion region; a similar analysis can be done for M_4 operating in the weak inversion region.

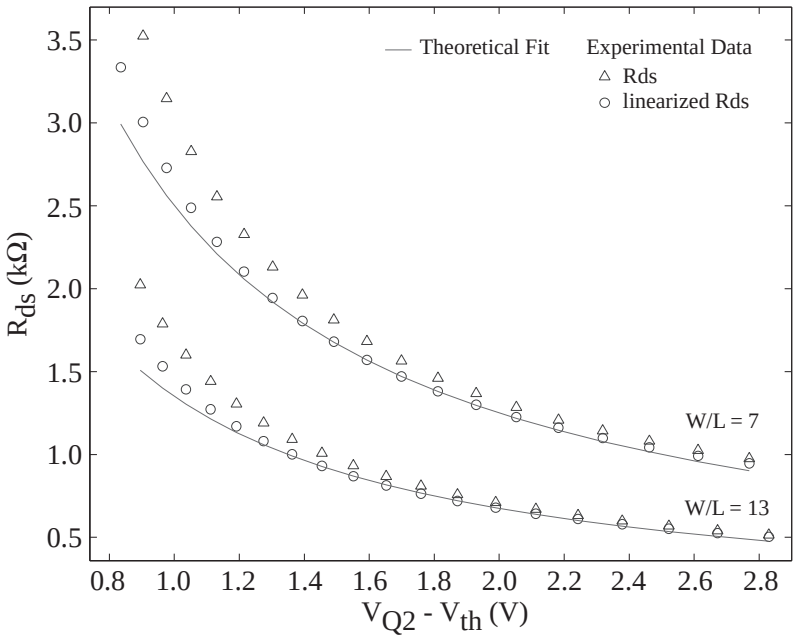
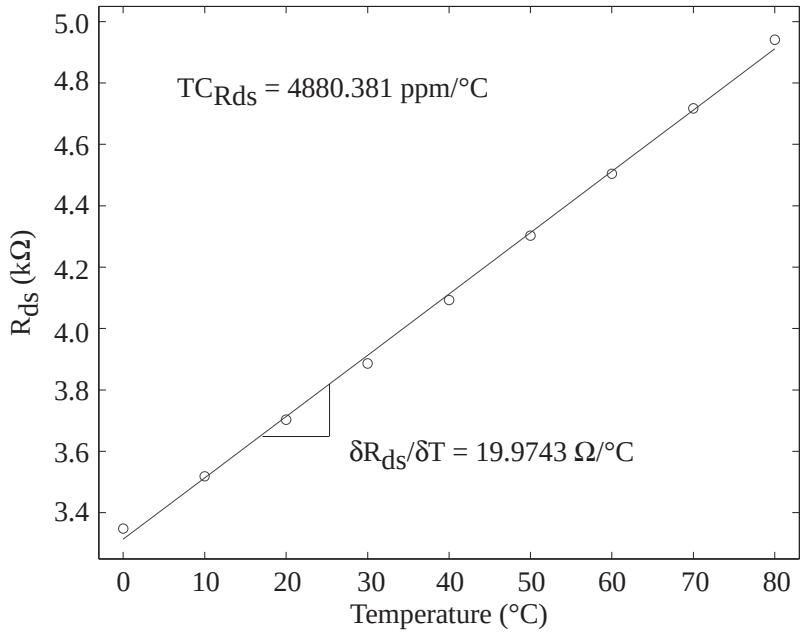


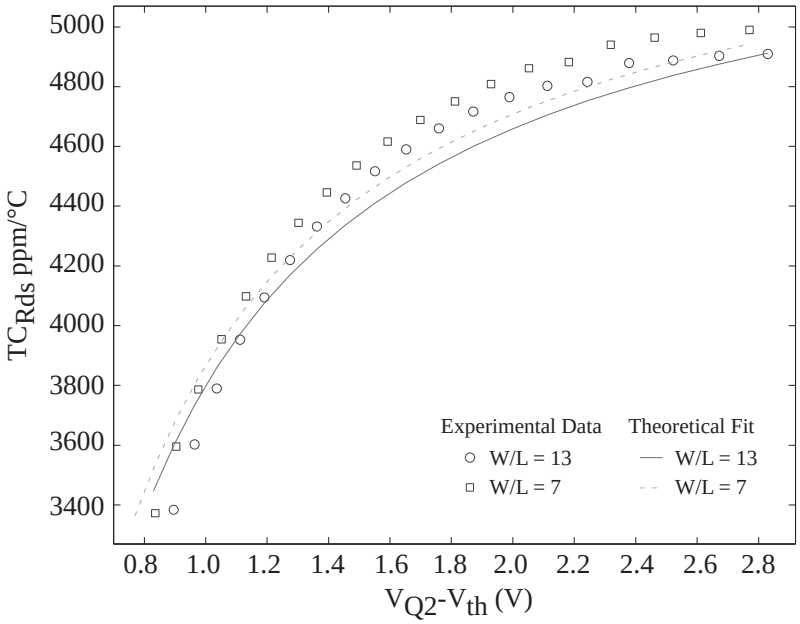
Figure 36. Ohmic Resistor Programming: Plot of R_{ds} for different $V_{Q_2} - V_{th_n}$ values.

R_{ds} can be modified with V_{Q_2} as seen in (60) and (61). For large enough V_{Q_2} values $\left(V_{Q_2} > \frac{T \cdot \alpha}{n} + V_{th_n}\right)$ a positive $TC_{R_{ds}}$ is obtained.

Figure 36 shows experimental data, along with a theoretical fit, of R_{ds} for different $V_{Q_2} - V_{th_n}$ values. As expected, the linearized version of R_{ds} [62] follows closely the behavior predicted by (59). Figure 37(a) shows the temperature behavior of R_{ds} over a temperature range of -60^oC to 140^oC. The ohmic resistor exhibits a strong linear dependence with temperature; higher-order temperature effects are due to mobility. A temperature coefficient of $+4880ppm/^oC$ was obtained for a $V_{Q_2} - V_{th_n}$ value of $1.8V$. Values of -1.65 and $-1.6mV/^oC$ were extracted for device parameters n and α respectively. The temperature coefficient of R_{ds} for different $V_{Q_2} - V_{th_n}$ values is shown in Figure 37(b). The experimental data follows closely the theoretical behavior predicted by (61). A small difference between the temperature coefficient behavior of different sized R_{ds} arise from device parameter mismatch.



(a)



(b)

Figure 37. Ohmic Resistor Temperature Behavior: (a) Plot of R_{ds} for a temperature range of -60^oC to 140^oC. (b) Plot of R_{ds} temperature coefficient for different $V_{Q_2} - V_{th_n}$ values.

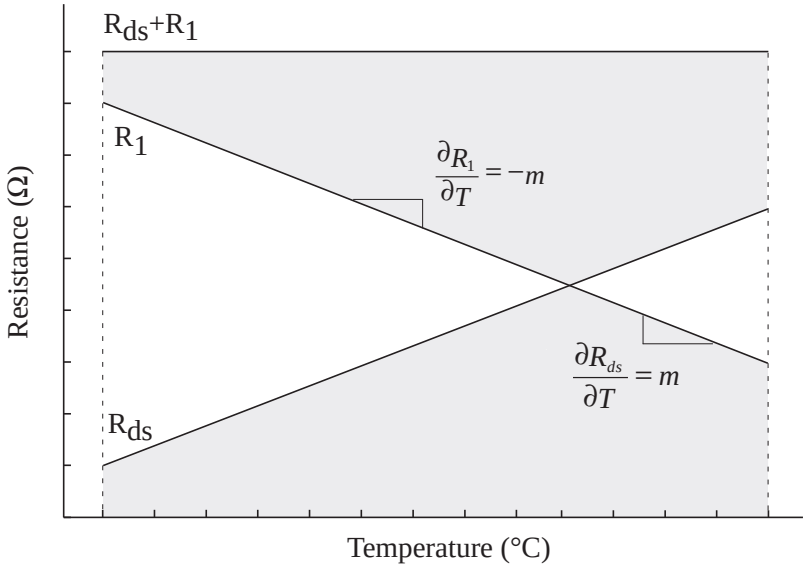


Figure 38. Temperature Sensitivity Cancelation: Graphical representation of the linear cancelation of the resistor temperature sensitivity.

5.3.2 A Low TC Resistor

Using (59), R can be written as

$$R = R_1 + R_{ds} = R_1 + \frac{1}{K_n \left(V_{Q_2} - V_{th_n}\right)} \tag{62}$$

where all the variables have their usual meaning. A first-order temperature variation of R, obtained by differentiating (62) against temperature, is given by

$$\frac{\delta R}{\delta T} = \frac{\delta R_1}{\delta T} + \frac{\delta R_{ds}}{\delta T} = R_1 \cdot TC_{R_1} + R_{ds} \cdot TC_{R_{ds}} \tag{63}$$

were $TC_{R_1} = \frac{1}{R_1}\frac{\delta R_1}{\delta T}$ is the temperature coefficient of R_1. Temperature sensitivity cancelation $\left(\frac{\delta R}{\delta T} = 0\right)$ can be achieved by satisfying

$$\frac{\delta R_1}{\delta T} = -\frac{\delta R_{ds}}{\delta T} \tag{64}$$

or

$$\frac{R_1}{R_{ds}} = -\frac{TC_{R_{ds}}}{TC_{R_1}}. \tag{65}$$

Temperature sensitivity cancelation is possible for resistors with opposite temperature behavior as seen in (64). Figure 38 shows a graphical representation of the proposed

approach. Linear cancelation of the positive temperature sensitivity of R_{ds} is possible with a resistor with negative temperature sensitivity.

Substituting (61) into (65), the TC cancelation can be achieved by properly sizing R_1 and M_4 according to

$$R_1 K_n \left(V_{Q_2} - V_{th_n}\right) = \left[\frac{n}{T} - \frac{\alpha}{V_{Q_2} - V_{th_n}}\right] \cdot \left[\frac{1}{TC_{R_1}}\right]. \tag{66}$$

Although this cancelation is totally dependent on device parameters, optimal TC cancelation can be obtained by modifying V_{Q_2} as seen in (66).

5.4 Proposed Current Reference

Figure 39 shows a detailed circuit diagram of the proposed current reference along with a pictorial representation of the temperature behavior of the different components. A temperature insensitive programmable current reference is obtained by combining the programmable current reference circuit presented in Section 5.2 with the temperature-compensated resistor circuit presented in Section 5.3.

The analytical expression for I_{ref}, obtained by substituting (62) in (57), is given as

$$I_{ref} = \frac{V_{Q_1}}{R} = V_{Q_1} \cdot \left[\frac{K_n(V_{Q_2} - V_{th_n})}{R_1 K_n(V_{Q_2} - V_{th_n}) + 1}\right]. \tag{67}$$

The temperature dependence of I_{ref} will depend directly on R as shown in (58). Modification of V_{Q_2} allows for optimal TC cancelation of R, as discussed in Section 5.3.2, while modification of V_{Q_1} allows for precise programming of I_{ref} to any arbitrary value. In contrast to other approaches [24, 25, 28, 33, 38, 41, 55], the TC cancelation is independent of the I_{ref} due to ${I_{ref}}'s$ direct proportionality to V_{Q_1}.

5.5 Charge Modification

On-chip programming of Q_1 and Q_2 was enabled with the use of a constant charge injection circuit, a high voltage charge pump, and a negative voltage charge pump. Figure 40 shows the circuit used to program the voltage reference part (Q_1) of the

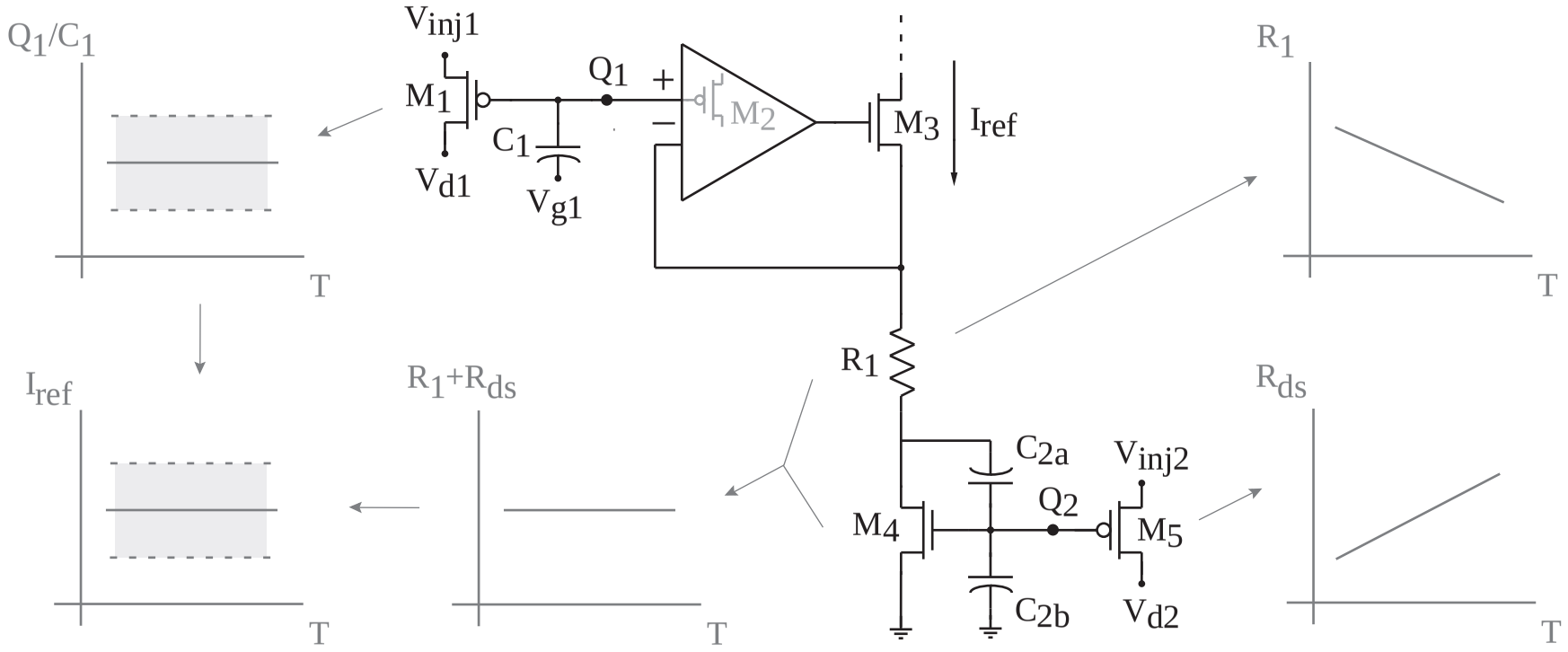


Figure 39. Low TC Programmable Current Reference: Schematic diagram of the proposed current reference along with a graphical representation of the temperature behavior of the different components.

