2004-03-12

# Design of a High Speed Mixed Signal CMOS Mutliplying Circuit

David Ray Bartholomew
*Brigham Young University - Provo*

# DESIGN OF A HIGH SPEED MIXED SIGNAL CMOS MULTIPLYING CIRCUIT

by

David Ray Bartholomew

A thesis submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirements for the degree of

Master of Science

Department of Electrical and Computer Engineering

Brigham Young University

April 2004

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

Of a thesis submitted by

David Ray Bartholomew

This thesis has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

| | |
|---|---|
| 11 March 2004 | David J Comer |
| Date | David J. Comer, Chair |
| Mar 11, 2004 | Donald T. Comer |
| Date | Donald T. Comer |
| 3/11/04 | Craig S Petrie |
| Date | Craig S. Petrie |

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the thesis of David Ray Bartholomew in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

11 March 2004
Date

David J. Comer
David J. Comer
Chair, Graduate Committee

Accepted for the Department

Michael A. Jensen
Michael A. Jensen
Graduate Coordinator

Accepted for the College

for Richard W. Christiansen
Douglas M. Chabries
Dean, College of Engineering and Technology

ABSTRACT

DESIGN OF A HIGH SPEED MIXED SIGNAL CMOS MULTIPLYING CIRCUIT

David Ray Bartholomew

Department of Electrical and Computer Engineering

Master of Science

This thesis presents the design of a mixed-signal CMOS multiplier implemented with short-channel PMOS transistors. The multiplier presented here forms the product of a differential input voltage and a five-bit digital code. A TSMC 0.18 μm MOSFET model is used to simulate the circuit in Cadence Design Systems. The research presented in this thesis reveals a configuration that allows the multiplier to run at a speed of 8.2 GHz with end-point nonlinearity less than 5%. The high speed and low nonlinearity make this circuit ideal for applications such as filtering and digital to analog conversion.

# ACKNOWLEDGMENTS

I would like to thank everybody that donated time and helped with this thesis. I would like to especially like to thank David Comer and everybody on my thesis committee for all their time and patience along with all they taught me during my time at BYU. I would also like to thank Intel for making this research possible. And of course none of this would have been possible without the love and support of my wife Julianne and my family.

**Contents**

**Abstract** vii

**Acknowledgments** ix

**1 Introduction** 1

1.1 Contributions....2

**2 Background Theory** 5

2.1 Linear V-I Conversion....6

2.2 Four-Quadrant Multiplication....14

**3 The Code Select Binary Weighting Circuit** 23

3.1 General Multiplier Configuration....24

3.2 Circuit Improvement....29

**4 Circuit Simulation and Optimization** 33

4.1 General Simulations....33

4.2 Device Width Simulations....38

4.3 Substrate Bias Simulations....44

4.4 Causes of Nonlinearity....46

4.5 Other Important Simulation Values....51

**5 Applications** 55

5.1 Linear Equalization....55

5.2 Digital To Analog Conversion....57

**6 Conclusions** **59**

6.1 Topics For Further Research....59

**A Appendix A** **61**

A.1 Notes on Cadence Design Systems....61

A.2 Correct Circuit Results....64

**Bibliography** **71**

**List of Figures**

Figure 1: Block diagram of two major circuit components. ..... 5

Figure 2: Schematic for single PMOS simulation. ..... 8

Figure 3: Output characteristics for 0.18 μm PMOS. ..... 8

Figure 4: Input linear V-I converters. ..... 9

Figure 5: Output drain current from V-I converters. ..... 10

Figure 6: Differential current $I_2$-$I_1$ from currents in Figure 5. ..... 11

Figure 7: Source voltages for the final circuit. ..... 12

Figure 8: Input converter devices with drain voltages added. ..... 13

Figure 9: Converter output current with nonlinearity. ..... 13

Figure 10: MOSFET four-quadrant multiplier. ..... 15

Figure 11: Multiplier with the n-bit code added. ..... 24

Figure 12: Actual multiplier schematic..... 27

Figure 13: Circuit output for all possible weighting codes. ..... 34

Figure 14: Transient output for all possible weighting codes. ..... 35

Figure 15: End-point nonlinearity for all possible weighting codes. ..... 36

Figure 16: End-point nonlinearity versus input transistor width. ..... 38

Figure 17: Percentage of target voltage reached versus input transistor width. ..... 40

Figure 18: End-point nonlinearity versus weighting transistor width. ..... 42

Figure 19: Percentage of target voltage reached versus weighting transistor width..... 43

Figure 20: Comparison of transient for substrate bias change. ..... 44

Figure 21: End-point nonlinearity comparison for substrate bias change. ..... 45

Figure 22: Simulation to determine input operating region. ..... 49

Figure 23: End-point nonlinearity versus differential scaling voltage ..... 54

Figure 24: Correct fast transient results. ..... 65

Figure 25: Substrate comparison including well capacitances. ..... 69

# CHAPTER 1

## INTRODUCTION

Multiplier circuits are used in several areas of electrical engineering such as neural networks, mixing, and filtering. An important subclass of filter circuits is that of linear equalizers. In high-frequency chip-to-chip digital communications, the digital signals degrade rapidly due to transmission line effects in the conductors that link the two chips. Linear equalization is often used to restore the transmitted data to a level that leads to accurate reception [1].

This type of equalization circuit requires the formation of the sum of products of voltages and weighting codes. Assuming $S_i$ is a sample of the input voltage at time $i$, $W_j$ is the $j^{th}$ weighting code, and $K$ is an arbitrary constant formed in the circuitry, the products formed in the equalizer are of the form:

$$I_{out} = K \times S_i \times W_j . \tag{1}$$

The multiplier discussed in this thesis forms the product in (1) where the weighting code, $W_j$, is an n-bit binary code. This thesis analyzes and suggests improvements in the design of this multiplier configuration through general analysis and simulation. In this multiplier, device characteristics and size are important in determining speed and

accuracy. For short channel devices the equations to predict operation are often simplified, inaccurate approximations or complicated models with hundreds of variables. Therefore, this thesis discusses equations for active region current approximations, but relies heavily on simulation of this circuit with TSMC 0.18 μm technology using Cadence Design Systems for accurate prediction of circuit behavior.

This thesis covers several aspects of the multiplier. Chapter 2 begins with equations and descriptions of the short-channel MOSFET. The basic configuration and circuit background are then presented and analyzed. In Chapters 3 and 4 the multiplier design, solutions for circuit implementation, and improvements made to the circuit design are discussed. Chapter 3 presents the actual multiplier configuration and operation, while Chapter 4 covers circuit simulation. The multiplier is studied in depth, including problems, solutions, and improvements resulting from simulation and research. In Chapter 5 various applications of the multiplier are briefly explored. The conclusion and suggestions for further research are found in Chapter 6.

## 1.1 CONTRIBUTIONS

The following contributions are a result of the work described in this thesis:

- Detailed analysis of the binary weighting multiplier using short channel CMOS device models.
- Derivation of equations to describe binary weighting multiplier operation.
- Development of an optimization technique for multiplier device size allowing higher speed and accuracy.

- Simulations characterizing the operation of the binary weighting multiplier using a 0.18 μm TSMC device model simulated in Cadence Design Systems.
- A new multiplier configuration showing improved speed with very little increase in nonlinearity.

## CHAPTER 2

## BACKGROUND THEORY

Figure 1 shows a block diagram of the multiplication circuit divided into its two major functions. The analog input voltage is first converted to a proportional current using a linear V-I converter. This current is then weighted according to a digital weighting code to form an output current that is proportional to the product of the input signal and the weighting code. For noise and nonlinearity minimization, this circuit uses differential inputs and a double-ended output.

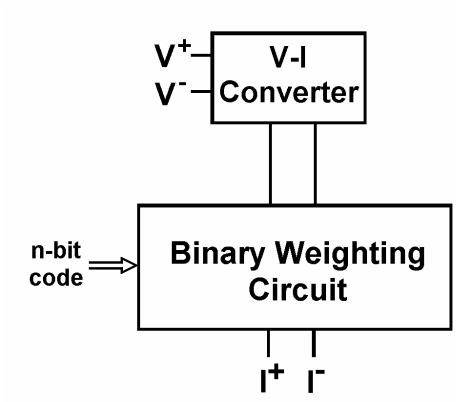


Figure 1: Block diagram of two major circuit components.

## 2.1 LINEAR V-I CONVERSION

The circuit built to accomplish the linear voltage to current conversion includes two common-source p-channel devices. In 0.18 μm devices, the relationship between output current and input voltage closely approximates a linear variation in the active region of conduction. Thus, the short-channel device itself becomes a linear V-I converter.

There are many short-channel effects that cause a linear V-I characteristic in short-channel MOS devices. In short-channel devices, the electric field across the channel can become sufficiently high that the velocity of the carriers approaches a constant value rather than increasing indefinitely with the electric field intensity. This constant is referred to as the scattering-limited velocity, and the electric field at which it occurs is called the critical electric field. The effects of velocity saturation on the drain current of a device in the active region have been derived in [2] as:

$$I_d = \frac{\mu C_{ox} W}{2L\left(1 + \frac{V_{gs} - V_t}{\varepsilon_c L}\right)} (V_{gs} - V_t)^2 . \tag{2}$$

In (2), μ is the carrier mobility, $C_{ox}$ is the oxide capacitance, $W$ is transistor channel width, $L$ is the transistor channel length, $V_t$ is the threshold voltage, $V_{gs}$ is the gate to source voltage, and $\varepsilon_c$ is the critical field. The critical field has a value of $6.98 \times 10^3$ V/cm for electrons and $1.80 \times 10^4$ V/cm for holes [3]. For longer channel devices, $L \geq 2$ μm, the product $\varepsilon_c L$ can be approximated as infinity and this equation leads to the common square-law expression of:

$$I_d = \mu C_{ox} \frac{W}{2L} (V_{gs} - V_t)^2 . \quad (3)$$

In the case of short-channel devices, the product $\varepsilon_c L$ becomes much smaller and, in the limit, the equation for drain current becomes:

$$I_d = \mu C_{ox} W \varepsilon_c (V_{gs} - V_t) . \quad (4)$$

A discussion of this approximation for the active region drain current is also found in [4].

Equation (4) does not describe all short-channel effects, but can be used as an approximation for conversion of the input voltage to current. Equation (4) , for instance, does not reflect the effects of channel modulation that lead to finite output impedance for the converter devices. In short-channel devices the Early voltage, used to represent this channel modulation, changes with input voltage and follows complicated models [5]. Some of the nonlinearity in the multiplier results from this finite output impedance, but can be minimized by attempting to keep the drain voltages of the input converter devices constant. The nonlinearity of the circuit due to finite output impedance is the limiting effect for the nonlinearity and is discussed later in this section.

The initial voltage to current conversion behavior is verified in Cadence with a single 0.18 μm gate length PMOS transistor with a width of 2 μm as shown in Figure 2. In Figure 3, $I_d$ of this 0.18 μm PMOS is plotted versus $V_{gs}$ with $V_{ds}$ = -1.6 V and $V_{gs}$ swinging from -1.6 V to -0.8 V.

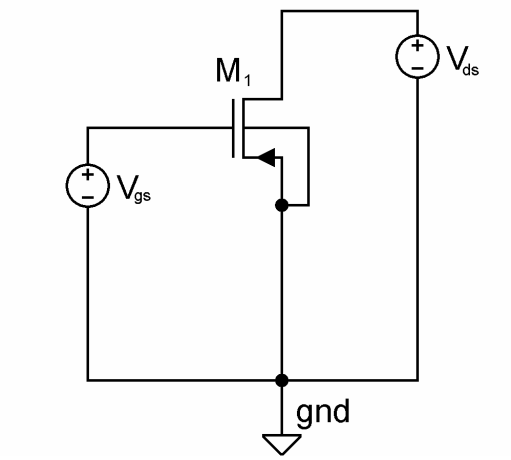


Figure 2: Schematic for single PMOS simulation.

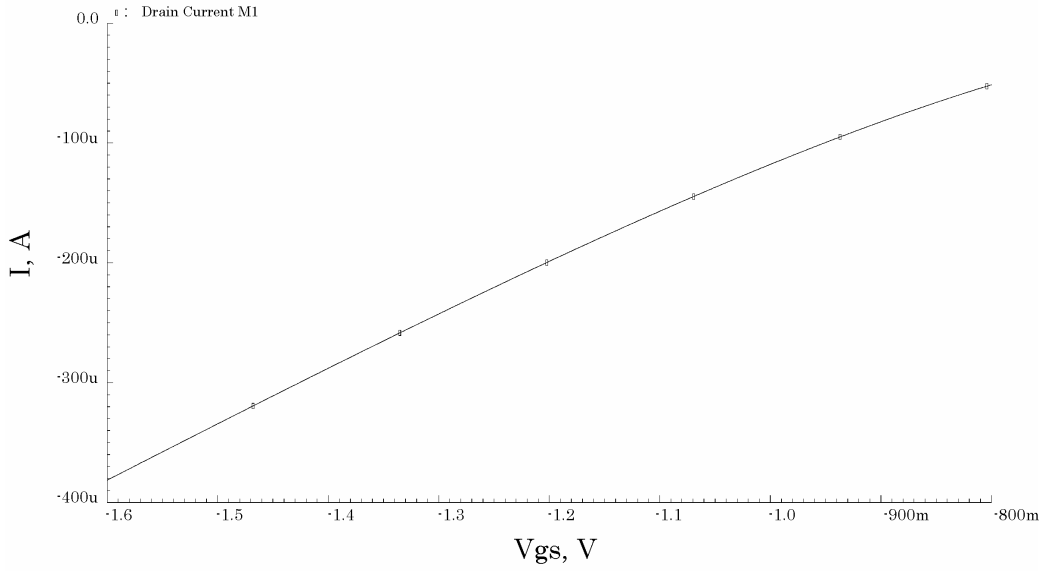


Figure 3: Output characteristics for 0.18 μm PMOS.

The active region current shown in Figure 3 is approximately linear. From Figure 3 it can be seen that although there are variations between (3) and (4) with an exponent varying between 2 and 1 [6], the 0.18 μm device current exhibits an exponent near unity

in the active region making (4) a good approximation for the drain current in the active region of conduction. Figure 4 shows the Cadence circuit used to simulate the differential configuration for the input V-I converters. It consists of two common source 16 μm wide PMOS devices with a 1.6 V power supply voltage applied to the source and bulk and two input voltages that linearly vary from 0.8 V to 0.2 V and from 0.2 V to 0.8 V on the gates of devices 1 and 2 respectively. The simulation used to test the input V-I converters is a slow transient. Each input voltage takes 12 μs to swing from its initial to its final value, and the output drain current is monitored for 12 μs as the inputs change. The slow transient approximates a DC solution for the input converters, since the speed of the circuit is in the GHz range.

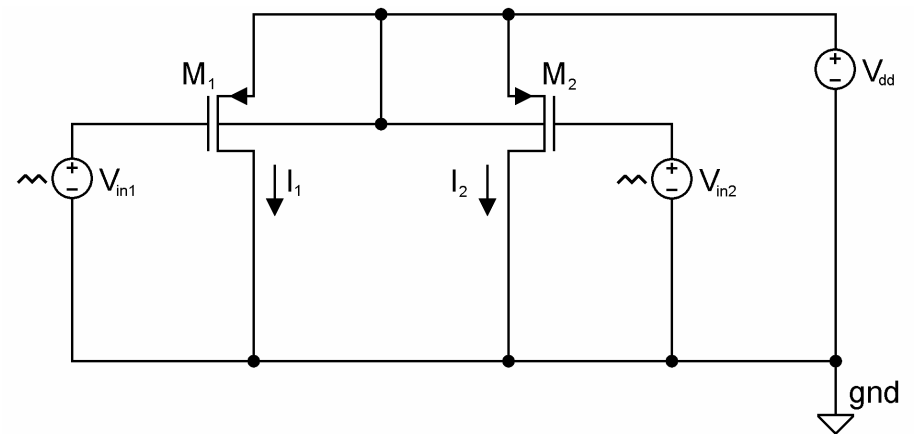


Figure 4: Input linear V-I converters.

