2-1-2011

Analog integrated circuit design in ultra-thin oxide CMOS technologies with significant direct tunneling-induced gate current

Eric Bohannon

ANALOG INTEGRATED CIRCUIT DESIGN IN ULTRA-THIN OXIDE CMOS TECHNOLOGIES WITH SIGNIFICANT DIRECT TUNNELING-INDUCED GATE CURRENT

by

ERIC BOHANNON

A DISSERTATION

Submitted in partial fulfillment of the requirements
For the degree of Doctor of Philosophy
in
Microsystems Engineering
at the
Rochester Institute of Technology

February 2011

Author: _____
Microsystems Engineering Program

Certified by: _____
P.R. Mukund, Ph.D.
Professor of Electrical Engineering

Approved by: _____
Bruce W. Smith, Ph.D.
Director of Microsystems Engineering Program

Certified by: _____
Harvey J. Palmer, Ph.D.
Dean, Kate Gleason College of Engineering

NOTICE OF COPYRIGHT

Analog Integrated Circuit Design in Ultra-Thin Oxide CMOS Technologies with Significant Direct Tunneling-Induced Gate Current

By

Eric Bohannon

Submitted by Eric Bohannon in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Microsystems Engineering and accepted on behalf of the Rochester Institute of Technology by the dissertation committee.

We, the undersigned members of the Faculty of the Rochester Institute of Technology, certify that we have advised and/or supervised the candidate on the work described in this dissertation. We further certify that we have reviewed the dissertation manuscript and approve it in partial fulfillment of the requirements of the degree of Doctor of Philosophy in Microsystems Engineering.

Approved by:

Dr. P.R. Mukund _____
(Committee Chair and Dissertation Advisor) Date

Mr. Clyde Washburn _____

Dr. James E. Moon _____

Dr. Sean Rommel _____

Dr. Dhireesha Kudithipudi _____

MICROSYSTEMS ENGINEERING PROGRAM
ROCHESTER INSTITUTE OF TECHNOLOGY
February 2011

ABSTRACT

Kate Gleason College of Engineering
Rochester Institute of Technology

Degree Doctor of Philosophy **Program** Microsystems Engineering

Name of Candidate Eric Bohannon

Title Analog Integrated Circuit Design in Ultra-Thin Oxide CMOS Technologies with Significant Direct Tunneling-Induced Gate Current

The ability to do mixed-signal IC design in a CMOS technology has been a driving force for manufacturing personal mobile electronic products such as cellular phones, digital audio players, and personal digital assistants. As CMOS has moved to ultra-thin oxide technologies, where oxide thicknesses are less than 3 nm, this type of design has been threatened by the direct tunneling of carriers though the gate oxide. This type of tunneling, which increases exponentially with decreasing oxide thickness, is a source of MOSFET gate current. Its existence invalidates the simplifying design assumption of infinite gate resistance. Its problems are typically avoided by switching to a high-κ/metal gate technology or by including a second thick(er) oxide transistor. Both of these solutions come with undesirable increases in cost due to extra mask and processing steps. Furthermore, digital circuit solutions to the problems created by direct tunneling are available, while analog circuit solutions are not. Therefore, it is desirable that analog circuit solutions exist that allow the design of mixed-signal circuits with ultra-thin oxide MOSFETs. This work presents a methodology that develops these solutions as a less costly alternative to high-κ/metal gate technologies or thick(er) oxide transistors. The solutions focus on transistor sizing, DC biasing, and the design of current mirrors and differential amplifiers. They attempt to minimize, balance, and cancel the negative effects of direct tunneling on analog design in traditional (non-high-κ/metal gate) ultra-thin oxide CMOS technologies. They require only ultra-thin oxide devices and are investigated in a 65 nm CMOS technology with a nominal V_{DD} of 1 V and a physical oxide thickness of 1.25 nm. A sub-1 V bandgap voltage reference that requires only ultra-thin oxide MOSFETs is presented ($T_C = 251.0$ ppm/°C). It utilizes the developed methodology and illustrates that it is capable of suppressing the negative effects of direct tunneling. Its performance is compared to a thick-oxide voltage reference as a means of demonstrating that ultra-thin oxide MOSFETs can be used to build the analog component of a mixed-signal system.

Abstract Approval: Committee Chair _____

Program Director _____

Dean KGCOE _____

ACKNOWLEDGMENTS

First and foremost, I would like to thank God for giving me all of the wonderful opportunities I've had in my life. I would also like to thank my beautiful wife Leanne, who supported me throughout this entire process. She quietly listened to all of my technical and non-technical ramblings even though she didn't know or care what I was talking about. Leanne, you helped give me the strength and focus I needed to finish. I love you.

To my family, I appreciate all that you have done for me throughout the years. I would like to thank the following family members for their support: Meme, Papa, Chuck, Tamie, Jeanne, Gianna, Mr. Stefano, Mrs. Stefano, Lisa, Theresa, Granny Bohannon, and Granny Betty. To Rindy, you are a wonderful person and I'm proud to have you as a sister. To John Fatcheric, I'm happy that you and your family are a part of my life. You are a true father figure to me and have had a special impact on my life and on my mother's life. To my mother, I want to thank you for the unconditional love you have shown me throughout my life. I can't tell you how blessed and honored I am to have you as my mother. None of this would have been possible without you. To my father, I hope one day you read this and that you know I love and support you. I thank you for everything that you have done for me and I pray that you are able to find peace.

I've been fortunate to have many great teachers in my life, but two in particular have had a special impact. The first is Billy Winks. Billy, you taught me to have a passion for what I do. You also taught me not to be afraid to learn new things. I thank you for this. The second great teacher in my life is Clyde Washburn. Clyde, this work is

as much yours as it is mine. Without your guidance, instruction, and endless patience, I would not be where I am today. You represent an answered prayer in my life. You are a true mentor and advisor. For that I will be forever grateful.

I would like to acknowledge and thank my committee for their contributions and support. Specifically, I would like to thank Mr. Washburn, Dr. Mukund, Dr. Moon, Dr. Rommel, Dr. Kudithipudi, and Dr. Smith. Dr. Moon, your work ethic is truly inspirational. You are an excellent teacher and professor. I'm honored to have been your student. Dr. Mukund, I appreciate all that you have done for me over the years. I wish you success in your future endeavors.

To Christopher Urban, thank you for going through this process with me. You were a constant source of comedic relief. I'll never forget the many memories we shared in the lab. You are a good person with a good heart. I wish you success and happiness in the future.

To John Pernasilice, you are a true best friend. You are like a brother to me. I thank you for your constant words of encouragement. To Jimmy Retzos, Scott Brown, James Barrett, Chris Nassar, Mark Pude, Daniel Fava, Luiz Freitas, and Gabriel Rinaldi, I thank you for your friendship over the years. To Tejasvi Das, Priya Das, and Sharmila Sridharan, I thank you for all that you taught me while I was a student in the lab.

To my colleagues, I would like to thank you for your patience and support. You have taught me a significant amount in a short period of time. Specifically, I would like to acknowledge Imre Knausz, Brian Mott, Clint Meyer, Murat Ozbas, John Lynch, Tom Quattrini, Jeff Lillie, and Chris Ludden.

TABLE OF CONTENTS

ABSTRACT iv
ACKNOWLEDGMENTS v
TABLE OF CONTENTS vii
LIST OF FIGURES xi
LIST OF TABLES xvii
LIST OF COMMONLY USED SYMBOLS AND ABBREVIATIONS xviii
CHAPTER 1 INTRODUCTION 1
CHAPTER 2 OBJECTIVES 8
CHAPTER 3 BACKGROUND 10
3.1 Analog Design in Nanoscale CMOS 10
3.1.1 Output Resistance Degradation 11
3.1.1.1 Channel Length Modulation (CLM) 12
3.1.1.2 Drain-Induced Barrier Lowering (DIBL) 14
3.1.1.3 Drain-Induced Threshold Shift (DITS) 16
3.1.1.4 Substrate Current-Induced Body Effect (SCBE) 18
3.1.2 Reductions in Supply Voltage 18
3.1.2.1 Supply Voltage Scaling 19
3.1.2.2 Reductions in Voltage Headroom 20
3.1.2.3 Reduced Transistor Stacks 21
3.1.2.4 Weak and Moderate Inversion Operation 21
3.1.2.5 Reduced SNR 23
3.1.3 Modeling Complexity and Process Variations 23
3.1.3.1 Modeling Complexity 24
3.1.3.2 Process Variations 25
3.1.4 Circuit Solutions 28
3.1.4.1 Body-Biased and Bulk-Driven Transistors 28
3.1.4.2 Sub-V_{TH} Operation 30
3.1.4.3 Self-Cascoding 32
3.2 Gate Current 36
3.2.1 Tunneling Background 36
3.2.2 Fowler-Nordheim Tunneling and Direct Tunneling 39
3.2.3 Modeling of Direct Tunneling 41
3.2.4 Impact of Direct Tunneling on Current Mirror Design 44
3.2.5 Comparing Direct Tunneling to Base Current 45
3.2.6 Impact of Direct Tunneling on Analog Device Performance 47

3.2.7 Existing Circuit Solutions to Gate Current 50
3.2.8 Direct Tunneling and High-κ/Metal Gates 51
3.3 Voltage References 53
3.3.1 Temperature Independence and Bandgap Voltage References 54
3.3.1.1 The Use of Vertical PNP BJTs in Bandgap Voltage References 57
3.3.2 Startup Circuits, Process Variations, and Supply Voltage Dependence 58
3.3.3 Traditional Bandgap Voltage References 60
3.3.4 All-MOSFET Voltage References 61
3.3.5 Sub-1 V Bandgap Voltage References 61
CHAPTER 4 APPROACH 65
4.1 Computing Resources 65
4.2 Gate Current Performance Metrics 66
4.2.1 Simulation Strategy 69
4.3 Impact of Body Biasing on Gate Current 70
4.3.1 Simulation Strategy 70
4.4 The Design of Ultra-Thin Oxide CMOS Current Mirrors 72
4.4.1 Self-Cascode Current Mirrors 72
4.4.2 Triple Self-Cascode Current Mirrors 74
4.4.3 Simulation Strategy 75
4.5 The Design of Ultra-Thin Oxide CMOS Differential Amplifiers 76
4.5.1 Amplifier Input Current 76
4.5.2 Gate Balancing 78
4.5.3 Input Current Cancellation 82
4.5.4 Simulation Strategy 84
4.6 The AC Simulation of Ultra-Thin Oxide CMOS Amplifiers 85
4.6.1 Simulation Strategy 88
4.7 Impact of Gate Current on Sub-1 V Bandgap Voltage References 88
4.8 The Design of an Ultra-Thin Oxide Sub-1 V Bandgap Voltage Reference 92
4.8.1 Self-Cascoding and Gate-Balancing 92
4.8.2 Startup 94
4.8.3 Impact of Amplifier Input Current 95
4.8.4 Power and Area Tradeoffs 96
4.8.5 Amplifier Compensation 99
4.8.6 Simulation Strategy 100
4.9 Topics Not Addressed in This Work 101
CHAPTER 5 RESULTS 103

5.1 Gate Current Performance Metrics 103
5.1.1 Impact of Gate Current on Diode-Connected Transistors 104
5.1.2 Impact of V_{DS} on Gate Current 107
5.1.3 Channel Length Selection Methodology 110
5.2 Impact of Body Biasing on Gate Current 113
5.2.1 Constant Terminal Voltages 113
5.2.2 Constant Drain Current 114
5.3 The Design of Ultra-Thin Oxide CMOS Current Mirrors 116
5.3.1 Current Mirror Comparison 116
5.3.2 Self-Cascode Current Mirrors 118
5.3.3 Self-Cascode Current Mirrors with a Helper Transistor 123
5.3.4 Triple Self-Cascode Current Mirrors 125
5.4 The Design of Ultra-Thin Oxide CMOS Differential Amplifiers 127
5.4.1 Gate Balancing 127
5.4.2 Amplifier Gain Comparison 128
5.4.3 Input Current Cancellation 129
5.5 The AC Simulation of Ultra-Thin Oxide CMOS Amplifiers 131
5.6 The Design of an Ultra-Thin Oxide Sub-1 V Bandgap Voltage Reference 132
5.6.1 Thick-Oxide Sub-1 V Bandgap Voltage Reference 132
5.6.2 Thick-to-Ultra-Thin Oxide Sub-1 V Bandgap Voltage Reference 135
5.6.3 Ultra-Thin Oxide Sub-1 V Bandgap Voltage Reference 136
5.6.3.1 General Design Strategy 136
5.6.3.2 Impact of Error Amplifier's PMOS Active Load 140
5.6.3.3 Impact of Error Amplifier's Input Pair 142
5.6.3.4 Impact of Gate Current Flowing into the Output 145
5.6.3.5 Monte Carlo and Process Corners Analyses 146
5.6.3.6 Startup Analyses 148
5.6.3.7 Transistor Loading 150
5.6.3.8 Sensitivity Analysis 152
5.7 Sponsored Fabrication 155
CHAPTER 6 CONCLUSION 161
APPENDIX A Low-Frequency Small-Signal Analysis of the Self-Cascode Amplifier 163
A.1. Derivation of G_M, R_{OUT}, and A_V of the Self-Cascode Amplifier 163
APPENDIX B Sub-1 V Voltage Reference Analyses 165
B.1. Analysis of Ideal Sub-1 V Bandgap Voltage Reference 165

B.2. Analysis of a Sub-1 V Bandgap Voltage Reference Including Offset Voltage, Input Bias Current, and Input Offset Current 167

REFERENCES 170

LIST OF FIGURES

Figure 1.1: Simulated β_{F_MOS} vs. V_{GS} and I_G vs. V_{GS} for an NMOS transistor with W = 20 μm and L = 5 μm. 4

Figure 3.1: Simplified cross section and symbol of an NMOS transistor. The gate, drain, body, and source represent input/output terminals. I_{DS} is the drain-to-source current. The drain and source regions are represented by heavily doped n-type regions (n^+) while contacts to the body are represented by a heavily doped p-type (p^+) region. The substrate is p-type. 11

Figure 3.2: (a) Simulated I_D vs. V_{DS} and r_O vs. V_{DS} for an NMOS transistor in the obtained 65 nm process. W = 1 μm, L = 50 nm, and V_{GS} = 0.3 V. (b) Simulated I_D vs. V_{DS} and r_O vs. V_{DS} for an NMOS transistor in the obtained 65 nm process. W = 10 μm, L = 1 μm, and V_{GS} = 0.3 V. 14

Figure 3.3: Simulated ΔV_{TH} vs. L for two NMOS transistors with different V_{DS} voltages in the obtained 65 nm process. Each transistor had W = 1 μm and V_{GS} = 0.3 V. The V_{DS} voltages were 0.1 V and 1.0 V. 15

Figure 3.4: Simulated V_{TH} vs. L of an NMOS transistor in the obtained 65 nm process. W = 1 μm and V_{DS} = 100 mV. $V_{GS} = V_{DS}$ = 0.3 V. 17

Figure 3.5: V_{DD} and V_{TH} vs. technology node. V_{TH} was extracted for an NMOS device with $V_{GS} = V_{DD}$, $V_{DS} = V_{DD}$, $V_{BS} = 0$, $L = L_{MIN}$, and $W = W_{MIN}$. L_{MIN} and W_{MIN} represent process minima for the channel length and channel width. 19

Figure 3.6: General behavior of the MOSFET threshold voltage mismatch slope vs. L in an ultra-thin oxide CMOS process [95]. W is held constant. 26

Figure 3.7: Simulated V_{TH} vs. V_{BS} for two NMOS transistors in the obtained 65 nm process. One transistor had W = 10 μm and L = 1 μm. The other transistor had W = 1 μm and L = 50 nm. Both transistors had $V_{GS} = V_{DS}$ = 0.3 V. 29

Figure 3.8: Simple example of a differential bulk-driven amplifier. M1 and M2 represent the bulk-driven input differential pair, V_{IN1} and V_{IN2} are the bulk input voltages, I_{BIAS} is the bias current, R_L is the load resistor, and V_{DD} is the supply voltage. 30

Figure 3.9: I_D vs. V_{GS} for an NMOS transistor in the obtained 65 nm process. W = 10 μm, L = 1 μm, and V_{DS} = 0.3 V. 31

Figure 3.10: Basic cascode structure. V_{IN} is the input voltage, V_{OUT} is the output voltage, V_{BIAS} is the bias voltage for M2, and I_{OUT} is the output current. M1 and M2 form the basic cascode structure. 32

Figure 3.11: Self-cascode structure [117]. V_{IN} is the input voltage, V_{OUT} is the output voltage, and I_{OUT} is the output current. M1 and M2 form the self-cascode structure. 33

Figure 3.12: Tunneling in a rectangular potential barrier [124]. $V(x)$ is the potential energy of the system and ε_k is the incident particle kinetic energy. V_0 is the barrier height and L is the barrier width. The carrier is described by its wave function, $\Psi(x)$. 37

Figure 3.13: Ideal energy band diagrams for: (a) Fowler Nordheim tunneling and (b) direct tunneling in an NMOS transistor. E_C and E_V are the conduction and valence bands, t_{ox} is the oxide thickness, X_B is the barrier height, V_{OX} is the voltage across the oxide, and e^- is the tunneling electron [16]. 39

Figure 3.14: Direct tunneling in an NMOS transistor. E_C and E_V are the conduction and valence bands, V_{OX} is the voltage across the oxide, t_{ox} is the oxide thickness, e^- and h^+ represent tunneling electrons and holes. X_{B_ECB}, X_{B_EVB}, and X_{B_HVB} represent the barrier heights for ECB, EVB, and HVB [14], [86]. 40

Figure 3.15: Components of direct tunneling in an NMOS transistor [136]. I_{GCS} and I_{GCD} flow into the channel, I_{GS} flows into the source overlap region, I_{GD} flows into the drain overlap region, and I_{GB} flows into the substrate. 42

Figure 3.16: DC components of direct tunneling in an NMOS transistor [137]. I_{GCS} and I_{GCD} flow directly into the source. I_{GCD} flows out of the source via the drain. I_{GD} can flow out of the source via the drain or out of the gate. I_{GB} flows out of the body. 43
Figure 3.17: Simple current mirror. V_{DD} is the supply voltage, I_{IN} is the input current, I_{OUT} is the output current, and V_{OUT} is the output voltage. M1 and M2 form the current mirror. 44
Figure 3.18: Logarithmic plot of β_{F_MOS} vs. L of an NMOS transistor in the obtained 65 nm process. W = 10 µm and $V_{GS} = V_{DS}$ = 1 V. 46
Figure 3.19: High-level circuit schematic of a bandgap voltage reference [44]. V_{DD} is the supply voltage, I_{BIAS} is the bias current, V_{CTAT} is the CTAT voltage, V_{PTAT} is the PTAT voltage, k is Boltzmann's constant, q is the electronic charge, T is the temperature, and K is a scale factor. ... 54
Figure 3.20: Cross section of a vertical PNP BJT made out of a PMOS transistor [165]. The base is formed from the body terminal. The emitter is formed from the source and drain terminals. The collector is formed from the substrate. 57
Figure 3.21: Example of different startup operating points that occur in bandgap voltage references 58
Figure 3.22: Simplified representation of the voltage reference in [116]. V_{DD} is the supply voltage, Q1 and Q2 are diode-connected PNP BJTs. I_1, I_2, and I_3 are voltage-controlled current sources. The error amplifier ensures $V_{EB1} = V_{EB2} + V_{R1}$. V_P and V_M represent the non-inverting and the inverting input voltages of the amplifier. R_1, R_2, R_3, and R_4 are resistors used to zero the temperature slope and set the output voltage, V_{REF}. 62
Figure 4.1: Schematic of circuits used to extract gate current performance metrics. V_{DD} is the supply voltage, I_{BIAS} is the bias current, V_{G1} is the gate voltage of M1, V_{D2} is the drain voltage of M2, and V_{G2} is the gate voltage of M2. V_{BIAS} is copied to the gates of M1 and M2 via VCVSs. 69
Figure 4.2: Schematic of circuit used to determine impact of body voltage on gate current with constant terminal voltages. V_{DD} is the supply voltage, I_{BIAS} is the bias current, V_{G1} is the gate voltage of M1, V_{D2} is the drain voltage of M2, V_{G2} is the gate voltage of M2, and V_{BODY2} is the body voltage of M2. V_{BIAS} is copied to the gates of M1 and M2 via VCVSs. 71
Figure 4.3: Schematic of circuit used to determine impact of body voltage on gate current with constant drain current. V_{DD} is the supply voltage, I_{BIAS} is the bias current, V_{G3} is the gate voltage of M3, and V_{BODY3} is the body voltage of M3. V_{BIAS} is copied to the gate of M3 via a VCVS. 71
Figure 4.4: Self-cascode current mirror. V_{DD} is the supply voltage, I_{IN} is the input current, V_{OUT} is the output voltage, I_{OUT} is the output current, and V_{BIAS} is the gate voltage of M1-M4. 73
Figure 4.5: Self-cascode current mirror with a helper transistor. V_{DD} is the supply voltage, I_{IN} is the input current, V_{OUT} is the output voltage, I_{OUT} is the output current, and V_{BIAS} is the gate voltage of M1-M4. M5 is the helper transistor. It is used to block I_{IN} from flowing into the gates of M1-M4. 74
Figure 4.6: (a) Triple self-cascode structure. V_{IN} is the input voltage, V_{OUT} is the output voltage, and I_{OUT} is the output current. M1, M2, and M3 form the self-cascode structure. (b) Triple self-cascode current mirror. V_{DD} is the supply voltage, I_{IN} is the input current, V_{OUT} is the output voltage, I_{OUT} is the output current, and V_{BIAS} is the gate voltage of M1-M6. M7 is a helper transistor. It is used to block I_{IN} from flowing into the gates of M1-M6. 75
Figure 4.7: Differential amplifier. M1 and M2 form the input pair. M3 is the tail current source. M4 and M5 form an active load. V_{DD} is the supply voltage, V_{IN1} and V_{IN2} are the common-mode input voltages, V_{DIO} is the diode-connected voltage of M4 and M5, V_{BIAS} is the gate-bias voltage of M3, and V_{OUT} is the output voltage. 77
Figure 4.8: Balanced differential amplifier. M1 and M2 form the input pair. M3, M7, M8, and I_{BIAS} form the bias network. M4 and M5 form an active load. V_{DD} is the supply voltage, V_{IN1} and V_{IN2} are the common-mode input voltages, V_{BIAS} is the gate-bias voltage for M3, V_{DIO} is

the diode-connected voltage of M4 and M5, V_{OUT} is the output voltage, and I_{OUT} is the output current. M6 is used to restore balance to the amplifier. M9 is used to force similar drain voltages between M4, M5, and M6. C_C is the compensation capacitor. 79

Figure 4.9: Two-stage self-cascode operational amplifier. SC1 and SC2 form the input pair. SC4 and SC5 form the active load. SC6 forms the second stage. SC3, SC7, SC8, and I_{BIAS} form the bias network. V_{DD} is the supply voltage. V_{IN1} and V_{IN2} are the common-mode input voltages. M9 is a diode-connected transistor used to force similar drain voltages between SC4, SC5, and SC6. V_{OUT} and V_{OUT}' are the output voltages of the first and second stages. C_C is the compensation capacitor. 80

Figure 4.10: Differential amplifier with input current cancellation. M1 and M2 form the input pair. M12 is the tail current source. M4 and M5 form an active load. M16 is a helper transistor. V_{DD} is the supply voltage, V_{COM} is the common-mode input voltage, V_{DIO} is the diode-connected voltage of M4 and M5, V_{BIAS} is the gate-bias voltage of M10-M13, I_{BIAS} is the bias current, and V_{OUT} is the output voltage. C_C is the compensation capacitor. The input current cancellation network is formed by the error amplifier, M3, M7-M9, M11, and M15. V_S is the source voltage of M15 and V_E is the output voltage of the error amplifier. 82

Figure 4.11: Transistor-level schematic of the error amplifier in Figure 4.10. M1 and M2 form the input pair. M3, M4, M5, and I_{BIAS} form the bias network. M7 and M8 form an active load. V_{DD} is the supply voltage, $V_{TAIL}$is connected to the tail voltage of M1 and M2 in Figure 4.10, V_S is connected to the source voltage of M15 in Figure 4.10, V_{BIAS} is the gate-bias voltage for M3, V_{DIO} is the diode-connected voltage of M7 and M8, V_E is the output voltage and is connected to the gate terminals of M7, M8, and M9 in Figure 4.10. M9 is the second stage of the amplifier. It is used to restore balance to the amplifier. C_C is the compensation capacitor. 84