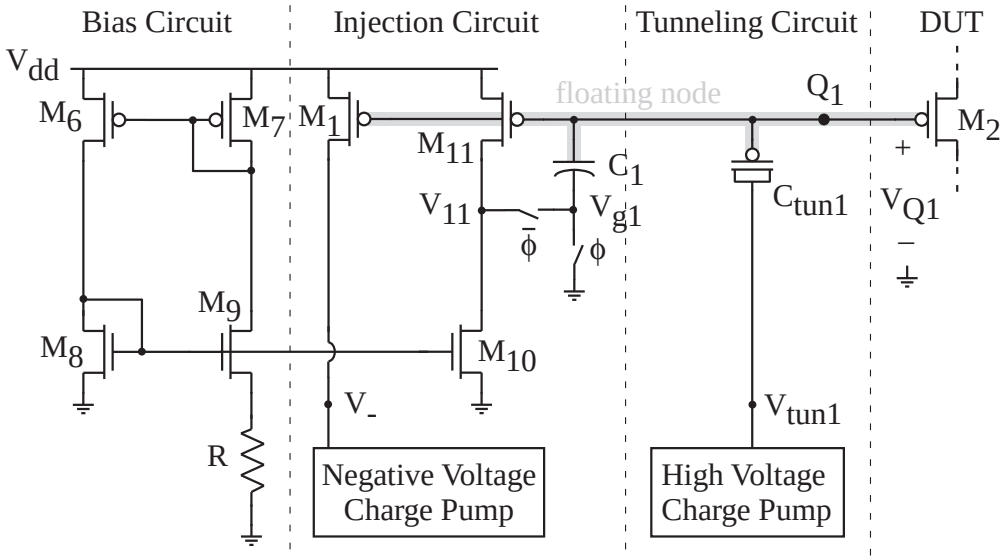


Figure 40. Constant Charge Injection Circuit: Schematic diagram of the circuit used to modify the charge Q_1 (see Figure 39) of the proposed current reference.

proposed reference. Transistors M_1 and M_2, and capacitor C_1 represent the same devices shown previously in Figure 39. The additional transistor M_{11}, connected to the floating node, is used for constant charge injection. Transistors $M_6 - M_{10}$, along with resistor R, form a bootstrap current source that bias M_{11}. A bias current of $1\mu A$ was used in this design, thus burning only an additional $3\mu A$ of current. An identical approach is used to program the charge Q_2 of the proposed resistor for temperature compensation.

During normal operation, $\phi = V_{dd}$, charge pumps are turned off, and V_{tun1} and V_{-} are set to gnd and V_{dd}, respectively. This ensures there is no coupling though C_{tun1} and M_1 is turned off. Transistor M_{11} will be on; it's region of operation will depend on the charge Q_1 available on the floating node. The value of the floating node voltage V_{Q_1} will be given by (56).

During programming, $\phi = gnd$, a feedback loop is established by the diode connected transistor M_{11}. The voltage V_{11} will ensure that the current set by M_{10} flows through M_{11}, independently of Q_1. This results in a constant current through M_1 as it will mirror the current of M_{11} (see Figure 40). For injection, a negative voltage

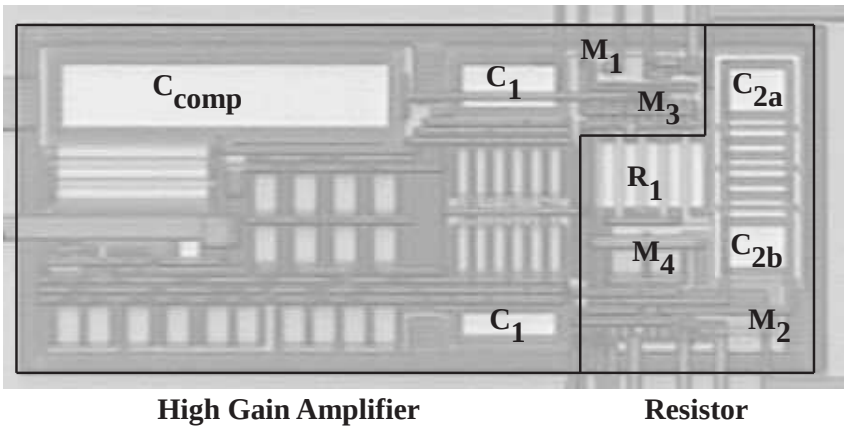


Figure 41. Current Reference Prototype Circuit: Chip micrograph of the prototype current reference in a $0.5\mu m$ CMOS process.

pulse is applied to the drain terminal of M_1 with the use of a negative charge pump. A constant charge modification will occur when injecting due to the fixed current through M_1. The change in charge will be a function of the bias current of M_1, the drain-source voltage applied to M_1, and the duration of the pulse. For tunneling, a high voltage pulse is applied to C_{tun1} with the use of a high voltage charge pump.

5.6 Experimental Results

A prototype chip was fabricated in $0.5\mu m$ CMOS process. A folded cascode topology was used to implement the high gain amplifier. The power consumption of the amplifier along with the bias circuitry was just $21\mu W$ at a V_{dd} of $3.3V$. Figure 41 shows the die micrograph of the prototype integrated circuit (charge pumps not included); the total area of the current reference is just $200\mu m$ x $75\mu m$. The charge pumps and the programming circuit occupy an additional area of $132\mu m$ x $342\mu m$.

Figure 42 shows an error plot of different programmed current reference values, from $200nA$ to $100\mu A$. A programming accuracy of $< 0.02\%$ was obtained for currents $> 5\mu A$. A degradation in accuracy at the lower currents occurred due to resolution limitations; the measurement equipment was set to a fix range of $200\mu A$ for the complete measurement.

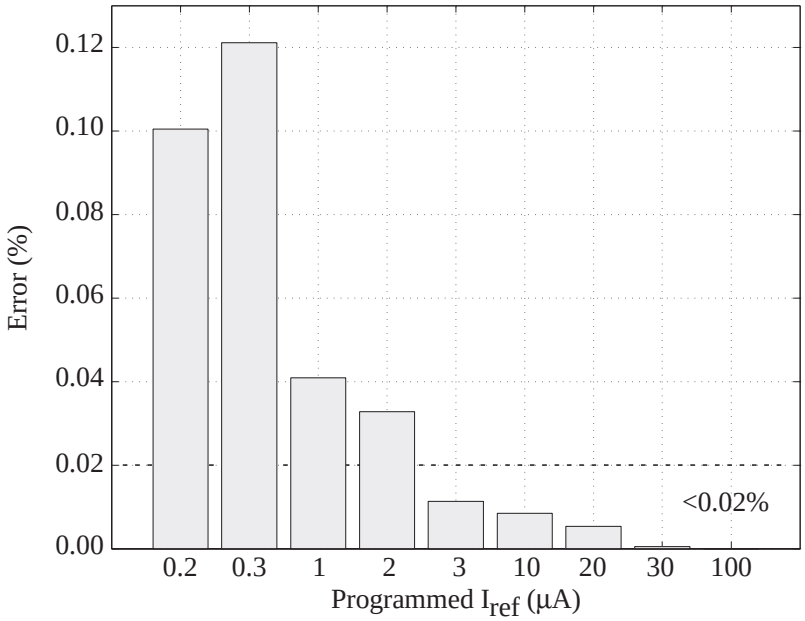


Figure 42. Current Reference Precision: Percentage error of several programmed I_{ref} values ranging from $200nA$ to $100\mu A$.

Measurements were conducted to characterize run-specific device parameters. Experimental results showed R_1 and TC_{R_1} to be $12.1k\Omega$ and $-1750ppm/^oC$ respectively, which results in $\frac{\delta R_1}{\delta T} = -21.2\Omega/^oC$. Optimal TC compensation was carried by measuring the temperature sensitivity of R_{ds} for different programmed values of V_{Q_2} as shown in Figure 43(a). The temperature sensitivity $\frac{\delta R_{ds}}{\delta T}$ was found to decrease with increasing $V_{Q_2} - V_{th_n}$, as expected from (60). An optimal V_{Q_2} of $2.51V$ was extracted at a temperature of 40^oC, which corresponds to an R_{ds} of $4.2k\Omega$.

Figure 43(b) shows the temperature sensitivity of the proposed current reference programmed at the optimal point. The parabolic shape of the curve, confirms the first-order TC cancelation; a temperature coefficient of $116ppm/^oC$ was obtained for a $40.78\mu A$ reference. Although higher-order temperature effects were expected due to the transistor mobility, it was found that the poly resistor introduced additional second order terms. Simulations predict a temperature coefficient of only $50ppm/^oC$ for a linear temperature dependent resistor.

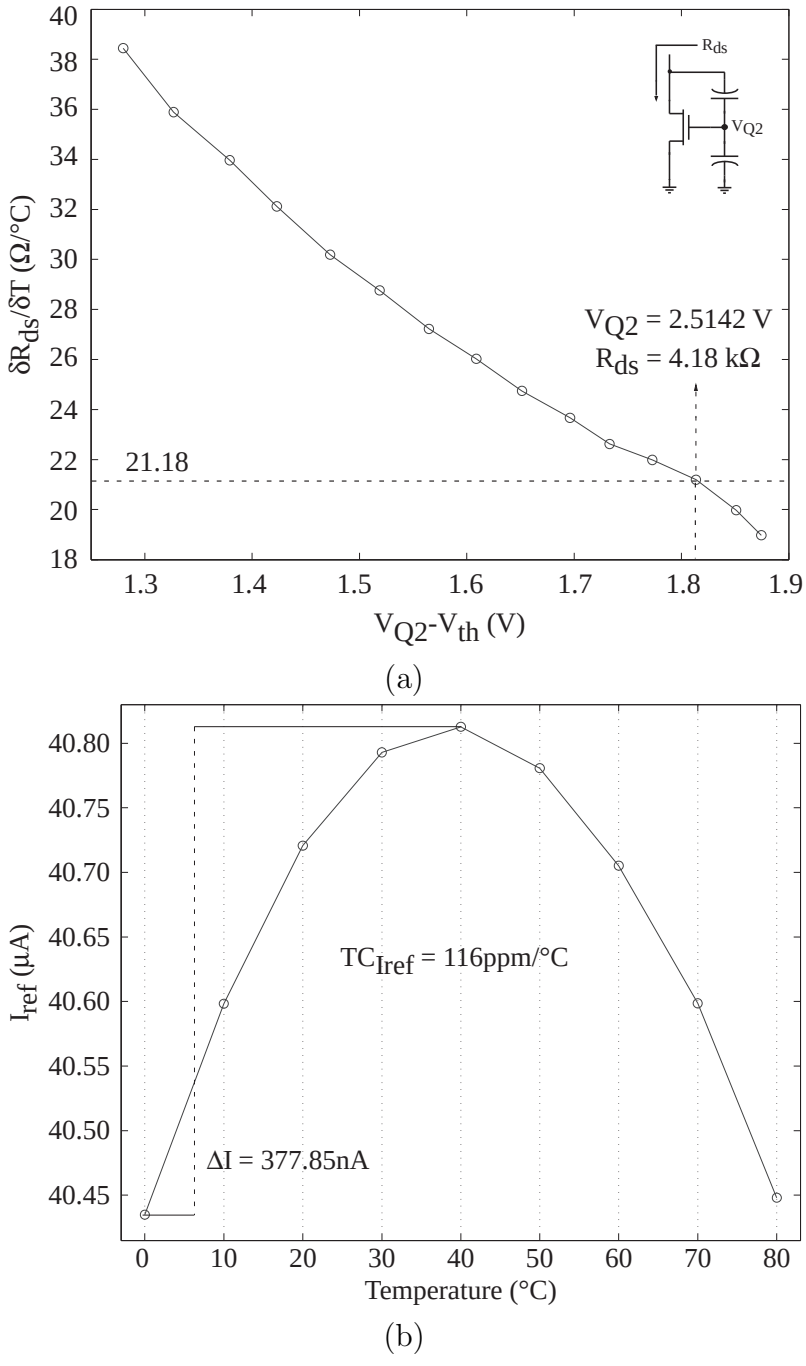
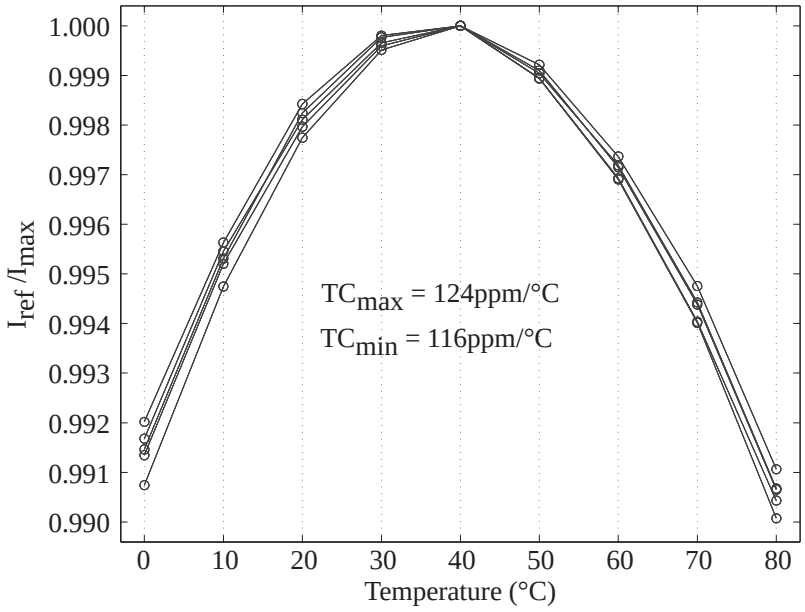