Figure 5 is a plot of the output currents $I_1$ and $I_2$ from each of the transistors of Figure 4 measured at the drain terminals.

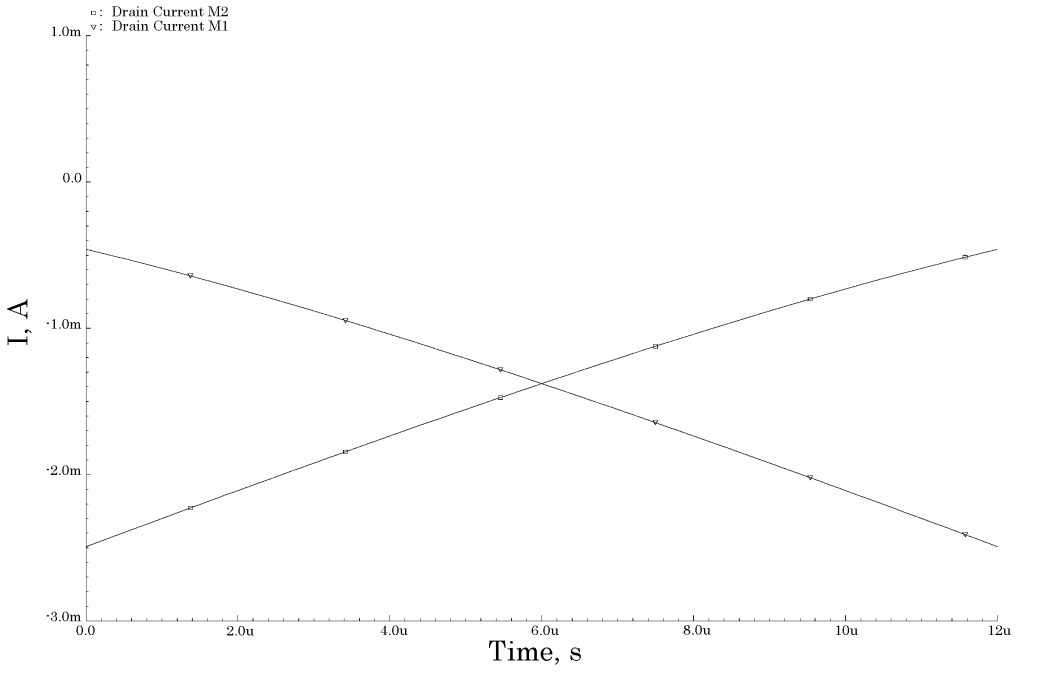


Figure 5: Output drain current from V-I converters.

The nonlinearity of each of the currents in Figure 5 is around 5% using end-point nonlinearity measurements. In these nonlinearity measurements a straight tangent line is drawn between the end-points of the signal. The tangent line and actual signal are then compared to obtain a maximum y-value deviation between the two lines. Finally, this deviation from the ideal line is divided by the total difference between the maximum and minimum y values of the signal, or the total y-value height of the signal. The total end-point nonlinearity measurement for the final output of the multiplier needs to be below 5% for the circuit to have a reasonable accuracy. Although an end-point nonlinearity of 5% for the input converter transistor outputs is too high for the multiplier, the differential

configuration of the devices cancels most of the nonlinearity. Figure 6 shows the difference of the two output currents of Figure 5 ($I_2$-$I_1$).

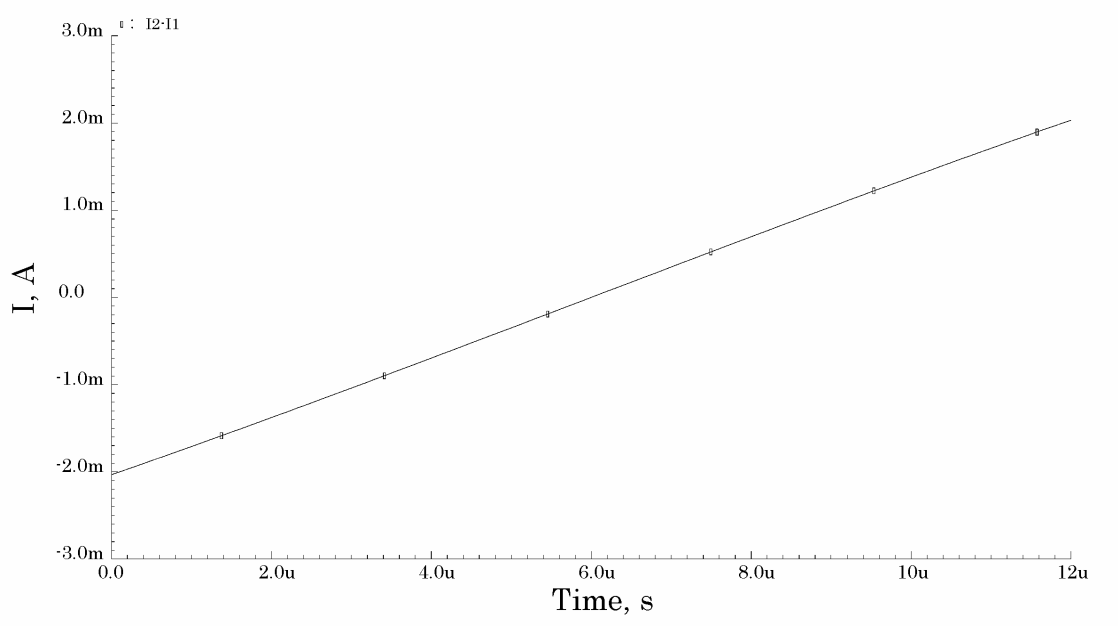


Figure 6: Differential current $I_2$-$I_1$ from currents in Figure 5.

The differential current of Figure 6 has an end-point nonlinearity of 0.59%. A measurement of 0.59% for the input shows that the differential input configuration is effective at canceling nonlinearity. Thus, these input V-I converters work well with constant drain voltages. In the actual multiplication circuits the drain voltage does not remain constant.

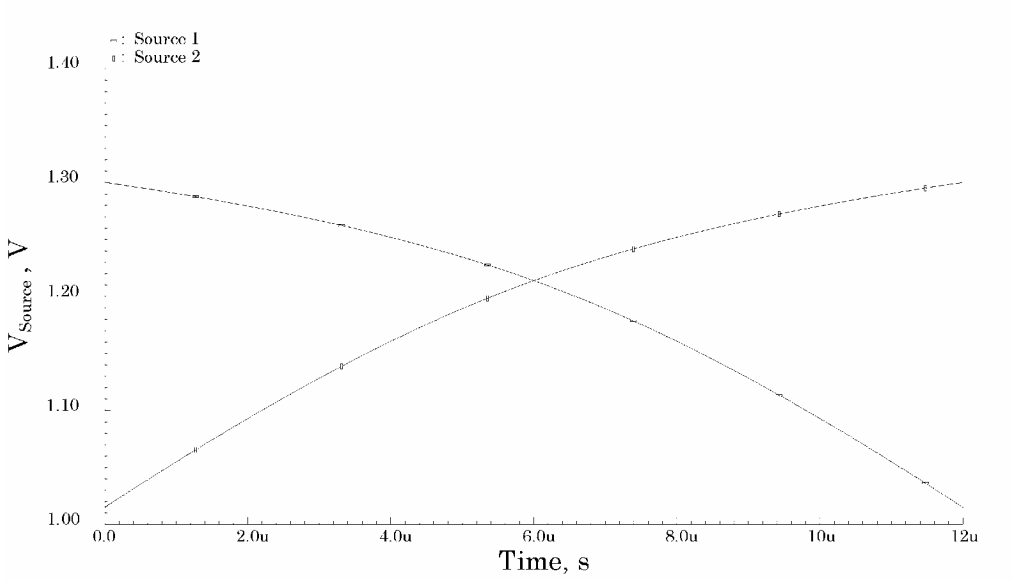


Figure 7: Source voltages for the final circuit.

Figure 7 shows the drain voltages at the input devices in the actual final circuit. The voltages are not exactly linear because of the nonlinear current conversion and the changing impedance of the input devices, but the drain voltages can be approximated as a linearly varying voltage on the drain of each input device. With this varying voltage on the drains of the input converter devices, they are driven are driven out of the active region, causing nonlinearity in the transition from the active to the triode region. This voltage swing also causes nonlinearity due to the finite output impedance of the input devices.

In order to account for the drain voltage present on the input converter devices, a simulation was carried out with a circuit similar to that of Figure 4, but with linearly varying voltage sources connected to the drain terminals of each input device. The circuit for this simulation is shown in Figure 8.

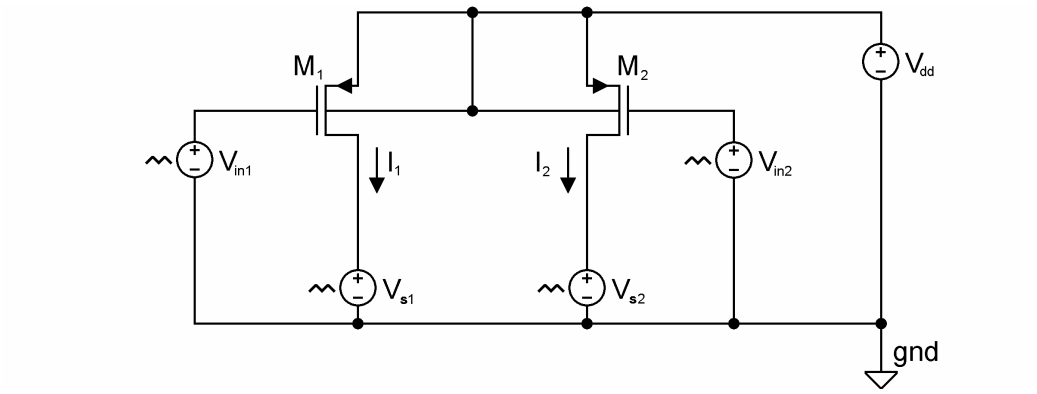


Figure 8: Input converter devices with drain voltages added.

The voltage on the drain of device M1 varies from 1 V to 1.3 V and the voltage on the drain of device M2 varies from 1.3 V to 1 V in 12 μs to approximate the voltages of Figure 7. Simulation of the input converters with the drain voltages added results in the differential current of Figure 9 .

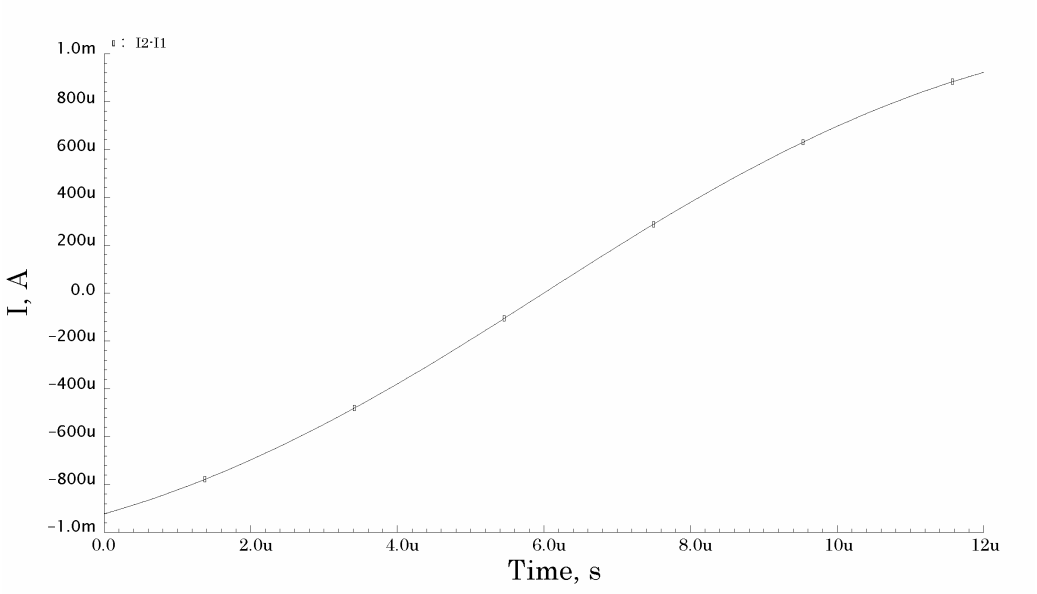


Figure 9: Converter output current with nonlinearity.

The current in Figure 9 has an end-point nonlinearity of 4.7%, which approaches the limit set for this circuit. As will be discussed in Chapter 4, the input device nonlinearity is the major cause of nonlinearity in the overall circuit. Because of the differential configuration of the weighting devices, some of this nonlinearity will cancel out in these devices resulting in a final nonlinearity below 4% for the final circuit output impedance as will be shown in Chapter 4. Therefore, these input voltage to current conversion devices are still linear enough to function for this circuit. Further discussion of the nonlinearity caused by finite output impedance and device operating region is found in Chapter 4.

## 2.2 FOUR-QUADRANT MULTIPLICATION

Figure 10 is a schematic of a basic four-quadrant transconductance multiplier. A similar configuration is used in the final circuit in this thesis. This configuration has many similarities to the configuration of the transconductance amplifier found in [7]. The two input devices function as input voltage to current converters, and the four lower devices are the differential weighting devices used for the multiplication of the current.

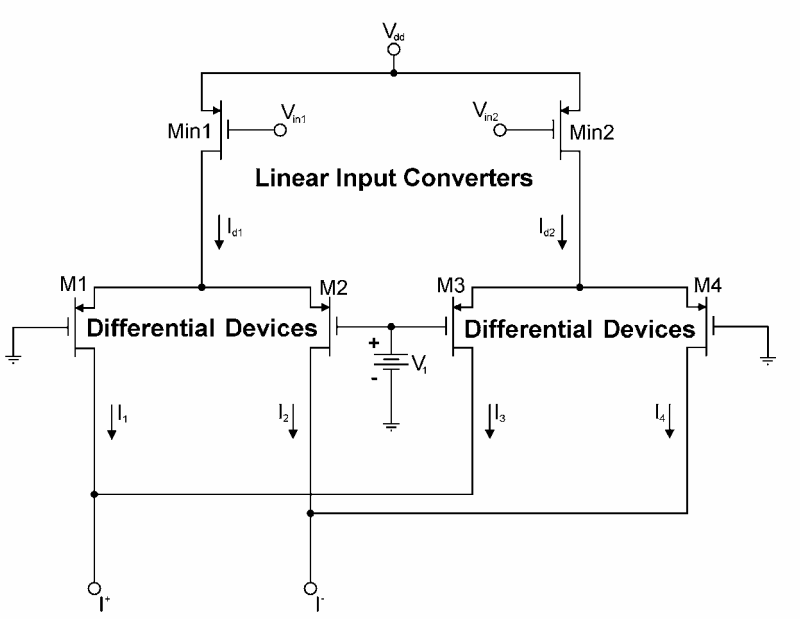


Figure 10: MOSFET four-quadrant multiplier.