Figure 4.12: Circuit technique used to maintain the DC bias point when performing amplifier AC simulations [48]. V_P and V_M represent the amplifier's non-inverting and inverting input voltages, V_{FB} is the feedback voltage, V_{IN} is the small-signal input voltage, R_{C_BIAS} and C_{BIAS} create a low-pass filter, and the VCVS is used to copy the DC component of V_{FB} to V_{BIAS}. 85

Figure 4.13: Circuit technique used to maintain the DC bias point when performing amplifier AC simulations in the presence of non-negligible amplifier input current. V_P and V_M represent the amplifier's non-inverting and inverting input voltages, V_{FB} is the feedback voltage, V_{IN} is the small-signal input voltage, R_{C_BIAS} and C_{BIAS} create a low-pass filter, and the VCVS is used to copy the DC component of V_{FB} to V_{BIAS}. I_{IN_A} and I_{IN_B} represent the amplifier input current. R_{L_BIAS} and L_{BIAS} from a low-pass filter that transfers the DC current component of I_{IN_A} via two CCCSs to V_{FB}. 87

Figure 4.14: Sub-1 V bandgap voltage reference including amplifier input offset voltage and amplifier input current. Q1 and Q2 are diode-connected PNP BJTs. I_1, I_2, and I_3 are voltage-controlled current sources. The error amplifier ensures $V_{EB1} + V_{OS} = V_{EB2} + V_{R1}$. V_P and V_M represent the non-inverting and inverting input voltages of the amplifier. R_1, R_2, R_3, and R_4 are resistors used to zero the temperature slope and set the output voltage, V_{REF}. I_{IN_B} and I_{OS} represent the input bias current and the input offset current of the amplifier. 90

Figure 4.15: Sub-1 V bandgap voltage reference that minimizes the effects of amplifier input current. Q1 and Q2 are diode-connected PNP BJTs. I_1, I_2, and I_3 are voltage-controlled current sources. The error amplifier ensures $V_{EB1} = V_{EB2} + V_{R1}$. V_P and V_M represent the non-inverting and inverting input voltages of the amplifier. R_1, R_2, R_3, and R_4 are resistors used to zero the temperature slope and set the buffer voltage, V_{BUFFER}. V_{BUFFER} is the voltage transferred by the buffer to output of the reference, V_{REF}. The buffer is added to drain the input current of the error amplifier out of I_3. M_L and I_{LOAD} represent the load transistor and load current. 91

Figure 4.16: Transistor level schematic of Figure 4.15. SC1-SC5, SC9, SC10, and M20 form the error amplifier. SC6-SC8 form I_1-I_3. SC13-S19 form the buffer amplifier. SC22, SC23, M24, M25, Q3 and Q4 form the startup circuit. M_L and I_{LOAD} form the load transistor and load

current. R_1, R_2, R_3, and R_4 are resistors used to zero the temperature slope and set the buffer voltage, V_{BUFFER}. C_{C1}, C_{C2}, and R_{C2} form the compensation networks for the error amplifier and the buffer amplifier. 93

Figure 4.17: High-level schematic of [116] with excessive resistors. V_{DD} is the supply voltage, Q1 and Q2 are diode-connected PNP BJTs. I_1, I_2, and I_3 are voltage-controlled current sources. The error amplifier ensures $V_{EB1} = V_{EB2} + V_{R1}$. V_P and V_M represent the voltages on the non-inverting and inverting terminals of the amplifier. R_1, R_2, R_3, and R_4 are represented by series or parallel combinations a unit resistor (R_U). V_{REF} is the output voltage..... 98

Figure 4.18: High-level schematic of [116] with combined resistors. V_{DD} is the supply voltage, Q1 and Q2 are diode-connected PNP BJTs. I_1, I_2, and I_3 are voltage-controlled current sources. The error amplifier ensures $V_{EB1} = V_{EB2} + V_{R1}$. V_P and V_M represent the voltages on the non-inverting and inverting terminals of the amplifier. R_1, R_2, R_3, and R_4 are represented by series or parallel combinations a unit resistor (R_U). V_{REF} is the output voltage..... 98

Figure 5.1: (a) β_{F_MOS} vs. I_{BIAS}. (b) β_{0_MOS} vs. I_{BIAS}. Both graphs refer to the circuit shown in Figure 4.1 (a). Transistor area was held constant at 100 μm^2. The legends specify L. 104

Figure 5.2: (a) $\beta_{0\_MOS}/\beta_{F\_MOS}$ vs. I_{BIAS}. (b) r_{π_MOS} vs. I_{BIAS}. Both graphs refer to the circuit shown in Figure 4.1 (a). Transistor area was held constant at 100 μm^2. The legends specify L. 105

Figure 5.3: (a) β_{F_MOS} vs. I_{BIAS}. (b) β_{0_MOS} vs. I_{BIAS}. Both graphs refer to the circuit shown in Figure 4.1 (a). $L = 1$ µm in both graphs. The legends specify W. 106

Figure 5.4: (a) I_G vs. V_{DS}. (b) α_{F_MOS} vs. V_{DS}. Both graphs refer to the circuit shown in Figure 4.1 (b). $L = 1$ µm and $W = 100$ µm for both graphs. The legends specify I_{BIAS}. 108

Figure 5.5: (a) β_{F_MOS} vs. V_{DS}. (b) r_{μ_MOS} vs. V_{DS}. Both graphs refer to the circuit shown in Figure 4.1 (b). $L = 1$ µm and $W = 100$ µm for both graphs. The legends specify I_{BIAS}. 109

Figure 5.6: Simulated $|\partial V_{TH}/\partial L|$ vs. L and β_{F_MOS} vs. L for NMOS and PMOS transistors with $W \cdot L = 100$ µm^2 and $I_D = 10$ µA..... 110

Figure 5.7: Simulated L_{MAX} vs. I_D for NMOS and PMOS transistors with $W \cdot L = 100$ µm^2 for $\beta_{F\_MOS\_MIN} = 100$..... 112

Figure 5.8: Simulated (a) β_{F_MOS} vs. $|V_{BS}|$ and (b) percent reduction in I_G vs. $|V_{BS}|$ for an NMOS transistor and a PMOS transistor under a constant voltage condition. Each transistor was sized with $W = 100$ µm and $L = 1$ µm and had an I_D of 16 µA at $|V_{BS}| = 0$ V. V_{BS} of the NMOS device and V_{SB} of the PMOS device were both kept greater than 0 V. 114

Figure 5.9: Simulated (a) β_{F_MOS} vs. $|V_{BS}|$ and (b) percent reduction in I_G vs. $|V_{BS}|$ for an NMOS transistor and a PMOS transistor under a constant current condition. Each transistor was sized with $W = 100$ µm and $L = 1$ µm and had an I_D of 16 µA at $|V_{BS}| = 0$ V. V_{BS} of the NMOS device and V_{SB} of the PMOS device were both kept greater than 0 V. 115

Figure 5.10: Basic Cascode Current Mirror. V_{DD} is the supply voltage, I_{IN} is the input current, V_{OUT} is the output voltage, I_{OUT} is the output current. V_{BIAS1} is the gate-bias voltage of M3 and M4. V_{BIAS2} is the gate-bias voltage of M1 and M2. M1-M4 form the basic cascode current mirror. 116

Figure 5.11: (a) A_i vs. V_{OUT} for the three types of current mirrors noted in the legend ($I_{IN} = 2$ µA). $W = 10$ µm and $L = 10$ µm for all devices in the simple and basic cascode current mirrors. The cascoded devices of the self-cascode current mirror were designed with $W = 10$ µm and $L = 10$ µm. The cascoding devices of the self-cascode current mirror were designed with $W = 30$ µm and $L = 3.33$ µm (b) R_{OUT} vs. V_{OUT} for a simple current mirror with $W = 10$ µm and $L = 10$ µm. The legend specifies I_{IN}. 117

Figure 5.12: (a) A_i vs. V_{OUT} and (b) R_{OUT} vs. V_{OUT} for a self-cascode current mirror with $I_{IN} = 2$ µA. Both graphs refer to Figure 4.4. The cascoded devices were designed with $W = 100$ µm and $L = 1$ µm. The cascoding devices were designed with $L = 0.25$ µm. The legends specify the width of the cascoding transistors. 119

Figure 5.13: (a) A_i vs. V_{OUT} and (b) R_{OUT} vs. V_{OUT} for a self-cascode current mirror with I_{IN} = 16 µA. Both graphs refer to Figure 4.4. The cascoded devices were designed with W = 100 µm and L = 1 µm. The cascoding devices were designed with L = 0.25 µm. The legends specify the width of the cascoding transistors. 120

Figure 5.14: (a) R_{OUT} vs. V_{OUT} for the self-cascode current mirror of Figure 4.4. The cascoded devices were designed with W = 100 µm and L = 1 µm. The cascoding devices were designed with W = 100 µm and L = 0.25 µm. The legend specifies I_{IN}. (b) $R_{OUT\_SC}/R_{OUT\_SIMPLE}$ vs. V_{OUT}. The simple current mirror was designed with W = 100 µm and L = 1 µm. The legend specifies I_{IN}..... 122

Figure 5.15: (a) A_i vs. V_{OUT} and (b) R_{OUT} vs. V_{OUT} for the self-cascode current mirror of Figure 4.4. The cascoded devices were designed with W = 20 µm and L = 5 µm. The cascoding devices were designed with W = 40 µm and L = 1.25 µm. I_{IN} was 2 µA. The legends specify the desired current gain..... 124

Figure 5.16: (a) A_i vs. V_{OUT} and (b) R_{OUT} vs. V_{OUT} for the self-cascode current mirror of Figure 4.5. The cascoded devices were designed with W = 20 µm and L = 5 µm. The cascoding devices were designed with W = 40 µm and L = 1.25 µm. I_{IN} was 2 µA. The helper transistor was designed with W = 5 µm, L = 0.5 µm. The legends specify the desired current gain. 124

Figure 5.17: (a) A_i vs. V_{OUT} for the triple self-cascode current mirror of Figure 4.6 (b). The cascoded devices of the triple self-cascode current mirror were designed with W = 100 µm and L = 1 µm. The middle cascoding devices of the triple self-cascode current mirror were designed with W = 100 µm and L = 0.5 µm. The top cascoding devices of the triple self-cascode current mirror were designed with W = 100 µm and L = 0.25 µm. The legend specifies I_{IN}. (b) $R_{OUT\_TRIPLE\_SC}/R_{OUT\_SC}$ vs. V_{OUT}. The cascoded devices of the self-cascode current mirror were designed with W = 100 µm and L = 1 µm. The cascoding devices of the self-cascode current mirror were designed with W = 100 µm and L = 0.25 µm. The legend specifies I_{IN}. 126

Figure 5.18: (a) $I_{D1} - I_{D2}$ vs. I_{BIAS} and $V_{DIO} - V_{OUT}$ vs. I_{BIAS} for the unbalanced amplifier of Figure 4.7. (b) $I_{D1} - I_{D2}$ vs. I_{BIAS} and $V_{DIO} - V_{OUT}$ vs. I_{BIAS} for the balanced self-cascode amplifier of Figure 4.9. V_{IN1} and V_{IN2} of both amplifier's were biased at 650 mV..... 128

Figure 5.19: A_V vs. Frequency for the balanced simple amplifier (Figure 4.8) and the balanced self-cascode amplifier (Figure 4.9). I_{BIAS} = 16 µA. The load capacitance was 1 pF. V_{IN1} and V_{IN2} of both amplifier's were biased at 650 mV. The intrinsic gain of M1 in Figure 4.8 was 27.87 dB..... 129

Figure 5.20: (a) I_G vs. V_{COM} and (b) A_V vs. Frequency for two self-cascode differential amplifiers. I_{G_Cancel} and A_{V_Cancel} refer to an amplifier with input current cancelation (Figure 4.10). I_G refers to an amplifier without input current cancellation. The amplifier without input current cancellations was the same as the amplifier with input current cancellation except that it did not have M3, M7-M9, M11, M15, and the error amplifier of Figure 4.10. 130

Figure 5.21: (a) V_{REF} vs. T and (b) R_{OUT} vs. frequency for the AC Simulation techniques described in Section 4.6. 131

Figure 5.22: Monte Carlo analysis of V_{REF} vs. T for the thick-oxide sub-1 V bandgap voltage reference presented in [116]. The graph shows 300 Monte Carlo runs across three different supply voltages (0.9 V, 1.0 V, and 1.1 V). Each supply voltage simulated 100 runs. . 134

Figure 5.23: Comparison of the Monte Carlo analyses of the thick-oxide sub-1 V bandgap voltage reference presented in [116] and the thick-to-ultra-thin sub-1 V bandgap voltage reference shown in [116]. The graph shows 300 Monte Carlo runs across three different supply voltages (0.9 V, 1.0 V, and 1.1 V). Each supply voltage simulated 100 runs..... 135

Figure 5.24: V_{REF} vs. T and V_{SG} of SC5 vs. T for V_{DD} = 0.9 V at the fast NMOS process corner and the fast PMOS process corner for the voltage reference of Figure 4.16. The cascoded transistors of the PMOS mirrors were sized with W = 400 µm and L = 0.25 µm. The cascoding transistors of the PMOS were sized with W = 800 µm and L = 0.25 µm..... 141

Figure 5.25: V_{REF} vs. T and V_{DS} of the error amplifier's input pair vs. T for V_{DD} = 0.9 V at the fast NMOS process corner and the slow PMOS process corner. The input pair was sized with W = 400 µm and L = 0.25 µm. 144
Figure 5.26: $I_{G2} - I_{G12}$ vs. T and $V_{GS2} - V_{GS12}$ (ΔV_{GS}) vs. T for V_{DD} = 1.1 V at the slow NMOS process corner and the slow PMOS process corner. 145
Figure 5.27: (a) Monte Carlo analysis of V_{REF} vs. T for the ultra-thin-oxide sub-1 V bandgap voltage reference shown in Figure 4.16. The graph shows 300 Monte Carlo runs across three different supply voltages (0.9 V, 1.0 V, and 1.1 V). Each supply voltage simulated 100 runs. (b) Comparison of the Monte Carlo analyses of the ultra-thin-oxide sub-1 V bandgap voltage reference shown of Figure 4.16 and the thick-oxide bandgap voltage reference presented in [116]. 146
Figure 5.28: Process Corners analysis of V_{REF} vs. T for the ultra-thin oxide sub-1 V bandgap voltage reference shown in Figure 4.16. 147
Figure 5.29: (a) Process Corners analysis of V_{REF} vs. V_{DD} for the ultra-thin oxide sub-1 V bandgap voltage reference shown in Figure 4.16. (b) V_{REF} vs. t for a V_{DD} rise time of 1 µs for the ultra-thin oxide sub-1 V bandgap voltage reference shown in Figure 4.16. 148
Figure 5.30: (a) V_{REF} vs. t for a V_{DD} rise time of 10 ms for the ultra-thin oxide sub-1 V bandgap voltage reference shown in Figure 4.16. (b) V_{REF} vs. t for a V_{DD} rise time of 10 s for the ultra-thin oxide sub-1 V bandgap voltage reference shown in Figure 4.16. 149
Figure 5.31: (a) Process Corners analysis of V_{REF} vs. T for the ultra-thin oxide sub-1 V bandgap voltage reference shown in Figure 4.16. (b) Process Corners analysis of I_G of the loading transistor vs. T for the ultra-thin oxide sub-1 V bandgap voltage reference shown in Figure 4.16. V_{REF} was loaded down with the gate of an NMOS transistor that had a PTAT current source connected to its source terminal (see M_L and I_{LOAD} in Figure 4.16). The current source had a temperature slope of 170 nA/°C and a value of 50 µA at T = 25 °C. Three loading transistor channel lengths were simulated: 0.5 µm, 1 µm, and 2 µm. The width of the loading transistor was set equal to 100 µm. 151
Figure 5.32: Sensitivity analysis of V_{REF} vs. T for the BSIM4 direct tunneling model parameter *aigc*. V_{DD} = 1.0 V. The process corner was TT. 152
Figure 5.33: Sensitivity analysis of V_{REF} vs. T for the BSIM4 direct tunneling model parameter *poxedge*. V_{DD} = 1.0 V. The process corner was TT. 153
Figure 5.34: Sensitivity analysis of V_{REF} vs. T for the BSIM4 direct tunneling model parameter *aigsd*. V_{DD} = 1.0 V. The process corner was TT. 154
Figure 5.35: Sensitivity analysis of V_{REF} vs. T for the BSIM4 direct tunneling model parameter *toxref*. V_{DD} = 1.0 V. The process corner was TT. 155
Figure 5.36: Layout of the standard ultra-thin oxide sub-1 V bandgap voltage reference of Figure 4.16 (202.165 µm by 198.1 µm). 157
Figure 5.37: Layout of the thick-oxide bandgap voltage reference of Figure 3.22 (183.17 µm by 187.69 µm). 158
Figure 5.38: Layout of the body-biased version of the standard ultra-thin oxide bandgap voltage reference of Figure 4.16 (248.265 µm by 209.43 µm). 158
Figure 5.39: Complete layout of the designed chip. 160
Figure A.1: Low-frequency small-signal equivalent of a self-cascode amplifier. 163
Figure B.1: Simplified representation of the sub-1 V bandgap voltage reference presented in [116]. 165
Figure B.2: Simplified representation of the sub-1 V bandgap voltage reference presented in [116]. The schematic includes input offset voltage, input bias current, and input offset current. 167

LIST OF TABLES

Table 5.1: Comparison of the simulated voltage references. 147

LIST OF COMMONLY USED SYMBOLS AND ABBREVIATIONS

A_E	[μm] BJT emitter area
A_i	Current gain
A_R	MOSFET aspect ratio
A_V	[dB] Voltage gain
C_{OX}	[F/cm^2] MOSFET oxide capacitance
I_B	[A] BJT base current
I_{BIAS}	[A] Amplifier and MOSFET drain bias current
I_{BODY}	[A] MOSFET body current
I_C	[A] BJT collector current
I_D	[A] MOSFET drain current
I_{DS}	[A] MOSFET drain-to-source current
I_E	[A] BJT emitter current
I_G	[A] Total MOSFET gate current due to direct tunneling
I_{GB}	[A] MOSFET gate-to-body direct tunneling current
I_{GCD}	[A] MOSFET gate-to-channel-to-drain direct tunneling current
I_{GCS}	[A] gate-to-channel-to-source direct tunneling current
I_{GD}	[A] gate-to-drain overlap region direct tunneling current
I_{GS}	[A] gate-to-source overlap region direct tunneling current
I_{IN}	[A] Current mirror input current
I_{IN_B}	[A] Amplifier input bias current
I_{OS}	[A] Amplifier input offset current
I_{OUT}	[A] Amplifier and current mirror output current
J_T	[A/cm^2] Direct tunneling current density
L	[μm] MOSFET channel length
P_T	Probability a carrier will directly tunnel through a potential barrier
r_O	[$M\Omega$] MOSFET small-signal output resistance
R_{OUT}	[$M\Omega$] Amplifier and current mirror output resistance
r_μ	[$M\Omega$] BJT small-signal collector-to-base resistance
r_{μ_MOS}	[$M\Omega$] MOSFET small-signal drain-to-gate resistance
r_π	[$M\Omega$] BJT small-signal base resistance
r_{π_MOS}	[$M\Omega$] MOSFET small-signal gate resistance
S_F	Ratio of aspect ratios between top and bottom transistors of a self-cascode
S_{F_T}	Ratio of aspect ratios between top and middle transistors of a triple self-cascode
T	[K] or [$°C$] Temperature
t_{ox}	[nm] MOSFET oxide thickness
V_B	[V] BJT base voltage
V_{BE}	[V] BJT base-to-emitter voltage
V_{BIAS}	[V] MOSFET gate-bias voltage
V_{BODY}	[V] MOSFET body voltage
V_{BS}	[V] MOSFET body-to-source voltage
V_D	[V] MOSFET drain voltage
V_{DD}	[V] Supply voltage
V_{DIO}	[V] Gate voltage of diode-connected active load in differential amplifier
V_{DS}	[V] MOSFET drain-to-source voltage
V_{DSsat}	[V] MOSFET V_{DS} voltage required for saturation

V_E	[V] BJT Emitter Voltage
V_{EB}	[V] BJT emitter-to-base voltage
V_G	[V] MOSFET gate voltage
V_{GB}	[V] MOSFET gate-to-body voltage
V_{GD}	[V] MOSFET gate-to-drain voltage
V_{GO}	[V] Bandgap voltage
V_{GS}	[V] MOSFET gate-to-source voltage
V_M	[V] Amplifier voltage at negative input terminal
V_{OS}	[V] Amplifier Offset Voltage
V_{OUT}	[V] Amplifier and current mirror output voltage
V_{OX}	[V] MOSFET voltage drop across the oxide
V_P	[V] Amplifier voltage at positive input terminal
V_{REF}	[V] Output voltage of a voltage reference
V_S	[V] MOSFET source voltage
V_t	[V] Thermal voltage
V_{TH}	[V] MOSFET threshold voltage
W	[μm] MOSFET channel width
α_F	BJT ratio of emitter current to collector current
α_{F_MOS}	MOSFET ratio of drain current to source current
β_0	BJT small-signal forward current gain
β_{0_MOS}	MOSFET small-signal forward current gain
β_F	BJT large-signal forward current gain
β_{F_MOS}	MOSFET large-signal forward current gain
λ	[V^{-1}] CLM coefficient
BSIM	Berkeley Short-Channel Insulated Gate Field-Effect Transistor Model
BJT	Bipolar Junction Transistor
CCCS	Current-Controlled Current Source
CLM	Channel Length Modulation
CMOS	Complementary Metal-Oxide-Semiconductor
CTAT	Complementary to Absolute Temperature
DIBL	Drain-Induced Barrier Lowering
DITS	Drain-Induced Threshold Shift
ECB	Electrons Tunneling from the Conduction Band
ESD	Electrostatic Discharge
EVB	Electrons Tunneling from the Valence Band
FF	Fast PMOS, Fast NMOS Process Corner
FS	Fast PMOS, Slow NMOS Process Corner
HVB	Holes Tunneling from the Valence Band
IC	Integrated Circuit
MOSFET	Metal-Oxide-Semiconductor Field-Effect Transistor
NMOS	n-type Metal-Oxide-Semiconductor
NPN	n-type p-type n-type
PMOS	p-type Metal-Oxide-Semiconductor
PNP	p-type n-type p-type
PTAT	Proportional to Absolute Temperature
SCBE	Substrate Current-Induced Body Effect
SCR	Silicon-Controlled Rectifier
SF	Slow PMOS, Fast NMOS Process Corner
SS	Slow PMOS, Slow NMOS Process Corner
TT	Typical PMOS, Typical NMOS Process Corner

VCVS	Voltage-Controlled Voltage Source
VCCS	Voltage-Controlled Current Source

CHAPTER 1
INTRODUCTION

The ability to do mixed-signal integrated circuit (IC) design[1] in a complementary metal-oxide-semiconductor (CMOS) technology has been a driving force for manufacturing personal mobile electronic products such as cellular phones, digital audio players, and personal digital assistants [1]. These products are notorious for being extremely compact while providing functionality comparable to that of a personal computer. Their demand has rapidly increased over the past ten years. For example, in 2000, the number of mobile subscribers was estimated at 650 million. This number rose to 5 billion in 2010 [2]–[3]. This type of growth fuels competition between businesses to release their next-generation products. Typical goals of these products include additional features and improved performance. With regard to the electronics that meet these goals, they are often implemented in a scaled CMOS technology [4]. To minimize the time to market and ease the design process, it is desirable that the mixed-signal design techniques used in previous product generations apply in these scaled technologies.

Over the past four decades, as CMOS has scaled, mixed-signal design techniques have been used in technologies with minimum channel lengths as large as 5 μm to as small as 22 nm. The main motivating factor for this scaling has been the reduction in cost obtained by the increase in component density [5]. Another motivating factor is the increase in device frequency response, which has allowed radio-frequency (RF) circuitry to be implemented on-chip [6]. A third motivating factor for scaled CMOS technologies

[1] A mixed-signal system is defined as a system that contains analog and digital components.

is that device functionality, ideally, remains constant. This translates into mixed-signal design techniques that can readily be applied to create system-on-chips (SoCs) and system-in-packages (SiPs) in any given technology [7]. Of course, in reality, device functionality is not independent of scaling. For example, when a device is scaled, problems arise that must be taken into account by process engineers and circuit designers. Process engineers solve these problems with novel fabrication techniques [8]. Circuit designers solve these problems with creative circuit architectures. These problems are often attacked with digital performance in mind because of the high demand for digital electronics. This explains why digital metrics like switching speed, packing density, and power consumption are often given as reasons to move from one generation of CMOS to the next.