Figure 43. Current Reference TC Cancelation: (a) Temperature sensitivity of R_{ds} as a function of $V_{Q_2} - V_{th}$. (b) Plot of I_{ref} against temperature for of optimal TC cancelation.

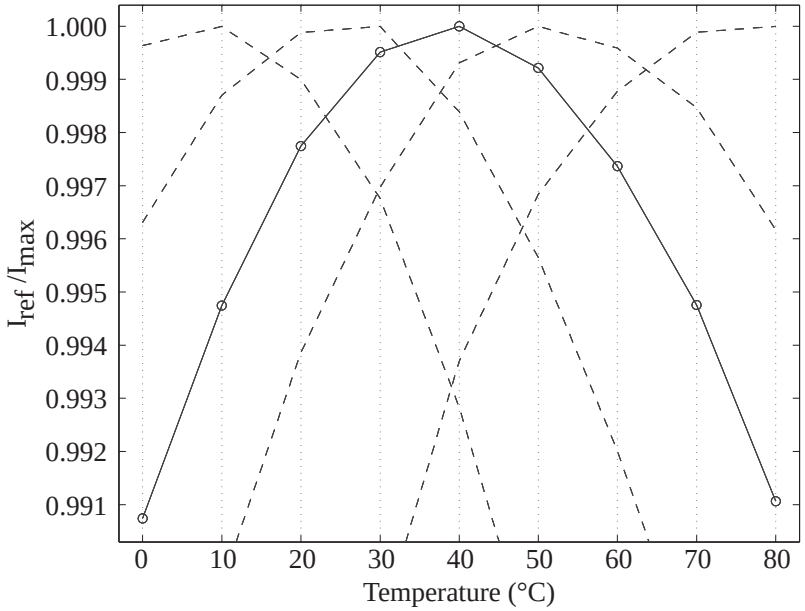
Figure 44(a) shows the current reference temperature behavior for five different prototypes. All five chips were programmed using the optimal point extrapolated from the first device. A maximum temperature coefficient of $124 ppm/^oC$ was obtained. Results indicated good temperature coefficient matching among chips. The direct influence of V_{Q_2} on the temperature sensitivity of the current reference can be observed in Figure 44(b), were the normalized temperature sensitivity of a single prototype is plotted for different V_{Q_2} values.

Characterization of the prototype over a wide range of currents was enabled by programming V_{Q_1} accordingly. Temperature sensitivities for current references ranging from $5\mu A$ to $53\mu A$ are shown in Figure 45(a). A maximum TC of $132 ppm/^oC$ was measured for a current range of $16\mu A$ to $53\mu A$ as seen in Figure 45(b). Degradation of the temperature coefficient at currents $< 16\mu A$ may be caused by the temperature dependence of the amplifier offset voltage. At this lower currents the offset voltage is no longer negligible since the reference voltage is $< 250mV$.

Figure 46(a) shows the line regulation for a current reference of $29.5\mu A$. A line regulation of $< 0.7\%/V$ was obtained for a supply voltage of 2.3 to $3.3V$. The reference exhibit a maximum line regulation of $1\%/V$ for a current range of $5\mu A$ to $53\mu A$ as shown in Figure 46(b).

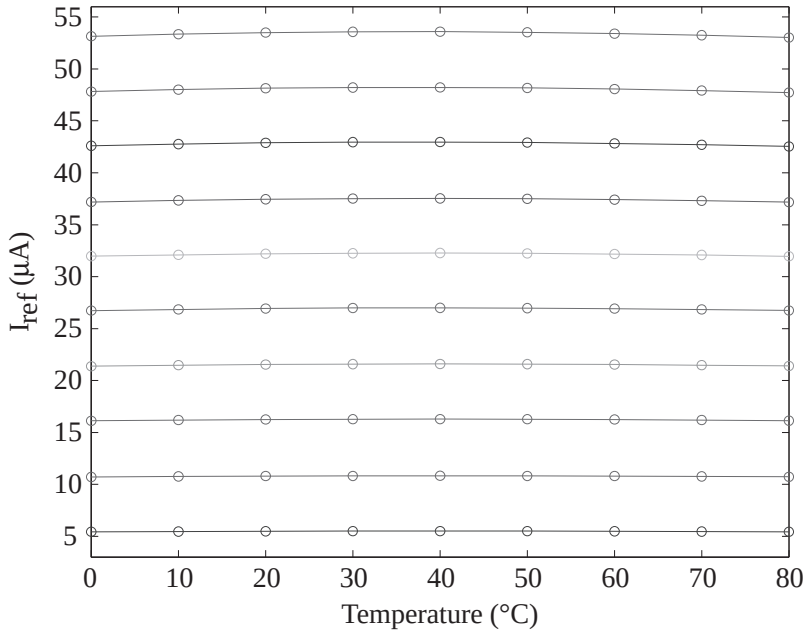


(a)

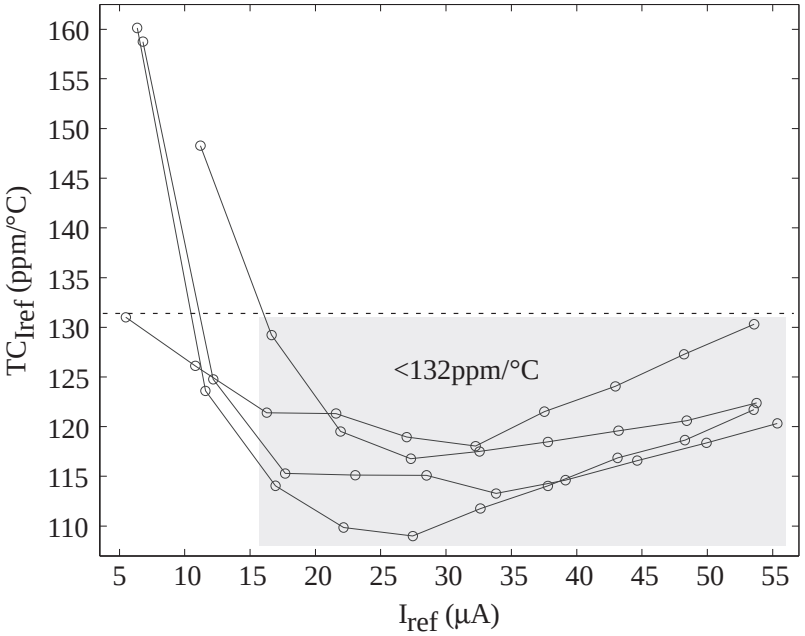


(b)

Figure 44. Current Reference Temperature Sensitivity: (a) Plot of the normalized current reference $\frac{I_{ref}}{I_{max}}$ against temperature for 5 different prototypes. (b) Plot of I_{ref} against temperature for different V_{Q_2} values.



(a)



(b)

Figure 45. Current Reference Temperature Coefficient: (a) Plot of I_{ref} against temperature for different programmed values. (b) Plot of the temperature coefficient obtained for different programmed I_{ref} values from 4 different prototypes.

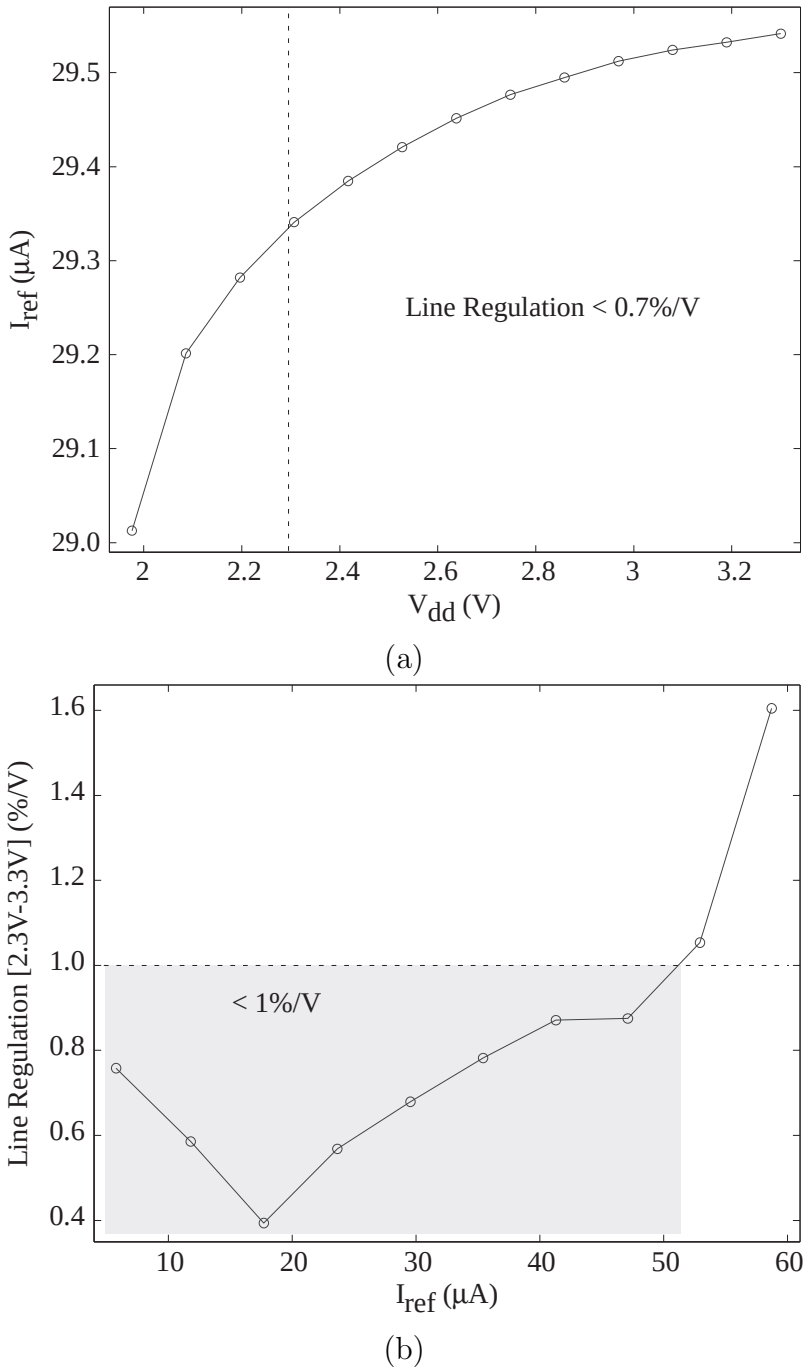


Figure 46. Current Reference Supply Sensitivity: (a) Plot of I_{ref} against power supply variation. (b) Power supply sensitivity for different programmed I_{ref} values.

Table 8. Summary of Current Reference Performance

Parameter	Experimental Results
I_{ref} Range	$16\mu A - 50\mu A$
I_{ref} Accuracy	$< 0.02\%$
Temperature Coefficient	$< 130ppm/^oC$
Temperature Range	$0^oC - 80^oC$
Line Regulation	$< 1\%/V$
Power Supply	$2.3V - 3.3V$
Additional Power Consumption	$21\mu W(V_{dd} = 3.3)$
Area	$200\mu m$ x $75\mu m$
Technology	$0.5\mu m$ $CMOS$

5.7 Summary

A programmable current reference based on a low TC resistor has been presented. This reference achieves first-order TC compensation by canceling the negative TC of an on-chip resistor with the positive TC of a transistor operating in the ohmic region. The proposed approach is robust against device parameter variation since the temperature coefficient can be set the optimal value through charge modification after fabrication. Temperature coefficients of $< 130ppm/^oC$ were obtained for a current range of $16\mu A - 50\mu A$ with a precision of $< 0.02\%$. A summary of the experimental results is shown in Table 8. Table 9 presents the performance comparison of the proposed current reference with some of the proposed architectures in the literature. The main advantages of the proposed approach over these techniques are: 1) it allows for a very accurate reference without the use of additional trimming circuitry, 2) unlike any other schemes, the reference value is no dictated by device parameters, it can be programmed to any arbitrary value, 3) it exhibit a relative low TC for a wide range of values.