The configuration of the mixed-signal multiplier of this thesis is based on the four quadrant multiplier of Figure 10, thus an understanding of its operation is important. The linear input converters Min1 and Min2 convert the applied differential voltage to a linear differential output current, $I_{diff} = I_{d1} - I_{d2}$. $I_{d1}$ and $I_{d2}$ are each applied to the source terminals of two differential weighting pairs consisting of devices M1, M2 M3, and M4. The differential weighting pairs are connected in the configuration of a Gilbert multiplication cell [8]. In a conventional Gilbert multiplier circuit, the voltage $V_1$ is used to control the gain of the differential stages and thereby may serve as a multiplicand. In the mixed-signal multiplier reported here, this voltage is set to a constant value and is unimportant to the operation of the multiplier except as a scale factor.

A derivation of the basic operation of the four-quadrant multiplier of Figure 10 is given below. One assumption made in this derivation is that the active region $I_d$ is found by the general expression suggested in [6] with a variable exponent between 2 and 1. The exact value of this exponent will then be discussed later in the derivation. This expression for $I_d$ is:

$$I_d = K(V_{eff})^{\alpha}. \tag{5}$$

By taking the derivative of (5) with respect to $V_{gs}$, we find an expression for the transconductance, *gm*. This expression is:

$$gm = \alpha K(V_{eff})^{(\alpha-1)}. \tag{6}$$

In (5) and (6), $K = \mu C_{ox} W \varepsilon_c$, $V_{eff} = (V_{gs} - V_t)$, and $\alpha$ is the general exponent found by fitting the curve of the drain current. For the TSMC 0.18 μm MOS device model the exponent $\alpha$ can be approximated at about 1.2 for the p-channel devices with $W$ = 16 μm. The exponent being greater than 1 explains part of the nonlinearity in the input voltage to current conversion. An interesting discovery in this research is that the operation of the four-quadrant multiplier depends on the active region current exponent being greater than one because of the approximations for *gm*. If the short channel devices truly had an exponent of one and could be represented as in (4), the equation for $g_m$ would no longer show any dependence on $V_{gs}$ and *gm* would become a constant. The constant *gm* would render the transconductance multiplier of Figure 10 useless as the output would no longer

be dependent on the input as can be seen in the derivations below. Thus (4) is useful as an approximation for the voltage to current conversion, but does not adequately represent all aspects of MOSFET operation necessary for the description of multiplier operation.

With the assumptions mentioned above, an expression for the four-quadrant multiplication can be derived. The output currents $I_{out} = I^+ - I^-$ are found to be:

$$I^+ = I_1 + I_3, \tag{7}$$

and

$$I^- = I_2 + I_4. \tag{8}$$

The expansions of these equations are:

$$I^+ = \frac{I_{in}^+}{2} + \frac{I_{in}^-}{2} + \frac{V_1 gm_1}{2} - \frac{V_1 gm_3}{2}, \tag{9}$$

and

$$I^- = \frac{I_{in}^+}{2} + \frac{I_{in}^-}{2} - \frac{V_1 gm_2}{2} + \frac{V_1 gm_4}{2}. \tag{10}$$

Where $\frac{I_{in}^{+}}{2}$ and $\frac{I_{in}^{-}}{2}$ are the drain currents coming from the input devices that are divided and sent down each leg of the lower weighting devices, and $\frac{V_1 gm}{2}$ in each of the equations represents the contribution to the output from the differential voltage. The *gm* of the lower differential weighting devices is determined by two different contributions. One contribution is that of the input device drain current coming through each lower device. This is shown as $gm_{n,in}$ for the n[th] weighting device. The other contribution to the weighting device *gm*, $\frac{KV_1}{2}$, is from half of the differential voltage applied to these devices dropping across each weighting transistor. The equations representing the two contributions to the weighting device *gm* are shown below:

$$gm_1 = K(V_{eff} + \frac{V_1}{2}) = gm_{1,in} + \frac{KV_1}{2}, \tag{11}$$

$$gm_2 = K(V_{eff} - \frac{V_1}{2}) = gm_{2,in} + \frac{KV_1}{2}, \tag{12}$$

$$gm_3 = K(V_{eff} - \frac{V_1}{2}) = gm_{3,in} + \frac{KV_1}{2}, \tag{13}$$

$$gm_4 = K(V_{eff} + \frac{V_1}{2}) = gm_{4,in} + \frac{KV_1}{2}. \tag{14}$$

In (11), (12), (13), and (14) $K$ is the drain current constant, $V_{eff}$ represents the contribution to the voltage from gate to source of the lower weighting devices caused by the current coming from the input devices minus the threshold voltage, and $V_1$ is the differential voltage. The distinctions between the two contributions allow us to represent the output more accurately. Equations (9)-(14) are combined to form:

$$I_{out} = \frac{V_1}{2}(gm_{1,in} + \frac{KV_1}{2} + gm_{2,in} - \frac{KV_1}{2}) - \frac{V_1}{2}(gm_{3,in} - \frac{KV_1}{2} + gm_{4,in} + \frac{KV_1}{2}). \quad (15)$$

The contributions in *gm* due to the differential voltage cancel out, and the result is:

$$I_{out} = V_1(gm_{1,in} - gm_{3,in}). \quad (16)$$

The last step in arriving at (16) from (15) comes from the assumptions that $gm_{1,in} = gm_{2,in}$ and $gm_{3,in} = gm_{4,in}$. For a general solution, an assumption was also made that the exponent for the drain current is $\alpha = 2$. This will be the long-channel general case. Another assumption is that the input transconductance is related to the transconductance of the differential weighting devices by the following expression: $gm_{in} = \sqrt{2}gm_{n,in}$. This assumption relates to the input current dividing almost evenly through the lower devices. With these assumptions, it follows that:

$$I_{out} = \frac{V_1}{2\sqrt{2}}(gm_{in}^{+} - gm_{in}^{-}) = \frac{K}{\sqrt{2}}V_1(V_{gs}^{+} - V_{gs}^{-}) = \frac{K}{\sqrt{2}}V_1(V_{in}^{+} - V_{in}^{-}). \quad (17)$$

In (17) $V_1$ is the differential voltage applied to the weighting devices and $K$ represents the MOSFET drain current constant, $K = \mu C_{ox} \frac{W}{2L}$. Therefore the multiplier built with long channel devices has the following relationship between input voltage and output current:

$$I_{out} = \frac{K}{\sqrt{2}} V_1 (V_{in}^+ - V_{in}^-) . \tag{18}$$

Equation (18) shows a perfectly linear relationship between current and voltage. The short-channel 0.18 µm devices with an exponent of $\alpha$ = 1.2 result in the following multiplication relationships:

$$gm_{input} = (2)^{-(\frac{0.2}{1.2})} gm_{weighting} , \tag{19}$$

and

$$gm_{weighting} \cong 1.2K(V_{gs} - V_t)^{0.2} . \tag{20}$$

A more realistic equation for output current can be derived using (19) and (20). An assumption about (20) needs to be made for this derivation. The assumption is that the contribution from the differential voltage to *gm* is ignored and the transconductance of the weighting devices is assumed to be determined solely by the current from the input devices dividing evenly through the weighting devices. Equation (19) can then be used

to relate the input device *gm* to the weighting device *gm*. Equation (18) now translates to:

$$I_{out} = (0.673)KV_1((V_{in}^{+} - V_{dd} - V_t)^{(0.2)} - (V_{in}^{-} - V_{dd} - V_t)^{(0.2)}). \quad (21)$$

Equation (21) shows that the output current resulting from the use of short-channel devices results in less linearity compared to the long-channel general case. In simulations comparing a 2 μm gate length to a 0.18 μm gate length with everything else remaining the same, the 2 μm gate length results in 1% less endpoint nonlinearity. The increase in nonlinearity with the decrease in gate length is obvious from the exponents in (21).

This equation is useful in seeing the relationships between output current and input voltage, but might not be as accurate as a dc solution because of the large variations in transconductance. A dc solution is very difficult to derive and may involve more accurate representations of the current equations for a MOSFET. Attempts to solve for a dc solution show that there is a direct relationship between the source voltages across the lower differential weighting devices and the output current [9], but the full dc relationship between input voltage and output current of the circuit of Figure 10 is much more difficult to derive. The conclusion drawn in this research is that the relationships of the output current to input voltage as derived in (18) and (21) are useful in getting a general idea of circuit operation, but anything accurate must be derived from computer simulation.

As will be seen in the simulations of Chapter 4, the 1% nonlinearity added by the short-channel multiplication equation does not affect the output of the large signal

version multiplier as much as the nonlinear V-I conversion. Therefore output current is linear enough that it is useful for many applications. Because of the complicated nature of the short-channel devices, actual simulations will be more helpful in fully exploring the nonlinearity of the circuit then equations. One important idea that can be drawn from the work presented in this chapter is that the transconductance multiplier configuration may become less and less useful as MOSFET miniaturization moves the active region current exponent closer to one. This is listed as a topic for further research, to find out if the dependence of short-channel *gm* on $V_{gs}$ actually becomes less with MOSFET miniaturization.

# CHAPTER 3

## THE CODE SELECT BINARY WEIGHTING CIRCUIT

The mixed-signal multiplier discussed in this thesis, referred to as the Code Select Binary Weighting Circuit, is a combination of the four-quadrant multiplier configuration discussed in Chapter 2 and general MOSFET properties. An understanding of the four-quadrant multiplier is useful, as the general multiplication terms remain the same for the new multiplier configuration, but there are key differences in the two configurations. The multiplier configuration discussed in Chapter 2 is modified from a small-signal four-quadrant multiplier to a large-signal two-quadrant multiplier. The input voltage signal was changed for the mixed-signal multiplier to vary over the full possible input range, where the input voltage of the multiplier of Chapter 2 was assumed to be a small signal quantity. The differential voltage, $V_1$, is no longer used as an input voltage, but is simply used as a scale factor. With these changes the input voltage can be positive or negative, but both the differential voltage and the binary number multiplicand are kept positive. This makes the final configuration a large-signal two-quadrant multiplier.

This new large-signal multiplier configuration was originally discovered at Intel [10], and the results of the research leading up to this thesis are summarized in [11]. This chapter first presents the circuit in general. Once the multiplier configuration has been presented, improvements and changes made as a result of research are discussed.

## 3.1 GENERAL MULTIPLIER CONFIGURATION

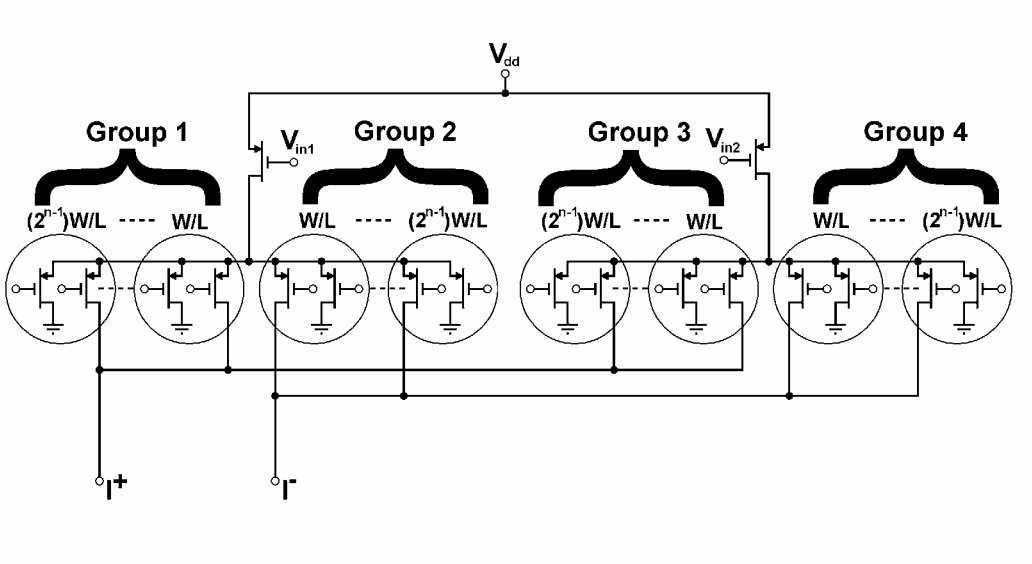


Figure 11: Multiplier with the n-bit code added.

The Code Select Binary Weighting Circuit is completed by replacing each of the lower differential weighting devices in Figure 10 with groups of transistors as shown in Figure 11. The weighting of the current from the input devices is accomplished by using binary-weighted transistor widths in each group of devices. These weighting devices occur in pairs which either send a portion of the total drain current from the input devices to the output or divert it to ground. The destination of the current entering the weighting device pairs is chosen by the voltage on the gate of each weighting transistor. The gates of the devices in groups 2 and 3 are connected either to $V_1$ to conduct current or $V_{dd}$ to turn the device off. Groups 1 and 4 use ground and $V_{dd}$ to turn the devices on and off respectively. One transistor in each pair conducts at any specified time. This preserves

the binary-weighting of the groups while allowing current to be diverted to ground or to the output.

The groups of weighting transistors added in Figure 11 result in a division and multiplication of the current coming from the input devices. Now instead of six devices in the multiplier, there is a number of devices represented by:

$$N_d = 2 + 8(n), \tag{22}$$

where $N_d$ represents the total number of devices in the multiplier and $n$ represents the number of bits in the weighting code. All except two of the transistors in (22) are found in the weighting groups, and half of the transistors in these weighting groups are always conducting current either to the output or to ground. This means that at any particular time one device in each pair of each weighting width in each group is conducting current. The conducting devices result in the current coming from the input devices being divided in a binary fashion through the branches of the weighting devices. This division changes with the number of bits in the weighting code.

The multiplication by the weighting code is determined by where the current in each pair of weighting devices is diverted. With a 11111 code on the multiplier, all the weighting devices with drain terminals connected to the output are conducting and all the devices with drain terminals connected to ground are turned off. The current at the output is then the same as the output current of the original multiplier configuration without the binary weighting circuit. As the weighting code is changed, a portion of the current is diverted to ground. The output current is a portion of the total expected

multiplier output current determined by the ratio of the multiplication and division caused in the weighting devices. The final differential output current is then found to be:

$$I_{out_{WeightingCircuit}} = I_{out_{FourQuadrant}} \times \frac{\sum_{i=1}^{n} b_{i-1} 2^{i-1}}{\sum_{i=1}^{n} 2^{i-1}}, \tag{23}$$

where $b_0, b_1, \cdots, b_{n-1}$ are the code bits. From (23) a more specific output current equation can be derived. The output of the original multiplier with ideal long-channel devices is given as:

$$I_{out_{FourQuadrant}} = KV_1V_{in}, \tag{24}$$

where $K$ is a constant formed in the circuitry, $V_1$ is the differential scaling voltage, and $V_{in}$ is the differential input voltage. The substitution of (24) into (23) results in:

$$I_{out} = KV_1V_{in} \times \frac{\sum_{i=1}^{n} b_{i-1} 2^{i-1}}{\sum_{i=1}^{n} 2^{i-1}}. \tag{25}$$

$K$ contains a negative sign because of the inverting configuration of the input devices. The output is an inverted version of the input, but reversing the differential output or including an inverter can counteract this negative if needed for the application. Equation

(25) for the final multiplication is verified using Cadence simulations in Chapter 4 of this thesis.