Unfortunately, this approach to scaling has made life difficult for analog IC design engineers. For example, given that processes are optimized for digital operation, analog performance metrics like supply voltage headroom, intrinsic gain, and signal-to-noise ratio (SNR), which often degrade with scaling, become secondary considerations [1], [9]. Process modifications are typically not made to mitigate these degradations out of fear they will disrupt digital operation. This significantly increases the complexity of analog design in scaled CMOS technologies. Fortunately, degradation in device performance is something analog designers have dealt with before. In certain aspects, performance has been degrading ever since the switch from bipolar junction transistors (BJTs) to metal-oxide-semiconductor field-effect transistors (MOSFETs) [10]. Designers overcame the switch to CMOS and its subsequent scaling by inventing circuit architectures that made mixed-signal design possible in scaled technologies [11]. This

trend will need to continue if future scaled CMOS technologies are going to be used to build next-generation electronics.

In sub-100 nm channel length CMOS technologies, many problems are caused by the thin insulating layer between the gate and silicon channel. This layer is often less than 3 nm thick [12]. In these so-called ultra-thin oxide technologies, carriers (electrons or holes) are able to tunnel directly through the oxide and conduct current. This type of current, which is proportional to device area, is referred to as direct tunneling and is a source of gate current in MOSFETs [13]–[15]. Other sources include Fowler-Nordheim (FN) tunneling and hot electron-induced gate current, which are typically considered negligible under normal operating conditions in processes with the supply voltage less than or equal to 1 V [16][2]. However, direct tunneling has become a major problem. In 2007, the International Technology Roadmap for Semiconductors (ITRS) cited the shrinking oxide, and the resultant performance degradations caused by direct tunneling, as a grand challenge to device scaling [17]. For example, Figure 1.1 plots the drain-to-gate current ratio ($\beta_{F\_MOS} \equiv |I_D/I_G|$) vs. V_{GS} and I_G vs. V_{GS} for an NMOS device with a channel width, W, of 20 μm and a channel length, L, of 5 μm in IBM's 10SF 65 nm technology (V_{DD} = 1 V, t_{ox} = 1.25 nm) [18]–[20]. The figure shows β_{F_MOS} values less than 20 and I_G values in the μA range. Compared to previous generations of CMOS, these results suggest direct tunneling-induced gate current is not negligible and must be considered when designing in traditional (non-high-κ/non-metal gate) ultra-thin oxide CMOS technologies [21].

[2]The terms gate current and direct tunneling will be used interchangeably throughout this work. It is understood that sources other than direct tunneling can contribute to gate current, notably FN tunneling and hot electrons. However, these currents are negligible in ultra-thin oxide CMOS technologies. Given that this work focuses on these technologies, these sources will not be considered, making direct tunneling the dominant source of gate current.

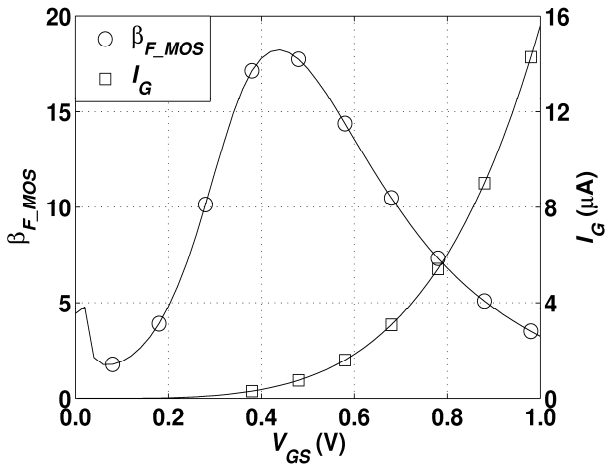

Figure 1.1: Simulated β_{F_MOS} vs. V_{GS} and I_G vs. V_{GS} for an NMOS transistor with W = 20 μm and L = 5 μm.

To overcome the challenge of direct tunneling, the ITRS called for a new gate stack to reduce its impact on circuit performance. This new gate stack is made up of a high-κ dielectric and a metal gate electrode [22]. There are several potential problems with this structure. First, compared to traditional CMOS technologies, this new gate stack comes with a significant increase in cost due to processing complexities [23]–[26]. This implies that traditional ultra-thin oxide technologies will have longer lives in the economic forefront than previous generations of CMOS. Second, high-κ/metal gate structures can result in threshold voltage pinning, mobility degradation, and phonon scattering [27]–[28]. Third, there is debate among the manufacturing community about whether the gate-first or the gate-last approach should be used when building the new gate stack [29]. Fourth, the high-κ/metal gate may not reduce direct tunneling to a point where it is negligible in analog design [30]–[31]. These problems suggest that circuit techniques are needed to minimize the negative effects of direct tunneling in existing and future ultra-thin oxide technologies.

The fact that non-negligible current can flow through the gate of a MOSFET invalidates the simplifying circuit design assumption of infinite gate resistance. This impacts analog and digital design. Typically, in digital applications, gate current is seen as a leakage source that contributes to overall power consumption. Digital techniques to minimize the negative effects of this current were presented in [16], [32]–[34]. The impact of gate current on analog circuit design was studied in [18]. It was shown that gate current can degrade matching, reduce frequency response, increase noise, and render long-channel devices practically useless. There have not been any published circuit techniques illustrating how these leaky devices can still be used for analog design. Instead, designers often opt for a set of complimentary thick(er) oxide devices, which have negligible gate current, to implement the analog component of a mixed-signal system [1]. By doing this, they increase cost and deviate from the true mixed-signal paradigm of designing an analog and digital system with a single set of complimentary devices. Therefore, given that digital solutions are available and that traditional ultra-thin oxide CMOS technologies will be revenue generators for an extended period of time, analog circuit solutions are needed to allow useful mixed-signal design using only ultra-thin oxide MOSFETs.

This work develops a methodology that allows the design of analog systems with ultra-thin oxide MOSFETs. This methodology focuses on transistor sizing, DC biasing, and the design of current mirrors and differential amplifiers. It attempts to minimize, balance, and cancel the negative effects of direct tunneling on analog design in traditional ultra-thin oxide CMOS technologies. The methodology requires only ultra-thin oxide devices and is investigated in IBM's 10SF 65 nm CMOS technology, which has a

nominal V_{DD} of 1 V and a physical oxide thickness of 1.25 nm. Theoretical analysis and simulation are used to develop the methodology. The methodology does not aggravate existing analog nanoscale CMOS problems such as reduced voltage headroom, decreased intrinsic gain, and reduced SNR. Note that the methodology focuses on low-frequency performance because the effects of direct tunneling have been shown to be negligible at higher frequencies [18].

A sub-1 V bandgap voltage reference is designed and implemented using the developed methodology in IBM's 10SF 65 nm process. It requires only ultra-thin oxide MOSFETs and its performance is used to illustrate that the negative effects of direct tunneling can be suppressed by following the techniques outlined in this document. A voltage reference was chosen because of its ubiquitous nature and due to the fact that it is a fundamental precision analog system designed to produce a voltage independent of variations in the power supply (V_{DD}), temperature (T), and process. Voltage references are widely used in mixed-signal systems, such as digital-to-analog converters (DACs), analog-to-digital-converters (ADCs), DC-DC converters, operational amplifiers, and linear regulators [35]. They are built using differential amplifiers and current mirrors, which are both sensitive to gate current [36]. The developed methodology presents techniques that overcome these sensitivities. Voltage references are also sensitive to mismatch between MOSFETs designed to be identical [37]. Given that gate current is proportional to device area, its negative effects seemingly limit the use of large-area transistors. However, this work shows that the tradeoff between gate current and mismatch can be minimized via informed device sizing. The voltage reference is used as a vehicle to prove that analog systems can be constructed with ultra-thin oxide

MOSFETs. Its performance is compared to a thick-oxide voltage reference as a means of demonstrating that ultra-thin oxide MOSFETs can achieve performance similar to that of more expensive thick(er) oxide MOSFETs.

This document is structured as follows[3]. Chapter 2 covers the main objectives this work strived to accomplish. Chapter 3 reviews the relevant background information relating to this work. Chapter 4 presents the approach that was taken to meet the objectives outlined in Chapter 2. Chapter 5 presents the results of this work and discusses their importance. Chapter 6 concludes the document.

[3] Discussions involving single transistors will be treated from the standpoint of an NMOS device unless otherwise noted.

CHAPTER 2
OBJECTIVES

The main goal of this work was to show that analog systems can be built using ultra-thin oxide MOSFETs. In order to accomplish this goal, three objectives were realized. These objectives are stated in the following three paragraphs.

The first objective was to demonstrate that gate current creates serious problems for analog device performance. This was accomplished by referencing existing literature and analyzing, via simulation, the effects of gate current on ultra-thin oxide MOSFETs in IBM's 10SF 65 nm CMOS technology (t_{ox} = 1.25 nm, V_{DD} = 1 V). Where appropriate, theoretical analysis was used to illustrate how gate current hinders device performance.

The second objective was to develop a methodology for implementing analog circuits with ultra-thin oxide MOSFETs. Given that gate current is not the only problem faced by analog designers in ultra-thin oxide technologies, it was desirable that the methodology not introduce new problems or aggravate existing problems. The developed methodology should coexist with other low-voltage techniques. The need for a methodology was motivated by showing, via simulation, the negative impact gate current can have on transistor sizing, DC biasing, and the design of current mirrors and differential amplifiers.

The third objective was to use the developed methodology to implement an analog system[4] using only ultra-thin oxide MOSFETs. The system chosen was a sub-1 V

[4] An analog system is defined as a circuit that makes use of fundamental building blocks such as amplifiers and current mirrors.

bandgap voltage reference. The voltage reference was designed, simulated, and laid out. Monte Carlo and process corners analyses were used to study its performance. The reference was compared to a thick-oxide version in the same technology as a means of demonstrating the developed methodology could produce results similar to those obtained using thick(er) oxide devices.

It was a goal of this work to fabricate and test the designed voltage reference in order to prove the effectiveness of the developed methodology. A sponsored fabrication was awarded based on technical merit via the MOSIS Educational Program [38]. The target technology was IBM's 10SF technology. The design, simulation, and layout of a 2 mm x 2 mm chip was completed and sent to MOSIS. However, for reasons beyond the author's control, this fabrication was delayed over 2 years. Therefore, fabrication results were unable to be included in this document. However, if fabrication does eventually occur after the publishing of this document, the results will be made available via a scholarly journal. Note that fabrication could have been pursued in a thick(er) oxide technology. However, to prove the value of this work, it is desirable that the proposed solutions function when the problems caused by gate current are at their worst. Therefore, larger, less expensive technologies with thicker oxides and negligible gate current are not applicable. This limits the fabrication of the voltage reference to expensive technologies with minimum channel lengths less than 100 nm, oxide thicknesses less than 2 nm, and nominal supply voltages less than 1 V. Note that even though fabrication did not occur, the Rochester Institute of Technology chose to patent the developed sub-1 V bandgap voltage reference [39].

CHAPTER 3
BACKGROUND

This chapter presents the relevant background for this work. It has three major sections. The first section reviews some of the difficulties involved in designing analog circuits in nanoscale CMOS technologies. The second section reviews the physical mechanisms of gate current and notes how previous work has treated its impact on circuit design. The last section reviews the fundamentals of voltage references, with a focus on those designed with supply voltages of 1 V or less.

3.1 Analog Design in Nanoscale CMOS

Nanoscale CMOS technologies are typically optimized for digital performance by providing faster speeds, lower power, and smaller area. These optimizations often pose significant problems to analog design, such as reduced output resistance, smaller supply voltages, and increased variability [18]. This section reviews these problems and the techniques used to cope with them. Its motivation stems from the main goal of this work, which is to show that analog systems can be built with ultra-thin oxide MOSFETs. In order to accomplish this goal, this work starts with established techniques that solve the aforementioned problems. These techniques include self-cascoding, sub-V_{TH} operation, and body-biasing. It is desired that these techniques, along with those developed in this work, be used in combination to show analog system design is possible using ultra-thin oxide MOSFETs.

3.1.1 Output Resistance Degradation

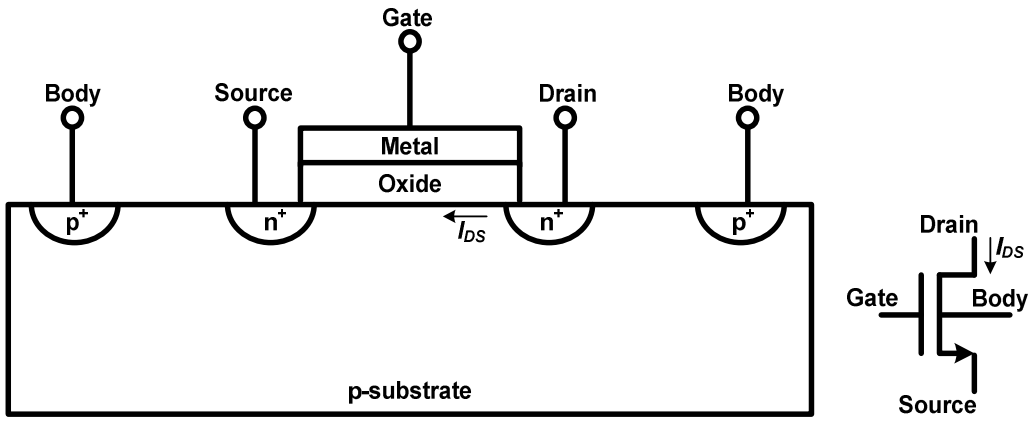

Figure 3.1: Simplified cross section and symbol of an NMOS transistor. The gate, drain, body, and source represent input/output terminals. I_{DS} is the drain-to-source current. The drain and source regions are represented by heavily doped n-type regions (n^+) while contacts to the body are represented by a heavily doped p-type (p^+) region. The substrate is p-type.

Figure 3.1 shows the cross section and circuit representation of an ideal long-channel n-type MOSFET (NMOS) device. The figure contains the gate, drain, source, and body terminals. It also contains the heavily doped n-type (n^+) drain/source regions along with a p-type substrate. The oxide and metal layers represent the gate stack. Perhaps the most important characteristic of this device is the current flowing from the drain terminal to the source terminal, labeled I_{DS} in Figure 3.1. The "square-law" approximation that is strictly valid only for long-channel devices is often used to model this current and is found in several textbooks on electronics and semiconductor devices [36], [40]–[48]. It serves as a basis for analog and digital circuit design, and is often the standard to which modern devices are held. It is formulated as:

$$I_{DS} = \frac{W\mu C_{ox}}{2L}(V_{GS} - V_{TH})^2 \tag{3.1}$$

where μ is the mobility, C_{OX} is the oxide capacitance per unit area, W is the channel width, L is the channel length, V_{GS} is the gate-to-source voltage, and V_{TH} is the threshold

voltage. This equation assumes the device is operated in the saturation region ($V_{GS} > V_{TH}$, $V_{DS} \geq V_{DSsat}$, $V_{DSsat} \approx V_{GS} - V_{TH}$). From an analog standpoint, one important outcome of this assumption is that I_{DS} is independent of V_{DS}, resulting in an infinite small-signal output resistance, r_O. This outcome is often used in textbooks to simplify amplifier and current mirror design [45]. However, in short-channel CMOS technologies, (3.1) is grossly inaccurate. This inaccuracy results from the fact that devices are not always operated in the saturation region and even when they are, they suffer from several non-ideal output-resistance-degrading short-channel effects. Modifications to circuit architectures must be made to account for these non-idealities. Some examples of short-channel effects include channel length modulation (CLM) [43], drain-induced barrier lowering (DIBL) [49], drain-induced threshold shift (DITS) [50], and substrate current-induced body effect (SCBE) [40].

3.1.1.1 Channel Length Modulation (CLM)

Channel length modulation typically occurs when a MOSFET is operated in the saturation region. Ideally, in this region, the concentration of inversion charge along the surface of the channel is constant. However, due to the varying potential difference between the gate and horizontal position in the substrate, it is not [43]. This causes the inversion charge concentration to decrease near the drain end of the channel. As this charge concentration decreases, the effective channel length of the device decreases, causing I_{DS} to increase ($I_{DS} \propto 1/L$). Therefore, I_{DS} is modulated by V_{DS} via the changes in the substrate surface potential.

Channel length modulation can be modeled in several different ways [43]. Most of these models involve the idea of the channel being "pinched off". For example, pinchoff can be modeled as the channel location at which the inversion charge goes to zero. This model requires the carriers to move at infinite speeds in order to travel through the depletion region [43]. The requirement of infinite speeds makes this approach physically implausible and mathematically intractable. A different approach involves pinchoff being modeled as the V_{DS} value at which carrier velocities saturate. In this model, instead of going to zero, the inversion charge becomes saturated at some point along the channel [43]. This provides an improved physical explanation of CLM, which allows its effects to be accurately captured in compact models [15].

Textbooks often model channel length modulation by introducing a multiplicative term into (3.1) [44]–[45]. This term contains λ, a constant, which is referred to as the channel length modulation coefficient. It represents a first-order model of the change in I_{DS} with V_{DS} and is similar to the Early voltage of a BJT [44]. This dependence is typically modeled as [44]:

$$I_{DS} = \frac{W\mu C_{ox}}{2L}(V_{GS} - V_{TH})^2(1 + \lambda \cdot V_{DS}) \tag{3.2}$$

where $(1 + \lambda \cdot V_{DS})$ is the added term. The small-signal output resistance of this equation is approximately, $r_O \approx 1/\ \lambda \cdot I_{DS}$, where it is assumed $\lambda \cdot V_{DS} << 1$. This output resistance model is predominantly used in textbooks when designing analog circuits [44]–[45]. Typically, to avoid the effects of channel length modulation, analog designers opt for long(er) channel devices. By doing this, they increase the length of the region that has a constant concentration of inversion charge along the channel. This effectively reduces

the λ term of (3.2), which implies an increase in r_O. For example, consider Figure 3.2 (a), which plots I_D vs. V_{DS} and r_O vs. V_{DS} for an NMOS transistor with a channel length of 50 nm. The device achieved a maximum output resistance of 259 kΩ. Its output resistance was well below 200 kΩ for $V_{DS} > 0.5$ V. Figure 3.2 (b) shows the same plot for an NMOS transistor with a channel length of 1 μm. It achieved a maximum output resistance of 384 kΩ and its output resistance was above 200 kΩ for $V_{DS} > 200$ mV. This simple example shows that increasing channel length can result in significant increases in device output resistance.

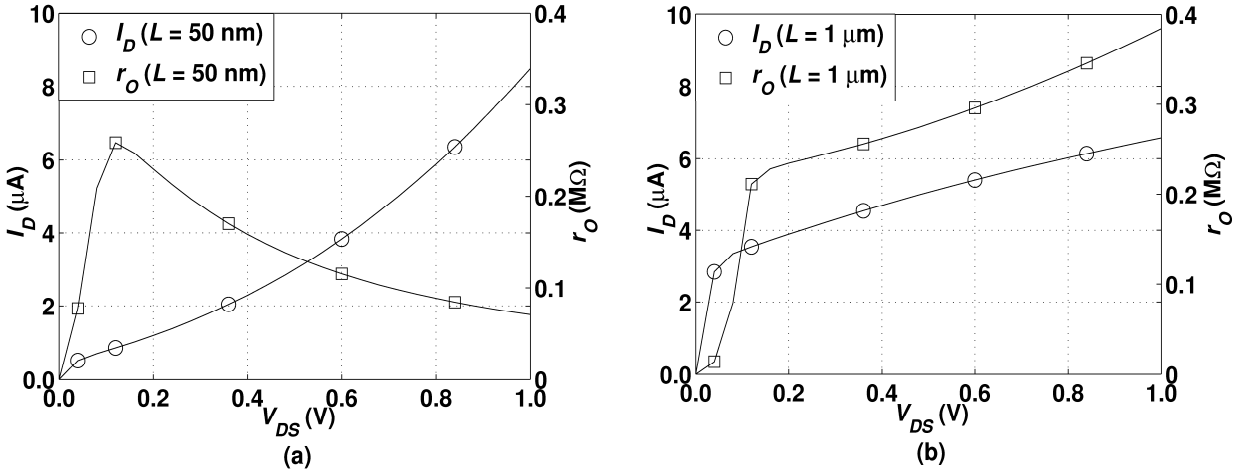

Figure 3.2: (a) Simulated I_D vs. V_{DS} and r_O vs. V_{DS} for an NMOS transistor in the obtained 65 nm process. W = 1 μm, L = 50 nm, and V_{GS} = 0.3 V. (b) Simulated I_D vs. V_{DS} and r_O vs. V_{DS} for an NMOS transistor in the obtained 65 nm process. W = 10 μm, L = 1 μm, and V_{GS} = 0.3 V.

3.1.1.2 Drain-Induced Barrier Lowering (DIBL)

Another source of output resistance degradation is drain-induced barrier lowering (DIBL), which takes place in all regions of operation. DIBL occurs when the potential barrier seen by electrons at the source terminal decreases due to increases in V_{DS} [43], [49]. It is dependent upon the source and drain depletion regions. The more of the channel these regions occupy, the more impact DIBL has on performance. This leads to DIBL impacting short-channel devices more than long-channel devices. When these

depletion regions occupy a significant portion of a short-channel device, the potential barrier seen by electrons at the source rapidly decreases with increases in V_{DS}. This implies that I_{DS} increases with increasing V_{DS} because more electrons can overcome the reduced barriers and contribute to current flow. This results in I_{DS} being dependent on V_{DS}, which reduces r_O. If V_{DS} becomes too large, a low-energy path is established from source to drain that is determined by V_{DS} rather than V_{GS}, causing punchthrough. Therefore, DIBL can be considered a precursor to punchthrough.

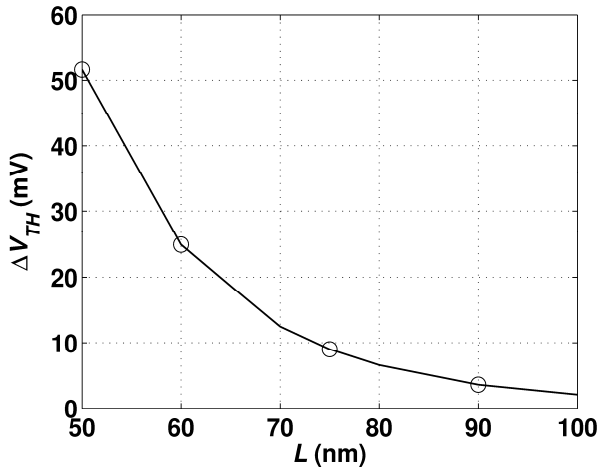

Figure 3.3: Simulated ΔV_{TH} vs. L for two NMOS transistors with different V_{DS} voltages in the obtained 65 nm process. Each transistor had W = 1 µm and V_{GS} = 0.3 V. The V_{DS} voltages were 0.1 V and 1.0 V.

DIBL is typically modeled as a shift in V_{TH} because it occurs over all regions of operation [43]. When DIBL occurs, transistors conduct more current than what would typically be expected for a given V_{GS}. For example, to measure DIBL, I_{DS} vs. V_{GS} plots are generated at different V_{DS} values. The V_{TH} of each plot is then extracted. The differences in V_{TH} between these plots are representative of the impact of DIBL on performance. For example, consider Figure 3.3, which plots ΔV_{TH} vs. L for two identically sized NMOS transistors with equal V_{GS} voltages but different V_{DS} voltages. The effects of DIBL can be seen at smaller channel lengths, with differences in threshold

voltages of up to 50 mV. Typically, with DIBL, I_{DS} increases with increasing V_{DS}, resulting in the effective V_{TH} being reduced. Therefore, I_{DS} is modeled as being a function of V_{DS} through V_{TH}, which reduces r_O.

One technique that can be used to minimize the effects of DIBL is to design with longer channel devices. This approach ensures that the drain and source depletion regions do not occupy a significant portion of the channel, restricting their impact on I_{DS}. Another technique is to design with smaller V_{DS} voltages. This technique is effective because the change in V_{TH} decreases with decreasing V_{DS}.

3.1.1.3 Drain-Induced Threshold Shift (DITS)

Another source of output resistance degradation is drain-induced threshold shift (DITS), which describes the effect of V_{DS} on I_{DS} in long-channel MOSFETs [50]–[52]. This is a relatively new phenomenon that occurs because of the halo and pocket implants [9], [18], [50]. These implants are designed to prevent punchthrough by adjusting V_{TH} of short-channel devices. Without these implants, V_{TH} decreases significantly with decreasing L, resulting in excessive sub-threshold leakage current in digital circuits [53]. The halo implant places two heavily doped p-type regions near the source and drain junctions. The rest of the channel has a doping concentration less than that of these regions. Therefore, as L decreases, the effective doping concentration of the channel increases because the higher concentration regions introduced by the halo implant occupy more of the channel. This causes V_{TH} to increase, which helps reduce sub-threshold leakage. For example, consider Figure 3.4, which plots V_{TH} vs. L for an NMOS transistor in the obtained 65 nm process. As L was swept from 1 μm to 50 nm, V_{TH} increased by

200 mV. The increases in V_{TH} with reductions in L are referred to as drain-induced threshold shift [50] or the reverse short-channel effect [43].