Table 9. Performance Comparison for Different CMOS Current References

	Unit	This Work	Sansen [65]	Chen [24]	Bendali [15]	De Vita [81]
Reference Current	μA	$16 - 50$ [a]	0.774	526	144	0.009
Reference Accuracy	%	< 0.02	2.5	–	7	6.5
Temperature Coefficient	$ppm/^oC$	< 130	375	50	185	44
Temperature Range	^oC	$0 - 80$	$0 - 80$	$0 - 110$	$0 - 100$	$0 - 80$
Min. Power Supply	V	2.3 [b]	3.5	1	1	1.5
Supply Regulation	$\%/V$	< 1	0.015	0.22	–	0.05
CMOS Technology	μm	0.5	3.0	0.18	0.18	0.35

[a] The reference can set to any arbitrary current value within this range while preserving the performance reported.
[b] Minimum power supply that meets the specified performance.

CHAPTER VI

A V_{TH} COMPENSATED DIGITAL-TO-ANALOG CONVERTER

This chapter presents a V_{th} compensated DAC. The proposed converter uses floating gate transistors to compensate for the intrinsic V_{th} mismatch of MOS transistors. This approach enables higher accuracy along with a substantial decrease of the die size.

6.1 Previous Work

Among several technology and architecture alternatives, CMOS current-steering DAC architectures are commonly used in applications such as video signal processing, digital signal synthesists, and wireless communications [40]. The evident cost and power consumption advantages in the integration with digital circuits, their inherent high speed, and their load driving capabilities make these architecture the preferred choice among designers.

The intrinsic accuracy of a current-steering DAC is dictated by device mismatch. A common design approach to improve the static linearity is the use of large devices along with layout techniques to compensate for gradient effects [13,27]. This approach results in an increase in die area and parasitics, thus affecting the dynamic performance of the converter. An improvement in accuracy can be accomplished by using special techniques such as laser or fuse trimming [56, 77], but this additional step is expensive. Others solution include self-calibration circuits [37] that perform the trimming during power-up or in fixed time intervals, and dynamical element matching (DEM) [58]. All these techniques improve linearity, but at the expense of die

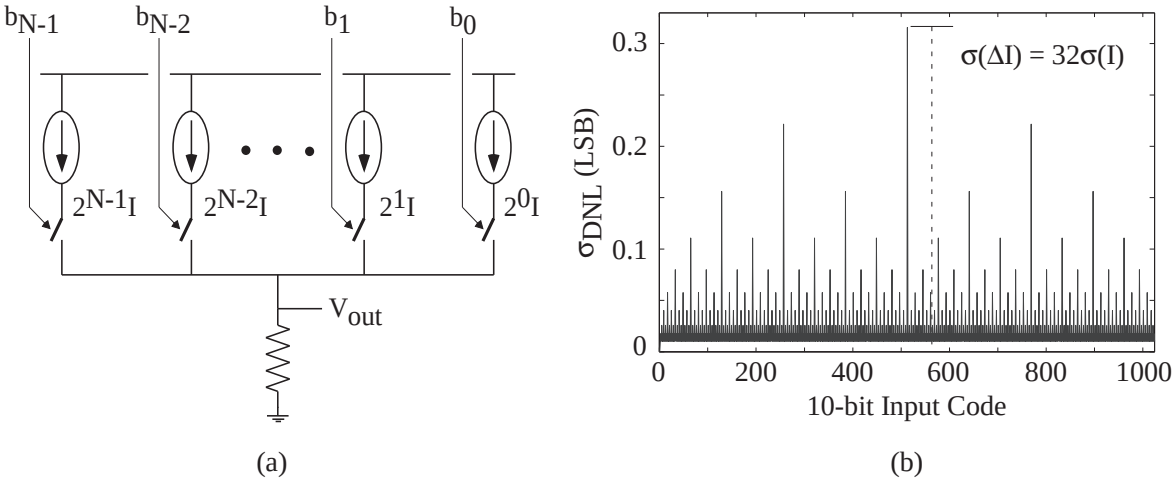


Figure 47. Binary Architecture: (a) Schematic diagram of a binary weighted DAC. (b) Monte-carlo simulation results of the DNL.

area, power dissipation, and/or dynamic performance.

6.2 DAC Architectures

Current-steering DACs are based on an array of matched current sources which are unity or binary weighted. The static and dynamic performance of the converter depends on the converter architecture. The differential non-linearity (DNL) and the dynamic behavior of a current-steering DAC are directly linked to the architecture while the integral non-linearity (INL) is architecture independent. Architectures variants, such as the binary, the unary, and the segmented architectures, are often used.

6.2.1 Binary Weighted DAC

As the name says, the binary architecture consists of an array of binary weighted current sources. An schematic diagram of such converter is shown in Figure 47(a). The digital code directly controls the switches, thus no decoding logic is necessary. The advantages of this architecture are its simplicity and the small silicon area required. Some of it's drawbacks are large DNL error and increased dynamic error. The DNL is most severe at the mid-code transition were all the current sources are switching

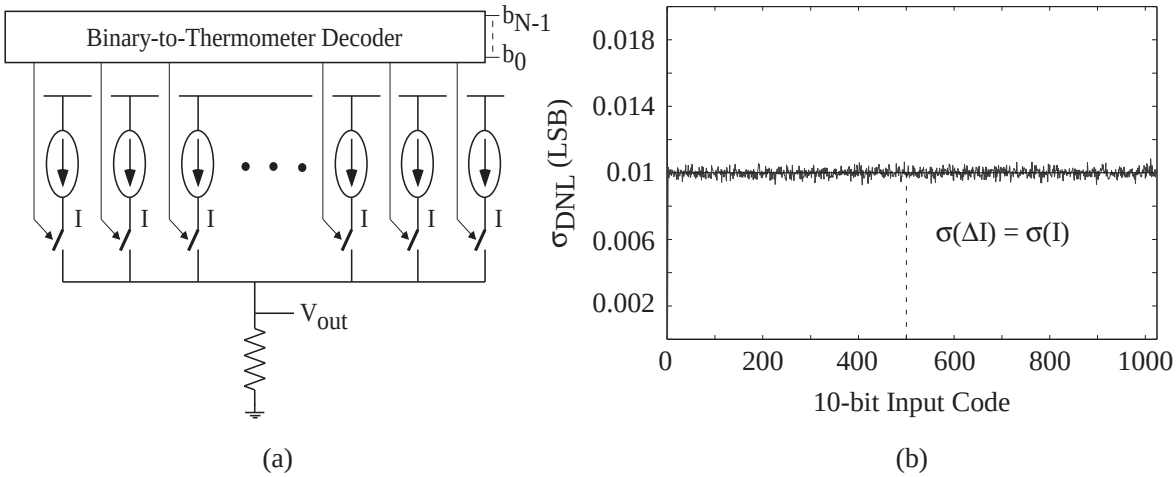


Figure 48. Unary Architecture: (a) Schematic diagram of a unary DAC. (b) Monte-carlo simulation results of the DNL.

at the same time. Errors caused by the dynamic behavior of the switches, such as charge-injection and clock feed-through, result in glitches in the output signal. These glitches contain highly nonlinear signal components and will manifest themselves as spurs in the frequency domain.

The DNL for this architecture, at the mid-code transition, is the difference between 2^{N-1} and $2^{(N-1)} - 1$ independent unit sources. Assuming a normal distribution for the unit current sources with a standard deviation $\sigma(I_{LSB})$, this step has a $\sigma(\Delta I)$ determined by

$$\sigma(\Delta I) = \sqrt{2^N - 1} \cdot \sigma(I_{LSB}) \tag{68}$$

where I_{LSB} is the least significant bit current (LSB) and $\sigma(\Delta I)$ is a good approximation for the DNL. Figure 47(b) shows DNL monte-carlo simulation results for a 10-bit binary weighted DAC. Here, DNL worsens as the number of switched current sources is increase. The mid-code DNL value is 32σ, which matches with the theoretical value described in (68).

6.2.2 Unary Decoded DAC

The unary architecture consists of an array of equally weighted current sources. An schematic diagram of such converter is shown in Figure 48(a). Here, every switch controls a single unit current source. This is achieved by converting the digital input code to a thermometer code that controls the switches. Such architecture has a good DNL error and a small dynamic switching errors because only one LSB current is switching at a given time. In contrasts to the binary weighted DAC, monotonicity is guaranteed when using this architecture. It's major drawbacks are area, complexity, and power consumption.

The DNL for this architecture is given by the error between any two consecutive codes. It can be approximated by

$$\sigma(\Delta I) = \sigma(I_{LSB}). \tag{69}$$

The matching requirements for DN$L < 0.5$LSB are much relaxed compared to the binary decoded DAC. Figure 48(b) shows DNL monte-carlo simulation results for a 10-bit unary weighted DAC. Here, DNL is constant for every input code; a DNL value of 1σ matches the theoretical value described in (69).

6.2.3 Segmented DAC

The segmented architecture merges the binary and the unary architectures thus achieving a balance between their advantages and drawbacks. An schematic diagram of such converter is shown in Figure 49(a). Here the LSBs are implemented using a binary DAC while the remaining most significant bits (MSBs) are implemented with an unary DAC. Compared to the unary architecture, the segmented DAC utilizes less area, consumes less power, and has a lower complexity. Compared to the binary architecture the converter has a better static and dynamic performance.

The maximum DNL error occurs at the transition of the two sub-DACs. The DNL

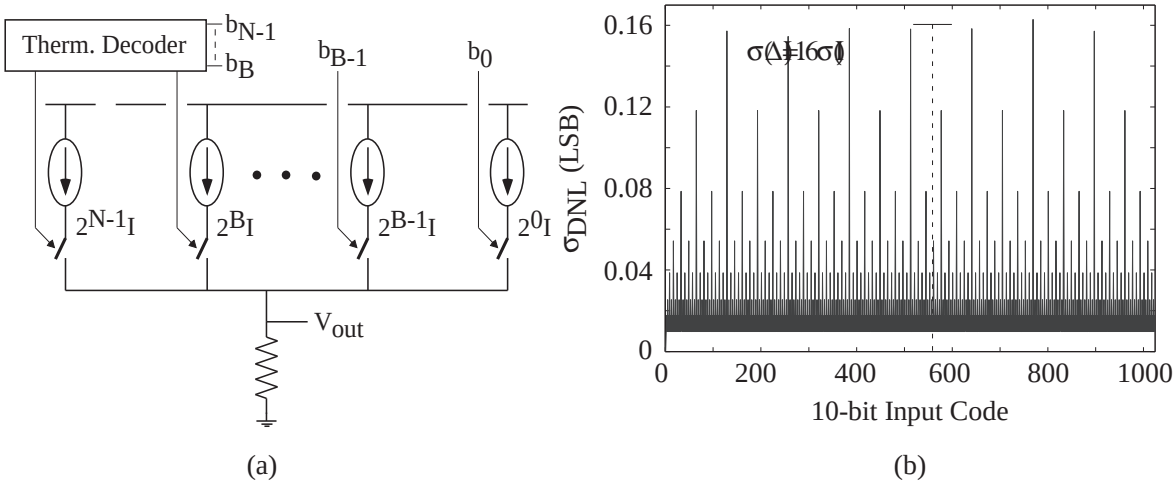


Figure 49. Segmented Architecture: (a) Schematic diagram of a segmented DAC. (b) Monte-carlo simulation results of the DNL.

can be estimated as

$$\sigma(\Delta I) = \sqrt{2^{B+1} - 1} \cdot \sigma(I_{LSB}) \tag{70}$$

where B is the number of bits implemented with a the binary architecture. Figure 49(b) shows DNL monte-carlo simulation results for a 10-bit segmented DAC (7-bit binary, 3-bit unary). Here, an improvement in the DNL is observed compared to the binary architecture. A DNL value of 16σ matches the theoretical value described in (70).

6.3 Basic DAC Design

Figure 50(a) shows the schematic diagram of a commonly used unit current source. Here, transistors M_1 and M_2 form the main current source while transistors M_3 and M_4 work as a differential switch. Matching of M_1 among the different unit current sources is important as it will determine the intrinsic accuracy of the converter. A higher output resistance is achieved with M_2 thus allowing the converter to drive low impedance loads without the use of a buffer. Finally, the differential pair M_3 and M_4 enables high speed current steering. The dynamic behavior of the converter will be affected by the control signals provided to M_3 and M_4.