In the actual circuit, a five bit code is used. It was determined that the end-point nonlinearity of the multiplier should be kept under 5% for use with the linear equalizer and general applications. This percentage needs to be determined and matched to each specific application, but a nonlinearity of 5% should act as an appropriate general approximation. The 5-bit code leads to a quantization error of $\frac{1}{2^{n+1}} = \frac{1}{64} = 1.5\%$, and this code length should be sufficiently accurate to limit the overall nonlinearity to 5%. A completed version of the circuit is shown in Figure 12. The voltage inputs are the gates of the input converter devices. The current from the input devices travels through the differential pairs which now consist of 4 groups with 10 devices each.

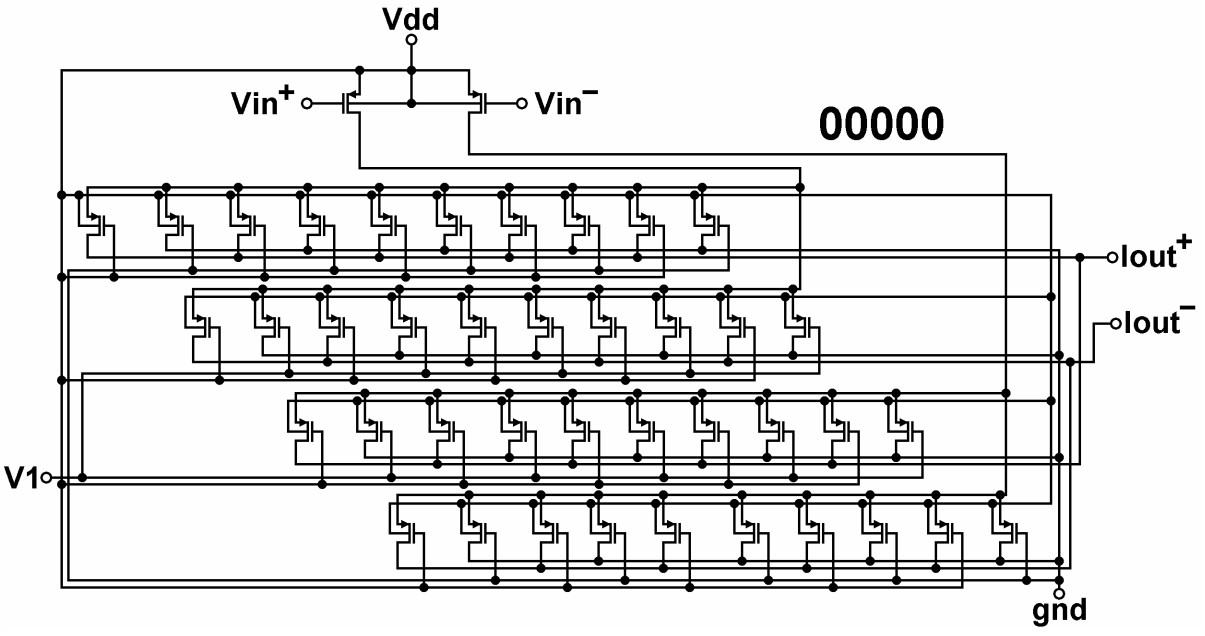


Figure 12: Actual multiplier schematic.

The binary code is set on the weighting devices by the voltage on the gate of these devices. As mentioned above, the gates are set to $V_1$ or ground to turn on the devices depending on which weighting group they are in, and $V_{dd}$ to turn them off. In Figure 12 the gates of the weighting devices are all hardwired to have a 00000 code on them. With a 00000 code there is no current at the output of the circuit as all of the current is diverted to ground. The gates of the weighting transistors are all hardwired for the simulations and implementations of the multiplier in this thesis. $V_1$ is used to scale the output and is kept fairly low to avoid large currents and voltages on the output of the multiplier. Ideal resistors with very low values (50 Ω) are used to sum the currents at the output of the multiplier for all simulations.

One of the most important features of this circuit is the constant capacitive and resistive loads driven by the input devices. Independent of the code on the weighting devices, there is always the same number of weighting devices conducting current at any time. With the same number of weighting devices with the same device width conducting current, the amount of capacitance and transconductance seen by the input devices remains constant independent of weighting code. The constant capacitive and resistive load allows the multiplier speed to remain constant over all weighting codes. The limiting capacitances for the circuit are the gate-to-source and bulk-to-source capacitances, $C_{gs}$ and $C_{bs}$, of all of the weighting devices. These capacitances see an equivalent impedance of all the parallel $\frac{1}{gm}$ impedances from all the weighting devices. With the capacitances and impedances remaining constant over all weighting codes, the

sources of the weighting devices present a low impedance, high speed node where the time constant remains the same independent of weighting code.

## 3.2 CIRCUIT IMPROVEMENT

All of the devices in the circuit have a gate length of 0.18 μm for the fastest possible transient response. The width of the devices is important to the performance of the circuit, and ends up being the parameter that affects circuit operation more than any other parameter. Through simulation it becomes apparent that there is an optimum width for both the input devices and the weighting devices. The pairs of weighting devices in each weighting group must be mathematically related to the width of the smallest pair of devices in a binary fashion as shown in Figure 11 to obtain the correct binary weighting. However, this does not fix the width of the devices in the weighting groups. The width of the smallest pair of transistors can be changed and all the transistor widths in the weighting groups can then be changed to match in a binary fashion. The optimum device widths are found in Chapter 4 through running multiplier simulations.

Beside the improvement to circuit operation with device size, there is another improvement to speed that can be made to the circuit as a result of research concerning substrate bias. The circuit analyzed in this thesis originally had a bulk to source connection for all the transistors. This work improves on this configuration by connecting all the bulk contacts to $V_{dd}$. This increases the speed of the multiplier, while not affecting the accuracy noticeably. The circuit in Figure 12 already has this improvement implemented.

This TSMC process uses a p-type substrate and therefore p-devices are fabricated in n-wells and n-devices are fabricated directly on the p-type substrate. This type of process forces the bulk contacts of the n-devices to be connected to ground, but there is a choice on where the bulk contacts of the p-devices are connected. The bulk of the p-devices can be connected to $V_{dd}$ or to the source contact depending on what is needed for the circuit. In this implementation, All of the bulk contacts of the transistors are wired to $V_{dd}$. The bulk to $V_{dd}$ connection introduces a positive bias, $V_{bs}$, that increases the switching speed of the transistors [12]. Although the threshold voltage increases with the bulk to $V_{dd}$ connection, the capacitance seen from the source terminal lowers enough to increase the speed.

The increased speed with this new configuration comes because the capacitance looking into the bulk terminal no longer adds into the capacitance node of the circuit, and the total capacitance of the overall circuit is lower. The limiting capacitance of the bulk to $V_{dd}$ configuration is $C_{gs}$, but the limiting capacitance of the original bulk to source circuit was the n-well to p-substrate junction capacitance. All devices in this new circuit are PMOS devices and are fabricated in n-wells on a p-substrate. This well to substrate capacitance depends directly on the area of the wells and the doping of the well and substrate areas. Therefore this circuit containing 42 PMOS devices will have well area associated with each device in the circuit. This means that the total area of n-wells associated with this circuit will be relatively large. The junction capacitance associated with these wells will also be large, limiting circuit speed to 4.8 GHz. In practice, this well capacitance is on the same order as the limiting capacitance of the bulk to $V_{dd}$

circuit, $C_{gs}$, and therefore the speed nearly doubles with the removal of this well capacitance.

It is significant to note that $C_{db}$ is also taken out of the circuit with the new configuration and the output impedance of the input devices is also lowered. These effects are minor in comparison to the well capacitance and only present about a 10-20% difference in speed compared with the doubling effect of the well capacitance. The capacitances are discussed more fully in the appendix, and the bulk to $V_{dd}$ connection is simulated along with the improvement in speed verified in Chapter 4.

# CHAPTER 4

# CIRCUIT SIMULATION AND OPTIMIZATION

This chapter describes simulations of the Code Select Binary Weighting Circuit to characterize its operation. These simulations are used to verify the findings in previous chapters, and to optimize circuit performance. The chapter includes simulations and descriptions of circuit non idealities and the causes of nonlinearity.

## 4.1 GENERAL SIMULATIONS

For the multiplier configuration of Figure 12 there are two main simulations which characterize general circuit operation. The first is a slow transient simulation that tests the low speed linearity and accuracy of the multiplier. The other simulation is a faster transient testing the fastest possible speed of the multiplier. Similar circuits to Figure 12 are hardwired with all the possible codes and simulated at the same time for comparison. Thus in every simulation all of the possible binary code configurations are accounted for. The differential voltage, $V_1$, is kept at 0.018 V for all the simulations. The voltages at the input of the multiplier are generated using piecewise linear voltage sources.

For the slow transient simulation the input voltage swings from 0.8 V to 0.2 V in 12 μs for the positive input and 0.2 V to 0.8 V in 12 μs for the negative input. A Cadence transient simulation monitors the output current for 12 μs and the differential current is

summed on the two 50 Ω resistors. This differential voltage across the two resistors is plotted versus time. Figure 13 is a plot of this simulation run for all possible binary codes.

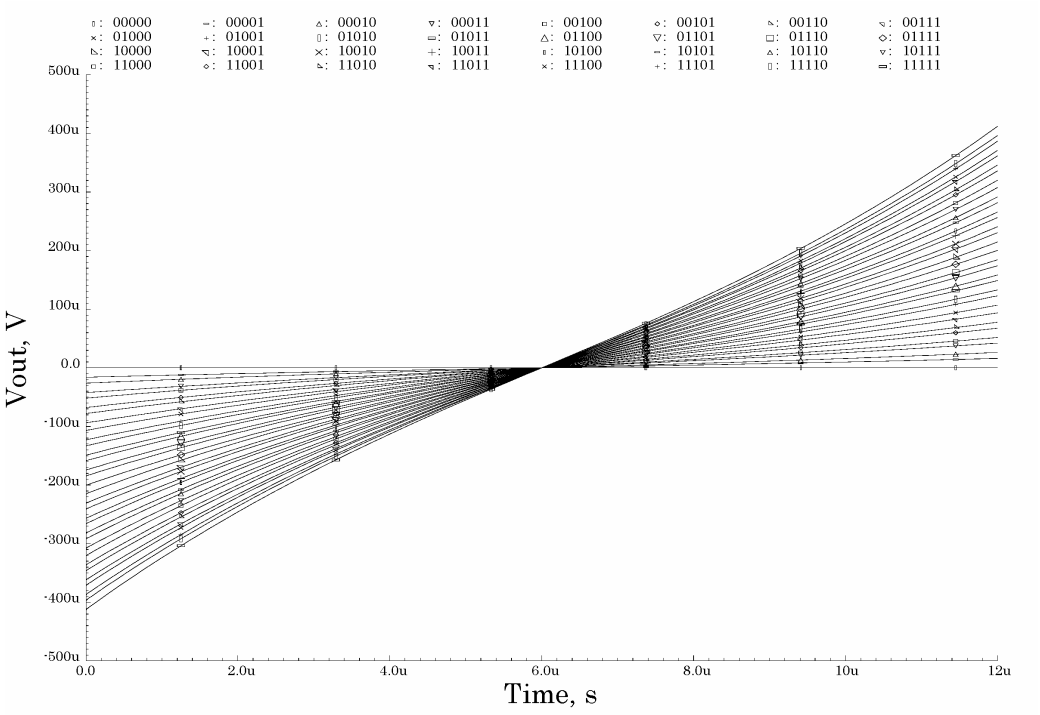


Figure 13: Circuit output for all possible weighting codes.

The results of the fast transient simulation are found in Figure 14. This simulation runs the multiplier at its fastest possible speed. The positive input for Figure 14 is a voltage that starts at 0.8 V and swings down to 0.2 V in 1 ps. The voltage remains at 0.2 V for 109 ps and then returns to 0.8 V in 1 ps. The negative input voltage consists of the opposite voltages at the same times. This creates a differential voltage at the input that swings from 0.6 V to -0.6 V in 1 ps, remains there for 109 ps, and then returns to 0.6 V in 1 ps. This pulse simulates a 9 GHz digital pulse at the input of the multiplier. The

transient simulation of Figure 14 is run for 500 ps so that the full positive and negative swings of the output current are included. The outputs are voltages summed on the output resistors.

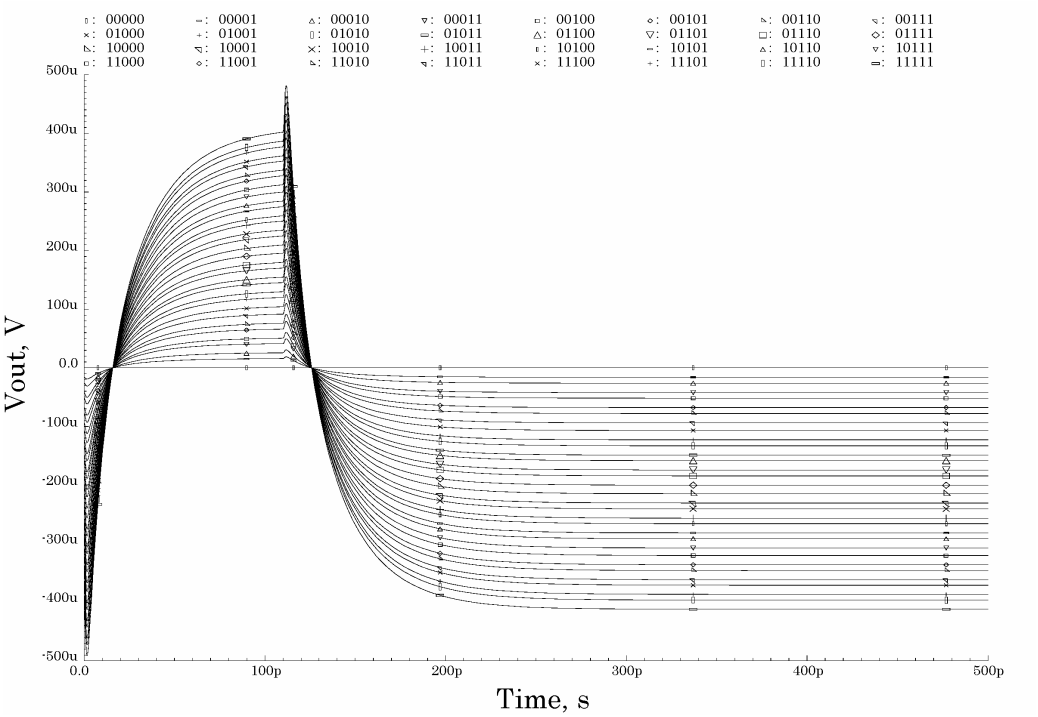


Figure 14: Transient output for all possible weighting codes.

In analyzing these simulations, there are two major concerns: the linearity and the speed. The output voltages in Figure 13 can be measured for end-point nonlinearity to determine the accuracy of the multiplication. The outputs in Figure 14 can be analyzed to determine the top speed of the multiplier.