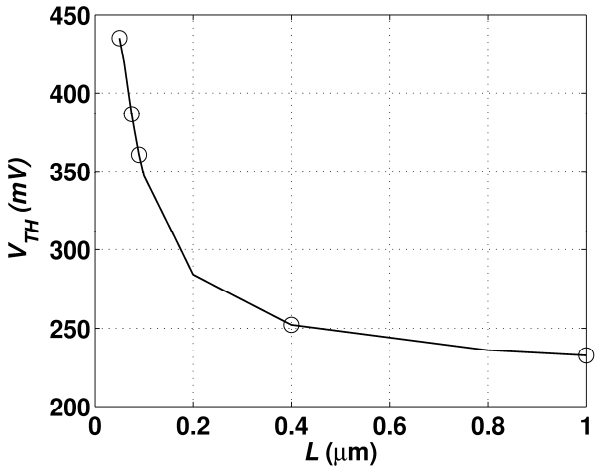

Figure 3.4: Simulated V_{TH} vs. L of an NMOS transistor in the obtained 65 nm process. W = 1 μm and V_{DS} = 100 mV. $V_{GS} = V_{DS}$ = 0.3 V.

The halo and pocket implants significantly impact the performance of long-channel devices. They form potential barriers at both the source and drain ends of the channel due to their higher doping concentration. Also, they make the channel look as if it has three different V_{TH} values: one at the source end, one at the drain end, and one in the middle portion of the channel. The threshold voltages at the source and drain ends are larger than the one in the middle because of the larger doping concentrations introduced at these ends. For a long-channel device, the overall V_{TH} is approximately equal to the middle V_{TH} because the doping concentration in the middle dominates the channel. As V_{GS} increases and eventually approaches the middle V_{TH}, the channel can be considered conductive because the conditions for inversion have been met. However, the potential barriers created by the halo implant still exist. Thus, as V_{DS} increases, these barriers are modulated and more current than would be expected can flow through the channel. This results in a DIBL-like mechanism for long-channel devices and is modeled

as shift in V_{TH}. This degrades r_O because I_{DS} is dependent upon V_{DS} via V_{TH}. DITS creates a significant problem for analog designers because short-channel devices suffer from DIBL and CLM, while long-channel devices suffer from DITS. Typically, analog designers use long(er) channels to maximize r_O and minimize the effects of DIBL and CLM [44]. In nanoscale CMOS this cannot be done because of DITS, which makes it difficult to obtain the r_O values realized in previous CMOS generations. Process solutions to this problem have been suggested. For example, in [54]–[56] it was shown that using a single-side halo significantly improves r_O while still providing the desired V_{TH} roll-up. However, this approach can increase the difficulty of layout because the devices are no longer symmetric. Therefore, for symmetric nanoscale devices, channel length selection plays a critical role in analog device performance.

3.1.1.4 Substrate Current-Induced Body Effect (SCBE)

Yet another source of output resistance degradation is the substrate current-induced body effect (SCBE) [15],[40]. It degrades r_O under high-voltage conditions. For example, if the applied drain voltage is too large, breakdown can occur in the pn junction formed by the drain and substrate. When this happens, avalanche multiplication becomes the dominant mechanism of current flow in the device. This results in an increase of current flowing from the drain terminal to the body terminal, effectively reducing I_{DS}. Thus I_{DS} decreases with increases in V_{DS}, which degrades r_O. SCBE can be minimized by ensuring that applied voltages are less than or equal to V_{DD}.

3.1.2 Reductions in Supply Voltage

Degradations in output resistance is not the only problem faced by analog designers. Reduced supply voltages also pose significant challenges [57]–[63].

Examples of these challenges include decreased supply voltage headroom, inability to stack transistors, forced operation into the weak and moderate inversion regions, and reduced signal-to-noise ratio (SNR).

3.1.2.1 Supply Voltage Scaling

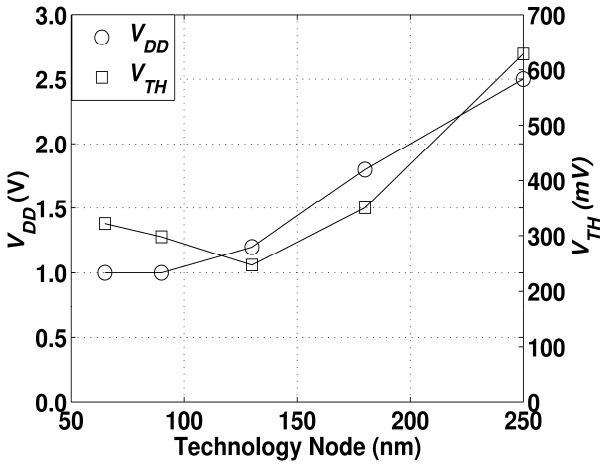

Figure 3.5: V_{DD} and V_{TH} vs. technology node. V_{TH} was extracted for an NMOS device with $V_{GS} = V_{DD}$, $V_{DS} = V_{DD}$, $V_{BS} = 0$, $L = L_{MIN}$, and $W = W_{MIN}$. L_{MIN} and W_{MIN} represent process minima for the channel length and channel width.

Figure 3.5 plots V_{DD} and V_{TH} vs. technology node for five different processes. It shows V_{DD} has reached a value of 1 V at the 65 nm node. One major motivation for reducing V_{DD} as technologies scale is to maintain electric field continuity [8], [12], [16], [43], [64]–[66]. This type of scaling is referred to as constant field scaling, where V_{DD} and V_{TH} are scaled at the same rate as W, L, and t_{ox}. This type of scaling ensures that internal electric fields remain unchanged, which helps maintain functionality and reliability. Figure 3.5 shows V_{DD} did not change between the 90 nm and 65 nm nodes. This is due to the impact of sub-threshold leakage on the performance of digital circuits [53], [67]–[69]. This off-state leakage increases with reductions in V_{TH}. The impact of sub-threshold leakage is monitored by the sub-threshold slope, S, which is defined as the V_{GS} required to change I_{DS} by a decade when operating in the sub-threshold region.

Ideally, S remains constant with scaling (60 mV/dec) [43]. Assuming this is true and assuming V_{DD} and V_{TH} scale at the same rate, it becomes increasingly difficult to turn devices on and off with scaling because the V_{GS} value needed to change I_{DS} by a decade is a larger percentage of V_{TH} and V_{DD}. This eats into digital noise margins, increases sub-threshold leakage, and makes it difficult to distinguish between weak and strong inversion. In digital circuits with millions of transistors, circuit techniques must be employed to reduce the effects of sub-threshold leakage [53]. The pocket and halo implants were introduced to minimize the impact of this leakage [51]. Considering that these implants cause V_{TH} to increase with reductions in L, it becomes more difficult to reduce V_{DD}. This can be seen by examining the V_{TH}/V_{DD} ratio for the different technologies in Figure 3.5. For example, at the 0.25 μm node, $V_{TH}/V_{DD} = 0.25$. However, at the 65 nm node, $V_{TH}/V_{DD} = 0.32$. This shows that with scaling V_{TH} is becoming a larger percentage of V_{DD}, which suggests that increasing V_{TH} to limit the impact of sub-threshold leakage will restrict further reductions in V_{DD}.

3.1.2.2 Reductions in Voltage Headroom

In order to understand the impact of V_{DD} reductions on analog supply voltage headroom, consider an amplifier designed in a technology with a nominal V_{DD} of 3.3 V. This amplifier may be able to meet specification with V_{DD} reduced to 2 V, giving it 1.3 V of voltage headroom. Now, consider an amplifier designed in a scaled technology with a nominal V_{DD} of 1 V. This amplifier may be able to meet specification with V_{DD} reduced to 0.9 V, giving it 100 mV of headroom. Compared to the 3.3 V amplifier, the 1 V amplifier has 1.2 V less headroom. Therefore, the 3.3 V amplifier is considered more robust to random V_{DD} shifts caused by power supply noise and electrostatic discharge

(ESD) events [48]. The differences in headroom stem from the voltage needed across each transistor to keep them in a desired region of operation and the number of transistors that must be stacked to achieve a certain level of performance.

3.1.2.3 Reduced Transistor Stacks

Decreased supply voltages also limit the number of transistors that can be stacked in circuit architectures. For example, assuming a device is in the saturation region and that $V_{DSsat} \approx V_{GS} - V_{TH}$, the total voltage needed across a stack of transistors to maintain saturation increases as the stack grows [36]. Therefore, in nanoscale technologies, where V_{DD} scales faster than V_{TH}, it becomes increasingly difficult to stack transistors without using a significant percentage of V_{DD}. One example of transistor stacking is cascoding, where devices are placed in series to enhance r_O [44]. In amplifiers, this technique leads to large voltage gains. If this technique is employed in nanoscale technologies, an amplifier's input common-mode range (ICMR) may be reduced [48]. ICMR is typically defined as the range of input common-mode voltages that maintain a constant voltage gain. A small ICMR limits an amplifier's input voltage swing, making it difficult to process a wide range of voltages. Also, if an amplifier is operated outside of this range, distortion could be introduced into the output because of changes in the amplifier's small-signal characteristics.

3.1.2.4 Weak and Moderate Inversion Operation

Traditionally, transistors are desired to operate in the saturation region ($V_{GS} > V_{TH}$, $V_{DS} \geq V_{DSsat}$). If V_{DD} and V_{TH} decrease at the same rate, scaling does not impact the voltage requirements for saturation. However, it was previously shown that V_{DD} scales at a faster rate than V_{TH}. Therefore, to achieve the same saturation condition in a scaled

technology, a larger percentage of V_{DD} is needed across the gate and source terminals. For example, consider a device operating with $V_{GS} = 1$ V in a technology with $V_{DD} = 3.3$ V and $V_{TH} = 0.7$ V. In this example, $V_{GS}/V_{DD} = 0.30$, which implies 30% of V_{DD} is being used between the gate and source terminal of the transistor. Now, consider a scaled device operating in a technology with $V_{DD} = 1$ V and $V_{TH} = 0.28$ V. To achieve the same overdrive voltage ($V_{GS} - V_{TH}$) as the non-scaled device, $V_{GS} = 0.58$ V. In this example, $V_{GS}/V_{DD} = 0.58$, which implies 58% of V_{DD} is being used from gate-to-source of this device. Compared to the less-scaled device, this is a 28% increase, which shows the extra voltage that must be used in the scaled technology to maintain a constant overdrive voltage.

To overcome this problem, devices can be operated in the weak and moderate inversion regions. This goes against the traditional textbook convention of operating all devices in strong inversion, specifically the saturation region. One motivation for using these regions is to remove the $V_{GS} > V_{TH}$ requirement for strong inversion, thus making it easier to stack devices and increase signal swing. To operate in weak inversion, V_{GS} must be significantly less than V_{TH}. In this region, MOSFETs function similar to BJTs and are dominated by diffusion current [43]–[44]. Because BJTs are well understood, device models exist that can accurately predict behavior in this region. Weak inversion is associated with small current densities because $V_{GS} << V_{TH}$ [43]–[44]. This leads to weak inversion being used to achieve high output resistance ($r_O \propto 1/I_{DS}$). It also results in reduced frequency response compared to strong inversion because g_m is reduced ($f_T \propto g_m$, where is f_T is the transition frequency and g_m is the gate transconductance) [44].

Therefore, weak inversion is suited for low-power, low-frequency, and high-gain applications.

If V_{GS} is within a few thermal voltages (kT/q) of V_{TH}, the transistor is said to be in the moderate inversion region [43]. This region is generally much more difficult to model than the weak or strong inversion regions because I_{DS} contains both drift and diffusion components. Interpolation is typically employed to model the moderate inversion region [43]. This results in fitting parameters and smoothing functions being used to ensure continuity of derivatives. Because of these modeling difficulties, caution must be exercised when operating devices in this region.

3.1.2.5 Reduced SNR

In [18] and [70], the impact of V_{DD} on SNR was investigated. It was shown that for a target SNR, the total power consumption must be increased if V_{DD} is decreased. This limits the achievable resolution of data converters designed for low-power applications [70]–[72]. For example, assuming converters are dominated by kT/C noise, on-chip capacitance must increase to achieve a desired SNR ($\text{SNR} \propto V_{DD}^2 \cdot C/kT$). Typically, this capacitance consumes a large amount of area. Therefore, to compensate the impact of V_{DD} reductions on SNR, power and area generally increase.

3.1.3 Modeling Complexity and Process Variations

As CMOS has moved into the nanometer regime, modeling complexity and process variations have become major concerns for circuit designers. These concerns stem from the atomistic dimensions of the devices and the limitations of the equipment fabricating them.

3.1.3.1 Modeling Complexity

Modeling transistor behavior has become increasingly difficult with scaling. This can be seen by comparing the relatively simple models used in older technologies [73] to the complex models used for nanoscale devices [15]. One reason is the introduction of new fabrication techniques, which include non-uniform doping profiles, stress/strain manipulation, salicide contacts, and source/drain extensions [66], [74]–[78]. These techniques alter device operation and must be accounted for by models, thus increasing their complexity. Another source of modeling difficulty stems from quantum mechanical effects caused by atomistic dimensions, large doping concentrations, and thin-oxides. Examples of quantum mechanical effects include energy quantization, direct tunneling, and the sub-surface inversion layer [43], [79]–[81]. These effects increase significantly as channel lengths drop below 100 nm, doping concentrations reach $10^{19}\,\mathrm{cm}^{-3}$, and oxide thicknesses approach 1 nm. They also limit achievable device performance, which magnifies the need to describe them accurately in compact models [82]–[86].

Leakage currents impose another complexity on device modeling. Examples of these currents include direct tunneling, reverse biased pn junction leakage, sub-threshold leakage, hot-carrier injection, gate-induced drain leakage (GIDL), and channel punchthrough current [16]. These currents are extremely important to power consumption in digital circuits and must be modeled accurately to gauge power profiles. [53], [67], [87]–[88].

Yet another source of modeling complexity stems from on-chip interconnects. In digital circuits, the impedance associated with these interconnects results in power supply

noise and ground bounce [89]–[90]. As supply voltages decrease and device densities increase, the contributions from noisy interconnect increases, resulting in a growing need for accurate interconnect models.

Novel fabrication techniques, quantum mechanical effects, and leakage currents greatly increase the difficulty of modeling modern MOSFETs. This implies that IC designers cannot rely on absolute values generated by simulators. Instead, they must have a working knowledge of these complexities such that circuit techniques can be used to minimize their effects.

3.1.3.2 Process Variations

Process variations result in electrical differences between devices designed to be identical. They stem from limited precision in fabrication equipment. There are two main types of variation: systematic and random [91]. Systemic variations occur between devices not close in on-chip proximity and are a result of on-chip gradients. They can be minimized by laying out transistors in a symmetrical pattern with multiple fingers or by using common-centroid techniques [91]. Other sources of systematic variations include the shallow trench isolation (STI) stress effect and the well-proximity effect [92]. These effects can be minimized by using dummy transistors to ensure that the devices desired to be matched are an acceptable distance away from trenches and wells [92].

Random variations, often called mismatch, occur between devices close in proximity. They are caused by statistical fluctuations in processing conditions or material properties [91]. Sources of random variation include random dopant fluctuations, oxide fluctuations, and edge roughness [93]–[94]. Ideally, the easiest way to

minimize random variations is to increase device area [37]. However, recent research has shown increases in device area may not always provide the expected improvements in device matching [95]–[97]. Therefore, measurements are needed to determine if matching improves with area. For example, Figure 3.6 shows the general behavior of the MOSFET threshold voltage mismatch slope vs. *L* in an ultra-thin oxide CMOS process. Ideally, the mismatch slope would remain constant with changes in *L*. However, the mismatch slope actually increases with increasing channel length and is largest for the devices with largest area, contradicting the expected results [95]. The impact of random variations on analog design has been studied extensively in literature. These variations result in amplifier input offset voltage and current mismatch. Along with increasing device area, circuit techniques like chopper stabilization, auto-zeroing, and correlated double sampling can be employed to minimize the effects of random variations on analog circuits [98].

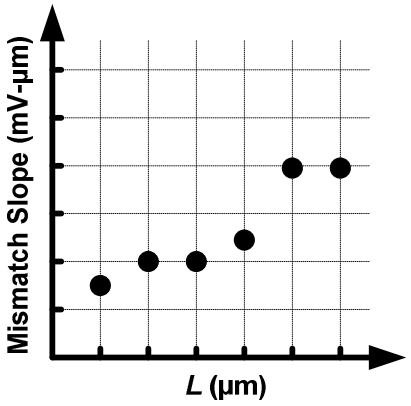

Figure 3.6: General behavior of the MOSFET threshold voltage mismatch slope vs. *L* in an ultra-thin oxide CMOS process [95]. *W* is held constant.

The impact of random variations on digital circuits has become extremely important in the nanoscale regime [99]–[102]. This is due to the atomistic dimensions of the transistors. For example, the number of expected channel dopants in a 65 nm device

is on the order of 100 [103]. However, due to process control limitations, this number can vary significantly. This greatly impacts the threshold voltage of minimum-length transistors. For example, the standard deviation of the difference in threshold voltages between two devices designed to be identical, $\sigma_{\Delta V_{TH}}$, can be as high as 45 mV in a 65 nm process [94]–[95], [104]–[105]. This makes it difficult to reliably predict circuit performance [106]–[108]. Increasing transistor area is a potential solution to this problem, but doing so negates the density advantage obtained by moving to a smaller process.

Process variations and complex models pose significant challenges to circuit designers in nanoscale CMOS technologies [109]. In [110], the term "designing for manufacturability" was used to describe the techniques that must be employed to overcome these challenges. The authors noted that many of the effects described in this section will continue to worsen with scaling. This implies designers can no longer rely on scaled processes to provide all-around superior performance. They must learn to cope with these problems by designing with established architectures, utilizing proper layout techniques, and seeking circuit solutions that provide balance while cancelling undesired effects. Also, Monte Carlo analyses must become an integral part of the design process [110]. A Monte Carlo analysis statistically evaluates performance in the presence of process, voltage, and temperature variations. Previously, in larger technologies, it has been used as a sort of "final check" before a chip is taped out. However, in nanoscale technologies, it can be used as a tool to understand the complex interactions between various devices within a circuit.

3.1.4 Circuit Solutions

Several potential solutions have been suggested to cope with the problems created by output resistance degradation and reduced supply voltages. Many of these are process solutions that require extra fabrication steps beyond what is needed for standard digital devices. Examples include single-side halo transistors, thick-oxide transistors, high-voltage transistors, and floating-gate transistors [1], [7], [11], [111]. Due to these extra fabrication steps, these devices are considered "process options", and are typically available in a standard process at an increased cost. This makes them less attractive from a monetary standpoint and motivates the need to seek circuit solutions using standard digital devices. The circuit solution approach was taken in this work, and for this reason, devices that represent process solutions were not considered. Existing circuit solutions that use standard digital devices include body-biased and bulk-driven transistors, sub-V_{TH} operation, and self-cascoding.

3.1.4.1 Body-Biased and Bulk-Driven Transistors

Body-biased transistors have been proposed as a solution to overcome the problems created from reduced supply voltages [88]. These transistors use the body terminal as a DC input. They manipulate the V_{TH} of a device by exploiting its dependence on V_{BS} [16], [67]. For example, consider Figure 3.7, which plots V_{TH} vs. V_{BS} for two transistors with different channel lengths in the obtained 65 nm process. The figure shows that V_{TH} can change up to 100 mV with changes in V_{BS}. In digital applications, this is done to improve frequency response or reduce power consumption. In analog applications, a reduction in V_{TH} could decrease the V_{DS} needed to achieve saturation, which could increase signal swing or allow more transistors to be stacked.

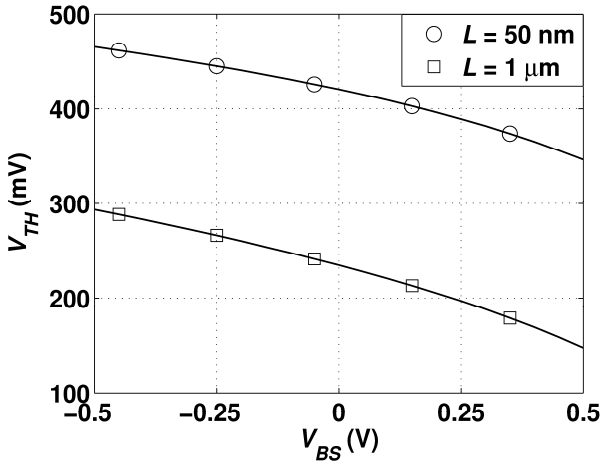

Figure 3.7: Simulated V_{TH} vs. V_{BS} for two NMOS transistors in the obtained 65 nm process. One transistor had W = 10 μm and L = 1 μm. The other transistor had W = 1 μm and L = 50 nm. Both transistors had $V_{GS} = V_{DS}$ = 0.3 V.

Bulk-driven transistors have also been proposed as a solution to overcome the problems created by reduced supply voltages [11], [112]–[113]. The main difference between these transistors and body-biased transistors is that bulk-driven transistors use the body terminal as both an AC and DC input. For example, Figure 3.8 shows a simple example of a differential bulk-driven amplifier. The body terminals of transistors M1 and M2 are used as inputs to the amplifier. Bulk-driven transistors rely on the body transconductance, g_{mb}, to obtain small-signal performance [112]. They have been touted as the solution to analog design in low-voltage CMOS processes [114]. However, because they typically operate with $V_{GS} = V_{DD}$, the V_{DS} required for saturation can become quite large [115]. This makes it difficult to achieve saturation, which limits their application. Also, they potentially suffer from decreased gain, increased area, reduced frequency response, reduced matching, and increased noise [11]. Given these potential problems, bulk-driven transistors were not considered in this work.

Body-biased and bulk-driven transistors require extra process steps to isolate the wells that make up their body terminals. Processes that perform this isolation are referred

to as twin-well or triple-well and are becoming standard with scaling [40]. These transistors are still considered standard digital devices even though extra steps are required to form their wells because the rest of their physical dimensions are equal to the physical dimensions of standard digital transistors.

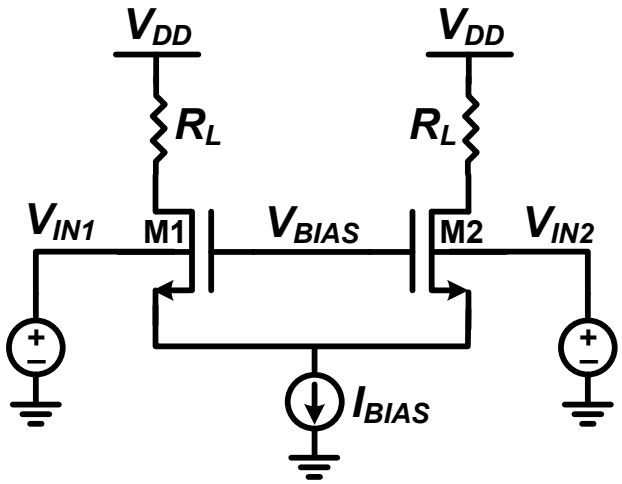

Figure 3.8: Simple example of a differential bulk-driven amplifier. M1 and M2 represent the bulk-driven input differential pair, V_{IN1} and V_{IN2} are the bulk input voltages, I_{BIAS} is the bias current, R_L is the load resistor, and V_{DD} is the supply voltage.