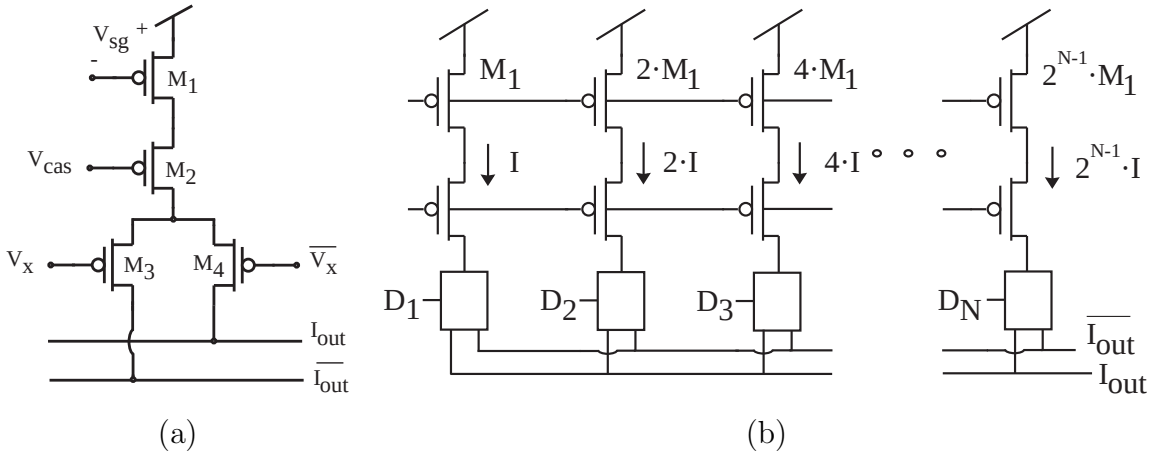


Figure 50. Basic DAC Architecture: (a) Schematic diagram of a unit current source. (b) Schematic diagram of a binary weighted current-steering DAC.

Typically, a N-bit current-steering DAC is built as an array of $2^N - 1$ unit current sources. The architecture selected will determine how each unit current source will be controlled (see Section 6.2). An schematic diagram of a binary weighted current-steering DAC is shown in Figure 50(b). Matching behavior of the current source transistors M_1 is one of the key issues in the design of high resolution CMOS current-steering DACs.

6.3.1 Intrinsic Accuracy

The INL specification of different DACs made in the same process technology will vary randomly due to mismatch. The relationship between the DAC INL and the matching properties of the used technology is given by INL yield, defined as the probability of the circuit complying with INL$< 0.5LSB$. A correct yield estimation can be obtained by performing Monte Carlo simulations [14]. This characterization is very time-consuming step in the design of high resolution DACs. An analytical relationship between the INL yield specification, the resolution, and the relative unit current standard deviation for a DAC has been developed in [26]. Using this formula, the unit current source mismatch allowed for a certain yield can be determined in a matter of seconds.

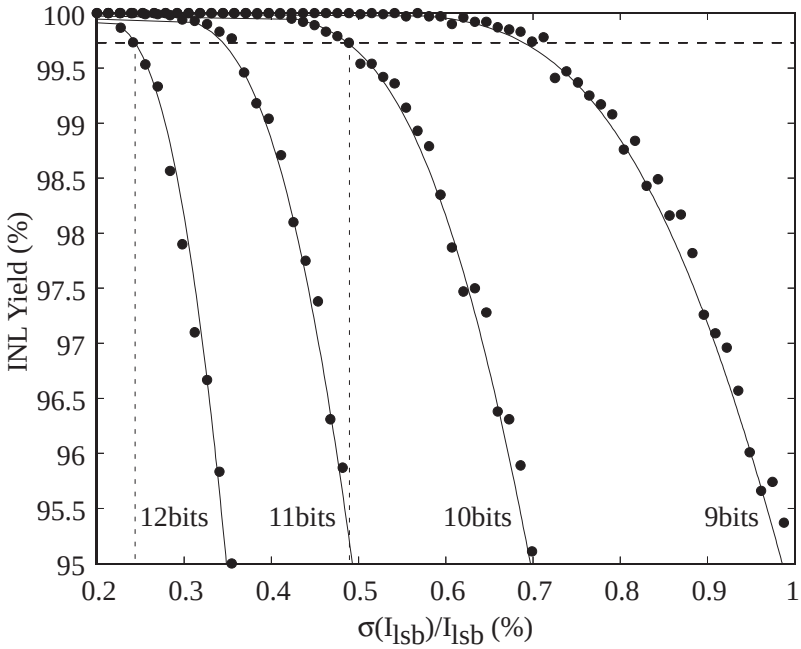


Figure 51. INL Yield: INL yield as a function of the relative current-source matching for the Monte Carlo simulations and the analytical formula described in [26]

Figure 51 shows the achievable INL yield as a function of the unit current source accuracy for different resolution DACs. The bubbles represent the Monte Carlo results while the solid lines represent results from the model developed in [26]. The model results in very accurate predictions without the need of time-consuming Monte Carlo simulations. To guarantee a DAC design with 10-bit or 12-bit of intrinsic accuracy and an INL yield of 99.7%, the unit current source mismatch needs to be $< 0.49\%$ and $< 0.25\%$ respectively.

6.3.2 Current Source Mismatch

From Section 1.1.1, the mathematical expression for the current mismatch between two identical designed transistors, obtained by substituting (1) and (2) into (3), is given by

$$\sigma^2\left(\frac{\Delta I}{I}\right) = \frac{{A_K}^2}{W{\cdot}L} + \left(\frac{2}{V_{gs}-V_{th}}\right)^2 \frac{{A_{V_{th}}}^2}{W{\cdot}L} \tag{71}$$

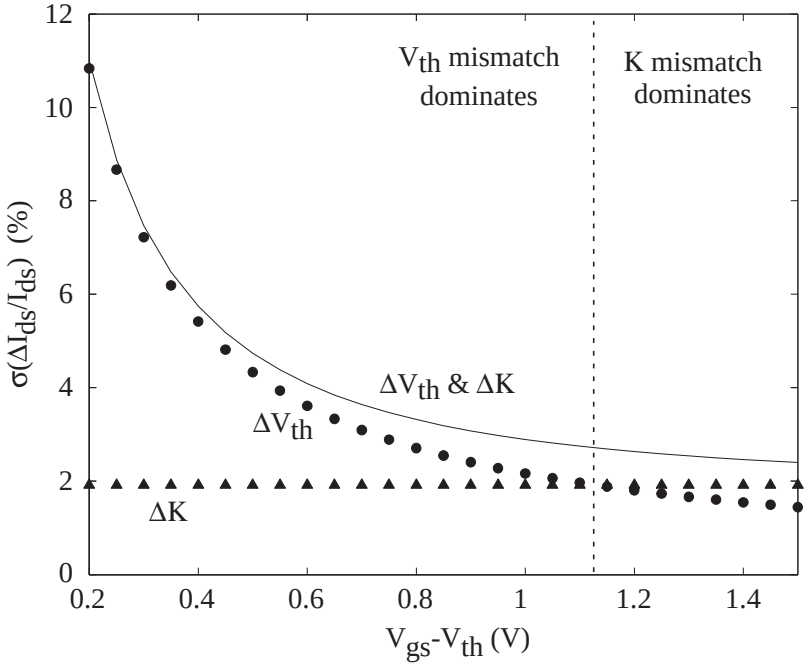


Figure 52. Current Source Mismatch: Drain-source current mismatch for a $1.2\mu m/1.2\mu m$ pMOS transistor in a $0.5\mu m$ CMOS technology with $A_{V_{th}} = 13\ mV\ \mu m$ and $A_K = 2.3\ \%\ \mu m$.

where A_{Vth} and A_K are mismatch technology parameters, $V_{gs} - V_{th}$ is the gate overdrive voltage of the current source transistor, and $\frac{\sigma(\Delta I)}{I}$ is the unit current source relative standard deviation. A graphical representation of (71) is shown in Figure 52. Here the individual contributions of V_{th} mismatch and K mismatch are also shown separately. It can be seen that the current source mismatch decreases as $V_{gs} - V_{th}$ increases. Also, for small $V_{gs} - V_{th}$ values the relative effect of the V_{th} mismatch dominates that of the K mismatch. Consequently, designers of CMOS current-steering DACs use relatively high $V_{gs} - V_{th}$ values to obtain lower current mismatch, thus keeping the size of the converter under control. This has a direct impact on the required supply voltage of the converter; minimum V_{dd} will be limited by the high $V_{gs} - V_{th}$ drop of the current source transistor. Also, the converter will have a small current span since the current mismatch worsens as $V_{gs} - V_{th}$ decreases.

A mathematical expression that relates the current source mismatch with the

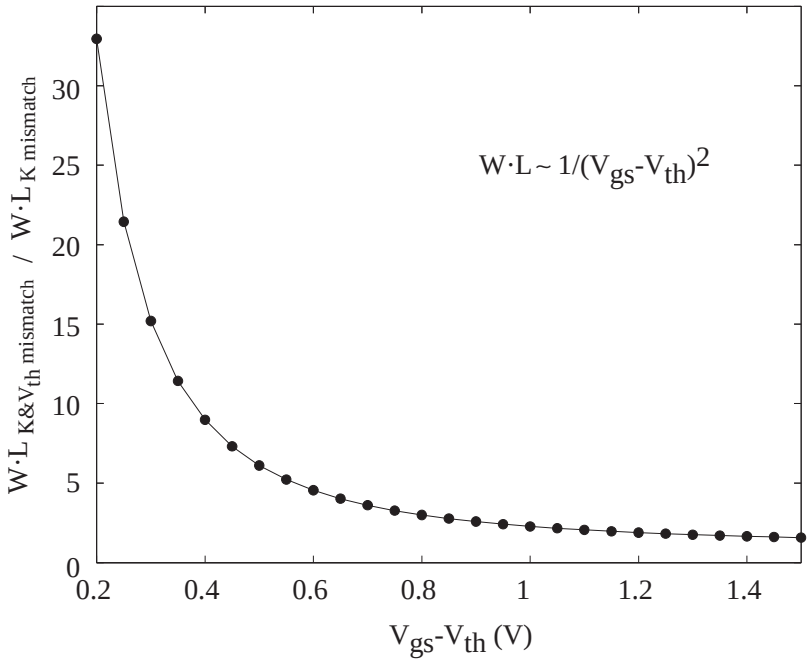


Figure 53. V_{th} **Mismatch Area Penalty: Ratio of the transistor area with** V_{th} **mismatch over the transistor area without** V_{th} **mismatch for 12-bit precision as a function of** $V_{gs} - V_{th}$**.**

gate-area $W{\cdot}L$ of the transistor can be obtained by solving (71) for $W{\cdot}L$, resulting in

$$W{\cdot}L = \left[{A_K}^2 + \frac{4{A_{V_{th}}}^2}{(V_{gs} - V_{th})^2} \right] / \left[\frac{\sigma(\Delta I)}{I} \right]^2. \tag{72}$$

Sizing the transistors according to (72) guarantees a current source mismatch specified by $\frac{\sigma(\Delta I)}{I}$. Using (72) along with the model described in Section 6.3.1, a current-steering DAC can be designed to meet a desired precision at the expense of die area. It can be estimated from (72) that for every additional bit the die area will increase $4\times$. This is on account of a $2\times$ increase do to mismatch requirements and a $2\times$ increase due to the additional number of current sources.

Figure 53 shows the ratio of the required gate-area of the unit current-source transistor with V_{th} and K mismatch over the required gate-area of the unit current-source transistor with only K mismatch. Operating at low $V_{gs} - V_{th}$ values come with a huge sacrifice in area mainly due to V_{th} mismatch (see Figure 52). In the absence of the V_{th} mismatch, an area improvement of $30\times$ could be obtained when operating

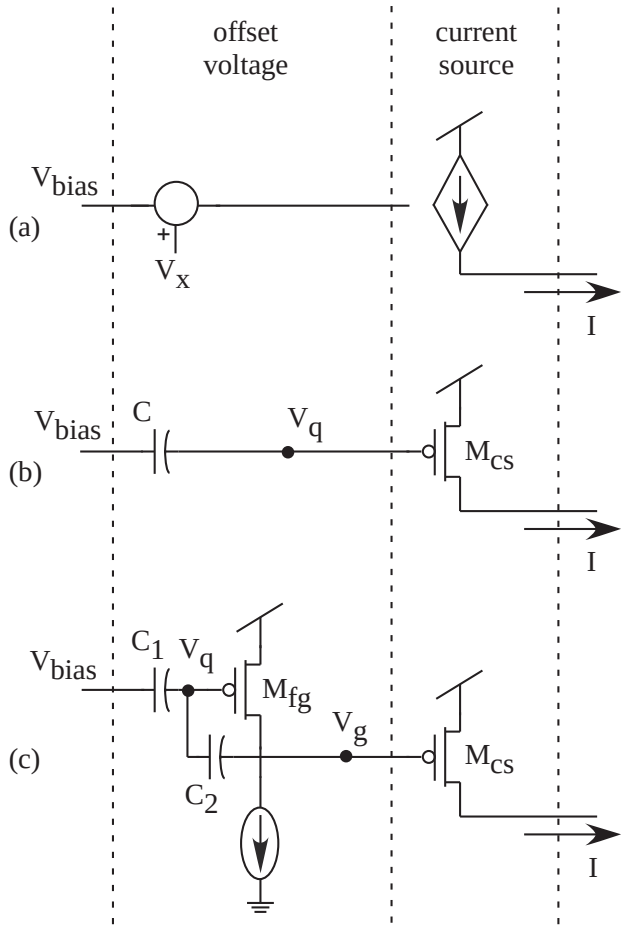


Figure 54. A V_{th} Compensated Current Source: (a) Conceptual representation of the proposed approach. (b) Schematic diagram of a simple implementaion of the proposed approach. (c) Schematic diagram of the V_{th} proposed compensated current source.

at $V_{gs} - V_{th} = 0.2$ as seen in Figure 53. Also, this will allow for a lower power supply operation and wider current span.