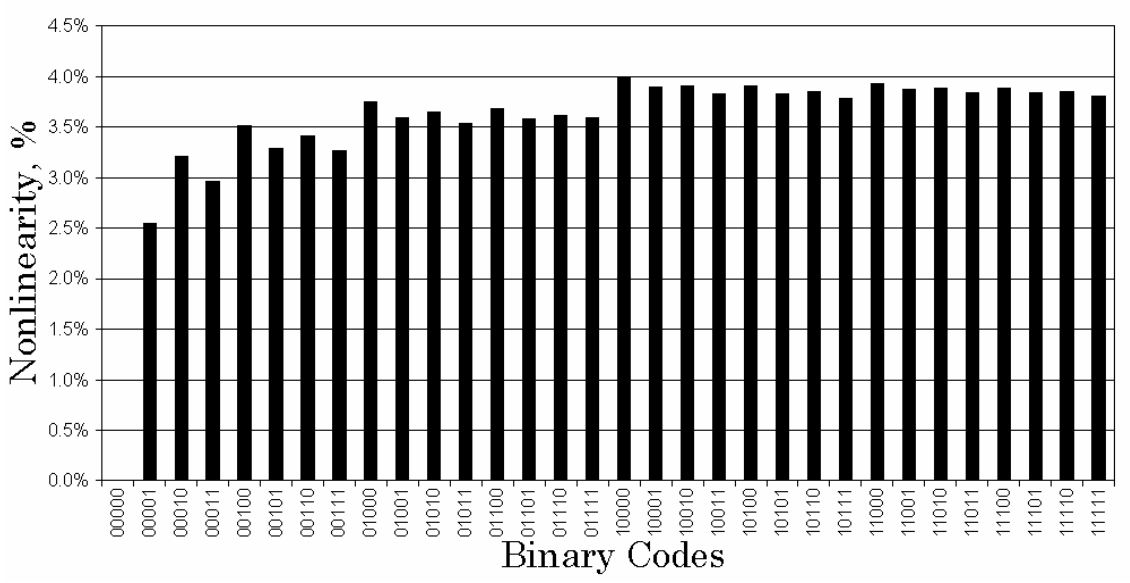


Figure 15: End-point nonlinearity for all possible weighting codes.

Figure 15 is a graph of the end-point nonlinearity for each of the simulations of Figure 13. An end-point nonlinearity measurement is taken for each output of Figure 13 in the same manner as described in Chapter 2. The outputs of this multiplier swing slightly below and above the ideal linear tangent line. None of the outputs for the binary codes in Figure 15 have end-point nonlinearity above 4%. It is possible to see that there is some difference in linearity over the weighting codes, but the difference is very minor, and comes from minor cancellation of nonlinearity with some combinations of weighting devices. The causes of this nonlinearity are discussed later in the chapter. These simulations assume that 5% end-point nonlinearity is the highest allowable nonlinearity measurement before the circuit operation is no longer accurate enough for most applications.

The speed that is derived from Figure 14 comes from measuring the percentage of the target voltage reached in a determined period of time. The traces of Figure 14 are all

charging to a specific value determined by the multiplication of the input and the binary code on the output. To consider the top speed of the multiplier to be 9 GHz, all of the traces must reach 98% of the final value that they are charging to within the 111 ps. In effect, the circuit must have a time constant small enough so that 4 time constants are included in the specified pulse width. All of the traces in Figure 14 reach at least 98% of their final target value in 111 ps, signifying 9 GHz operation. This 9 GHz speed is a result of the time constant at the source node of the weighting devices. The $C_{gs}$ and $C_{bs}$ of the weighting devices see the parallel $\frac{1}{gm}$ from all the weighting devices in parallel with the output impedance of the input converter devices, $r_{ds}$. Assuming that $r_{ds}$ of the input devices is large, the impedance at the source of the weighting devices is just a sum of the transconductances of all the weighting devices. The equation to predict device cut-off frequency, $f_c$, can then be approximated as:

$$f_c \cong \frac{1}{2\pi(C_{gs}+C_{bs})(\sum \frac{1}{gm_{WeightingDevices}})}. \tag{26}$$

Equation (26) should apply to either side of the multiplier consisting of one differential input device, and the weighting devices connected to it. Both sides of the multiplier should have roughly the same speed. Further investigation of multiplier speed is found in the appendix.

## 4.2 DEVICE WIDTH SIMULATIONS

The widths of the devices affect the linearity and speed of the output. The input devices can have a width from 0.22 µm to 100 µm in this TSMC process. The weighting device widths all depend on the width of the smallest transistor in each group. The smallest transistor width can vary between 0.22 µm and 6.25 µm before the width of the largest device, which is 16 times the width of the smallest device, reaches the process limit of 100 µm. The ideal widths chosen for the simulations of Figure 13 and Figure 14 were chosen through a series of simulations demonstrated in the next four figures.

Figure 16 is a plot of the end-point nonlinearity of the output versus the width of the two linear conversion devices at the input. This simulation predicts the linearity of all the possible outputs at all the possible input widths.

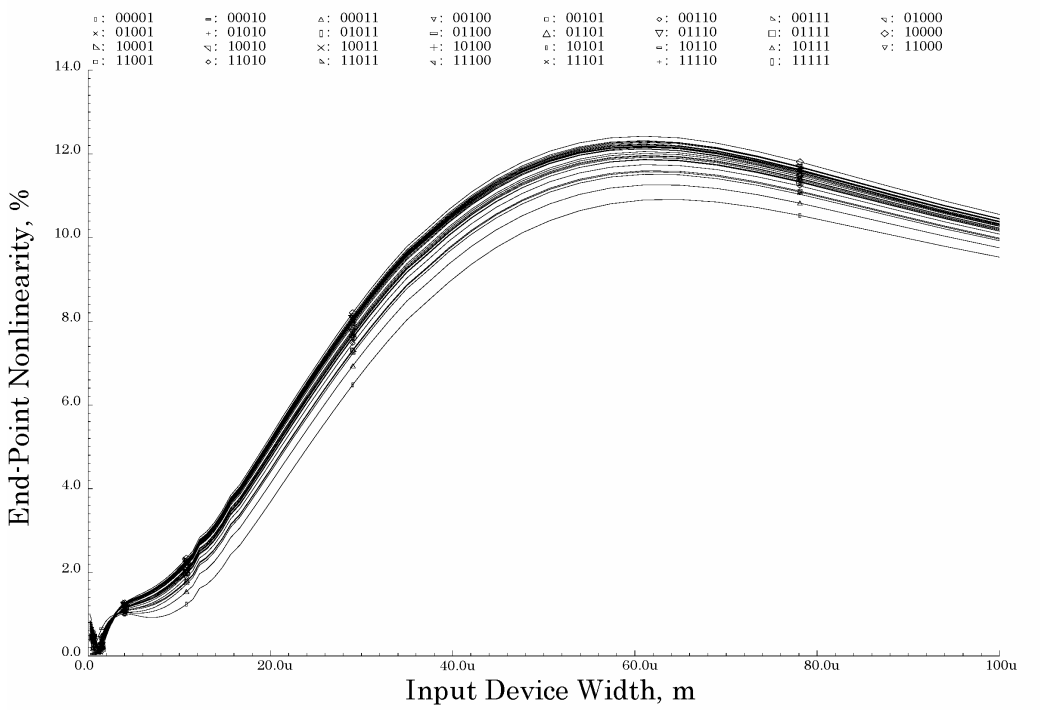


Figure 16: End-point nonlinearity versus input transistor width.

The general trend of Figure 16 shows the circuit becoming more nonlinear as the size of the input transistors increase. As the size of the input devices increases, the current feeding the weighting devices increases. This comes because of the direct dependence of drain current on the width of a device as can be seen in (4). As the current feeding the weighting transistors becomes larger, the voltage developing on the sources of the weighting devices becomes larger. This source voltage is the drain voltage seen by the input devices. When this drain voltage increases, the input devices are driven further and further out of the active region. The input voltage range becomes limited, and after about 60 μm or larger for the input devices, the nonlinearity measurements taper off and stop growing. This comes because the input voltage range for the input converter devces is limited to the point that the nonlinearity measurements are no longer meaningful, as the current converters are no longer usable with the resulting small input range. There is one dip in the nonlinearity around W = 1 μm, caused by a combination of input converter size and weighting transistor size being ideal for all devices to operate in the active region of conduction, but such a small input device size is not desirable for the higher speeds. The simulation of Figure 16 suggests that the input transistors need to be roughly 20 μm or smaller to keep the end-point nonlinearity below 5%.

Figure 17 is a plot of the percentage of the output final target voltage reached versus input transistor width in a transient simulation. Figure 17 shows that the width of the input devices must be larger than 10 μm to ensure that all of the outputs reach 98% of their final voltage. Increasing the input device width increases the current in the weighting devices, and therefore the resistance of the weighting devices decreases and speed increases. The simulations of Figure 16 and Figure 17 suggest that for optimum

circuit operation at 9 GHz, the two input devices should have a width between 10 and 20 μm. A device width of 16 μm is chosen for the input transistors and allows 9 GHz operation of the circuit with acceptable nonlinearity.

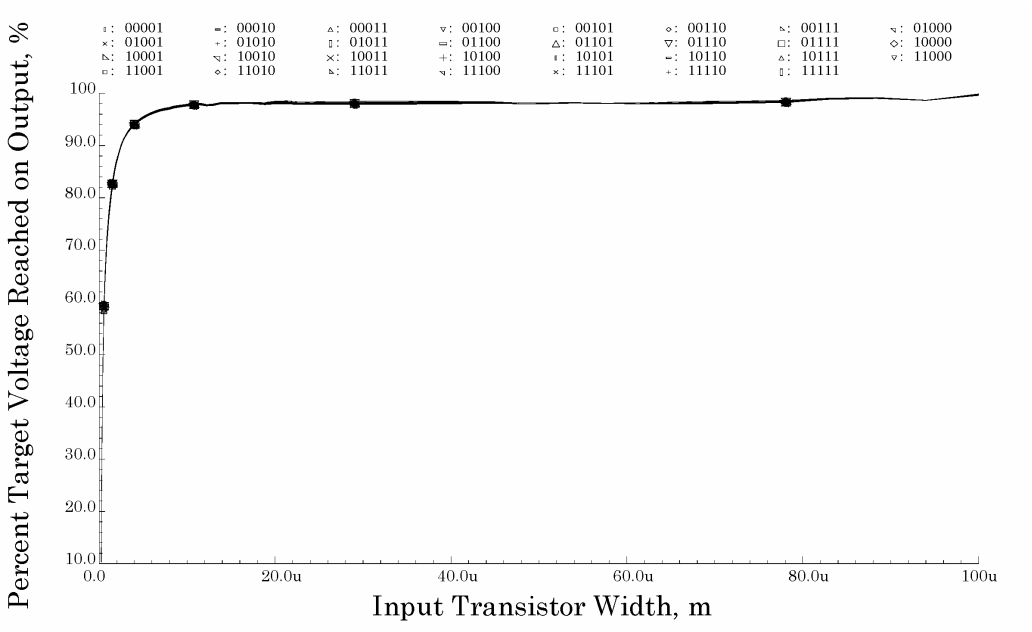


Figure 17: Percentage of target voltage reached versus input transistor width.

The width of the weighting devices is also a good way to improve circuit performance. The simulations of Figure 18 and Figure 19 vary the smallest transistor size from 0.22 μm to 6.25 μm, since the width of the rest of the weighting transistor is related in a binary fashion to the smallest. At 6.25 μm for the smallest transistor, the largest transistor is 100 μm wide.

Figure 18 shows a plot of the end-point nonlinearity versus the width of the smallest of the 40 weighting devices. Figure 18 shows a minimized nonlinearity at around 3 μm, while the nonlinearity increases in either direction. While none of the nonlinearities are above 5%, the nonlinearity increases rapidly as the devices move

toward the minimum size. The decrease in width causes the transconductance to become smaller and therefore the impedance at the source nodes is higher. This higher impedance causes more voltage to develop on the drains of the input converter devices. This voltage will contribute to nonlinearity in the voltage to current conversion at the input. The minimum in nonlinearity at 3 μm shows an ideal device width for accuracy and nonlinearity cancellation. As the width increases past 3 μm, there is a slight increase in nonlinearity due to the decreasing bias voltage developing across the weighting transistors for a given current coming from the input devices. Since the drain current is fixed, the increasing width results in less voltage developing across the weighting devices This decreasing bias could reach the point where the weighting devices leave the active region of conduction if the width were able to go any smaller.

The minimum size device was chosen for the smallest weighting devices because of speed, but is not usually desirable, as the small devices are difficult to match. As the device width shrinks, the contacts become wider that the device itself. These wider contacts result in a "dog bone" layout causing more capacitance at the drain and source junctions. In this case, the capacitance for the junctions is not the limiting capacitance, and the smallest devices can be used. If more accuracy were needed for a specific application, increasing the width of the weighting devices would improve the linearity of the output current.

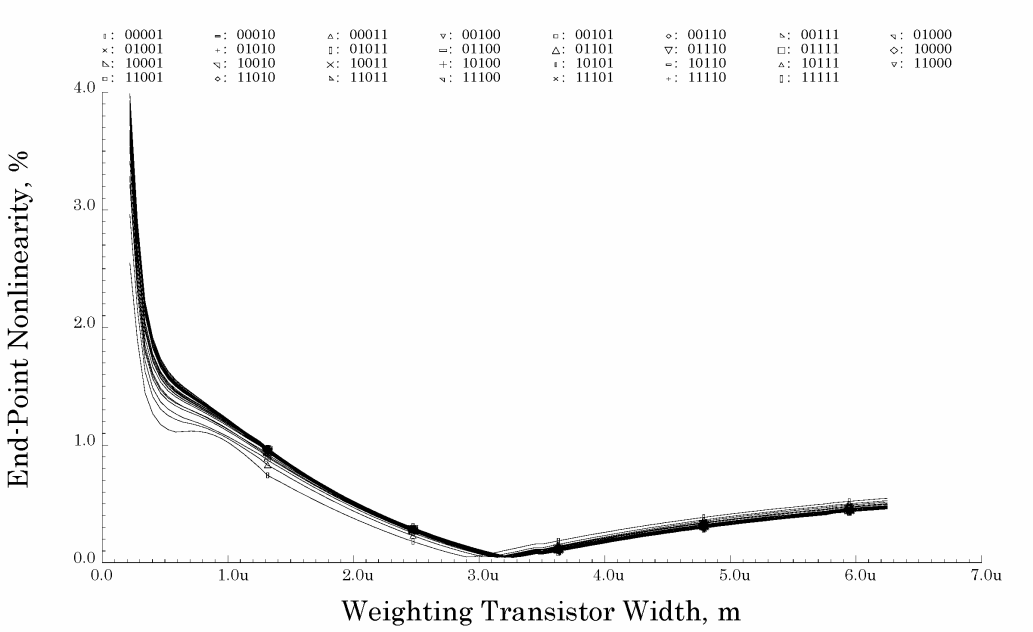


Figure 18: End-point nonlinearity versus weighting transistor width.

In Figure 18 the end-point nonlinearity never reaches 5% or above. The more important simulation for the binary weighting transistor width is found in Figure 19. It shows that the speed of the circuit increases with smaller device widths. The percentage of the target output voltage reached by the circuit in the allotted time approaches 100% as the width decreases. This means that the only limitation on the width of the weighting devices is the process design rules. At 0.22 μm, the smallest width for the TSMC process, the end-point nonlinearity is below 4% and the speed is at 9 GHz. This trend in speed and the width of the weighting devices is to be expected since making the devices smaller lowers their capacitances, and this circuit requires the lowest capacitances and time constants possible for the weighting devices.