3.1.4.2 Sub-V_{TH} Operation

Another potential solution to the shrinking supply voltage is sub-V_{TH} design [11], [36]. This technique, which requires $V_{GS} < V_{TH}$, goes against traditional saturation region design. If $V_{GS} << V_{TH}$, the device is said to operate in the weak inversion region, where it is dominated by diffusion current and functions similar to a BJT. It can be shown that I_{DS} in the sub-threshold region is formulated as [44]:

$$I_{DS} = \frac{W}{L} I_t e^{(V_{GS}-V_{TH})/(nV_t)}\left(1 - e^{-V_{DS}/V_t}\right) \tag{3.3}$$

where I_t is a current related to the diffusion constant (D_n), V_t, and the equilibrium concentration of electrons in the substrate (n_{po}). If $V_{DS} > 3V_t$, I_{DS} is approximately independent of V_{DS}. This implies that r_O is infinite. This is an important result because it

solves many of the voltage headroom and r_O problems in nanoscale design. However, as mentioned previously, because $V_{GS} << V_{TH}$, I_{DS} is extremely small. This restricts weak inversion operation to nano-power circuitry. For example, consider Figure 3.9, which plots I_{DS} vs. V_{GS} for an NMOS transistor with W = 10 µm, L =1 µm, and V_{TH} = 384 mV. At V_{GS} voltages slightly less than V_{TH}, I_{DS} drops below 10 µA and at V_{GS} voltages less than 0.3 V, I_{DS} drops below 1 µA. These current levels may be undesirable because they may have to be generated via large on-chip resistors [116]. Therefore, larger current values may be desired to reduce resistor area. Also, large current values may be desired to increase frequency response. This equates to increases in V_{GS}, which forces the transistor to exit weak inversion and enter moderate inversion. In this region, I_{DS} is made up of drift and diffusion components, which, as mentioned previously, complicates modeling [43]. In this region there is a degradation of the large r_O and small V_{DS} values obtained in weak inversion. However, compared to strong inversion, the moderate inversion region still provides adequate output resistance at smaller V_{DS} values. This potentially allows for larger signal swings and the ability to stack transistors.

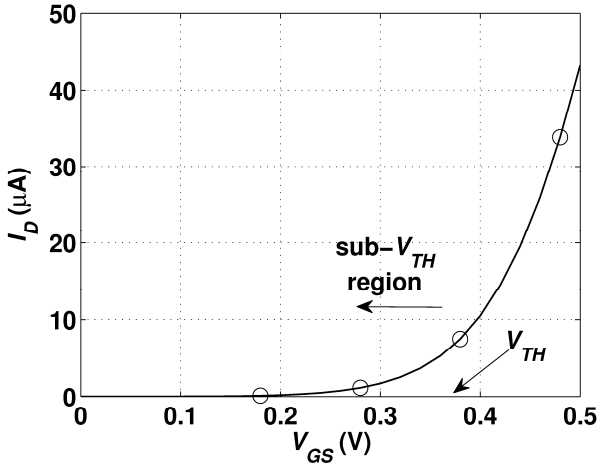

Figure 3.9: I_D vs. V_{GS} for an NMOS transistor in the obtained 65 nm process. W = 10 µm, L = 1 µm, and V_{DS} = 0.3 V.

3.1.4.3 Self-Cascoding

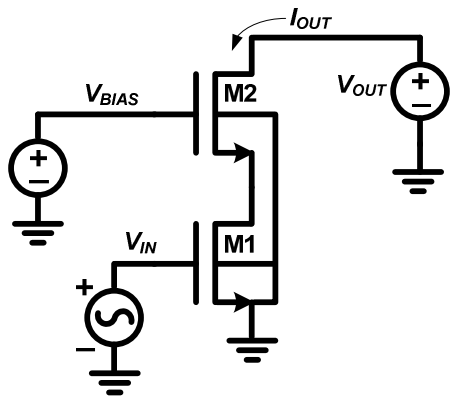

Figure 3.10: Basic cascode structure. V_{IN} is the input voltage, V_{OUT} is the output voltage, V_{BIAS} is the bias voltage for M2, and I_{OUT} is the output current. M1 and M2 form the basic cascode structure.

Cascoding is a technique used to address degradations in r_O [36], [40], [44], [48]. An example of cascoding is shown in Figure 3.10, where M1 is the device being cascoded and M2 is the cascoding device. The input voltage, V_{IN}, drives the gate of M1. V_{BIAS} is used to set the DC bias point on the gate of M2. V_{OUT} is the output voltage and I_{OUT} is the output current. This structure is heavily covered in textbooks and analyzed as a common source amplifier (M1) in series with a common gate amplifier (M2). An equation for the small-signal DC output resistance of this structure can be written as [36]:

$$R_{out_B} = r_{O1} + r_{O2} + (g_{m2} + g_{mb2})r_{O1}r_{O2}. \tag{3.4}$$

Assuming g_{mb2} and $r_{O1} + r_{O2}$ are negligible, this equation can be approximated as $r_{O1}(g_{m2}r_{O2})$. Compared to a single device, which has an output resistance of r_O, this is a significant improvement, and is one of main reasons cascode structures are used in analog design. The small-signal DC voltage gain of this structure is [36]:

$$A_{V_B} = -[g_{m1}r_{O1} + (g_{m2} + g_{mb2})g_{m1}r_{O1}r_{O2}]. \tag{3.5}$$

This equation shows that a cascode structure is capable of producing an approximate voltage gain of $-g_{m1}r_{O1}\cdot g_{m2}r_{O2}$ (ignoring the first term of (3.5) and assuming

g_{mb2} is negligible). This is significantly greater than the intrinsic voltage gain of a single device ($g_m r_O$).

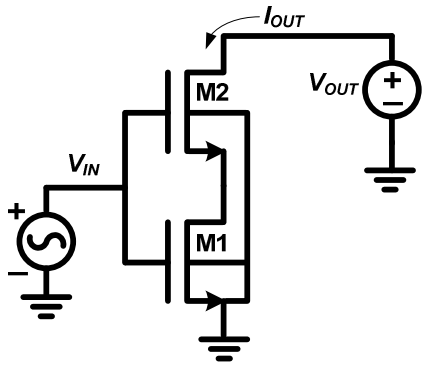

Figure 3.11: Self-cascode structure [117]. V_{IN} is the input voltage, V_{OUT} is the output voltage, and I_{OUT} is the output current. M1 and M2 form the self-cascode structure.

One disadvantage of the basic cascode is that a large V_{OUT} may be required to obtain the output resistance enhancement. If the saturation region of operation is assumed and the body effect ignored, then V_{OUT} must be greater than $2 \cdot V_{DSsat}$ ($V_{DSsat} = V_{GS} - V_{TH}$). In technologies with a V_{DD} of 1 V, this minimum voltage requirement can eat into available headroom and limit signal swing. Another disadvantage of this structure is the potential need for extra circuitry to generate V_{BIAS}. This extra circuitry increases power compared to a single device.

Self-cascoding is a potential solution to these problems [11], [113], [117]–[118]. An example of self-cascoding is shown in Figure 3.11. The main difference between this structure and the basic cascode is that the input voltage, V_{IN}, drives the gates of M1 and M2. Intuitively, this may seem incorrect. For example, ignoring the body effect, and assuming that both devices operate in the saturation region, have equal dimensions, infinite r_O values, and drain currents equal to I_{OUT}, V_{GS1} would have to equal V_{GS2} to supply I_{OUT}. This can only occur if $V_{DS1} = 0$, which implies M1 is turned off and no

current flows through the structure. Therefore, it appears self-cascoding does not work. However, if the aspect ratio of M2 is made larger than the aspect ratio of M1 ($W_2/L_2 > W_1/L_1$), the required V_{GS2} to supply I_{OUT} is less than the required V_{GS1}. Considering that $V_{GS1} = V_{G2}$, this can only occur if V_{S2} increases. Therefore, increasing the aspect ratio of M2 relative to M1 results in V_{DS1} increasing because $V_{S2} = V_{DS1}$. Under these conditions, M1 turns on, allowing the self-cascode structure to function. This analysis shows that the ratio of device aspect ratios is an important parameter. A scale factor, S_F, can be defined to characterize this relationship [118]:

$$S_F = \frac{W_2/L_2}{W_1/L_1}. \tag{3.6}$$

As S_F increases, V_{GS2} decreases, and the V_{DS2} value needed to saturate M2 also decreases (assuming $V_{DSsat2} = V_{GS2} - V_{TH2}$). This implies that the V_{OUT} needed to place the self-cascode structure into saturation is smaller than the basic cascode. This results in a savings of voltage headroom, which allows for larger signal swings. Also, because the gates of M1 and M2 are tied together, no extra bias circuitry is needed for M2. This results in a savings of power and area compared to the basic cascode. Therefore, from a DC biasing standpoint, the self-cascode has several advantages over a basic cascode.

The low-frequency small-signal performance of a self-cascode is equal to or better than that of a basic cascode. For example, (3.7) shows that the low-frequency small-signal output resistance of a self-cascode is equal to that of the basic cascode [117] (see Appendix A). Equation (3.8) shows that the low-frequency voltage gain of the

self-cascode is greater than that of the basic cascode by an additional term of $g_{m2}r_{o2}$ [117] (see Appendix A).

$$R_{out_{SC}} = R_{out_B} = r_{O1} + r_{O2} + (g_{m2} + g_{mb2})r_{O1}r_{O2}. \tag{3.7}$$

$$A_{V_{SC}} = -[g_{m1}r_{O1} + g_{m2}r_{O2} + (g_{m2} + g_{mb2})g_{m1}r_{O1}r_{O2}]. \tag{3.8}$$

Depending on the values of S_F and I_{OUT}, M1 or M2 could be forced into the sub-V_{TH} region [117]. For example, S_F could be large enough to force M2 in the sub-V_{TH} region or I_{OUT} could be small enough to force both devices into the sub-V_{TH} region. As stated previously, these regions of operation can potentially provide large r_O at small V_{DS} values. Thus, the small-signal characteristics of each device may be improved by operating in these regions. Note that care should be taken to ensure S_F is not large enough to turn off M1 or M2. One potential disadvantage of the self-cascode structure is the increased Miller capacitance from its input to the drain of M2. This increased capacitance occurs because of the structure's increased gain and it could create a second undesired dominant low-frequency pole when used in an amplifier configuration.

Several other cascoding techniques have been proposed to help improve output resistance [113]. Examples include active cascoding, folded cascoding, gain-boosting, and wide-swing cascode structures. These techniques use more devices than the basic or self-cascode structures. For example, wide-swing cascode current mirrors require two reference currents and gain-boosted current mirrors require the use of an amplifier [36]. As a result, they consume more area and power. Therefore, these techniques were not considered in this work.

3.2 Gate Current

The gate resistance of a MOSFET is often assumed to be infinite [36], [40], [45], [48]. This simplifying assumption allows the device to be analyzed as if no DC current is flowing through its gate terminal, which greatly simplifies circuit analysis and design. As CMOS has scaled to technologies with oxide thicknesses less than 3 nm, this assumption no longer holds, as significant amounts of carriers directly tunnel through the gate insulation. These carriers contribute to a source of gate current, referred to as direct tunneling, that fundamentally changes MOSFET operation [119]–[120]. This section investigates these changes by reviewing the physical mechanisms behind direct tunneling and its impact on MOSFET modeling. It also compares direct tunneling to base current of a BJT and notes its impact on current mirror design, frequency response, matching, noise, MOSFET capacitance, and temperature-sensitive circuits. It concludes by discussing the use of high-κ dielectrics and metal gate electrodes as a solution to minimizing the impact of direct tunneling on circuit performance. The terms gate current and direct tunneling are used interchangeably, even though direct tunneling is not the only source of gate current in CMOS technologies. Specifically, Fowler-Nordheim (FN) tunneling and hot electrons can contribute to gate current [121]–[123]. However, in CMOS technologies with $t_{ox} < 3$ nm and $V_{DD} \leq 1$ V, these sources are often considered negligible under normal operating conditions [16]. Therefore, direct tunneling was assumed to be the dominant source of gate current.

3.2.1 Tunneling Background

Tunneling is a quantum mechanical phenomenon in which carriers can penetrate into and through a potential barrier. It typically occurs between two conducting materials

separated by an insulator. The insulator creates a potential barrier between the conducting materials. If certain electrical and physical requirements are met, carriers can flow between the conducting materials by tunneling through this barrier. Classically, this is impossible because these barriers represent a point at which the total system energy is completely potential. Ideally, when classical carriers encounter these barriers, they are reflected. If they were to overcome them, their potential energy would have to become more than that of the barrier itself. For this to happen and to ensure conservation of energy, the kinetic energy of the carriers would have to be negative. Negative kinetic energy violates the laws of classical physics and is one of the fundamental reasons why quantum mechanics is used to explain tunneling [124].

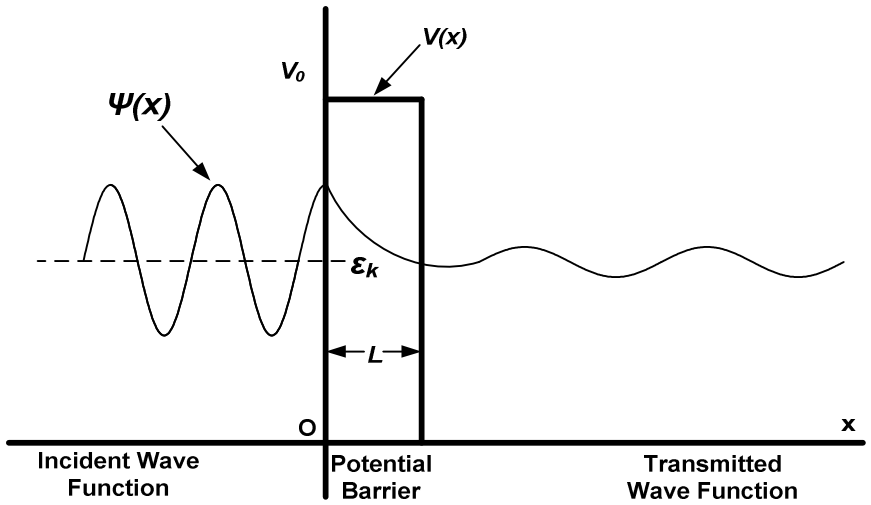

Figure 3.12: Tunneling in a rectangular potential barrier [124]. $V(x)$ is the potential energy of the system and ε_k is the incident particle kinetic energy. V_0 is the barrier height and L is the barrier width. The carrier is described by its wave function, $\Psi(x)$.

Quantum mechanically, carriers are described by their wave functions, which are continuous and used to determine the probability of finding a particle at a specific time and position [124]. When a carrier encounters a potential barrier, its wave function remains continuous, but has an exponential decay inside the barrier. On the other side of

the barrier, the wave function is still continuous, which results in a finite probability that the carrier will tunnel through it. For example, consider Figure 3.12, which shows quantum mechanical tunneling through a rectangular potential barrier [124]. $V(x)$ is the potential energy of the system and ε_k is the incident particle kinetic energy. The barrier is described by its height, V_0, and its width, L. The carrier is described by its wave function, $\Psi(x)$. The figure shows that $\Psi(x)$ exponentially decays upon entering the barrier, but remains continuous and eventually makes it to the other side. The probability of this occurring for the condition of $\varepsilon_k < V_0$ is [124]:

$$P_T = \frac{1}{1 + \frac{V_0^2 sinh^2 \beta L}{4\varepsilon_k (V_0 - \varepsilon_k)}} \tag{3.9}$$

where $\beta^2 = 2m(V_0 - \varepsilon_k)/\hbar^2$, $\hbar = 1.055 \text{ x } 10^{-34}$ J-sec is Planck's constant, and m is the carrier's mass. This equation shows that P_T increases with decreases in barrier height and width. In physical systems, the barrier width is related to the thickness of an insulating material. The barrier height is related to the physical and electrical properties of the insulating and conducting materials. Interestingly, P_T is not guaranteed to be one when $\varepsilon_k > V_0$, which implies some carriers will be reflected even though they possess more energy than the barrier [124].

Potential barriers are not always rectangular. For example, they can be triangular or trapezoidal [16]. The Wentzel-Kramers-Brillouin (WKB) approximation is typically employed when deriving P_T for these types of barriers [125]. Generally, once P_T is known, the tunneling current density, J_T, between the two conducting materials can be calculated using the following equation [49]:

$$J_T = \frac{qm^*}{2\pi^2\hbar^3}\int F_1 N_1 P_T (1 - F_2) N_2 dE \tag{3.10}$$

where F_1, F_2, N_1, and N_2 are the Fermi-Dirac distributions and density of states functions of the two conducting materials (material 1 and material 2), m^* is the effective mass, and q is the electronic charge ($1.602\text{x}10^{-19}$ C). This equation shows that J_T is dependent upon the product of the number of available carriers originating from material 1 and the number of empty states in material 2 [49].

3.2.2 Fowler-Nordheim Tunneling and Direct Tunneling

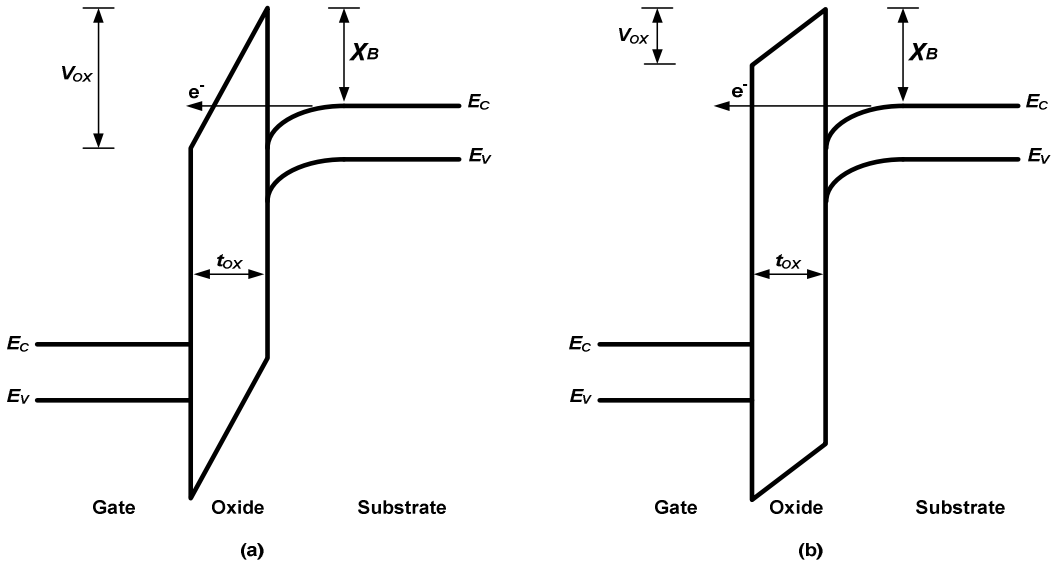

Figure 3.13: Ideal energy band diagrams for: (a) Fowler Nordheim tunneling and (b) direct tunneling in an NMOS transistor. E_C and E_V are the conduction and valence bands, t_{ox} is the oxide thickness, X_B is the barrier height, V_{OX} is the voltage across the oxide, and e^- is the tunneling electron [16].

In MOSFETs, tunneling is typically analyzed as occurring between two pieces of silicon separated by a thin layer of silicon dioxide (S_iO_2). One of the pieces of silicon represents the heavily doped gate electrode and the other piece represents either the silicon channel or the heavily doped source/drain junction. Carriers can tunnel through the S_iO_2 via two different mechanisms: Fowler-Nordheim (FN) tunneling and direct tunneling. The difference between these two types of tunneling is the shape of the

potential barrier the carriers must tunnel through. In FN tunneling the potential barrier is triangular, while in direct tunneling it is trapezoidal [16], [126]–[127]. An example of these two types of tunneling is shown in Figure 3.13. E_C and E_V are the conduction and valence bands, t_{ox} is the oxide thickness, X_B is the barrier height, V_{OX} is the voltage across the oxide, and e^- is the tunneling electron. The shape of the potential barrier depends on V_{OX}. If $V_{OX} > X_B$, as shown in Figure 3.13 (a), a triangular potential barrier is formed, and FN tunneling is possible. Note that FN tunneling is sometimes used to build flash electrically erasable read-only memory (Flash EEPROM) [128]. However, it is considered negligible in nanoscale CMOS because the supply voltages are typically much less than the barrier heights.

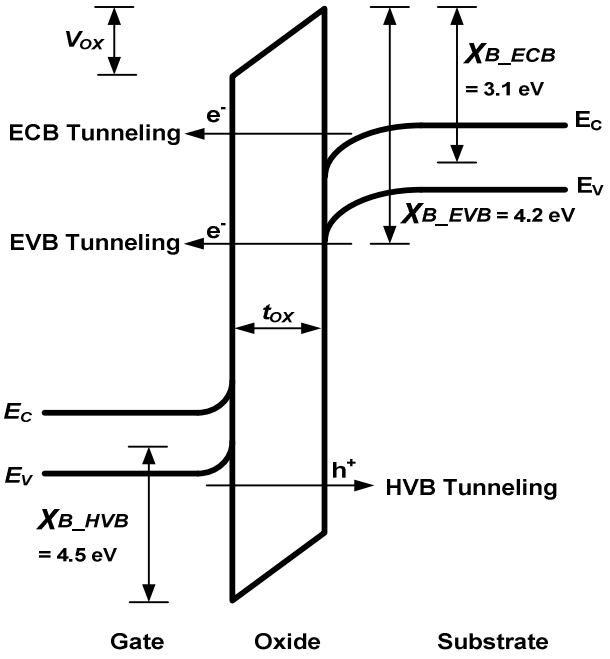

Figure 3.14: Direct tunneling in an NMOS transistor. E_C and E_V are the conduction and valence bands, V_{OX} is the voltage across the oxide, t_{ox} is the oxide thickness, e^- and h^+ represent tunneling electrons and holes. X_{B_ECB}, X_{B_EVB}, and X_{B_HVB} represent the barrier heights for ECB, EVB, and HVB [14], [86].

If $V_{OX} < X_B$, as shown in Figure 3.13 (b), a trapezoidal potential barrier is formed, and direct tunneling is possible. Direct tunneling is exponentially dependent upon t_{ox} and

becomes a non-negligible source of gate current in technologies with $t_{ox} < 3$ nm [49], [119]–[120], [129]–[130]. It represents a fundamental limitation to the scaling of CMOS technologies [8], [12], [64], [66], [85], [107]–[108], [131]. There are three major types of direct tunneling in MOSFETs. They are shown in Figure 3.14 [14], [86]. The first is electrons tunneling from the conduction band (ECB). The second is electrons tunneling from the valence band (EVB). The third is holes tunneling from the valence band (HVB). ECB and EVB are typically associated with NMOS devices and HVB is associated with p-type MOSFETs (PMOS). The associated barrier heights for these three types of direct tunneling are $X_{B\_ECB} = 3.1$ eV, $X_{B\_EVB} = 4.2$ eV, and $X_{B\_HVB} = 4.5$ eV [132]. Ignoring differences in threshold voltages and carrier mobilities, the direct tunneling current for a PMOS device will typically be less than that of an NMOS device because the barrier height for HVB is greater than the barrier heights for ECB and EVB. This suggests that circuits should be designed with PMOS devices, or, more specifically, the device with the larger barrier height, to minimize direct tunneling currents.

3.2.3 Modeling of Direct Tunneling

Several attempts have been made at modeling direct tunneling in CMOS technologies [13], [14], [31], [132]–[136]. All of these models emphasize the exponential dependence of J_T on t_{ox} and attempt to model direct tunneling over a broad range of terminal voltages and device sizes. In [13] and [136], direct tunneling was partitioned into five components: I_{GCS}, I_{GCD}, I_{GS}, I_{GD}, and I_{GB}. These components are shown in Figure 3.15. They flow simultaneously and their summation yields an equation for the total amount of gate current due to direct tunneling, I_G:

$$I_G = I_{GCS} + I_{GCD} + I_{GS} + I_{GD} + I_{GB}. \tag{3.11}$$

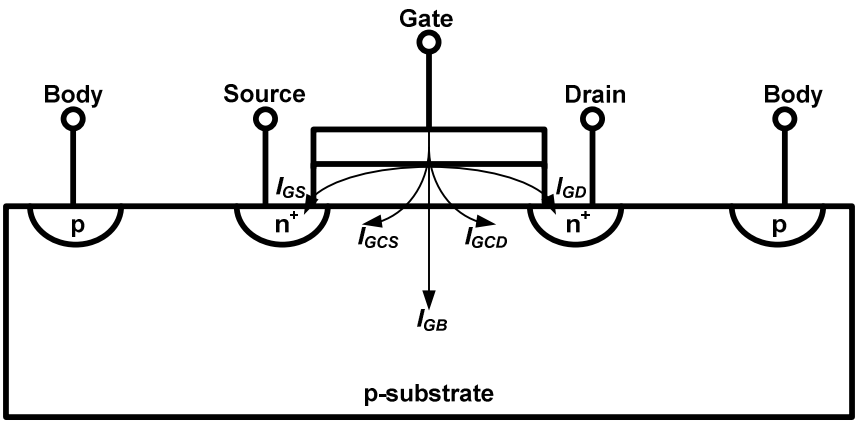

Figure 3.15: Components of direct tunneling in an NMOS transistor [136]. I_{GCS} and I_{GCD} flow into the channel, I_{GS} flows into the source overlap region, I_{GD} flows into the drain overlap region, and I_{GB} flows into the substrate.