6.4 A V_{th} Compensated Current Source

Figure 54(a) shows the conceptual representation of the proposed approach in designing a unit current source. Here, the current source mismatch due to ΔV_{th} is compensated by adding an offset voltage V_x, equal in magnitude but with opposite sign, to the bias voltage V_{bias}. This technique allows for mismatch compensation independent of the bias current, as the mismatch will no longer depend on $V_{gs} - V_{th}$

(see Section 6.3.2). This is not possible with commonly used approaches were an additional current source is used for compensation.

A simple implementation of the proposed approach is shown in Figure 54(b) with the use of a single floating-gate transistor. Here, the charge stored on the floating node will produce a voltage V_q, equivalent to V_x in Figure 54(a). The addition of V_{bias} is done through the capacitive coupling. Although simple and effective, this method is not suitable for binary current sources due to area constraints. As the number of bits increases, the size of M_{cs} increases resulting in larger parasitic capacitances, thus affecting the capacitive coupling. This mismatch in capacitive coupling will manifest as a degradation in the temperature behavior. To obtain the same coupling for all the current sources the input capacitor has to increase accordingly; the last bit of a 10-bit converter will require an input capacitance around 1000× bigger than the first bit.

Figure 54(c) shows a different implementation of a V_{th} compensated current source. Here the current source transistor M_{cs}, no longer a floating-gate transistor, is preceded by an offset circuit. An increase in the parasitic capacitance of M_{cs} will not be a problem since the capacitive coupling is no longer used at the gate node of this transistor. The offset circuit is composed of a common source amplifier with a capacitive coupled input C_1 and a capacitive feedback C_2. The input voltage V_{bias}, obtained from a bias circuit common to all current sources, will generate the appropriate voltage V_g to bias M_{cs}. The minimum current change ΔI will be given by

$$\Delta I = g_m \frac{\Delta Q}{C_2} \tag{73}$$

were g_m is the transconductance of M_{cs} and ΔQ is the change in charge on the floating-node of M_{fg}. For a desired resolution, appropriate compensation can be achieved by satisfying

$$\frac{\Delta Q}{C_2} < \frac{V_{od}}{2^{N+1}} \tag{74}$$

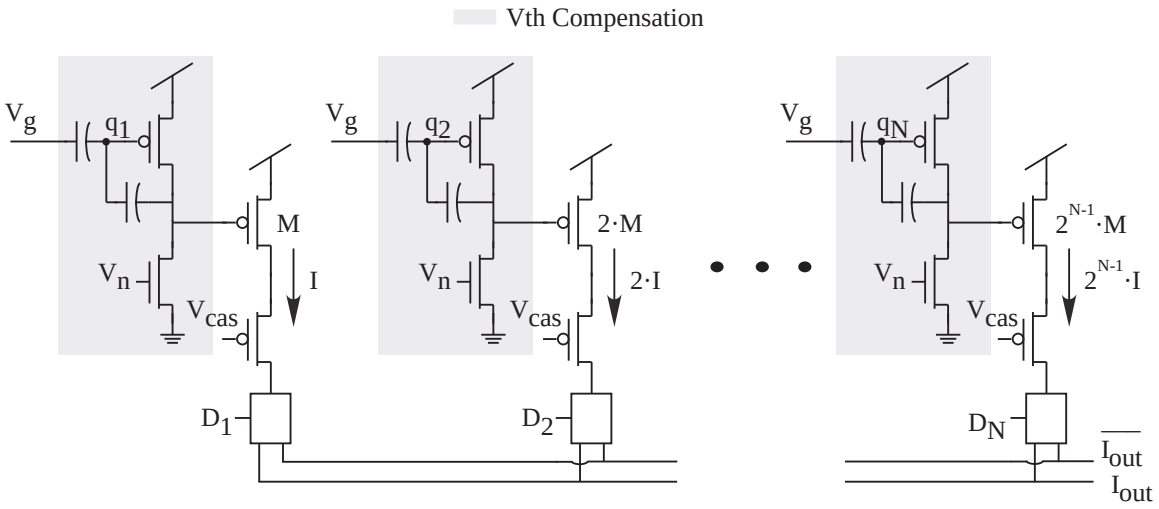


Figure 55. A V_{th} Compensated DAC: Schematic diagram of the proposed V_{th} compensated current-scaling DAC.

were N is the number of bits and V_{od} is the overdrive voltage of M_{cs}.

6.5 Proposed Digital-to-Analog Converter

Figure 55 shows the schematic diagram of the proposed current-steering V_{th} compensated DAC. The proposed approach uses the V_{th} compensation circuit presented in Section 6.4 (see Figure 54(c)) to compensate for mismatch errors. Compensation of the V_{th} mismatch allows for a reduction in the total size of the converter, enables a lower supply voltage operation, and allows for a wider current range.

Figure 56 shows the block diagram of the proposed DAC. The DAC has a $9 + 3$ segmented architecture; 9-bit binary decoded and 3-bit unary decoded. A double common centroid current matrix implements all current sources. The DAC utilizes 16 V_{th} compensation circuits: 9 for the binary bits and 7 for the unary bits. Latches are used to synchronize the switching of the current sources. A current reference was included on-chip.

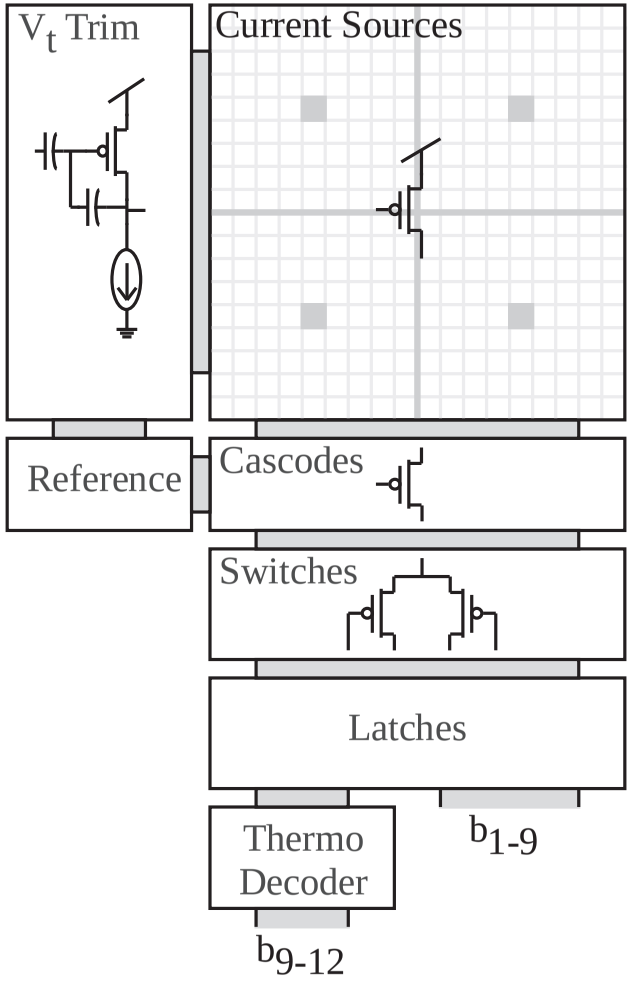


Figure 56. DAC Block Diagram: Block diagram of the proposed current-steering DAC.

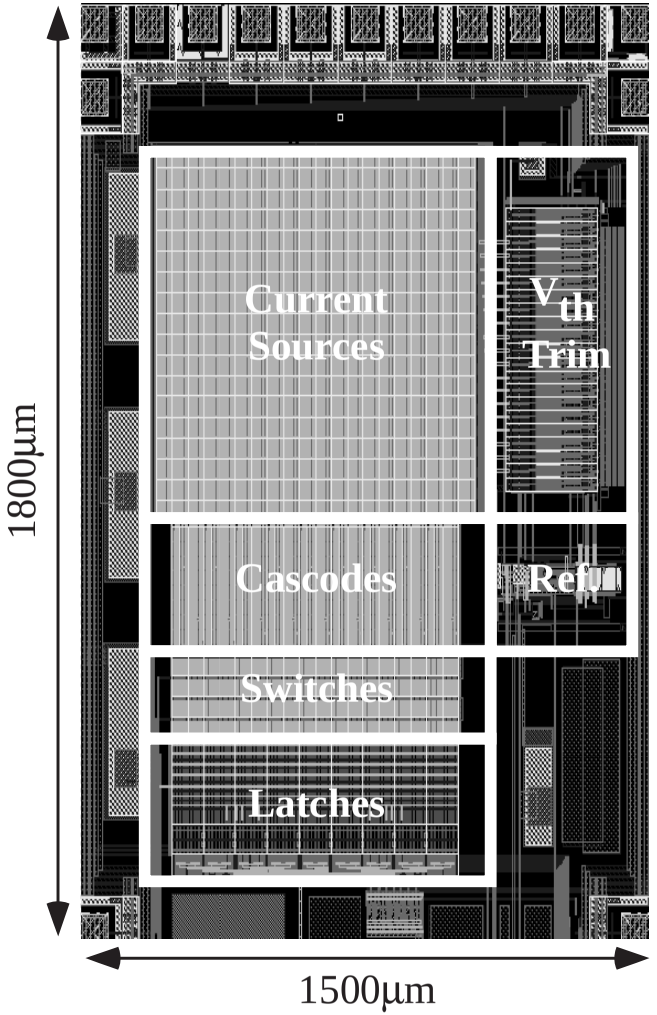


Figure 57. Chip Layout: Layout view of the fabricated prototype.

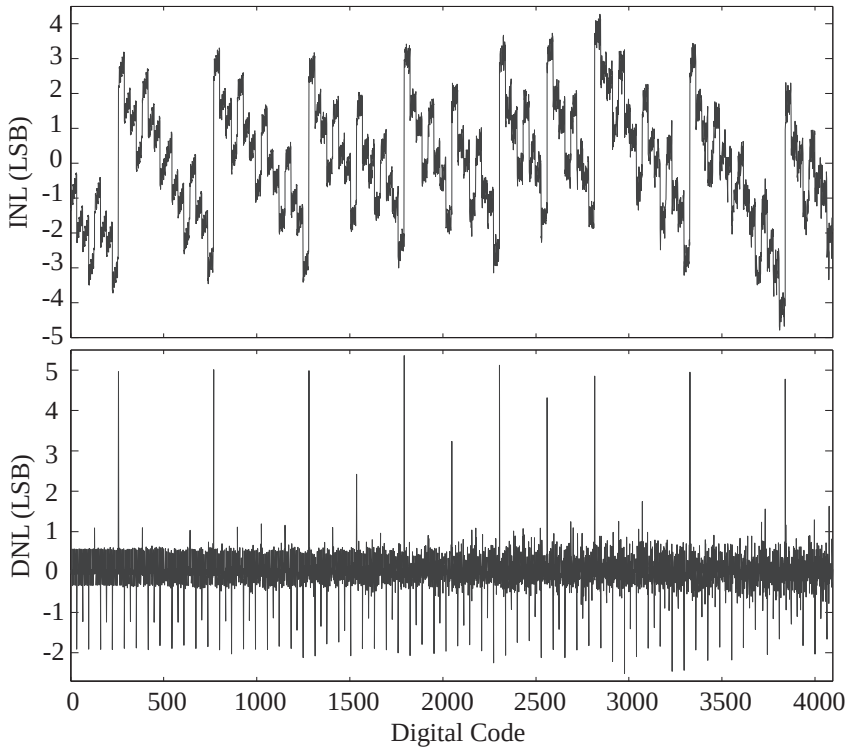


Figure 58. Uncompensated DAC: INL and DNL results of the proposed DAC without compensation.

6.6 Experimental Results

A prototype amplifier was fabricated in a $0.5\mu m$ CMOS process. The converter was designed for a full scale current of $10mA$, a resistor load of 100Ω, and a power supply of $3.3V$. The chip layout of the proposed DAC is shown in Figure 57. The V_{th} compensation circuit occupies an area of $570\mu m \times 310\mu m$ while the total area of the converter is $1.8mm \times 1.5mm$.