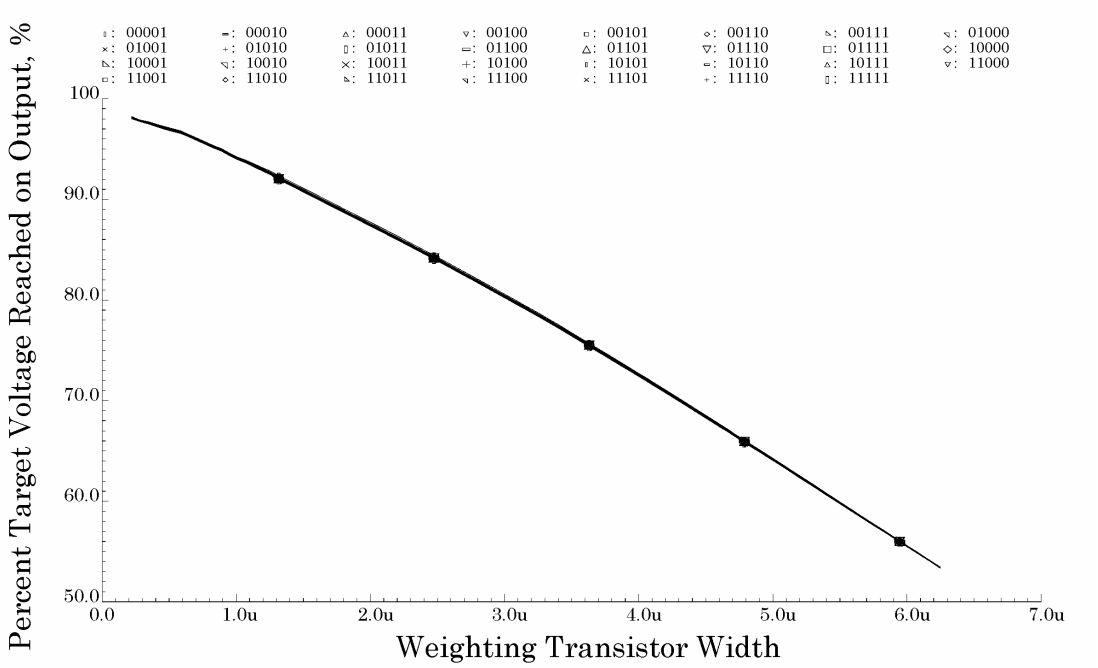


Figure 19: Percentage of target voltage reached versus weighting transistor width.

The previous four simulations suggest that the best device widths are 16 µm for the two input devices, and 0.22 µm, 0.44 µm, 0.88 µm, 1.76 µm and 3.52 µm for the pairs of weighting devices. All device gate lengths are held at 0.18 µm to preserve the speed of the transistor process. With the 0.18 µm process it appears that the fastest possible speed is 9 GHz, but speed improves with smaller gate lengths and transistor widths.

## 4.3 SUBSTRATE BIAS SIMULATIONS

The $V_{dd}$ connection for the bulk ends up being more desirable than connecting the bulk to source because the $V_{dd}$ configuration increases circuit speed from 6.7 GHz with bulk connected to source on all transistors to 9.2 GHz with the current configuration using a 11111 code. The increase in speed is significant while the increase in nonlinearity is less than a percent. Figure 20 shows the improvement in the fast transient simulation. Figure 21 shows the slight degradation in linearity with the substrate bias change for all the possible weighting codes.

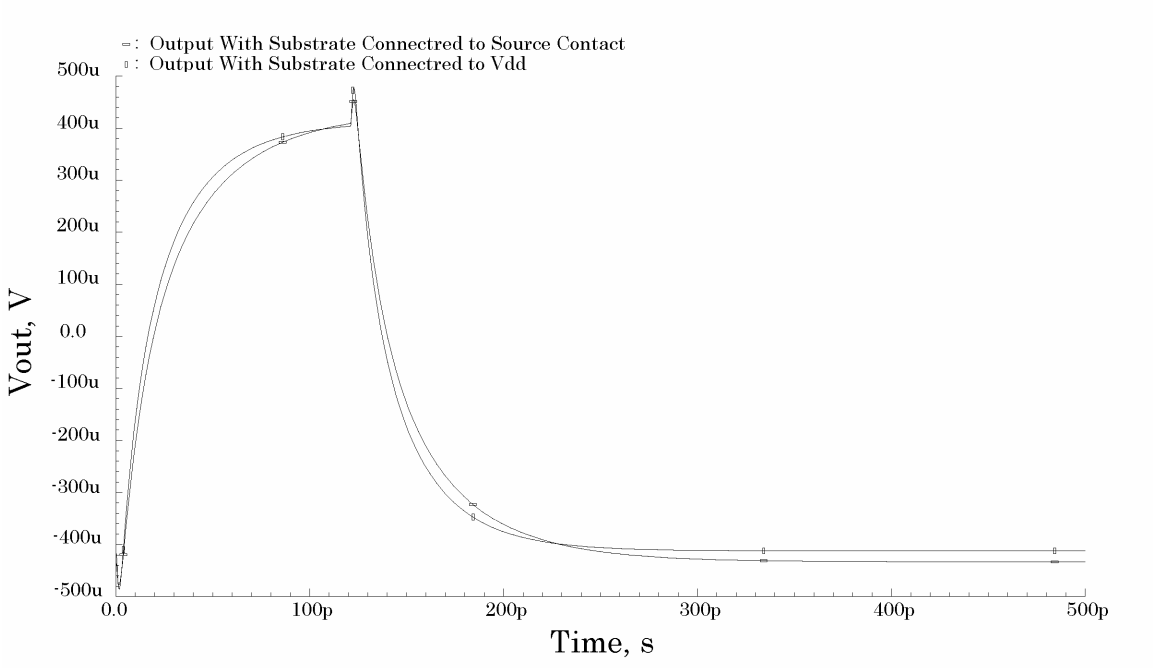


Figure 20: Comparison of transient for substrate bias change.

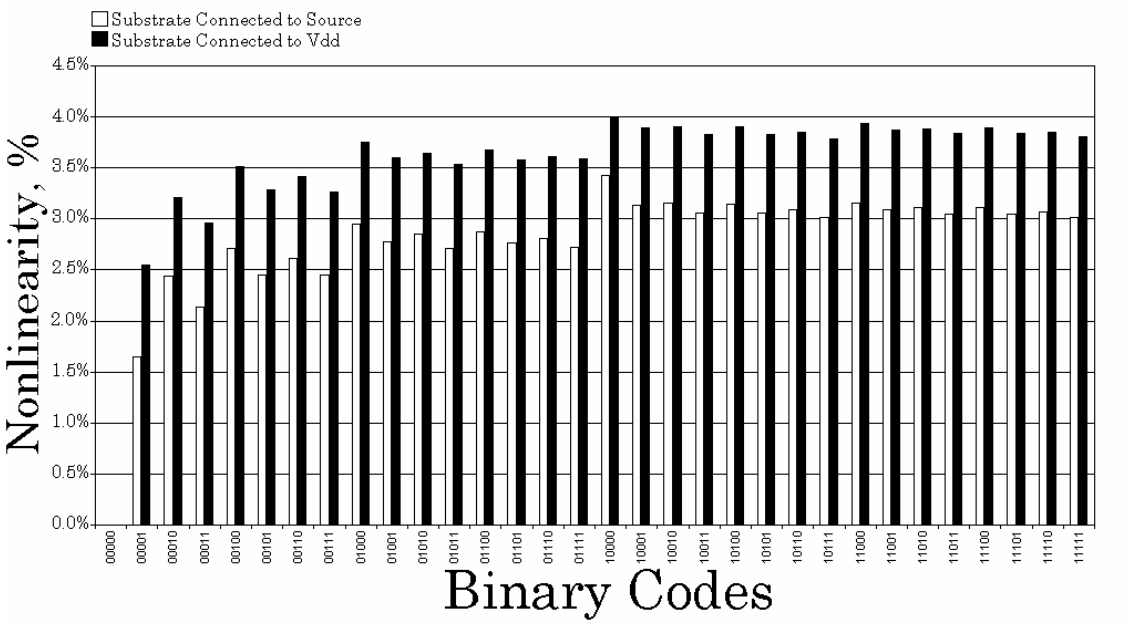


Figure 21: End-point nonlinearity comparison for substrate bias change.

As discussed before, the increase in speed comes from the well-to-substrate and drain-to-bulk capacitance being taken out of the circuit and the decrease in the output impedance of the input devices. In the simulation above, the well capacitance is not modeled. The well-to-substrate capacitance simulation is covered in the appendix. The increase in nonlinearity comes from the nonlinearity of the varying threshold voltage now that the body effect has been introduced in the weighting devices. This body effect increases the threshold voltage with the increasing bulk-to-source voltage, $V_{bs}$. This increased threshold voltage causes a larger drain voltage for the input converter devices, and therefore more nonlinearity in the voltage to current conversion.

## 4.4 CAUSES OF NONLINEARITY

Since linearity is one of the major measurements of circuit performance, the causes of nonlinearity merit more detailed discussion. There are three main causes of nonlinearity that affect circuit operation. The first cause is the nonlinear voltage to current conversion of the input devices. The second cause of nonlinearity is the nonlinearity of the short-channel devices. And the third major cause of nonlinearity is the operating region of the input devices. Each of these causes will be briefly explored in this section.

The output impedance of the input devices has a nonlinear effect on the voltage to current conversion as demonstrated in Chapter 2. This effect causes most of the overall linearity of the multiplier. The current generated by the input devices drives the impedance of the weighting devices, $\frac{1}{gm}$, causing a voltage to develop on the drains of the input transistors. This voltage increases with increasing current, and therefore this voltage changes with the input devices. This means that the drain to source voltage, $V_{ds}$, of the input devices changes with the current coming from the input devices. This changing $V_{ds}$ causes a change in the current of the output devices according to the output impedance of the input devices. The less output impedance a device has, the more the drain current will vary with $V_{ds}$, and the more impedance the input devices have, the less the drain current will be affected by the changing $V_{ds}$.

As the current of the input devices varies with the input voltage and $V_{ds}$ at the same time, a nonlinear curve develops in the output current of the input devices. This curve will become more pronounced as the impedance of the input devices decreases. The drain current coming from the input devices will become more linear if $r_{ds}$ increases.

The linearity of the multiplier could be greatly increased if this impedance could be increased. One solution is adding source resistors to the input devices, but this would take away from the possible input range, and resistors are often not very accurate in VLSI processes. Another solution might be to increase the length of the input devices, as resistance is proportional to device length. This will also increase the time constant of the circuit by increasing the device capacitance and impedance, but there is a trade-off between linearity and speed. Since none of the input capacitances are limiting capacitances there may be some room for improvement with minor changes in circuit speed. For the simulations in this circuit, the lengths of all the devices are left at the minimum length for the fastest possible speed.

As discussed in Chapter 2, the 0.18 μm MOSFET devices are not completely linear in voltage to current conversion in the active region of conduction. The devices have an exponent of 1.2 in the active region for the drain current. As discussed also in Chapter 2, the multiplier configuration equation as shown in (21) shows the final effect this exponent has on the output. The output is nonlinear. This device nonlinearity is partially cancelled by the differential configuration as can be seen in the differences in nonlinearity in the currents of Figure 3 and Figure 6. Subtracting nonlinear currents as done in Figure 3, the very linear differential current of Figure 6 is produced. As can be seen in Figure 15, the active region current exponent nonlinearity does not become a major factor because of the very linear Gilbert cell multiplier configuration.

An important assumption for the operation of the mixed-signal multiplier of this thesis is that all transistors that are conducting current operate in the active region. This assumption will always hold true for the weighting devices when they are turned on and

conducting current. This is because of the very low impedance summing node assumed for the output of the circuit. By keeping the output currents small, the output current of many devices can be summed without developing large voltages on the drain terminals of the weighting devices. The differential voltage, $V_1$, is kept low, 0.018 V, to make sure that the output current remains low enough to keep the weighting transistors in the active region.

The nonlinearity caused by devices leaving the active region is a problem in the input linear converter devices. The source voltage that develops across the weighting devices becomes very important. As was stated in Chapter 2, this source voltage, which directly changes $V_{ds}$ for the input devices, must be kept constant. Not only does holding this voltage constant reduce the small-channel effect DIBL, which will affect the output resistance for the input device, but $V_{ds}$ must be kept large enough that the input devices do not enter the triode region. The triode region and the active region both have an approximately linear voltage to current relationship but if the input devices start switching between the two regions, the current conversion will be nonlinear.

One problem that has come up with the new mixed-signal multiplier configuration is that the bulk to $V_{dd}$ connection causes the threshold voltage to increase from 0.5 V to around 0.6 V in the weighting devices. This increase of 0.1 V causes an increase in the weighting device source voltage. The increase from the original 4 bit investigations to the current 5 bit configuration also added up to 0.4 V to the weighting device source voltage. With all these effects totaled up, the source voltage is around 1 V to 1.3 V, and there is about a 0.3 V change in the source voltage. This high source voltage with a significant change adds to the nonlinearity. The change in source voltage also causes the

input devices to change from the active to triode region during the swing of the input voltage. Thus for the slow transient simulation, the input converter devices stay in the active region for about 4 µs during the 12 µs swing, and after 4 µs the input devices change into the triode region and remain there for the rest of the simulation.

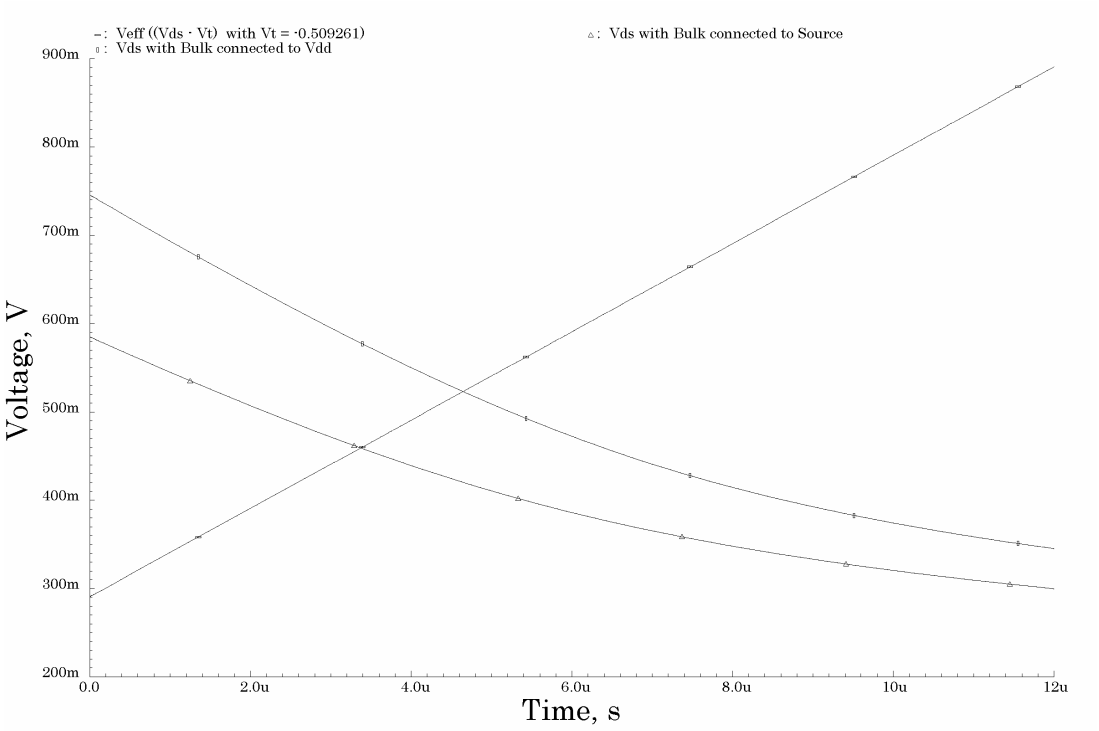


Figure 22: Simulation to determine input operating region.