The I_{GCS} and I_{GCD} components are typically ECB, flow into the silicon channel and go to the source (I_{GCS}) and drain (I_{GCD}). The I_{GS} and I_{GD} components are ECB and flow into the source (I_{GS}) and drain (I_{GD}) overlap regions. I_{GB} can be ECB or EVB and it flows to the body terminal. It is important to understand how each component functions under different terminal voltages. Figure 3.16 can be used to aid in this understanding [137]. The figure shows I_{GCS} flowing into the source terminal and I_{GCD} flowing into the drain terminal. For an NMOS device, these components are strong functions of V_{GS} and weak functions of V_{DS} [136]–[137]. Therefore, because V_{GS} is typically positive, these currents can be assumed to be flowing in the direction shown in Figure 3.16.

Figure 3.16 shows I_{GS} flowing into the source terminal and I_{GD} flowing into the drain terminal. I_{GS} is a strong function of V_{GS} and I_{GD} is a strong function of V_{GD} [136]–[137]. Similar to I_{GCS} and I_{GCD}, I_{GS} can be assumed to be flowing in the direction shown in Figure 3.16. However, this assumption cannot be made when analyzing I_{GD} because V_{GD} may be positive or negative. If V_{GD} is a large positive value, I_{GD} flows in the direction shown in Figure 3.16. If V_{GD} is a large negative value, I_{GD} flows opposite to

what is shown in Figure 3.16. I_{GB} is a strong function of V_{GB}, which is typically positive, and can be assumed to flow in the direction shown in Figure 3.16. I_{GB} is often considered be negligible in nanoscale technologies [16].

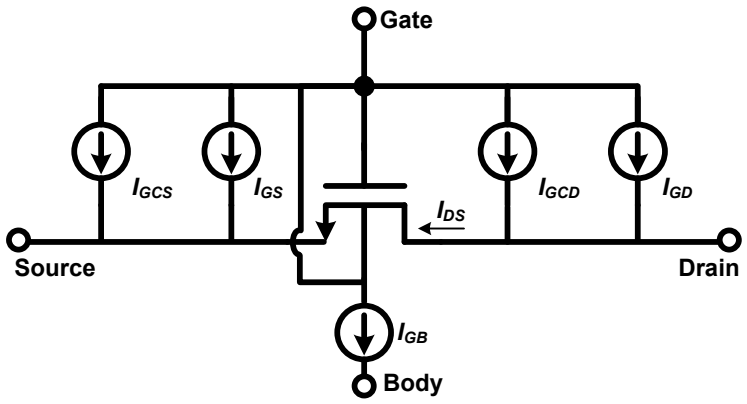

Figure 3.16: DC components of direct tunneling in an NMOS transistor [137]. I_{GCS} and I_{GCD} flow directly into the source. I_{GCD} flows out of the source via the drain. I_{GD} can flow out of the source via the drain or out of the gate. I_{GB} flows out of the body.

The previous paragraph noted that I_{GD} can be bidirectional under normal operating conditions. This significantly impacts I_G. For example, if $V_{GS} = 0$ and $V_{DS} = V_{DD}$, then $V_{GD} = -V_{DD}$. Because $V_{GS} = 0$, I_{GCS}, I_{GCD}, and I_{GS} can be considered negligible. Assuming I_{GB} is also negligible, I_{GD} becomes the dominant component of I_G. However, because $V_{GD} = -V_{DD}$, I_{GD} is a large negative value, which results in I_G becoming a large negative value. This implies I_G is flowing out of the gate terminal of an NMOS device, instead of into it. On the other hand, if $V_{GS} = V_{DD}$ and $V_{DS} = V_{DD}$, then $V_{GD} = 0$ V. In this example, I_{GCS}, I_{GS}, and I_{GD} dominate and result in I_G flowing into the gate. These relationships show that the directionality and magnitude of I_G is heavily dependent upon bias voltages.

Device sizing also plays a critical role in determining I_G. In [13]–[14], and [136], it was shown that I_{GCS}, I_{GCD}, and I_{GB} are proportional to $W{\cdot}L$ and I_{GS} and I_{GD} are

proportional to $W \cdot \Delta L_{OV}$, where ΔL_{OV} is the overlap length between the source/drain and oxide. In long-channel devices, I_{GCS} and I_{GCD} dominate because $L >> \Delta L_{OV}$. However, in short-channel devices, I_{GS} and I_{GD} are comparable in magnitude to I_{GCS} and I_{GCD} because the difference between L and ΔL_{OV} is reduced. Therefore, L and ΔL_{OV} play an important role in determining which components factor into I_G.

3.2.4 Impact of Direct Tunneling on Current Mirror Design

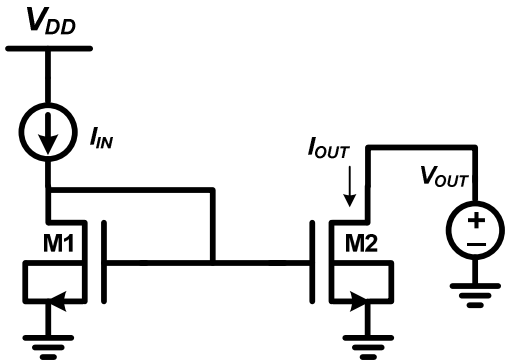

Figure 3.17: Simple current mirror. V_{DD} is the supply voltage, I_{IN} is the input current, I_{OUT} is the output current, and V_{OUT} is the output voltage. M1 and M2 form the current mirror.

The impact of gate current on analog design was studied in [18], [59]–[60], [138]–[141]. Each of these references notes that gate current presents significant challenges. In [141], the authors explained how gate current impacts simple current mirrors. A simple current mirror is shown in Figure 3.17. I_{IN} is the input current, I_{OUT} is the output current, V_{OUT} is the output voltage, and M1-M2 are the MOSFETs used to form the mirror. The current gain of this mirror, including gate current, can be written as [141]:

$$\frac{I_{OUT}}{I_{IN}} = \frac{I_{D2}}{I_{D1} + I_{G1} + I_{G2}} \tag{3.12}$$

where I_{D1}, I_{D2}, I_{G1} and I_{G2} are the drain and gate currents of M1 and M2. This equation shows that gate current degrades the current gain from its ideal value of I_{D2}/I_{D1}.

Specifically, if I_{G1} and I_{G2} are large and positive, the current gain is much less than desired, which results in I_{OUT} being less than I_{IN}. As the current gain A_i decreases, the DC bias point of both transistors may change because I_{IN} supplies more gate current to M1 and M2. This changes V_{GS1} and V_{GS2} to being less than what they normally would have been if $I_{IN} = I_{D1}$. This change in bias point could impact the small-signal performance and frequency response of the mirror.

If finite device output resistance is included in (3.12), the current gain depends on r_{o1} and r_{o2} along with I_{G1} and I_{G2}, which increases its complexity [141]. Considering that degradations in output resistance occur with scaling, current mirror design in ultra-thin oxide technologies must overcome gain degradations caused by gate current and reduced device output resistances. This makes the design of current mirrors, which are fundamental building blocks of analog circuits, more difficult in these technologies.

3.2.5 Comparing Direct Tunneling to Base Current

In [18] and [60] the authors compared MOSFET gate current, I_G, to the base current, I_B, of a BJT. Both can be thought of as input currents; I_G typically flows into the gate of an NMOS and I_B typically flows into the base of an npn BJT. Also, both currents are generally undesirable and degrade device performance. If it can be shown that I_G functions similar to I_B, perhaps established BJT circuit techniques can be used to minimize the negative effects of I_G. This is a major motivating factor for comparing I_G to I_B.

The forward current gain, β_F, of a BJT is defined as I_C/I_B, where I_C is the collector current. It is used to compare the undesired current, I_B, to the desired current, I_C, and has

implications for circuit design [44]. For example, I_B could be considered negligible if $\beta_F > 1000$. However, if $\beta_F = 10$, I_B should be taken into account. Typically, it is desired that β_F be as large as possible. Note that β_F is ideally independent of the DC bias point and is set by process parameters such as the emitter/base doping concentration and the base width [44]. Thus, from a circuit design standpoint, β_F is often treated as a constant value, which greatly simplifies analysis. Note that β_F does roll off at very high and very low currents.

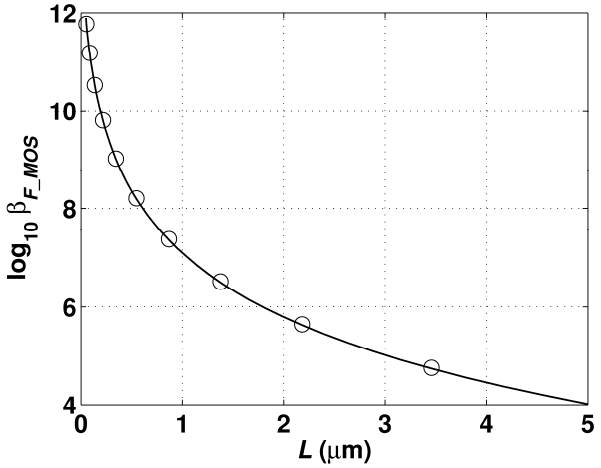

Figure 3.18: Logarithmic plot of β_{F_MOS} vs. L of an NMOS transistor in the obtained 65 nm process. W = 10 µm and $V_{GS} = V_{DS}$ = 1 V.

Applying this analogy to MOSFETs results in $\beta_{F\_MOS} \equiv |I_D/I_G|$, where I_D is the desired current and I_G is the undesired current. This was done in [18], where I_D/I_G was used as a performance metric to determine the impact of I_G on MOSFETs. The authors noted that I_D/I_G is a strong function of the DC bias point. This is important because it shows that I_D/I_G cannot be treated as a constant value. Assuming square law operation, using a simplified model for I_G, and ignoring the dependence of I_G on V_{DS}, the authors showed that I_D/I_G is roughly proportional to $1/L^2$ ($I_D \propto W/L$, $I_G \propto W \cdot L$, $I_D/I_G \propto 1/L^2$). This suggests that long-channel devices operate less like a traditional MOSFET because

they consume larger amounts of gate current relative to drain current, which implies that there is a point of diminishing returns when it comes to using long-channel devices. For example, consider Figure 3.18, which plots the base 10 logarithm of β_{F_MOS} vs. L for an NMOS transistor with W = 10 µm in the obtained 65 nm process. The figure confirms that β_{F_MOS} is dramatically reduced with increases in channel length. This suggests that long-channel devices should be avoided in ultra-thin oxide CMOS technologies. However, L is often increased to improve output resistance ($\lambda \propto 1/L$, see Section 3.1.1.1). If L is increased to a value where the device no longer operates like a MOSFET, the increased output resistance is meaningless. This suggests that there is a direct tradeoff between device r_O and β_{F_MOS}. Also, in [18], I_D/I_G was only shown to be proportional to $1/L^2$ under constant terminal voltages (the saturation region of operation was assumed). The authors suggested increasing W as a means to increase I_D without impacting I_D/I_G. However, if W is increased with constant drain current, I_D/I_G is dependent on W because the terminal voltages and region of operation may change. This implies that increases in W with constant I_D may reduce I_D/I_G. This could cause the negative effects of I_G to become more pronounced.

3.2.6 Impact of Direct Tunneling on Analog Device Performance

The impact of I_G on MOSFET gate impedance was studied in [18] and [139]. The authors derived a frequency, f_{gate}, which can be used as a metric to characterize the gate impedance. For signal frequencies larger than f_{gate}, the gate impedance was said to be capacitive and the device was said to behave like a traditional MOSFET. Below f_{gate}, the gate impedance was said to be mainly resistive and dominated by gate current. The authors noted that f_{gate} for a 65 nm technology was approximately 1 MHz. This is an

important result because it shows that gate current significantly impacts the low-frequency performance of ultra-thin oxide MOSFETs and that its effects on high-frequency performance are negligible.

The authors in [18] also studied the impact of I_G on drain current mismatch. They showed that drain current mismatch is dependent upon I_D/I_G, which limits achievable matching. Typically, to ensure a constant aspect ratio and improve matching, W and L are linearly scaled. However, following this approach in a technology with significant gate current can result in matching becoming worse as area increases. For example, in [18], it was shown that linearly scaling W and L in a 65 nm technology resulted in an optimal matching point at a device area of approximately 10^3 μm^2. Beyond this area, matching actually became worse. However, the authors noted that matching could be improved if L is kept constant and W is scaled. This approach increased power consumption because the aspect ratio increased and terminal voltages were kept constant. The impact of I_G on drain current mismatch was not a major concern in this work because the area at which matching began to degrade was far greater than what was used. However, because gate current is proportional to area, matching improves at its expense. This suggests that matching and gate current trade off with each other. Also, if there is a large difference in drain voltages between transistors designed to be identical, the I_{GD} contributions from each device could be different. This could lead to different amounts of gate current flowing through each device, which implies they are not electrically matched. This suggests that care should be taken to ensure that devices which are designed to be identical have similar terminal voltages and similar areas such that their gate and drain currents are matched.

The impact of gate current on noise performance was studied in [44], [138], [142]. It was shown that direct tunneling results in a shot noise component with a spectral density of $S_{IG} = 2qI_G$. This noise component is similar to the shot noise associated with base current in BJTs. Also, a 1/*f* noise component has been observed with direct tunneling [142]. Combined, these two components have been shown to create a noise corner frequency around 20 kHz. Both of these noise sources have been shown to be less than the traditional thermal and 1/f noise sources associated with MOSFETs. The fact that direct tunneling results in additional noise sources only magnifies the differences between conventional and ultra-thin oxide MOSFETs. Perhaps the best approach to reducing the impact of these noise sources is to treat gate current itself as a noise source and minimize it as much as possible. If this is accomplished, the device operates more like a conventional MOSFET and implies that traditional circuit techniques can be used to design analog circuits in ultra-thin oxide technologies.

Degraded MOSFET capacitor (MOSCAP) performance is another consequence of gate current. These capacitors use the gate and a shorted source/drain as terminals. They are typically designed to take advantage of C_{OX} [43]. In [18], it was shown that gate current can seriously degrade the performance of circuits designed with MOSCAPs. For example, a track and hold circuit designed using MOSCAPs in a 65 nm technology must be read within a few nanoseconds if the drop on a sampled value is to be limited to 1 mV [18]. This places severe restrictions on sampling frequencies and forces the use of other types of capacitors for track-and-hold circuits. MOSCAPs are also used to decouple high frequency power supply noise in digital circuits [143]. However, if large amounts of DC gate current are flowing through them, they may actually introduce low-frequency noise

into the circuit. Therefore, extreme caution must be exercised when using ultra-thin oxide MOSCAPs as decoupling capacitance.

Direct tunneling has been shown to be relatively independent of temperature under constant terminal voltages [144]–[145]. This has implications for temperature-sensitive circuits such as voltage references. Intuitively, one may assume that because gate current is independent of temperature, it would not impact the performance of voltage references. However, if device terminal voltages change with temperature, gate current could also change with temperature. This change impacts reference performance and was investigated in this work.

3.2.7 Existing Circuit Solutions to Gate Current

It is important to note that very few circuit techniques exist in the literature to minimize the negative effects of direct tunneling on analog performance. In [146] and [147], the authors attempted to use gate leakage as a means to reduce amplifier offset. The downside of these techniques was the requirement of thick-oxide transistors. This approach was not considered in this work because thick-oxide transistors represent a process solution.

Several techniques exist to minimize the impact of gate current on digital performance [16], [34], [148]–[149]. In [16], the authors suggested the use of supply voltage scaling as a means of reducing gate leakage. In [148], it was shown that pin reordering and NOR-based logic can be used to help minimize gate leakage. In [149], it was stated that digital circuits designed in the presence of gate leakage will be able to meet noise margin as long as $t_{ox} \geq 1.1$ nm. In [34], the use of PMOS-based logic was

promoted over the use of NMOS-based logic because PMOS devices have larger barrier heights and thus contribute less gate current. Given that these solutions exist for digital circuits, analog techniques are necessary if mixed-signal design is to be performed using ultra-thin oxide MOSFETs.

3.2.8 Direct Tunneling and High-κ/Metal Gates

Direct tunneling has become so problematic that changes to the gate dielectric and gate electrode must be made [12], [22], [31], [131], [150]–[153]. These changes represent a fundamental shift in CMOS technology because SiO_2 and polysilicon have been used as the gate stack for many generations of CMOS technology. SiO_2 is targeted to be replaced by a high-κ dielectric and polysilicon is targeted to be replaced by a metal. This new gate stack is often referred to as the high-κ/metal gate.

High-κ dielectrics are used to replace SiO_2 because of their increased dielectric constant. Compared to SiO_2, they can be made thicker to achieve the same amount of capacitance per unit area. This increased thickness results in reduced direct tunneling probability. For example, if the high-κ capacitance, $C_{hi\text{-}\kappa}$, is to be equal to the SiO_2 capacitance, C_{SiO_2}, the thickness of the high-κ material, $t_{hi\text{-}\kappa}$, would need to be [153]:

$$t_{hi-\kappa} = \frac{\kappa_{hi-\kappa}}{\kappa_{ox}} t_{ox} \tag{3.13}$$

where $\kappa_{\text{hi-}\kappa}$ and κ_{ox} are the dielectric constants of the high-k material and SiO_2. This equation shows that for a desired $t_{ox,}$ $t_{hi\text{-}\kappa}$ is dependent upon the ratio $\kappa_{\text{hi-}\kappa}$ and κ_{ox}. Therefore, to limit direct tunneling, it is desired to have $\kappa_{hi\text{-}\kappa}$ be as large as possible. Several high-κ materials have proposed as a possible replacement of SiO_2. These include silicon nitride (Si_3N_4), oxynitride (SiO_xN_y), zirconium oxide (Z_rO_2), hafnium oxide

(HfO_2), aluminum oxide (Al_2O_3), and lanthanum oxide (La_2O_3) [154]. Also, it has been shown that many materials with a dielectric constant > 20 cannot be used because they have extremely small barrier heights [152]. This prevents their use because direct tunneling, as previously explained, is a strong function of the barrier height. Materials with a dielectric constant between 8 and 20 have been shown to provide the thicknesses and barrier heights needed to significantly reduce direct tunneling [152], [154].

V_{TH} pinning, mobility degradation, and phonon scattering are a major concern when selecting a high-κ dielectric [12], [22], [28]. Because of these problems, an interfacial layer of SiO_2 has been proposed to be sandwiched between the high-κ dielectric and the silicon channel. This layer takes advantages of the good bonding properties between Si and SiO_2, resulting in less trapped charge and interface states. The ability to control the thickness of the SiO_2 layer is extremely difficult, which results in variability concerns [22]. In [28], a high-κ/metal gate process was presented that did not use an interfacial layer of SiO_2.

The use of high-κ materials will not totally eliminate direct tunneling [30]–[31]. As technologies scale and high-κ materials become thinner, the problems created by direct tunneling will return. This suggests that analog circuit techniques to minimize the negative effects of direct tunneling need to be developed.

Metal gate electrodes are used to minimize the effects of poly-gate depletion [131]. The metal used must be compatible with the high-κ material such that their interface has minimal defects [155]. This greatly increases the complexities involved in fabricating the high-κ/metal gate structure. In [155], a 45 nm CMOS process with a

high-κ/metal gate was presented. It was noted that the metal used for the gate of the NMOS device was different than that of the PMOS device. This hints at some of the difficulties involved in fabricating high-κ/metal gate structures.

Ideally, migration to these new technologies would occur quickly because of improved performance and increased density. However, heavy migration may be delayed by rising manufacturing costs and increased design complexities [23]–[26]. This implies that traditional (non-high-κ /non-metal gate) ultra-thin oxide technologies will have longer lives in the economic forefront than previous generations of CMOS. Therefore, given that digital solutions are available and that traditional ultra-thin oxide CMOS technologies will be revenue generators for an extended period of time, analog circuit solutions are needed to allow useful mixed-signal design using only ultra-thin oxide MOSFETs.

3.3 Voltage References

Voltage references are precision analog circuits designed to produce a voltage independent of variations in temperature, process, and supply voltages [156]. They are used in several analog applications, such as DACs, ADCs, DC-DC converters, operational amplifiers, and linear regulators [35]. This widespread use shows their importance to analog design and motivates the study of problems that may impact their performance. This section reviews the fundamentals of voltage references and the problems encountered when designing them in nanoscale CMOS technologies. Also, it notes that no techniques exist to compensate the negative effects of gate current on their performance.

3.3.1 Temperature Independence and Bandgap Voltage References

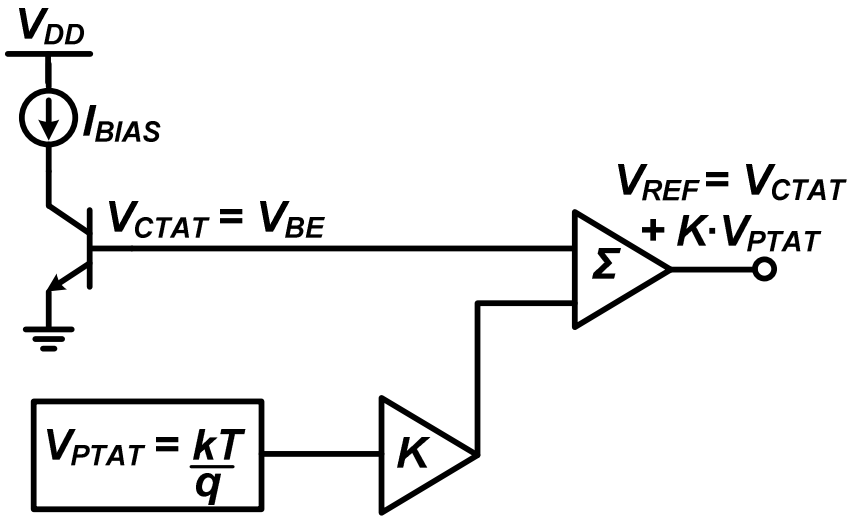

Figure 3.19: High-level circuit schematic of a bandgap voltage reference [44]. V_{DD} is the supply voltage, I_{BIAS} is the bias current, V_{CTAT} is the CTAT voltage, V_{PTAT} is the PTAT voltage, k is Boltzmann's constant, q is the electronic charge, T is the temperature, and K is a scale factor.

Temperature independence is typically the most difficult specification for a voltage reference to achieve. This difficulty stems from the fact that most electrical parameters vary with temperature [157]. To account for this variance, voltage references often attempt to sum two voltages; one that changes proportionally to absolute temperature (PTAT) with one that changes complementary to absolute temperature (CTAT). Ideally, these two voltages would have equal but opposite temperature slopes such that their sum results in a voltage independent of temperature. However, these voltages rarely have equal and opposite slopes, which necessitates the need to scale one of them by a constant. Mathematically, this can be written:

$$V_{REF} = V_{CTAT} + K \cdot V_{PTAT} \tag{3.14}$$

where V_{CTAT} is the CTAT voltage, V_{PTAT} is the PTAT voltage, K is the constant, and V_{REF} is the output voltage. Figure 3.19 shows a high-level circuit schematic of a bandgap

voltage reference [44]. In order to achieve temperature independence, (3.14) is differentiated with respect to temperature, set equal to zero, and solved for K:

$$K = \frac{-\partial V_{CTAT}/\partial T}{\partial V_{PTAT}/\partial T}. \tag{3.15}$$

Given that K is a constant, the temperature slopes of V_{PTAT} and V_{CTAT} must also be constant if V_{REF} is to be independent of temperature. This implies that V_{CTAT} and V_{PTAT} vary linearly with temperature. Physically, it may seem highly improbable that a voltage naturally varies linearly with temperature. However, to a first-order approximation, the voltage across a forward-biased diode varies linearly with temperature and has a slope of approximately -1.8 mV/°C [36], [40], [44]. For this reason, diodes are often used as CTAT voltage sources in voltage references. If temperature-dependent non-idealities are included, the diode voltage can be written as [44], [158]:

$$V_{DIODE} = V_{G0} - V_t[(\gamma - \alpha)lnT - ln(E \cdot G)] \tag{3.16}$$

where V_{DIODE} is the diode voltage, V_{GO} is the bandgap voltage of the material being used, V_t is the thermal voltage, T is the temperature, E and G are temperature-independent constants, γ is related to the current flowing through the diode, and α is related to the carrier mobility. This equation shows the true behavior of V_{DIODE} with temperature and provides physical insights as to why its temperature slope is not constant [44]. It also shows that V_{DIODE} is directly dependent upon the bandgap voltage of the material being used. This dependence results in a special type of voltage reference, referred to as the bandgap voltage reference.