Figure 58 shows the INL and DNL plots of the proposed DAC without compensation. Here, the intrinsic DAC linearity is limited by device mismatch. The DAC exhibits a maximum INL and DNL error of $4.8LSB$ and $5.2LSB$ respectively. An improvement in the linearity of the converter was achieved by properly programming each V_{Th} compensation circuitry. Figure 59(a) shows the relative improvement of the INL error for a 9-bit accuracy. Here, the maximum INL error of the uncompensated

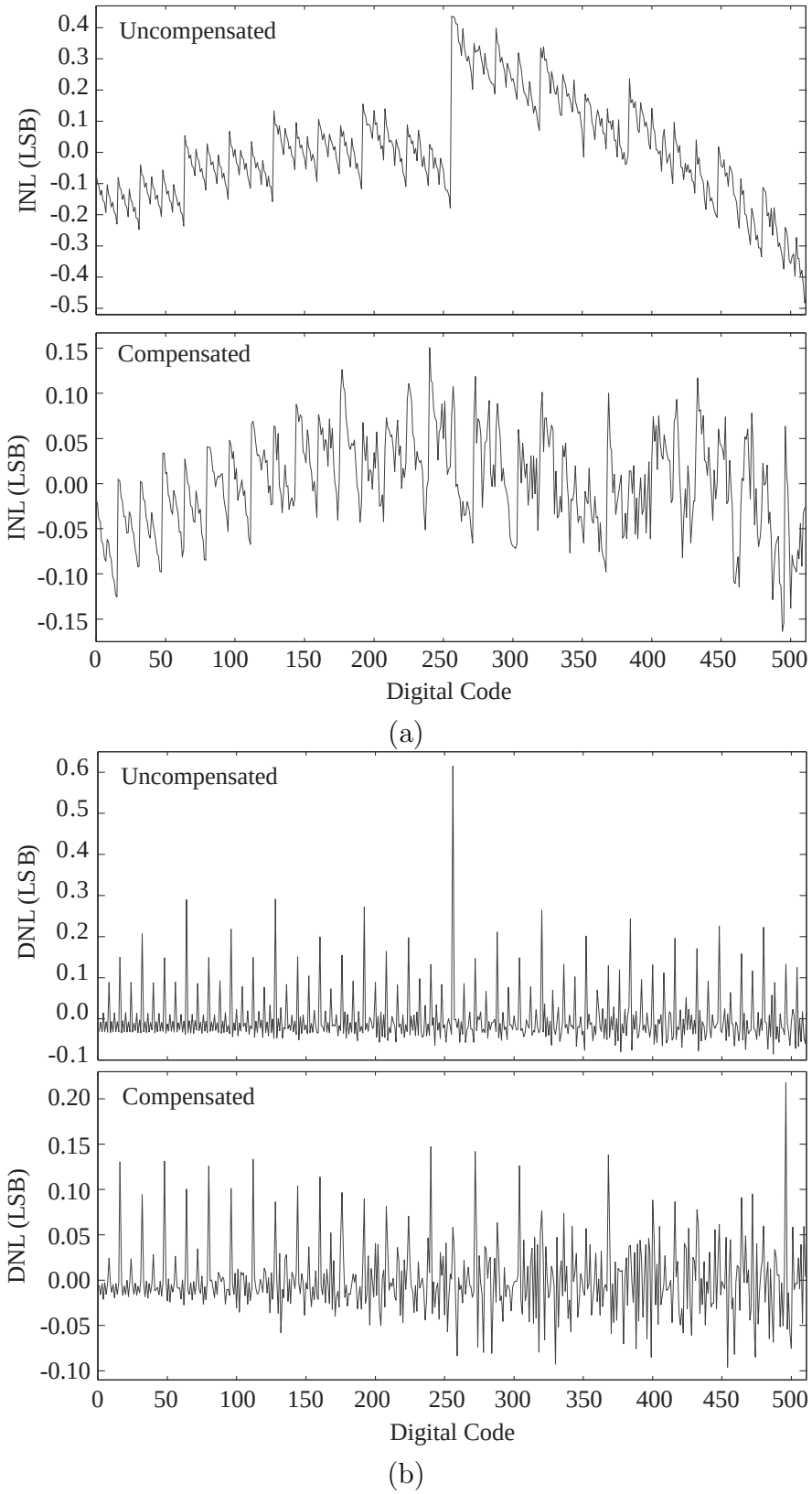


Figure 59. 9-bit DNL Compensation: (a) INL results of the proposed DAC before and after compensation for a 9-bit resolution. (b) DNL results of the proposed DAC before and after compensation for a 9-bit resolution.

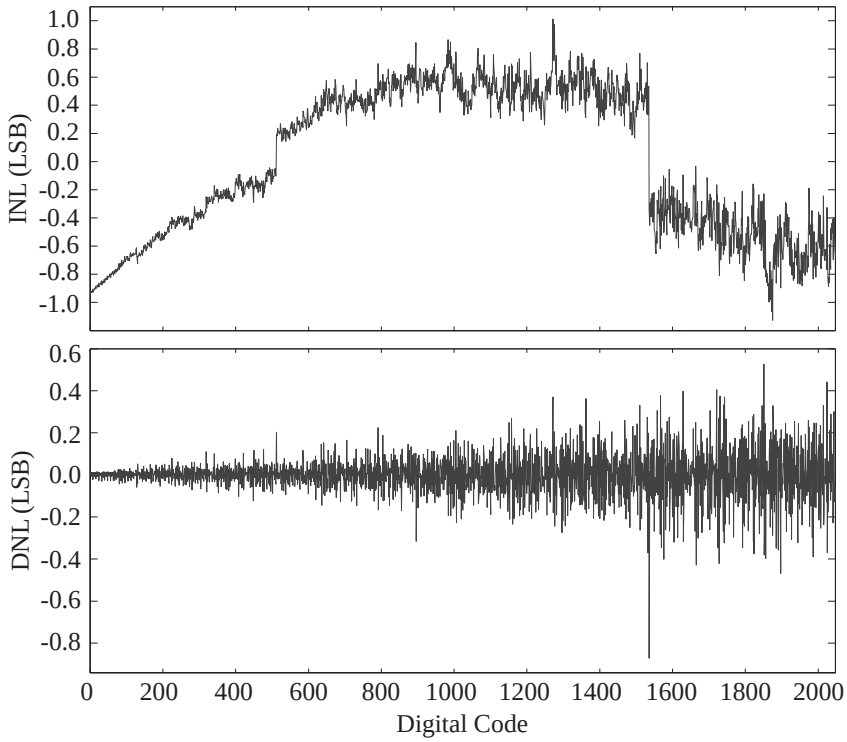


Figure 60. 11-bit Compensated DAC: (a) INL and DNL results of the proposed DAC after compensation for an 11-bit resolution.

DAC was decreased from $0.5LSB$ to $0.15LSB$. Similar results were obtained for the DNL error as seen in Figure 59(b). A maximum DNL error of $0.2LSB$ was obtained after compensation from an initial error of $0.6LSB$.

Figure 60 shows the INL and DNL results of the V_{th} compensated DAC for an 11-bit resolution. The DAC exhibits a maximum INL and DNL error of $1LSB$ and $0.8LSB$ respectively. The primary limitation for these results was the noise introduced by the reference. This can be inferred from the envelope shape shown by the DNL results in Figure 60.

6.7 Summary

A V_{th} compensated current-steering DAC has been presented. In order to obtain a desired linearity, the proposed converter uses floating-gate transistors to compensate for the intrinsic device mismatch. A V_{th} compensation circuit that allows for a direct

modification of each current source has been described. Experimental results showed that the intrinsic linearity errors can be improved by a factor of 3×. For example, the $0.6LSB$ DNL error of a 9-bit uncompensated DAC was reduced to $0.2LSB$. Also, maximum linearity errors of $1LSB$ and $0.8LSB$ were obtained for the INL and DNL of a 11-bit V_{th} compensated DAC. The main advantages of the proposed approach are: 1) it allows for an improve in linearity without the need of an increase in die area, 2) unlike any other schemes, linearity compensation can be obtained independently of the bias current and the temperature, 3) it allows for a low supply voltage operation.

CHAPTER VII

CONCLUSIONS

In this chapter, main contributions along with key milestones that have been achieved in this work are summarized.

7.1 Main Contributions

There has been some reluctance toward utilizing floating-gate transistors as analog memory in industry like applications. Issues such as programming accuracy, charge retention, and scalability appear to be the greatest concern for the designers. This work takes an important step towards proving that floating-gate technique is reliable and can be implemented in a commercial application. Ground work for reliability and programmability of floating-gate devices has been presented. Feasibility of this technique has been shown throughout this work with the implementation of high performance circuits such as a low offset amplifier, a voltage reference, a current reference, and a digital-to-analog converter. Also, this work portraits the use of floating-gate transistors as another tool in analog design, and not just as a trimming solution.

7.2 Research Summary

The successful use of floating-gate transistors in analog circuits depends on understanding certain key aspects of floating-gate transistors. A system that allows a fast and accurate programming of floating gates has been developed [69]. A predictive algorithm that allows programming of a target current within 0.2% error, in $7-8$ pulses has been introduced [11, 12]. Also, a theoretical analysis that relates programming precision with device parameters has been presented and proven experimentally

[72]. Charge retention in floating-gate transistors was measured through accelerated lifetime tests in $0.5\mu m$ and $0.35\mu m$ CMOS process [72].

Mismatch in analog circuitry is a critical issue that most commonly manifests itself as offset voltages in operational amplifiers. A floating-gate based offset voltage cancelation scheme [72, 74] has been proposed in this work. The offset voltage of the prototype amplifier was reduced to 25μV and exhibit a temperature drift of 130μV over a 170^oC temperature range. Overall, the proposed approach offers comparable offset cancelation with other techniques in a compact and low-power fashion while offering continuous-time amplifier operation.

Typically, trimming circuits are needed for optimal temperature compensation of voltage reference circuits. An alternate approach was presented in this work with the use of floating-gate transistors. The proposed architecture obtains the voltage reference as the difference in charge between two floating-gate transistors [73]. Temperature coefficients of $< 30ppm/^oC$ were obtained for a voltage range of $0.9V - 2.7V$ with $40\mu V$ precision. The key advantages of the proposed work include programmability, high initial accuracy, and low temperature dependence.

Temperature compensated and accurate current reference circuits are difficult to obtain due to device parameter variations. This work addressed this problem with a programmable current reference based on a low TC resistor [70]. Temperature coefficients of $< 130ppm/^oC$ were obtain for a current range of $16\mu A - 50\mu A$ with a precision of $< 0.02\%$. The key advantages of the proposed work include programmability, high initial accuracy, and low temperature dependence.

Mismatch between identical transistors limit the performance and dictate the size of current-scaling digital-to-analog converters. This work uses a floating compensation circuit to account for threshold voltage mismatch, thus reducing the size area of the converter while preserving its performance. Initial investigations [67, 68] have shown the feasibility of this approach. Experimental results showed that the intrinsic

linearity errors can be improved by a factor of 3×. Also, maximum linearity errors of $1LSB$ and $0.8LSB$ were obtained for the INL and DNL of a 11-bit V_{th} compensated DAC. The key advantages of the proposed work include decrease in die area, wider current range, and lower power supply.

REFERENCES

[1] "IEEE standard definitions and characterization of floating gate semiconductor arrays," in *IEEE Std.*, pp. 1005–1998, June 1998.

[2] A. Buck, C. McDonald, S. L. and Viswanathan, T. R., "A CMOS bandgap reference without resistors," in *IEEE International Solid-State Circuits Conference*, Feb. 2000.

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

[4] Ahuja, B. K., Vu, H., Aber, C. L., and Owen, W., "A 0.5μa precision CMOS floating-gate analog reference," *Proceddings of the International Solid State Circuits Conference*, vol. 48, pp. 286–287, Feb. 2005.

[5] Allen, P. and Holberg, D., eds., *CMOS Analog Circuit Design, Second Edition*. Oxford University Press, 2002.

[6] Allen, P. and Holdberg, *Hearing: An introduction to psychological and physiological acoustics*. New York: Marcel Dekker, Inc., 1981.

[7] an B. J. Sheu, B. W. L. and Yang, H., "Analog floating-gate synapses for general-purpose VLSI neural computation," in *IEEE Transactions on Circuits and Systems I*, vol. 38, pp. 654–658, June 1991.

[8] B. K. Ahuja, H. Vu, C. L. and Owen, W., "A 0.5μA precision CMOS floating-gate analog reference," in *IEEE International Solid-State Circuits Conference*, pp. 286–287, Feb. 2005.

[9] Banba, H., Shiga, H., Umezawa, A., Tanzawa, T., Atsumi, S., and Sakui, K., "A CMOS Bandgap Reference Circuit with Sub-1-V Operation," *IEEE Journal of Solid-State Circuits*, vol. 34, pp. 670–674, May 1999.

[10] Bandyopadhyay, A., Lee, J., Robucci, R. W., and Hasler, P., "MATIA: a programmable 80 /spl mu/w/frame CMOS block matrix transform imager architecture," *IEEE Journal of Solit-State Circuits*, vol. 41, pp. 663–672, Mar. 2006.

[11] Bandyopadhyay, A., Serrano, G., and Hasler, P., "Programming analog computational memory elements to 0.2% accuracy over 3.5 decades using a predictive method," in *IEEE International Symposium on Circuits and Systems*, vol. 3, pp. 2148–2151, May 2005.

[12] Bandyopadyay, A., Serrano, G., and Hasler, P., "Adaptive algorithm using hot-electron injection for programming analog computational memory elements within 0.2% of accuracy over 3.5 decades," in *IEEE Journal of Solid-State Circuits*, vol. 41, pp. 2107–2114, Feb. 2006.

[13] Bastos, J. and et al., "A 12-bit intrinsic accuracy high-speed CMOS DAC," in *IEEE Journal of Solid-State Circuits*, vol. 33, pp. 1959–1969, Dec. 1998.

[14] Bastos, J., Steyaert, M. S. J., and Sansen, W., "A high yield 12-bit 250-MS/s CMOS D/A converter," in *IEEE Custom Integrated Circuits Conference*, pp. 431–433, May 1996.

[15] Bendali, A. and Audet, Y., "A 1-V CMOS current reference with temperature and process compensation," in *IEEE Transactions on Circuits and Systems I*, vol. 54, pp. 1424–1429, July 2007.