Figure 22 shows $V_{ds}$ and $V_{eff}$ for the input devices. The voltage starting in the lower left and ending up in the upper right portion of the graph is $V_{eff}$. This simulation is done with the slow transient simulation, and the transistor being measured is the positive input transistor for the multiplier with a 11111 code on the weighting devices. Therefore the input gate voltage swings from 0.8 V to .2 V in 12 µs, and the source voltage is kept at $V_{dd}$. This configuration results in a $V_{gs}$ swinging from 0.8 V to 1.4 V. With a

Threshold voltage of about 0.5 V, the $V_{eff}$ ends up swinging from about 0.3 V to 0.9 V. The other two curves of Figure 22 are $V_{ds}$ measurements for two different cases. The top curve represents $V_{ds}$ for the multiplier with the bulk connected to the source, and the bottom curve represents $V_{ds}$ for the multiplier with the bulk connected to $V_{dd}$. This plot serves to demonstrate two things. The first is that the input converter devices operate in the active region until the $V_{ds}$ voltages cross the $V_{eff}$ voltage. The input devices only operate in the active region for between 3 and 4 μs of the 12 μs slow transient simulation, and then they pass into the triode region. Second, the plot shows the input devices would stay in the active region longer if the weighting devices had the bulk terminals connected to the source terminals.

In investigating this problem, several solutions present themselves, but the best solution may be to leave the devices as they are. The input range could be limited to 0.1 V instead of 0.6 V. The limited voltage would keep the input devices in the active region, but would significantly limit the operation of the multiplier. The gate voltage input range of each input converter transistor has already been limited to 0.6 V instead of 0.8 V in an attempt to limit the nonlinearity. Another solution might be to increase the source voltage by adding more weighting bits or increasing the current coming from the input devices. If the source voltage could be increased and controlled enough to keep the input devices in the triode region, the voltage to current conversion may become more linear. While both of the solutions above may increase the linearity of the multiplier somewhat, the current configuration uses differential devices that cancel out much of the nonlinearity. The nonlinearity measurements suggest that the current configuration can achieve very linear results even with the nonlinear operation of the input devices.

## 4.5 OTHER IMPORTANT SIMULATION VALUES

Linearity and speed are the most important simulation quantities that need to be verified, but there are a couple of other simulation values that should be mentioned. The first is the power consumption of the circuit. The other is the common mode current at the output. These two quantities are important to know when determining circuit functionality. There is also a simulation included in this section to verify the choice of differential voltage, as this voltage helps determine the output current magnitude and linearity relationships.

With VLSI circuits it is important to know how much power a circuit consumes. With the Code Select Binary Weighting Circuit the power supply is a 1.6 V power supply. This power supply must supply each multiplier with an average of 1.87 mA of current. During the cycle of the multiplier, the current varies between 1.6 mA and 1.9 mA. This signifies an average power consumption of 3 mW of power for each multiplier. The current may vary as much as 0.4 mA according to the weighting code present on the weighting devices, but the current generally stays around 2 mA for the multiplier. From the current measurements it is apparent that this multiplier is not meant to be used in low power situations. If the general configuration of the multiplier needed to use less current, then the input devices can be scaled down along with the speed of the multiplier. As is common with many circuits, the power consumption and speed are directly proportional to each other.

The common mode current at the output can cause the detection of the differential signal to become difficult. If the differential signal is very small, but the common mode voltage is large, a detection circuit may be overdriven by the common mode signal. This

may result in a signal that is undetectable. Like many other circuit quantities, this becomes more of an application specific problem. The detector chosen for the application would have to be analyzed to determine if common mode current is a significant problem. The common mode detection problem would depend on the architecture of the summing or detection circuit.

With the Code Select Binary Weighting Circuit there is a significant amount of common mode current at the output of the circuit. The input devices create this current when they convert the voltage to a current. The common mode current is then scaled by the weighting code. Measurements have been taken to measure the common mode current at the output using a slow 12 μs transient simulation with the differential input voltage Swinging from 0.6 V to -0.6 V in 12 μs. The output of the multiplier with a 11111 code, or the full amount of current from the input converter devices flowing to the output, shows a differential output current that varies from -8.24 μA to 8.24 μA. With this differential output current, there is a 932.2 μA average common mode current. This current is present for all weighting codes, but it is scaled by the weighting code. This means that the common mode current is 113 times larger than the differential current for all weighting codes. While this could cause significant error in the detection of the differential output, the common mode current is small enough that a detector could most likely detect it without overdriving the input.

One solution for the common mode current problem that could be used if needed to minimize common mode current is a pair of matched NMOS current mirrors on the two drain terminals of the input linear converter devices. The current mirrors subtract an equal amount of current from each side of the differential signal. While this would

minimize the common mode voltage, it would also add another capacitive node that might slow the speed of the multiplier down significantly. The decreased current in the weighting devices would also increase the resistance of the weighting devices, and this would significantly decrease the speed. This correction is probably only appropriate if it is really needed for the particular summing or detection circuit on the output of the multiplier.

The differential voltage, $V_1$, was set for 0.018 V for all of the simulations in this chapter. This value was chosen to keep the output current small enough that the voltage developing on the drain terminals due to any summing impedance does not become a factor in nonlinearity. If the drain voltage for the weighting devices becomes large, the finite impedance of the weighting devices would cause nonlinearity in passing the current from the input converter devices to the output node. The differential voltage also causes a larger drain voltage for the input devices. As discussed earlier, this larger drain voltage for the input voltage to current converters may cause them to leave the active region of conduction, and also causes more nonlinearity due to the finite output impedance of these input devices.

Figure 23 shows a plot of the end-point nonlinearity as a function of increasing differential scaling voltage, $V_1$. $V_1$ is swept from 5 mV to 500m V causing the nonlinearity to range from below 4% to 8%. The plot shows a direct increase in nonlinearity with increasing $V_1$. There is one slight discontinuity at around 20 or 30 mV caused by unequal spacing of the simulation points chosen, but the general trend is obvious. The increasing $V_1$ increases the drain voltage for the input converter devices. This increased drain voltage increases the nonlinearity due to the finite output impedance

of the input devices as discussed earlier in the chapter. To keep nonlinearity to below 5% the differential voltage should be kept in the range of 10 mV to around 100mV. Below 10 mV, the output current of the multiplier will be very small and probably unusable, and above 100 mV the nonlinearity begins to approach the limits chosen for this circuit. 18 mV is an appropriate choice for general simulation, but larger values could be chosen in practice.

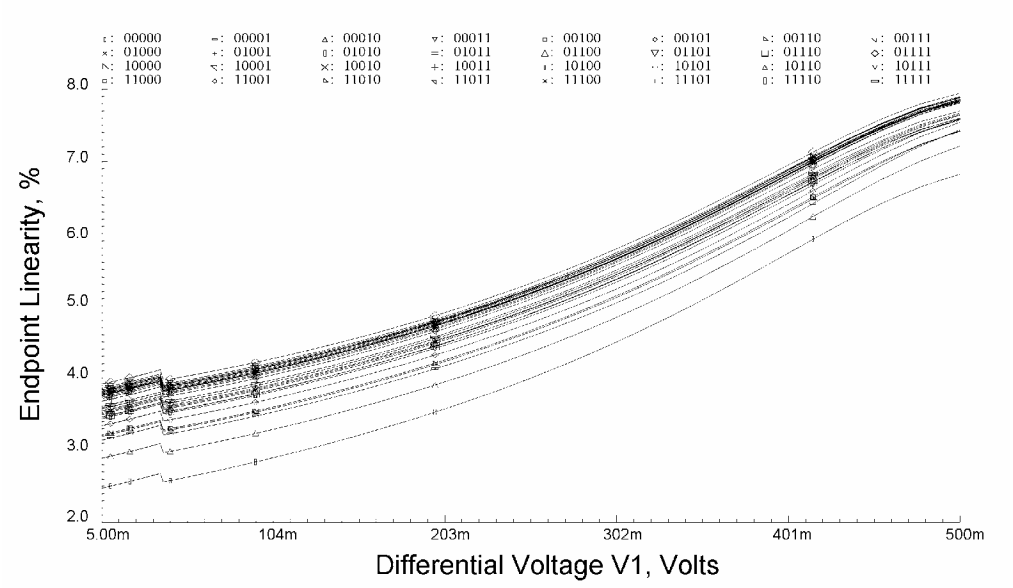


Figure 23: End-point nonlinearity versus differential scaling voltage.

# CHAPTER 5

## APPLICATIONS

Two main applications are mentioned in this chapter. The simulation and verification of applications of the multiplier are outside the scope of this thesis, as this thesis characterizes the general multiplier operation. Therefore the applications here are presented as ideas for further research and experimentation. The idea of linear equalization was mentioned in Chapter 1 and is expanded on here. The multiplier also performs the function of digital to analog conversion. These two applications are probably the most obvious and useful for the Code Select Binary Weighting Circuit.

## 5.1 LINEAR EQUALIZATION

An important subclass of filter circuits is that of linear equalizers. This type of circuit requires the formation of the sum of the products of voltages. For example, if $\{X\} = \{X_1, X_2, X_3, X_4, X_5\}$ is an input signal sampled at five different times and $\{W\} = \{W_1, W_2, W_3, W_4, W_5\}$ is a weighting vector with five values, the linear equalizer might form a result $Y$ given by:

$$Y = K \times (W_5 \times X_1 + W_4 \times X_2 + W_3 \times X_3 + W_2 \times X_4 + W_1 \times X_5). \qquad (27)$$

In (27) $K$ represents any constants formed in the circuitry. Multiplier circuits are necessary to form the product terms prior to summation. The multiplier discussed in this thesis performs the multiplication and provides an output current. The current mode output is more easily summed, and there is a circuit that has been proposed for this purpose in[13]. The proposed summing circuit has very low input impedance as needed to keep the weighting devices of the multiplier in the active region of conduction, and should work well in conjunction with the Code Select Binary Weighting Circuit.

As mentioned in Chapter 1, high-frequency digital communication signals degrade rapidly due to transmission line effects in the conductors that link circuits and chips. Some form of filtering or linear equalization is often used to restore the transmitted data to a level that leads to accurate reception. The linear equalizer may be required to form products in short times. A 10 GHz system will need to form the products in less than 100 ps. The multiplier in this thesis is directed toward applications with data rates of 9 GHz by forming a product in 111.11 ps or less.

In a linear equalizer, the weighting vector may be determined by earlier measurements or it might be continuously measured and updated at appropriate intervals. Often the vector is stored as a digital code. In the multiplier of this thesis, a digital code provides a corresponding weighting value that multiplies a sampled analog input value. This weighting vector is chosen to counteract the degrading response of the channel the signal is traveling on using mathematical algorithms. With the correct filter coefficients, the linear equalizer can completely get rid of all ISI (Inter-Symbol Interference) caused in the channel. While full ISI cancellation may be desirable, the response of the linear equalizer often amplifies high frequency noise making detection of the signal difficult.

There are different possible solutions to this problem that include different equalization techniques such as decision-feedback equalization and linear equalization that minimizes the mean-squared error instead of matching the channel response exactly[14].

It should be noted that while the multiplication circuit discussed here is designed primarily with equalization in mind, there is still research needed before implementing the equalizer circuit. Designing linear equalization and decision-feedback equalization circuits is not a trivial problem. Further research may show whether an implementation of a linear equalizer is realistic using the mixed-signal multiplier analyzed here.

## 5.2 DIGITAL TO ANALOG CONVERSION

As a topic for further research, the Code Select Binary Weighting Circuit may also be found useful as a digital to analog converter (DAC). The configuration of the multiplier is that of a multiplying DAC and could readily be converted to perform this function. There are several aspects of this multiplier that should be noted in respect to the DAC functions. The device size and configuration would need to be optimized with the DAC function in mind, and the limitations on speed and linearity might change.

The optimization performed for the thesis is done with the highest speed in mind while keeping the linearity high enough for basic performance. DACs often require very high linearity and so a re-optimization would need to be performed. The nonlinearity may cause significant errors in the digital to analog conversion. The Code Select Binary Weighting Circuit can achieve much lower nonlinearity but the speed will lower significantly. As is common among circuit architectures, the speed of this multiplier is

inversely proportional to the accuracy. These trade-offs need to be researched for the specific DAC applications where the multiplier is needed.

Also it is important to note that the number of bits in the DAC is very limited with this multiplier configuration. 4 or 5 bits may end up being the limit for this multiplier configuration, as the input devices are already somewhat limited in their input swing and active region operation. This configuration may limit the linearity possible with the circuit. Further research may show that this device is suitable for some limited DAC applications.

# CHAPTER 6

# CONCLUSIONS

This thesis introduces a mixed-signal CMOS multiplier with high speed capabilities, and good linearity. The Code Select Binary Weighting Circuit runs fast enough to function in an 8.2 GHz linear equalizer application with end-point nonlinearity below 5%. The circuit uses 0.18 μm technology to achieve this speed and linearity, but the circuit speed and linearity depend on device size. This leaves room to improve in speed with each new semiconductor process.

## 6.1 TOPICS FOR FURTHER RESEARCH

- Explore the implementation of a linear equalizer using the multiplier of this thesis.
- Explore the implementation of the multiplier as a digital to analog converter.
- Analyze and improve linearity and speed of this multiplier with smaller device sizes and different configurations.
- Design and analyze a current to voltage converter for the output of this circuit.
- Build and test the multiplier in a VLSI chip.

# APPENDIX A

## SIMULATIONS

Due to some possible confusion about Cadence Design Systems and analog simulations, the simulation setup for the circuits in this thesis is briefly discussed. The analog simulations done in Cadence Design Systems are accurate enough to include physical meaning and reality in the simulations, but for this to happen an understanding of Cadence is needed. Cadence is very difficult to master, but a small section of information is included here to help with understanding the simulations of this thesis. Also, a section is included below with updates adding more physical meaning to the simulations done in the thesis.

### A.1 NOTES ON CADENCE DESIGN SYSTEMS

The tools in Cadence Design Systems allow for many different types of analog models and simulations. The simulations in this thesis were done using Spectre analog simulations. There are three important parts necessary to complete a physically true and accurate simulation using Spectre SPICE. One is an accurately extracted model, and another is correct simulator setup. Along with the model and simulator setup, an understanding of the accuracy and meaning of the simulation results is needed to correctly interpret any meaningful information.

The model used throughout the thesis is a TSMC extracted model. The model was provided through MOSIS and installed on the Brigham Young University Cadence system for the analog research group. This model is a binned BSIM 3v3 model extracted from actual TSMC manufactured devices. This model was the smallest and most accurate MOSFET model available for BYU research at the time this thesis was written. Much time and energy was put into researching accurate ways to simulate this circuit, and it was found that Cadence Design Systems is widely used and trusted. Therefore the model and simulation software used in this thesis were the most accurate possible at the time the simulations were completed.

The second part necessary for accurate simulation results is the correct setup for the simulator. Cadence will extract 20 different parasitic capacitances for each transistor if the correct setup is used. It is these parasitic capacitances along with the MOSFET impedances that determine the final physical speed for any circuit if physically fabricated and tested. Most of the capacitances are determined by the parameters in the model along with the size of each transistor. The accurate determination of these capacitances, especially the junction capacitances associated with source to bulk and drain to bulk junctions, need to have the dimensions of the physical layout of the devices. These dimensions, namely source and drain perimeters and areas, are not determined from the SPICE circuit, but need to be added to the SPICE circuit in the setup of the simulation.