A PTAT voltage can be generated using two diodes with different emitter areas. For example, consider two diode-connected p-type/n-type/p-type (PNP) transistors, Q1 and Q2. If the emitter area of Q2 is *N* times the emitter area of Q1 and they operate at the same current, an equation for $V_{EB2} - V_{EB1}$ can be written as [44]:

$$V_{EB2} - V_{EB1} = \Delta V_{EB} = V_t \ln(N). \tag{3.17}$$

This equation shows that ΔV_{EB} is PTAT ($\partial \Delta V_{EB}/\partial T = \ln(N)\, k/q$) and dependent upon V_t and N. If this equation is substituted into V_{PTAT} of (3.14) and (3.16) is substituted into V_{CTAT} of (3.14), the following equation is obtained [44]:

$$V_{REF} = V_{G0} - V_t[(\gamma - \alpha) lnT - ln(E \cdot G)] + V_t \ln(N) \cdot K. \tag{3.18}$$

This equation shows that V_{REF} is directly dependent upon V_{GO}. More specifically, if $[(\gamma - \alpha)\ln T - \ln(E \cdot G)] = \ln(N) \cdot K$, $V_{REF} = V_{GO}$. Therefore, the ideal output voltage is V_{GO}. This explains why references that use diodes in this manner are referred to as bandgap voltage references. However, this output can only occur at a single temperature because N and K are constants while γ and α are functions of temperature. The weak dependence of γ and α on temperature explains why bandgap voltage references often have a non-zero temperature coefficient. References that attempt to compensate for this slope are referred to as curvature-compensated [159]–[160]. In most bandgap voltage reference architectures, K is set by resistor ratios. This is important because it reduces the impact of resistor tolerances and resistor temperature coefficients. Examples of bandgap voltage references can be found in [44], [161]–[164].

3.3.1.1 The Use of Vertical PNP BJTs in Bandgap Voltage References

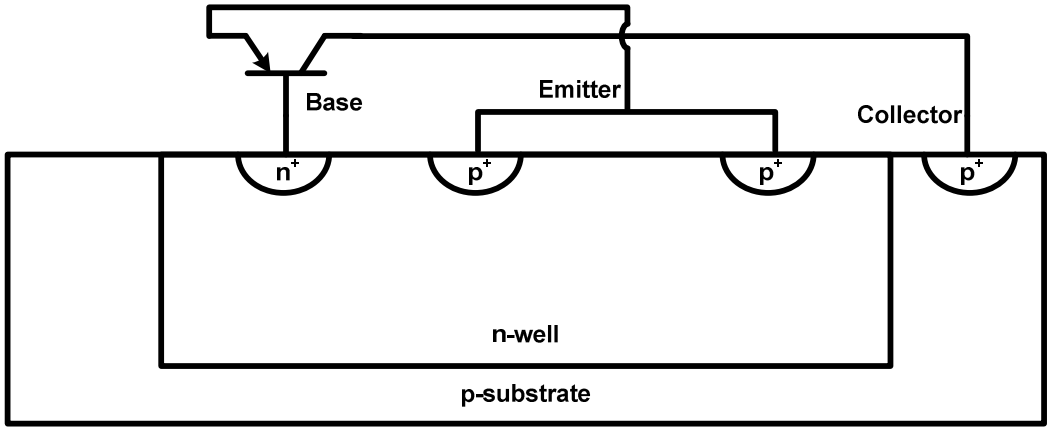

Figure 3.20: Cross section of a vertical PNP BJT made out of a PMOS transistor [165]. The base is formed from the body terminal. The emitter is formed from the source and drain terminals. The collector is formed from the substrate.

The previous subsection showed that the voltage across a forward-biased diode can be used as a CTAT voltage source. In modern CMOS technologies, this diode is typically created using a vertical PNP BJT [165]. A cross section of this device is shown in Figure 3.20. It can be made using a PMOS transistor. The emitter terminal is formed by shorting the source and drain terminals, the base terminal is formed by the body terminal (well contact), and the collector terminal is formed by the substrate. The device is unable to act like a MOSFET because of the shorted source and drain. These PNPs typically exhibit poor BJT characteristics and generally cannot be used in circuit architectures where the collector would be used as an input or output. However, if the base and collector are tied to the same potential, a diode is formed between the emitter and base [165]. This diode provides the temperature behavior described in the previous subsection, which implies that it can be used in the construction of voltage references. Many modern CMOS technologies characterize and model these devices for the sole purpose of voltage reference design [165].

3.3.2 Startup Circuits, Process Variations, and Supply Voltage Dependence

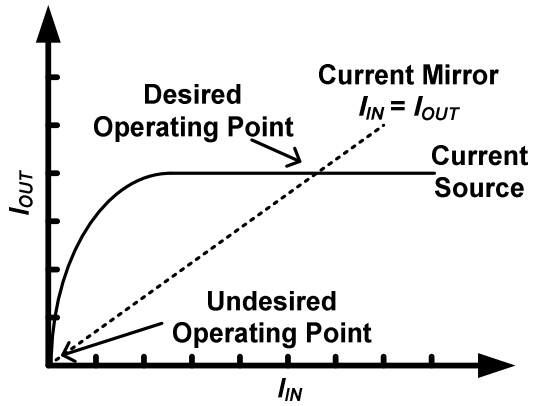

Figure 3.21: Example of different startup operating points that occur in bandgap voltage references

Bandgap voltage references require startup circuits which force them into the proper region of operation [40], [44], [166]. They are needed because a feedback loop exists within the reference, which creates two regions of operation; one at an undesired negligible current and the other at a desired current. For example, consider Figure 3.21, which shows the different startup operating points that can occur in bandgap voltage references. The startup circuit operates by injecting a small current, which triggers the feedback loop and sets the reference to its desired operating point. After this is done, the startup circuit turns off such that it does not further impact performance.

Process variations also play an important role in the design of voltage references [35], [167]–[172]. In [35], current mirror mismatch, resistor mismatch, resistor tolerance, MOSFET mismatch, and BJT mismatch were shown to be the main source of performance degradation. Of these sources, BJT mismatch and voltage offsets due to MOSFET mismatch are most important [35]. In [37], it was shown that matching can be improved by increasing device area. Therefore, in voltage references, devices are made relatively large to minimize the impact of mismatch on performance. This can be

understood by examining drain current mismatch, $\frac{\sigma_{\Delta I_D}}{I_D}$, and V_{GS} mismatch, $\sigma_{\Delta V_{GS}}$. In [173], it was shown that for devices biased in the saturation region and assuming square law operation, drain current mismatch between two devices designed to be identical can be written as:

$$\frac{\sigma_{\Delta I_D}}{I_D} = \sqrt{\left(\frac{\sigma_{\Delta\beta}}{\beta}\right)^2 + \left(\frac{g_m \cdot A_{V_{TH}}}{I_D \cdot \sqrt{WL}}\right)^2} \tag{3.19}$$

where $\frac{\sigma_{\Delta I_D}}{I_D}$ is the standard deviation of the difference in drain currents between the two devices divided by I_D, $\beta = \mu C_{OX} W/L$, μ is the carrier mobility, C_{OX} is the oxide capacitance per unit area, g_m is the gate transconductance, and A_{VTH} is a technology-dependent parameter [37]. The β term of this equation is often assumed to be negligible. Therefore, for a given A_{VTH} and g_m/I_D, current mismatch can be reduced by increasing device area, reducing g_m or increasing I_D. Increasing device area can also be applied to reduce V_{GS} mismatch. For example, an equation for the standard deviation of the difference in V_{GS} voltages between two devices designed to be identical can be written as [173]:

$$\sigma_{\Delta V_{GS}} = \sqrt{\left(\frac{\sigma_{\Delta\beta}}{\beta} \cdot \frac{I_D}{g_m}\right)^2 + \left(\frac{A_{V_{TH}}}{\sqrt{WL}}\right)^2}. \tag{3.20}$$

Assuming the β term is negligible, this equation shows $\sigma_{\Delta V_{GS}}$ can also be reduced by increasing device area. Therefore, current and voltage matching can generally be improved by increasing device area.

Voltage references are also designed to maintain performance independent of V_{DD}. This functionality is tested by sweeping V_{DD} and measuring the output voltage. For a given V_{DD}, the reference is said to function correctly if the output voltage is within an acceptable tolerance. The maximum V_{DD} is typically set by device breakdown voltages. The minimum V_{DD} is typically set by transistor headroom requirements. Avoiding transistor stacks is one technique used to ensure that references operate over a wide range of supply voltages.

3.3.3 Traditional Bandgap Voltage References

Traditional bandgap voltage references have an ideal output equal to the bandgap voltage of the material being used. In CMOS, this usually equates to the bandgap of silicon, which is approximately 1.205 V [44]. A minimum V_{DD} of 1.4 V is needed for these references because of transistor headroom requirements. Therefore, if V_{DD} is less than 1.4 V, traditional bandgap voltage references cannot be used. Many nanoscale CMOS technologies have a $V_{DD} \leq 1$ V. Therefore, a different type of reference is needed in these technologies. These references are referred to as low-voltage references. Several different low-voltage architectures have been proposed. Some are all-MOS while others are based on the bandgap approach. Interestingly, several of these bandgap voltage references have not been shown to function with a $V_{DD} < 1.1$ V [174]–[184]. This may be due to reduced voltage headroom. Therefore, these references were not considered in this work. References that function with $V_{DD} \leq 1$ V are referred to as sub-1 V voltage references.

3.3.4 All-MOSFET Voltage References

MOSFET-only voltage references attempt to achieve temperature independence by balancing the temperature behavior of a MOSFET's threshold voltage with the temperature behavior of carrier mobility [157]. Several examples of this type of reference exist in literature [185]–[191]. One potential problem with this approach is the reliance on the temperature slope of V_{TH}. As CMOS has scaled, significant changes in device dimensions and channel doping have resulted in V_{TH} becoming a strong function of parameters such as L and V_{DS}. This results in the V_{TH} properties of nanoscale transistors differing from transistors of previous generations, which could make it difficult to port these references between technologies. For this reason, only bandgap voltage references were considered in this work.

3.3.5 Sub-1 V Bandgap Voltage References

The basic idea behind a sub-1 V bandgap voltage reference is to force the output to be dependent upon a summation of PTAT and CTAT currents instead of a summation of PTAT and CTAT voltages. For example, consider the sub-1 V bandgap voltage reference in [116]. A high-level schematic of this reference is shown in Figure 3.22. Q1 and Q2 are diode-connected PNP BJTs. I_1, I_2, and I_3 are voltage-controlled current sources (VCCSs). They are designed to be equal. V_P and V_M represent the voltages on the non-inverting and inverting input terminals of the error amplifier. R_1, R_2, R_3, and R_4 are resistors used to zero the temperature slope. V_{REF} is the output voltage. Diode-connected transistor Q2 has N times the emitter area of diode-connected transistor Q1 ($A_{E2} = N \cdot A_{E1}$). The amplifier is used to ensure $V_{EB1} = V_{EB2} + V_{R1}$. If this is true, V_{R1} is equal to the voltage difference of two forward-biased diodes with different emitter

areas operating at the same current. Previously, in (3.17), it was shown that this results in a PTAT voltage. This implies V_{R1} is PTAT, which results in I_{R1} being PTAT. Because I_1, I_2, and I_3 are equal, this implies they all supply I_{R1}, which results in a PTAT current flowing into R_4. The CTAT current is generated by I_{R2} and I_{R3}. These currents, which are ideally equal, flow into R_4 and are CTAT because they depend upon the forward-biased voltage of a diode. As explained in [116] and [192], I_{R4} has a PTAT and CTAT current component, which allows R_2, R_3, and R_4 to be chosen such that V_{REF} is ≤ 1 V and independent of T.

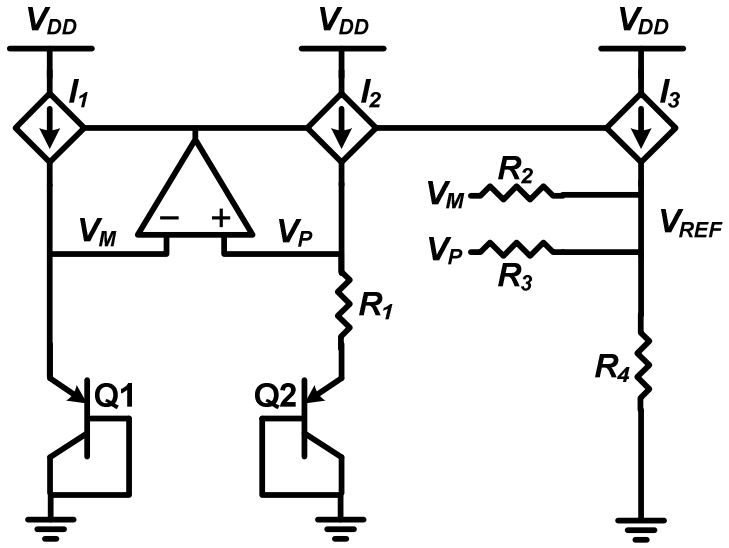

Figure 3.22: Simplified representation of the voltage reference in [116]. V_{DD} is the supply voltage, Q1 and Q2 are diode-connected PNP BJTs. I_1, I_2, and I_3 are voltage-controlled current sources. The error amplifier ensures $V_{EB1} = V_{EB2} + V_{R1}$. V_P and V_M represent the non-inverting and the inverting input voltages of the amplifier. R_1, R_2, R_3, and R_4 are resistors used to zero the temperature slope and set the output voltage, V_{REF}.

The reference in Figure 3.22 is similar to the reference presented in [193]. The main difference between these references is the location of R_2 and R_3. In [193], R_2 and R_3 are in parallel with Q1 and Q2. In this configuration, the effects of resistor tolerance and resistor mismatch have a significant impact on the current flowing through R_2 and R_3. Any variations of R_2 and R_3 directly changes the CTAT current they produce, which

modifies the absolute value of V_{REF} and its temperature slope. In [116], R_2 and R_3 are tied to V_{REF}. As explained in [116] and [192], if V_{REF} is chosen to be equal to V_{EB1} at a desired temperature, the current through R_2 and R_3 is approximately zero at that temperature. This effectively nulls the contributions of R_2 and R_3 at that temperature, reducing the impact of their variation on performance. If this temperature is chosen wisely (i.e., room temperature, the middle of the temperature range, or the temperature at which the IC will be most used), the impact of R_2 and R_3 on performance is minimized. Therefore, the reference in [116] has a significant advantage over [193].

Several other sub-1 V bandgap voltage references can be found in literature. Compared to [116], these references require extra circuitry to achieve the same performance [194]–[197]. This extra circuitry comes in the form of amplifiers, current mirrors, resistors, and diodes. These elements increase power and area. Therefore, these references were not considered in this work. Instead, the reference in [116] was used as a starting point for designing a sub-1 V bandgap voltage reference with ultra-thin oxide MOSFETs.

Sub-1 V bandgap voltage references are generally designed in one of two ways; in a technology with a nominal $V_{DD} > 1$ V or with thick-oxide devices. When designed in technologies with a $V_{DD} > 1$ V, sub-1 V performance is claimed by measuring the reference output with $V_{DD} \leq 1$ V [194], [198]–[199]. One potential problem with this approach is portability. For example, a sub-1 V reference that works in a 0.5 µm technology (nominal $V_{DD} = 3.3$ V) may not be able to be ported to a 65 nm technology (nominal $V_{DD} = 1$ V) because transistor performance between the two technologies is

drastically different. Non-ideal effects, such as gate current and degraded device output resistance may not have been addressed in the reference designed in the 0.5 μm technology.

When designed with thick-oxide devices, sub-1 V bandgap voltage references can be used in technologies with $V_{DD} \leq 1$ V. However, these references are avoiding problems caused by gate current instead of using circuit techniques to solve them. More importantly, there is no existing literature that addresses the problems presented to voltage references by gate current. Given that large area devices are used in voltage references and that gate current is proportional to area, significant amounts of gate current could flow through a poorly designed ultra-thin oxide sub-1 V voltage reference. This work presents a methodology that accounts for this tradeoff. The methodology is used to design and develop a sub-1 V bandgap voltage reference that is capable of functioning in the presence of gate current.

CHAPTER 4
APPROACH

This chapter specifies the approach that was taken to achieve the objectives outlined in Chapter 2. It is broken into nine sections. The first section reviews the computing resources used in this work. The second section presents BJT-like performance metrics that were used to determine the impact of gate current on the analog performance of ultra-thin oxide MOSFETs. The third section motivates the use of body-biasing as a means of reducing the relative impact of gate current on analog design. The fourth and fifth sections describe the approach that was taken to minimize the negative effects of gate current on current mirrors and differential amplifiers. The sixth section describes the AC simulation of amplifiers designed with ultra-thin oxide MOSFETs. The seventh section studies the impact of gate current on sub-1 V bandgap voltage references. The eighth section makes use of the previous seven sections as a methodology to develop an ultra-thin oxide MOSFET-only sub-1 V bandgap voltage reference. The ninth section discusses topics that were not addressed in this work. A simulation strategy subsection is provided in sections two, three, four, five, six, and eight. This subsection outlines the simulations that were performed to test the hypotheses of this work. The results of these simulations are discussed in Chapter 5.

4.1 Computing Resources

The computing resources required for this work included circuit simulation software, a process design kit (PDK) of an ultra-thin oxide CMOS technology with significant gate current, and a device model. Cadence was chosen as the circuit

simulation software. Within Cadence, Virtuoso was used to construct circuit schematics and Spectre was used as the circuit simulator. Analog Design Environment (ADE), which is a component of Cadence, was used to handle the inputs and outputs of Spectre. The PDK used in this work was IBM's 65 nm standard logic (10SF) PDK. This PDK completely describes IBM's 65 nm standard logic process, which has a nominal V_{DD} of 1 V and a t_{ox} of 1.25 nm. The fourth version of the Berkeley Short-channel Insulated Gate Field Effect Transistor Model (BSIM4) was chosen as the device model because of its common use within the analog IC design community [15]. Also, it provides a model for gate current that shows excellent correlation with physical measurement over device dimensions, terminal voltages, and temperature [136].

4.2 Gate Current Performance Metrics

The previous chapter showed that gate current fundamentally degrades MOSFET behavior. This degradation was characterized using the drain current to gate current ratio ($\beta_{F\_MOS} \equiv |I_D/I_G|$), which is similar to the forward current gain ($\beta_F = I_C/I_B$) of a BJT. This work proposes four new metrics to further characterize the impact of gate current on device performance. These metrics are rooted in BJT theory and extend the analogy between gate current and base current. They were used as a guide on how to size and bias ultra-thin oxide MOSFETs.

The first metric, α_{F_MOS}, is defined as I_D/I_S, where I_S is the current through the source terminal. It is analogous to the BJT metric α_F, which is defined as I_C/I_E, where I_E is the current through the emitter terminal [44]. In forward-biased BJTs, it is typically assumed that $\alpha_F \leq 1$, which implies $I_E \geq I_C$. Assuming gate current is similar to base current, α_{F_MOS} should also be ≤ 1, implying $I_S \geq I_D$. This assumption was made in [18],

where the impact of V_{DS} on gate current was assumed to be negligible. However, in Section 3.2.3, it was shown that the I_{GD} component of gate current, which is a function of V_{GD}, can impact the directionality of I_G. For example, consider an NMOS device with a large negative V_{GD}. This negative V_{GD} results in a negative I_{GD}. Because the total gate current, I_G, is a summation of five different components ($I_G = I_{GS} + I_{GD} + I_{GCS} + I_{GCD} + I_{GB}$), the negative contribution of I_{GD} could force I_G to become negative. This would result in I_G flowing out of the gate of an NMOS device, which implies $I_D > I_S$ and $\alpha_{F\_MOS} > 1$. This is analogous to I_B flowing out of the base of an npn BJT, which typically does not occur in the forward active region of operation. This suggests that I_G is not similar to I_B under all operating conditions, which implies that some BJT techniques used to compensate for I_B may not be applicable to ultra-thin oxide MOSFETs.

The second metric, r_{π_MOS}, is defined as $(|\partial I_G/\partial V_{GS}|)^{-1}$. It is analogous to the BJT small-signal resistance r_π, which is defined as $(\partial I_B/\partial V_{BE})^{-1}$. For BJTs in the forward active region, r_π is used to characterize the input resistance of single transistor amplifiers [44]. In MOSFETs, r_{π_MOS} is ideally infinite because $I_G = 0$. In [18], the authors analyzed g_g/I_G, where $g_g = 1/r_{\pi\_MOS}$. It was noted that r_{π_MOS} is finite and g_g/I_G can be studied similarly to g_m/I_D [18]. However, values for r_{π_MOS} were not given. This work provides values for r_{π_MOS} and compares these values to the small-signal output resistance, r_O. If these two values are comparable in magnitude, then r_{π_MOS} may need to be considered when analyzing the small-signal performance of CMOS amplifiers.

The third metric, β_{0_MOS}, is defined as $|i_d/i_g|$, where i_d and i_g are the small-signal drain and gate currents. An equation for β_{0_MOS} can be written as:

$$\beta_{0\_MOS} = \frac{\partial I_D}{\partial V_{GS}} \cdot \left(\frac{\partial I_G}{\partial V_{GS}}\right)^{-1} = g_m \cdot r_{\pi\_MOS}. \tag{4.1}$$

This equation shows that β_{0_MOS} is equal to the product of g_m and r_{π_MOS}. It is analogous to the small-signal current gain β_0 of a BJT, which is defined as i_c/i_b. For BJTs in the forward active region, β_0 is ideally equal to β_F [18]. This equality is due to the fact that β_F is ideally set by process parameters like the base width and emitter/base doping concentration, making it independent of the bias point. β_{0_MOS} is used to inspect the small-signal current gain of ultra-thin oxide MOSFETs. Unlike BJTs, β_{0_MOS} and β_{F_MOS} were not expected to be equal because gate current is a dynamic function of bias point and device dimensions. However, it was expected that β_{0_MOS} and β_{F_MOS} follow the same trends.

The fourth metric, r_{μ_MOS}, is defined as $(|\partial I_G/\partial V_{DS}|)^{-1}$. It is analogous to the BJT small-signal resistance r_μ, which is defined as $(\partial I_B/\partial V_{CE})^{-1}$. For BJTs in the forward active region, r_μ is often assumed to be infinite, causing it to be ignored in circuit analysis [44]. In MOSFETs, r_{μ_MOS} is ideally infinite because $I_G = 0$. This work investigated r_{μ_MOS} to determine if it needs to be considered in ultra-thin oxide design. It was desired that there was a region of V_{DS} values where r_{μ_MOS} was large enough to be ignored. This region would represent an ideal DC bias point to minimize the small-signal impact of V_{DS} on I_G.

4.2.1 Simulation Strategy

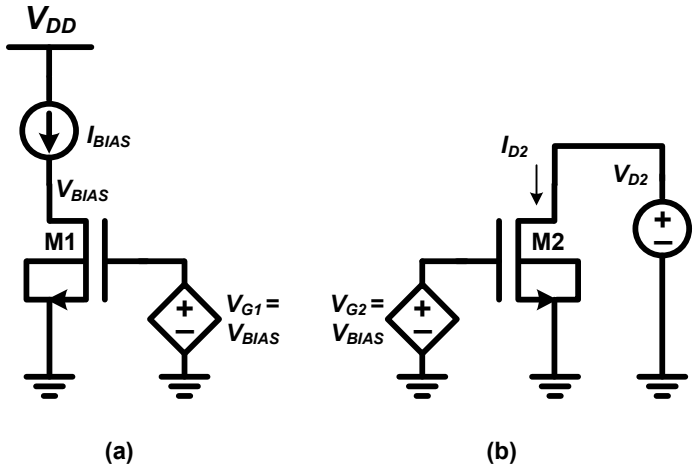

Figure 4.1: Schematic of circuits used to extract gate current performance metrics. V_{DD} is the supply voltage, I_{BIAS} is the bias current, V_{G1} is the gate voltage of M1, V_{D2} is the drain voltage of M2, and V_{G2} is the gate voltage of M2. V_{BIAS} is copied to the gates of M1 and M2 via VCVSs.

The preceding metrics were extracted via simulation using the circuit shown in Figure 4.1. Figure 4.1 (a) shows a transistor, M1, biased with a voltage-controlled voltage source (VCVS) and a DC current source, I_{BIAS}. I_{BIAS} was used to force a desired amount of current into the drain of M1. The VCVS forced $V_{G1} = V_{BIAS}$ without stealing any of I_{BIAS} into the gate of M1. The VCVS was responsible for supplying gate current to M1. Therefore, all of I_{BIAS} went into the drain of M1. This circuit is representative of a diode-connected transistor because $V_{G1} = V_{D1} = V_{BIAS}$. This type of transistor is commonly used in current mirrors. Because $V_{GD1} = 0$, the impact of I_{GD1} was negligible. Therefore, this circuit was used to study I_{G1} without considering the effects of I_{GD1}. β_{F_MOS}, β_{0_MOS}, and r_{π_MOS} were extracted using the circuit in Figure 4.1 (a).