[16] Berg, Y., Lande, T. S., Naess, O., and Gundersen, H., "Ultra-low-voltage floating-gate transconductance amplifiers," in *IEEE Transactions on Circuits and Systems II*, vol. 48, pp. 37–34, Jan. 2001.

[17] Blauschild, R. A., Tucci, P. A., Muller, R. S., and Meyer, R. G., "A new NMOS tempertarue-stable voltage reference," in *IEEE Journal of Solid-State Circuits*, vol. 13, pp. 767–774, Dec. 1978.

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

[19] Brokaw, A. P., "A Simple Three Terminal IC Bandgap Reference," *IEEE Journal of Solid-State Circuits*, pp. 388–393, Dec. 1974.

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

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

[22] Cellere, G., Pellati, P., Chimenton, A., Wyss, J., Modelli, A., Larcher, L., and Paccagnella, A., "Radiation effects on floating-gate memory cells," in *IEEE Transactions on Nuclear Science*, vol. 48, pp. 2222–2228, Dec. 2001.

[23] Chawla, R., Twigg, C. M., and Hasler, P., "An analog modulator/demodulator using a programmable arbitrary waveform generator," *IEEE International Symposium on Circuits and Systems*, vol. 6, pp. 6106–6109, May 2005.

[24] Chen, J. and Shi, B., "1V CMOS current reference with $50ppm/^oC$ temperature coefficient," in *IEEE Electronics Letters*, vol. 39, pp. 209–210, Jan. 2003.

[25] Dehghani, R. and Atarodi, S. M., "A new low voltage precision CMOS current reference with no external components," in *IEEE Transactions on Circuits and Systems II*, vol. 50, pp. 928–932, Dec. 2003.

[26] den Bosch, A. V., Borremans, M. A. F., Steyaert, M. S. J., and Sansen, W., "A 10-bit 1-Gsample/s nyquist current-steering CMOS D/A converter," in *IEEE Journal of Solid-State Circuits*, vol. 36, pp. 315–324, Mar. 2001.

[27] der Plas, G. V. and et al., "A 14-bit intrinsic accuracy Q random-walk CMOS DAC," in *IEEE Journal of Solid-State Circuits*, vol. 34, pp. 1708–1718, Dec. 1999.

[28] Deval, Y., Tomas, J., Begueret, J., Dugalleix, S., and Dom, J., "1-Volt ratiometric temperature stable current reference," in *IEEE International Symposium on Circuits and Systems*, vol. 3, pp. 1984–1987, June 1997.

[29] Duffy, C. and Hasler, P., "Modeling hot-electron injection in PFET's," *Journal of Computational Electronics*, vol. 2, pp. 317–322, Dec. 2003.

[30] Duffy, C. and Hasler, P., "Modeling hot-electron injection in pfet's," *Journal of Computational Electronics*, vol. 2, pp. 317–322, Jan. 2001.

[31] Enz, C. C. and Temes, G. C., "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.

[32] Filanovsky, I. M. and Allam, A., "Mutual compensation of mobility and threshold voltage temperature effects with applications in CMOS circuits," *IEEE Transactions on Circuits and Systems I*, vol. 48, pp. 876–884, July 2001.

[33] Fiori, F. and Crovetti, P. S., "A new compact temperature-compensated CMOS current reference," in *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 52, pp. 724–728, Nov. 2005.

[34] Fowler, R. H. and Norheim, L., "Electron emission in intense electric fields," in *Proceedings of the Royal Society of London*, vol. A119, pp. 173–181, 1928.

[35] Graham, D., Farquhar, E., Degnan, B., Gordon, C., and Hasler, P., "Indirect programming of floating-gate transistors," in *IEEE International Symposium on Circuits and Systems*, vol. 3, pp. 2172–2175, May 2005.

[36] Gray, P., Hurst, P. J., Lewis, S. H., and Meyer, R. G., *Analysis and Design of Analog Integrated Circuits*. Wiley, 4 ed., 2001.

[37] Groeneveld, D. and et al., "A self-calibration technique for monolithic high-resolution D/A converters," in *IEEE Journal of Solid-State Circuits*, vol. 24, pp. 1517–1522, Dec. 1989.

[38] Hao, Z., Bo, Z., Zhao-Ji, L., and Ping, L., "A new CMOS current reference with high order temperature compensation," in *IEEE International Coference on Comunications, Circuits, and Systems*, vol. 4, pp. 2189–2192, June 2006.

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

[40] Hendriks, P., "Specifying communication *DAC*'s," in *IEEE Spectrum*, pp. 58–69, July 1997.

[41] Hua, M., Mingyan, Y., Yizheng, Y., and Ma, J., "A novel current reference with low temperature coefficient in a large temperature range," in *IEEE International Conference on Solid-State and Integrated Circuit Technology*, pp. 1730–1732, 2006.

[42] Jiang, Y. and Lee, E. K. F., "Design of Low-Voltage Bandgap Reference Using Transimpedance Amplifier," *IEEE Transactions on Circuits and Systems II*, vol. 47, pp. 552–555, June 2000.

[43] Kahng, D. and Sze, S. M., "A floating-gate and its application to memory devices," in *The Bell System Technical Journal*, vol. 46, pp. 1288–1295, 1967.

[44] Kessel, H. V., "A new bipolar reference current source," in *IEEE Journal of Solid-State Circuits*, vol. 21, pp. 561–567, Aug. 1986.

[45] Kim, W. J., Sompur, S., and Kim, Y. B., "A novel digital controlled technique for operational amplifier compesation," in *IEEE Midwest Symposium on Circuits and Systems*, pp. 211–214, Aug. 2001.

[46] Kinget, P. R., "Device mismatch and tradeoffs in the design of analog circuits," in *IEEE Journal of Solid-State Circuits*, vol. 40, pp. 1212–1224, June 2006.

[47] Klimach, H., Arnaud, A., Galup-Montoro, C., and Schneider, M. C., "MOSFET mismatch modeling: a new approach," in *IEEE Design and Test of Computers*, vol. 23, pp. 20–29, Jan. 2006.

[48] Krawzsenek, D., Hsu, P., Anthony, H., and Land, C., "Single event effects and total ionizing dose results of a low voltage eeprom," in *IEEE Radiation Effect Data Workshop*, pp. 64–67, 2000.

[49] Lai, S., "Flash memories: where we were and where we are going," in *IEEE International Electron Devices Meeting*, pp. 971–974, 1998.

[50] Lanzoni, M., Tondi, G., Galbiati, P., and Ricco, B., "Automatic and continuous offset compensation of MOS operational amplifiers using floating-gate transistors," in *IEEE Journal of Solid State Circuits*, vol. 33, pp. 287–290, Feb. 1998.

[51] Larcher, L., Cellere, G., Paccagnella, A., Chimenton, A., Candelori, A., and Modelli, A., "Data retention after heavy ion exposure of floating-gate memories: Analysis and simulation," in *IEEE Transactions on Nuclear Science*, vol. 50, pp. 2176–2183, Dec. 2003.

[52] Lenzlinger, M. and Snow, E., "Fowler-Nordheim tunneling into thermally grown S_iO," in *Journal of Applied Fhysics*, vol. 40, pp. 278–283, 1969.

[53] Leung, K. N. and Mok, P. K. T., "A CMOS voltage reference based on weighted ΔV_{GS} for CMOS low-dropout linear regulators," in *IEEE Journal of Solid-State Circuits*, vol. 38, pp. 146–150, Jan. 2003.

[54] Leung, K. N. and Mok, P. T., "A sub-1-V $15 - ppm/°C$ CMOS bandgap voltage reference without requiring low threshold voltage device," in *IEEE Journal of Solid-State Circuits*, vol. 37, pp. 526–530, Apr. 2002.

[55] Liu, Y., Liu, G., and et al., "A novel CMOS current reference with low temperature and supply dependence," in *IEEE International Coference on Comunications, Circuits, and Systems*, vol. 4, pp. 2201–2204, June 2006.

[56] Mercer, D., "A 16bit D/A converter with increased spurious-free dynamic range," in *IEEE Journal of Solid-State Circuits*, vol. 29, pp. 1180–1185, Oct. 1994.

[57] Mok, P. K. T. and Leung, K. N., "Design considerations of recent advanced low-voltage low-temperature-coefficient CMOS bandgap voltage reference," in *IEEE Custom Integrated Circuits Conference*, pp. 635–642, Apr. 2004.

[58] Moyal, M., Groepl, M., and Blon, T., "A 25-kft 768-kb/s CMOS transceiver for multiple bit-rate DSL," in *IEEE International Solid-State Circuits Conference*, pp. 244–245, 1999.

[59] Nguyen, D. N., Guertin, S. M., Swift, G. M., and Johnston, A. H., "Radiation effect on advanced flash memories," in *IEEE Transactions on Nuclear Science*, vol. 46, pp. 1744–1750, Dec. 1999.

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

[61] Oguey, H. J. and Gerber, B., "MOS voltage reference based on polysilicon work function," in *IEEE Journal of Solid-State Circuits*, vol. 15, pp. 264–269, June 1980.

[62] Ozalevli, E. and Hasler, P., "Design of a CMOS floating-gate resistor for highly linear amplifier and multiplier applications," in *IEEE Custom Integrated Circuits Conference*, pp. 735–738, Sept. 2005.

[63] Pelgrom, M. J. M., Duinmaijer, A. C. J., and Welbers, A. P. G., "Matching properties of MOS transistors," in *IEEE Journal of Solid-State Circuits*, vol. 24, pp. 1433–1439, Oct. 1989.

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

[65] Sansen, W. M., Eynde, F., and Steyaert, M., "A CMOS temperature-compensated current reference," in *IEEE Journal of Solid-State Circuits*, vol. 23, pp. 821–824, June 1988.

[66] Schwartz, H. R., Nichols, D. K., and Johnston, A. H., "Single-event upset in flash memories," in *IEEE Transactions on Nuclear Science*, vol. 44, pp. 2315–2324, Dec. 1997.

[67] Serrano, G. and Hasler, P., "A floating-gate *DAC* array," in *IEEE International Symposium on Circuits and Systems*, vol. 1, pp. 357–360, May 2004.

[68] Serrano, G., Kucic, M., and Hasler, P., "Investigating programmable floating-gate digital-to-analog converter as single element or element arrays," in *IEEE Midwest Symposium on Circuits and Systems*, vol. 1, pp. 75–77, Aug. 2002.

[69] Serrano, G., Smith, P., Lo, H. J., Chawla, R., Hall, T., Twigg, C., and Hasler, P., "Automatic rapid programming of large arrays of floating-gate elements," in *IEEE International Symposium on Circuits and Systems*, vol. 1, pp. 373–376, May 2004.

[70] Serrano, G. J. and Hasler, P., "A precision low TC wide range CMOS current reference," in *IEEE Journal of Solid-State Circuits*, accepted.

[71] Song, B. S. and Gray, P. R., "A Precision Curvature-Compensated CMOS Bandgap Reference," *IEEE Journal of Solid-State Circuits*, pp. 634–643, Dec. 1983.

[72] Srinivasan, V., Serrano, G., Gray, J., and Hasler, P., "A precision CMOS amplifier using floating-gate transistors for offset cancellation," in *IEEE Journal of Solid-State Circuits*, vol. 42, pp. 280–291, Feb. 2007.

[73] Srinivasan, V., Serrano, G., Twigg, C., and Hasler, P., "A compact programmable CMOS reference with 40uV accuracy," in *IEEE Custom Integrated Circuits Conference*, pp. 611–614, Sept. 2006.

[74] Srinivasan, V., Serrano, G. J., Gray, J., and Hasler, P., "A precision *CMOS* amplifier using floating-gates for offset cancellation," in *IEEE Custom Integrated Circuits Conference*, pp. 739–742, Sept. 2005.

[75] Sweeney, J. and Geiger, R., "Very high precision analog trimming using floating gate MOSFETS," in *IEEE European Conference on Circuit Theory and Design*, pp. 652–655, Sept. 1989.

[76] Tam, S. and et. al., "Lucky electron model of channel hot-electron injection in MOSFETs," in *IEEE Transactions on Electron Devices*, vol. 31, pp. 1116–1125, 1984.

[77] Tesh, B. and Garcia, J., "A low-glitch 14-bit 100-MHz D/A converter," in *IEEE Journal of Solid-State Circuits*, vol. 32, pp. 1465–1469, Sept. 1997.

[78] Thomsen, A. and Brooke, M., "A temperature stable current reference source with programmable output," in *IEEE Midwest Symposium on Circuits and Systems*, vol. 2, pp. 831–834, Aug. 1992.

[79] Tsividis, Y., *Operation and Modeling of the MOS Transistor*. McGraw Hill, New York, 1999.

[80] Twigg, C. M. and Hasler, P., "A large-scale reconfigurable analog signal processor (RASP) IC," *IEEE Custom Integrated Circuits Conference*, pp. 5–8, Sept. 2006.

[81] Vita, G. D. and Iannaccone, G., "A 109nW, 44ppm/^{o}C CMOS current reference with low sensitivity to process variations," in *IEEE International Symposium on Circuts and Systems*, pp. 3804–3807, May 2007.

[82] Wong, Y. L., Cohen, M. H., and Abshire, P. A., "A floating-gate comparator with automatic offset adaptation for 10-bit data conversion," in *IEEE Transactions on Circuits and Systems I*, vol. 52, pp. 1316–1326, July 2005.