Cadence has a feature that will automatically determine the size of a minimum layout device based on the length and width entered for a transistor and enter the drain and source perimeter and area automatically. One caution that should be taken here is that the device sizes that Cadence chooses automatically are the minimum possible sizes,

and this may not be desired for every circuit. The other caution is that if a parameter such as W or L is used to replace the transistor length and width in the circuit setup, then the Cadence feature to automatically calculate device sizes will not work, and the drain and source sizes must be entered in manually.

The size of the drain and source help determine the junction capacitances accurately. Without specifying these parameters, or making sure that Cadence specifies them automatically, the source to bulk and drain to bulk capacitances will be calculated as lower than they really are. While many simulations involving low speed accuracy and other aspects of circuit design will not necessarily be affected by these parameters, transient simulations usually will. Care should to be taken to accurately calculate these parameters, as a simulation will not have any realistic physical meaning without them. These sizes and capacitances are some of the most important parameters to make sure that a SPICE simulation will reflect the operation of a circuit once it is laid out and fabricated.

The last, and most important, part to extracting realistic results from simulations is an understanding of what a circuit simulator is actually doing. There are many books written to provide deeper understanding of BSIM3v3 [15], and it is necessary to consult these resources and understand which results correspond to physical reality, and which results correspond to simulator error or theory only.

Beside the error of not including the source and drain areas and perimeters, there was another discrepancy in the results of the simulations and what might happen in the actual physical circuit. Cadence models most of the parasitic capacitances associated with the intrinsic MOSFET devices, but there are several parasitic capacitors that are not

modeled in Spectre. One such capacitance is the well-to-substrate capacitance for the PMOS devices used to construct the circuit of this thesis. Other capacitances include metal trace capacitances and any other parasitic capacitances that will depend on the layout of the circuit. These other parasitic capacitances will be small and application specific, but the well-to-substrate capacitance is the largest parasitic capacitance associated with this circuit, and needs to be accounted for. This can be done manually in Spectre by calculating a predicted capacitance value and adding it as a diode or a capacitor. This improved simulation is included in the next section.

## A.2 CORRECT CIRCUIT RESULTS

As mentioned above, the area of the source and drain of each MOSFET along with the perimeters of the source and the drain need to be specified to correctly represent all intrinsic parasitic capacitances accurately. In the simulations done in this thesis parameters were used to represent the transistor length and width, and therefore the sizes of the drain and source were not correctly specified. This was due to a lack of understanding about the Spectre simulator which was later resolved. Once the drain and source parameters were completely specified, the simulations were redone to verify the circuit operation. The well-to-substrate capacitance is also simulated in this section. It was neglected in the original thesis research simulations, but it is of major importance to the configuration of the multiplier and needs to be characterized to fully understand the behavior of the circuit.

The circuit showed no detectable difference in linearity, or any of the other simulations except for the fast transient speed simulation. This difference in speed is

because of the increase in source to bulk and drain to bulk capacitance. With this increase in capacitance, there is a 9% lowering in the multiplier speed. The corrected simulations suggest that if the multiplier circuit were to be laid out and fabricated, the speed would be around 8 GHz instead of the 9 GHz speed recorded in the thesis. These simulations neglect any transmission line effects that may come about because of traces in the actual circuit, but most traces should be fairly short, and the fabricated circuit should operate close to 8 GHz while exhibiting the other characteristics described in the main body of the thesis. The simulations in the thesis should be accurate enough to provide an approximate representation of physically accurate results.

Figure 24 is the corrected fast transient simulation that accounts for the parasitic capacitances correctly. It shows 8.2 GHz operation for the multiplier.

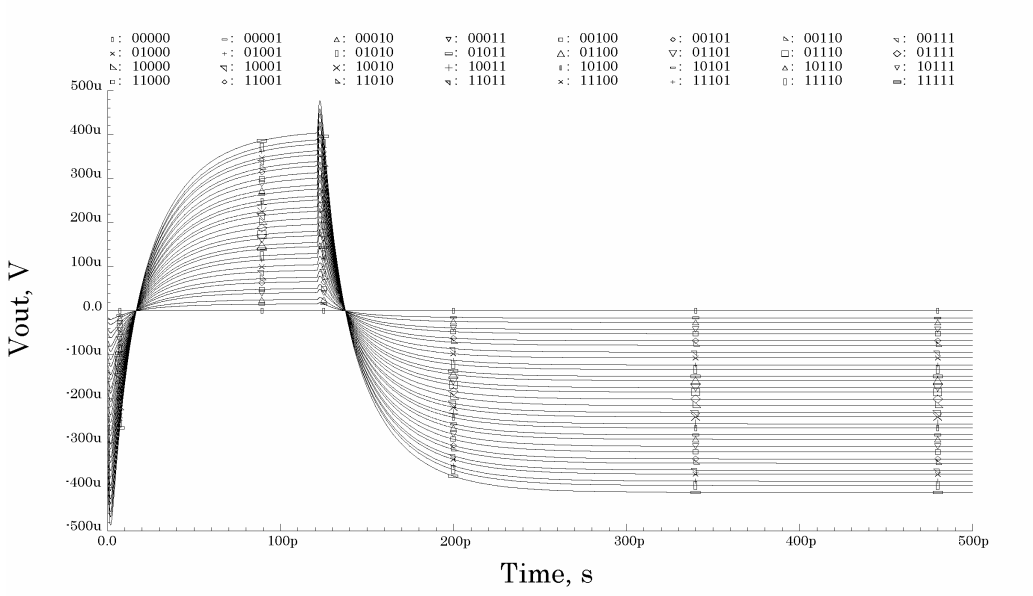


Figure 24: Correct fast transient results.

Simulation of the well-to-substrate capacitance involves the calculation of an approximate value, and the manual introduction of a capacitor to the circuit. The calculation of the approximate capacitance can be accomplished with an estimate of the n-well areas and the doping levels in the substrate and well. This particular multiplier involves only PMOS devices, and therefore the devices could all be lumped into several n-wells. The number of n-wells will depend on whether the bulk contacts of the weighting devices are connected to $V_{dd}$ or to the source contacts. In the worse case scenario, the bulk contacts are tied to the source terminals for the weighting devices, and the circuit would require three n-wells: two for each half of the weighting devices and one for the input converter devices. Either way, most of the area associated with the n-well-to-substrate capacitance is the bottom area, and therefore the total junction capacitance area can be approximated as the bottom area of the n-wells.

The bottom area for the n-wells can be divided into two halves to calculate a capacitance value for each half of the weighting devices connected to a separate input device. The area of the input device connected to each half of the weighting devices is included in the area of that half. Including the input device area is not accurate for the bulk to source connected configuration, but this area makes only a small difference in the overall area and it is correct for the bulk to $V_{dd}$ connected configuration. The areas are assumed to be equal here for either configuration to allow a better comparison in the simulation results. The n-well area for each half is calculated to be $8.554 \times 10^{-11}\ m^2$.

The doping is needed for the n-well and the substrate to be able to calculate this capacitance correctly. These quantities are not specified explicitly in the device model parameters, but a common default value for doping in these two regions is

$N_d^+ = N_a^- = 6 \times 10^{22}\ 1/m^3$. Although the doping values will not normally be equal for the n-well and the p-substrate, the values should be similar. In this case we will assume equal doping. With the junction area and doping, an approximate junction capacitance value can be calculated with the following expression [16]:

$$C_j = \frac{C_{jo}}{\sqrt{1 + \frac{V_R}{\Phi_0}}}, \tag{28}$$

where $C_{jo}$ is found as:

$$C_{j0} = \sqrt{\frac{q\varepsilon_s N_d^+ N_a^-}{2\Phi_0 (N_d^+ + N_a^-)}}, \tag{29}$$

and

$$\Phi_0 = 0.026 \ln(\frac{N_d^+ N_a^-}{n_i^2}). \tag{30}$$

In (28) $C_j$ is the junction capacitance per unit area, $C_{jo}$ is the zero bias junction capacitance per unit area, $V_R$ is the reverse bias voltage across the junction, and $\Phi_0$ is the built-in junction voltage. In the case of the multiplier in this thesis, $V_R = 1.6\,V$. In, (29) and (30) $q$ is the charge of an electron, $\varepsilon_s$ is the permittivity of silicon, $N_d^+$ is the doping

of the n-well, $N_a^-$ is the doping of the p-substrate, and $n_i$ is the intrinsic concentration of carriers in silicon. With these values, the well-to-substrate capacitance for each side of the multiplier is $2.763\times10^{-14}\ f$ connected from the bulk terminals of each half of the weighting devices to ground.

Rather than simulating all values for the circuit again, it is important to note that this capacitance will not affect any of the slow transient or DC simulations. It is also key to note that in the improved new configuration of this thesis, this capacitance will not affect the speed of the circuit. When the bulk contacts of the weighting devices are connected to $V_{dd}$, this well capacitance will be connected from small-signal ground to small-signal ground. In the original bulk to source contact configuration, this capacitance is connected directly from the signal path to ground, and is the largest of the parasitic capacitances. With these observations, the most relevant simulation is a fast transient simulation comparing the two connections for the bulk of the weighting devices like that of Figure 20. The simulation is carried out in the exact same manner except for the capacitances from bulk to ground.

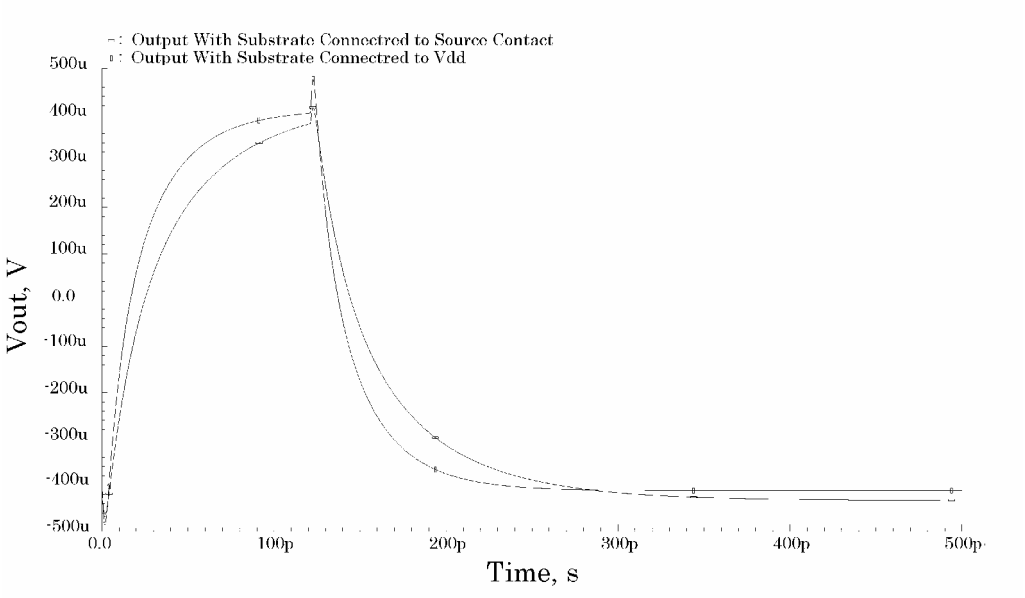


Figure 25: Substrate comparison including well capacitances.

Figure 25 shows that the circuit with the bulk to $V_{dd}$ connection for the weighting devices is not affected by the addition of the well capacitances, while the original circuit configuration is affected significantly. The speed of the multiplier with the weighting bulk contacts connected to the weighting source contacts runs at a speed of 4.8 GHz. An important conclusion can be drawn that in a high speed CMOS circuit it is advantageous to connect the bulk contacts of any PMOS devices to $V_{dd}$ to avoid the large well-to-substrate parasitic capacitance. The improved speed can then be weighed against any nonlinearity this connection may introduce.

**BIBLIOGRAPHY**


[1] A. Martin, B. Casper, J. Kennedy, J. Jaussi, and R. Mooney, "Eight Gb/s differential simultaneous bidirectional link with 4 mV, 9 ps waveform capture diagnostic capability," *Proceedings of the International Solid-State Circuits Conference*, Feb. 2003.

[2] P. Gray, S. Lewis, and R. Meyer, *Analysis and Design of Analog Integrated Circuits Fourth Edition*. New York: John Wiley & Sons, Inc., 2001, pp. 59-63.

[3] H. C. Casey, *Devices For Integrated Circuits.* New York: John Wiley & Sons, Inc., 1999, pp. 108-110.

[4] A. M. Shreeve, "A high speed wide range linear voltage to current converter," Masters Thesis, Brigham Young University, 2002, pp 5-8.

[5] J. H. Huang, Z. H. Liu, M. C. Jeng, P. K. Ko, and C. Hu, "A physical model for MOSFET output resistance," *Electron Devices Meeting, 1992. Technical Digest., International,* 13-16 Dec 1992, pp. 569-572.

[6] T. Sakurai and R. Newton, "Alpha-power law MOSFET model and its applications to CMOS inverter delay and other formulas," *IEEE Journal of Solid-State Circuits*, Vol. 25, No. 2, April 1990, pp. 584-594.

[7] R. Sarpeshkar, R. F. Lyon, and C. A. Mead, "An analog VLSI cochlea with new transconductance amplifiers and nonlinear gain control," *Proceedings of the IEEE International Symposium on Circuits and Systems*, Vol. 3, May 1996, pp 292-296.

[8] B. Gilbert, "A precise four-quadrant multiplier with subnanosecond response," *IEEE Journal of Solid-State Circuits*, Vol. 3, No. 4, December 1968, pp. 365-373.

[9] J. Ramirez-Angulo, R. G. Carvajal, and J. Martinez-Heredia, "1.4 V supply, wide swing, high frequency CMOS analogue multiplier with high current efficiency," *Proceedings, ISCAS 2000 Geneva,* Vol. 5, May 2000, pp. 533-536.

[10] D. J. Comer, A. K. Martin, and J. E. Jaussi, "Multiplier using MOS channel widths for code weighting," Patent applied for by Intel Corp. 2002.

[11] D. R. Bartholomew, D. J. Comer, D. T. Comer A. K. Martin, and J. E. Jaussi, "A 9 GHz mixed signal multiplying circuit in 0.18 μm CMOS," in review.

[12] M. J. Deen and O. Marinov, "Effect of Forward and Reverse Substrate Biasing on Low-Frequency Noise in Silicon PMOSFETs," *IEEE Transactions On Electron Devices*, Vol. 49, No. 3, March 2002, pp. 409-413.

[13] D. J. Comer, D. T. Comer, A. K. Martin, and J. E. Jaussi, "A high-frequency CMOS current summing circuit," *Analog Integrated Circuits and Signal Processing*, Vol. 36, 2003, pp. 215-220.

[14] E. A. Lee and D. G. Messerschmitt, *Digital Communication*. Boston: Kluwer Academic Publishers, 1988, pp. 289-371.

[15] W. Liu, *MOSFET Models for SPICE Simulation including BSIM3v3 and BSIM4.* New York: John Wiley & Sons, Inc., 2001, pp. 101-398.

[16] D. A. Johns and K. Martin, *Analog Integrated Circuit Design.* Toronto: John Wiley & Sons, Inc. 1997, pp. 1-77.