Figure 4.1 (b) shows a transistor, M2, biased with a VCVS and a voltage source. This circuit was used to determine the impact of V_{GD} and V_{DS} on gate current. The VCVS was used to copy V_{BIAS} from Figure 4.1 (a) to the gate of M2. This forced an equal gate-bias point between Figure 4.1 (a) and Figure 4.1 (b). V_{D2} of Figure 4.1 (b) was

swept to determine the impact of V_{GD} and V_{DS} on I_G. β_{F_MOS}, $\alpha_{F\ \_MOS}$, and r_{μ_MOS} were extracted using the circuit in Figure 4.1 (b).

4.3 Impact of Body Biasing on Gate Current

The impact of the MOSFET body voltage, V_{BODY}, on gate current was also studied. This study was motivated by the probability of a carrier directly tunneling through the oxide. In [13], it was shown that this probability is a function of the voltage across the oxide, V_{OX}, and can be approximated as:

$$P_T \approx \frac{V_{OX}}{t_{OX}} e^{-\frac{3}{2}\frac{B_C \cdot t_{OX}}{\chi_B}} \cdot e^{\frac{3}{8}\frac{B_C \cdot t_{OX}}{\chi_B^2} V_{OX}} \tag{4.2}$$

where t_{ox} is the oxide thickness, χ_B is the barrier height, and B_C is a physical constant [13]. The tunneling probability approaches zero as V_{OX} goes to zero ($\lim_{V_{OX}\to 0} P_T = 0$). Therefore, if V_{OX} can be written as a function of V_{BODY}, P_T could be potentially controlled by V_{BODY}. V_{OX} can be expressed as [16]:

$$V_{OX} = V_{GB} - V_{FB} - V_{POLY} - \psi_S \tag{4.3}$$

where V_{GB} is the gate-to-body voltage, V_{FB} is the flatband voltage, ψ_S is the surface potential, and V_{POLY} is the voltage drop due to poly-gate depletion. This equation shows that V_{OX} is dependent upon V_{BODY} through V_{GB} ($V_{GB} = V_G - V_{BODY}$) [200]. Therefore, the probability of a carrier directly tunneling through the oxide is a function of V_{BODY} through V_{GB}.

4.3.1 Simulation Strategy

The dependence of I_G on the body voltage was investigated using the circuits shown in Figure 4.2 and Figure 4.3. In both of these figures, I_{BIAS} was a DC bias current.

In Figure 4.2, V_{BIAS} was copied from the gate of M1 to the gate of M2 via a VCVS. V_{D2} was held at a constant value. V_{BODY} was then swept. This figure simulated the impact of V_{BODY} on M2 under constant terminal voltages. In Figure 4.3, V_{BIAS} was copied from the drain of M3 to the gate of M3 via a VCVS. V_{BODY} was then swept. This figure simulated the impact of V_{BODY} on M3 under constant drain current. β_{F_MOS} and the percentage reduction in I_G was extracted using the circuits in Figure 4.2 and Figure 4.3.

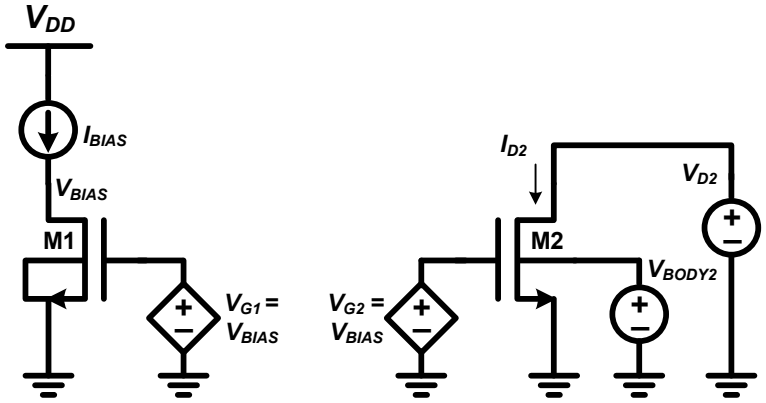

Figure 4.2: Schematic of circuit used to determine impact of body voltage on gate current with constant terminal voltages. V_{DD} is the supply voltage, I_{BIAS} is the bias current, V_{G1} is the gate voltage of M1, V_{D2} is the drain voltage of M2, V_{G2} is the gate voltage of M2, and V_{BODY2} is the body voltage of M2. V_{BIAS} is copied to the gates of M1 and M2 via VCVSs.

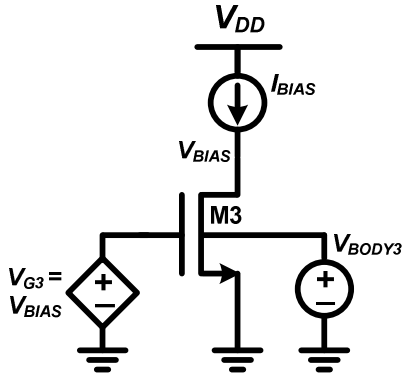

Figure 4.3: Schematic of circuit used to determine impact of body voltage on gate current with constant drain current. V_{DD} is the supply voltage, I_{BIAS} is the bias current, V_{G3} is the gate voltage of M3, and V_{BODY3} is the body voltage of M3. V_{BIAS} is copied to the gate of M3 via a VCVS.

4.4 The Design of Ultra-Thin Oxide CMOS Current Mirrors

This section describes the approach that was taken to minimize the negative effects of gate current on current mirrors. Note that gate current was not the only problem to consider when designing these circuits. The previous chapter showed that degradation of device output resistance and reduced supply voltages also pose significant challenges to current mirrors. Therefore, it was desired that the techniques used to minimize the effects of gate current do not aggravate these pre-existing problems. This section is broken into three subsections. The first subsection describes the design strategy for self-cascode current mirrors. The second subsection describes the design strategy for triple self-cascode current mirrors. The third subsection presents the simulation strategy.

4.4.1 Self-Cascode Current Mirrors

Figure 4.4 shows a self-cascode current mirror. I_{IN} is the input current, V_{OUT} is the output voltage, I_{OUT} is the output current, V_{BIAS} is the bias voltage, and M1-M4 form the mirror. This architecture was used as a starting point for studying the impact of gate current on current mirrors. The motivation for using this circuit comes from the previous chapter, where it was shown that self-cascode structures can achieve large output resistances with minimal voltage overhead. Also, they are able to achieve this type of performance in the saturation and sub-threshold regions of operation [117]. Ideally, the current gain for this structure is $A_i = I_{D4}/I_{D3}$. However, when including gate current, it becomes:

$$A_i = \frac{I_{D4}}{I_{D3} + I_{G1} + I_{G2} + I_{G3} + I_{G4}}. \tag{4.4}$$

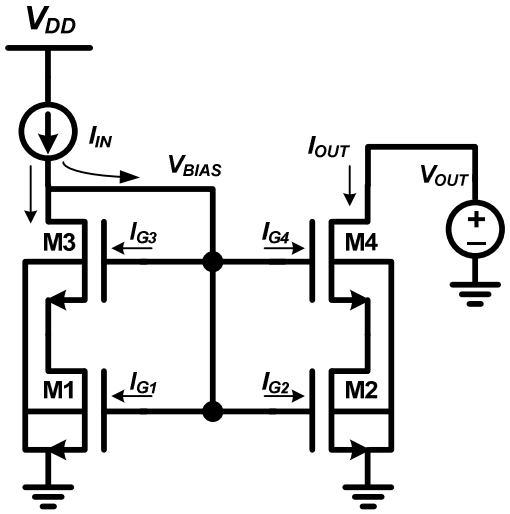

Figure 4.4: Self-cascode current mirror. V_{DD} is the supply voltage, I_{IN} is the input current, V_{OUT} is the output voltage, I_{OUT} is the output current, and V_{BIAS} is the gate voltage of M1-M4.

This equation shows that the current gain is degraded by I_{G1}-I_{G4}. To reduce the impact of I_{G1}-I_{G4} on the current gain, transistors M1-M4 have to be sized and biased such that the amount of total gate current flowing through them is minimized. The metrics described in the previous section were used as an aid for this purpose. One concern of this structure was the gate-to-drain voltage of M4, V_{GD4}. If $V_{GD4} << 0$ and V_{BIAS} is small, I_{G4} could flow out of the gate of M4. This implies that I_{OUT} is supplying gate current to M1-M3, which may not be desired because it could degrade R_{OUT}. This suggests that V_{GD} should be minimized by ensuring that V_{OUT} is not significantly larger than V_{BIAS}.

The circuit shown in Figure 4.5 can be used to further minimize the impact of gate current on current mirrors. This figure is similar to Figure 4.4 except for the addition of a helper transistor, M5 [44]. This additional transistor was used to supply some of the gate current needed by M1-M4. Assuming that M5 has a negligible amount of gate current, most of I_{IN} should go into the drain of M3. This implies that I_{OUT} should mirror I_{IN}

because M1-M4 have equal gate voltages and benefit from the high output resistance provided by the self-cascode structure. Specifically, the current gain of Figure 4.15 is:

$$\frac{I_{D4}}{I_{D3} + I_{G5}} \approx \frac{I_{D4}}{I_{D3}}. \tag{4.5}$$

M5 should be designed with a much smaller area compared to M1-M4 to ensure that its gate current is negligible. Also, if the aspect ratio of M5 is large, V_{GS5} is relatively small, which helps reduce I_{G5} and V_{GD3}. If V_{GD3} is small, this implies I_{GD3} is small, which prevents I_{G3} from flowing out of the gate of M3.

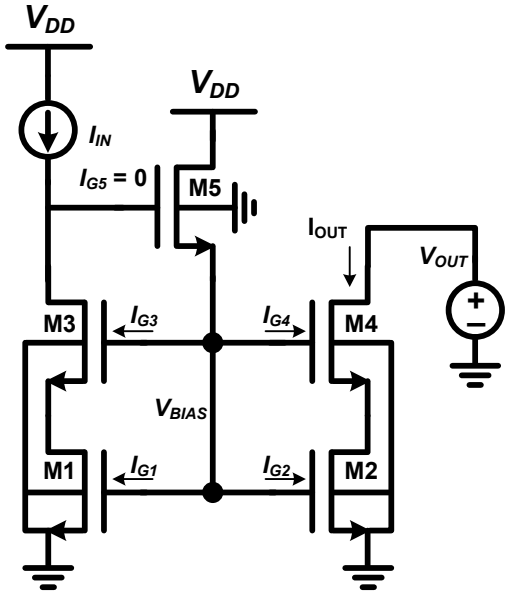

Figure 4.5: Self-cascode current mirror with a helper transistor. V_{DD} is the supply voltage, I_{IN} is the input current, V_{OUT} is the output voltage, I_{OUT} is the output current, and V_{BIAS} is the gate voltage of M1-M4. M5 is the helper transistor. It is used to block I_{IN} from flowing into the gates of M1-M4.

4.4.2 Triple Self-Cascode Current Mirrors

Note that multiple devices could be placed in series to increase the output resistance of the self-cascode structure shown in Figure 3.11 [37]. Individual scale factors would need to be defined between each pair of devices. Ideally, the bottom device of the structure would have the longest channel length and the top device of the

structure would have the shortest channel length. The middle devices would have channel lengths in between those of the bottom device and the top device. Device widths would be chosen such that the scale factor for each pair of series devices is greater than one. Although these types of structures may increase output resistance, they also increase area and could potentially increase gate current, which may limit their practical use. An example of a triple self-cascode structure (three devices in series) and a triple self-cascode current mirror are shown in Figure 4.6. Note that M7 in Figure 4.6 (b) is a helper transistor that serves the same purpose as M5 in Figure 4.5.

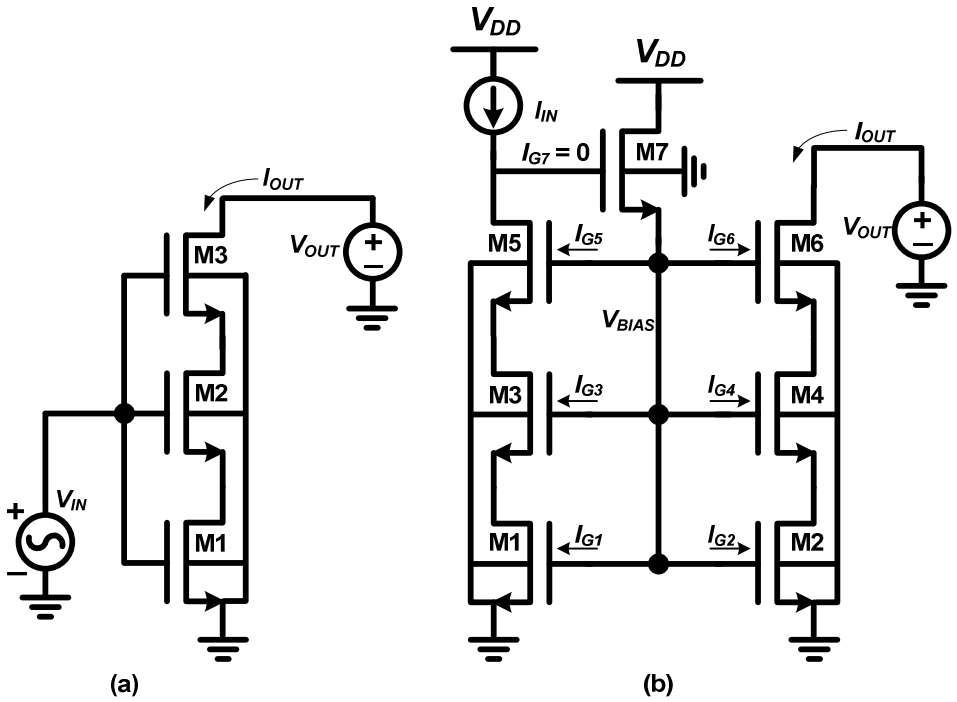

Figure 4.6: (a) Triple self-cascode structure. V_{IN} is the input voltage, V_{OUT} is the output voltage, and I_{OUT} is the output current. M1, M2, and M3 form the self-cascode structure. (b) Triple self-cascode current mirror. V_{DD} is the supply voltage, I_{IN} is the input current, V_{OUT} is the output voltage, I_{OUT} is the output current, and V_{BIAS} is the gate voltage of M1-M6. M7 is a helper transistor. It is used to block I_{IN} from flowing into the gates of M1-M6.

4.4.3 Simulation Strategy

The circuits in Figure 4.4 and Figure 4.5 were simulated to determine if low-voltage current mirrors with large current gains and high output resistances can be

designed with ultra-thin oxide MOSFETs. The output resistance of the triple self-cascode current mirror of Figure 4.6 (b) was simulated and compared to the output resistance of the self-cascode current mirror of Figure 4.5.

4.5 The Design of Ultra-Thin Oxide CMOS Differential Amplifiers

Like current mirrors, differential amplifiers are fundamental building blocks of analog circuit design. Gate current can have a significant impact on their performance. This section describes the approach that was taken to minimize the negative effects of gate current on amplifiers. It was a goal that this approach not aggravate existing problems such as degraded device output resistance and reduced supply voltages. This section is broken into three subsections. The first subsection describes the relationship between gate current and amplifier input current. The second subsection describes the gate-balancing technique. The third subsection presents a circuit technique that can be used to cancel amplifier input current. The fourth subsection presents the simulation strategy.

4.5.1 Amplifier Input Current

Figure 4.7 shows a differential amplifier. M1 and M2 form the input pair, M3 is the tail current source, M4 and M5 form an active load, V_{DD} is the supply voltage, V_{IN1} and V_{IN2} are the common-mode input voltages, V_{BIAS} is the bias voltage of M3, V_{DIO} is the diode-connected voltage between M4 and M5, and V_{OUT} is the output voltage. By inspection, gate current flows into the gate terminals of the input pair, M1 and M2. This fact invalidates the common assumption of negligible MOSFET amplifier input current and is important because differential input pairs are often made large to minimize input offset voltage [39]. Considering that gate current is proportional to device area, this

suggests that input offset voltage and ultra-thin oxide MOSFET amplifier input current trade off with each other.

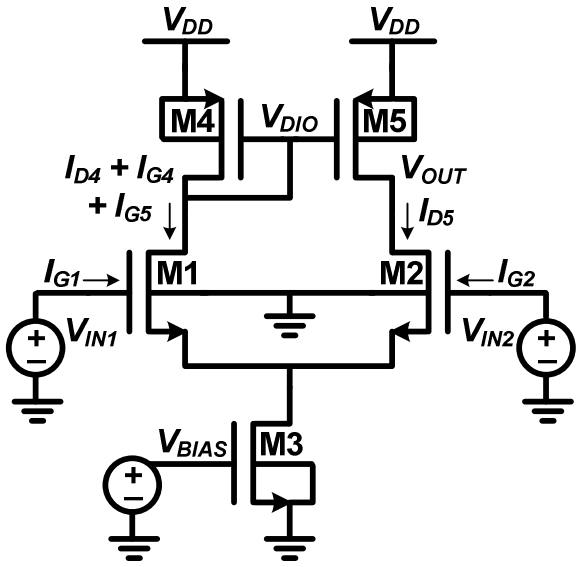

Figure 4.7: Differential amplifier. M1 and M2 form the input pair. M3 is the tail current source. M4 and M5 form an active load. V_{DD} is the supply voltage, V_{IN1} and V_{IN2} are the common-mode input voltages, V_{DIO} is the diode-connected voltage of M4 and M5, V_{BIAS} is the gate-bias voltage of M3, and V_{OUT} is the output voltage.

The amplifier input current can be quantified by two components: input bias current (I_{IN_B}) and input offset current (I_{OS}) [27]. I_{IN_B} is defined as the average current flowing into the gates of M1 and M2: $(I_{G1} + I_{G2})/2$. I_{OS} is defined as the difference in current flowing into the gates of M1 and M2: $I_{G1} - I_{G2}$. Perhaps the best way to minimize the impact of these currents is to minimize their absolute value. This can be accomplished by properly sizing M1 and M2 such that I_{G1} and I_{G2} are minimized (see Section 5.1.3). To ensure that I_{G1} and I_{G2} are similar, biasing techniques could be employed such that V_{IN1} and V_{IN2} have similar common-mode voltages. Note that the body biasing technique described in Section 4.3 could be potentially used to minimize I_{IN_B} and I_{OS} while still allowing for large area devices to decrease the input offset voltage.

4.5.2 Gate Balancing

Gate current also creates imbalance in differential amplifiers. For example, in Figure 4.7, I_{D1} is ideally equal to I_{D2} when $V_{IN1} = V_{IN2}$. This current equality is a direct result of V_{OUT} ideally equaling V_{DIO}. These ideal equalities are fundamental to the balance of differential amplifiers. However, this balance is disrupted by gate current. For example, if gate current flows out of M4 and M5, as shown in Figure 4.7, $I_{D1} = I_{D4} + I_{G4} + I_{G5}$. By inspection, $I_{D2} = I_{D5}$. Therefore, for I_{D1} to equal I_{D2}, I_{D5} must equal $I_{D4} + I_{G4} + I_{G5}$. This equality is unlikely because I_{D4} and I_{D5} are similar and largely set by V_{DIO}. However, if this equality were to occur, V_{OUT} would need to be smaller than V_{DIO} such that I_{D5} increased to compensate for I_{G4} and I_{G5} flowing into M1. This action would disrupt the voltage balance of the amplifier because V_{DIO} would no longer equal V_{OUT}. Therefore, under normal operating conditions, $V_{DIO} \neq V_{OUT}$ and $I_{D1} \neq I_{D2}$ because M2 is not being supplied the same amount of gate current as M1.

One approach to correct the amplifier imbalance of Figure 4.7 is to size M4 and M5 such that I_{G4} and I_{G5} are negligible. However, this may not be possible in technologies with physical oxide thicknesses less than 2 nm or if large area devices are needed to meet matching requirements. Therefore, another approach is needed. One possibility is the gate-balancing technique shown in Figure 4.8, where V_{OUT} ideally drives an equal amount of gate area as V_{DIO}. For example, V_{DIO} drives the gates of M4 and M5 while V_{OUT} drives the gate of M6. If $L_4 = L_5 = L_6$, $W_4 = W_5$, and $W_6 = 2 \cdot W_4$, the gate area driven by V_{DIO} is equal to the gate area driven by V_{OUT}. Therefore, V_{DIO} and V_{OUT} drive the equivalent of two M4 transistors. Assuming $I_{G4} = I_{G5}$, I_{G6} would ideally equal $2 \cdot I_{G4}$.

This implies that equal of amounts of gate current will be flowing into the drains of M1 and M2, thus restoring the amplifier's balance.

The gate-balancing technique of Figure 4.8 assumes V_{D6} is similar to V_{OUT} and V_{DIO}. If these voltages are not similar, the gate-to-drain overlap current of M6, I_{GD6}, may cause I_{G6} to be different than $2 \cdot I_{G4}$ [19]–[20]. This could disrupt the balance of the amplifier. Diode-connected transistors (M9 in Figure 4.8) or resistors can be used as voltage drop elements to force V_{D6} to be similar to V_{OUT} and V_{DIO}. However, these elements must be used with caution. They may reduce the amplifier's output voltage swing. For example, referring to Figure 4.8, more voltage will be required across V_{OUT}' to keep M9 in the desired region of operation.

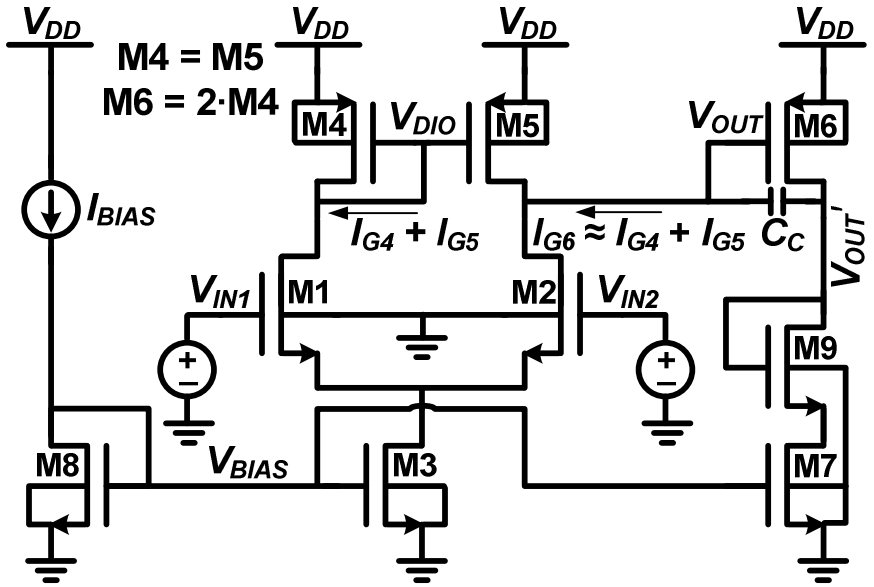

Figure 4.8: Balanced differential amplifier. M1 and M2 form the input pair. M3, M7, M8, and I_{BIAS} form the bias network. M4 and M5 form an active load. V_{DD} is the supply voltage, V_{IN1} and V_{IN2} are the common-mode input voltages, V_{BIAS} is the gate-bias voltage for M3, V_{DIO} is the diode-connected voltage of M4 and M5, V_{OUT} is the output voltage, and I_{OUT} is the output current. M6 is used to restore balance to the amplifier. M9 is used to force similar drain voltages between M4, M5, and M6. C_C is the compensation capacitor.

The gate-balancing technique is not restricted to the amplifier architecture shown in Figure 4.8. M6 could be a dummy transistor or, more generally, it could be a transistor

that is driven to obtain some desired functionality. It does not always have to be the input transistor of a second amplifier stage. For example, in [193], a single-ended differential amplifier drives the gates of multiple transistors to create a sub-1 V bandgap voltage reference. The gate balancing technique could be applied to such a circuit to aid in the creation of a sub-1 V bandgap voltage reference that accounts for gate current (see Section 4.8.1). Also, referring to Figure 4.8, V_{OUT} could drive the gates of two separate transistors, both sized equally to M4 and M5. Furthermore, if V_{OUT} drives three transistors, each sized equally to M4 and M5, V_{DIO} could drive the gate of a dummy or biasing transistor with dimensions equal to M4 and M5. This technique is general in nature and can be used where necessary to correct gate current-induced amplifier imbalance.

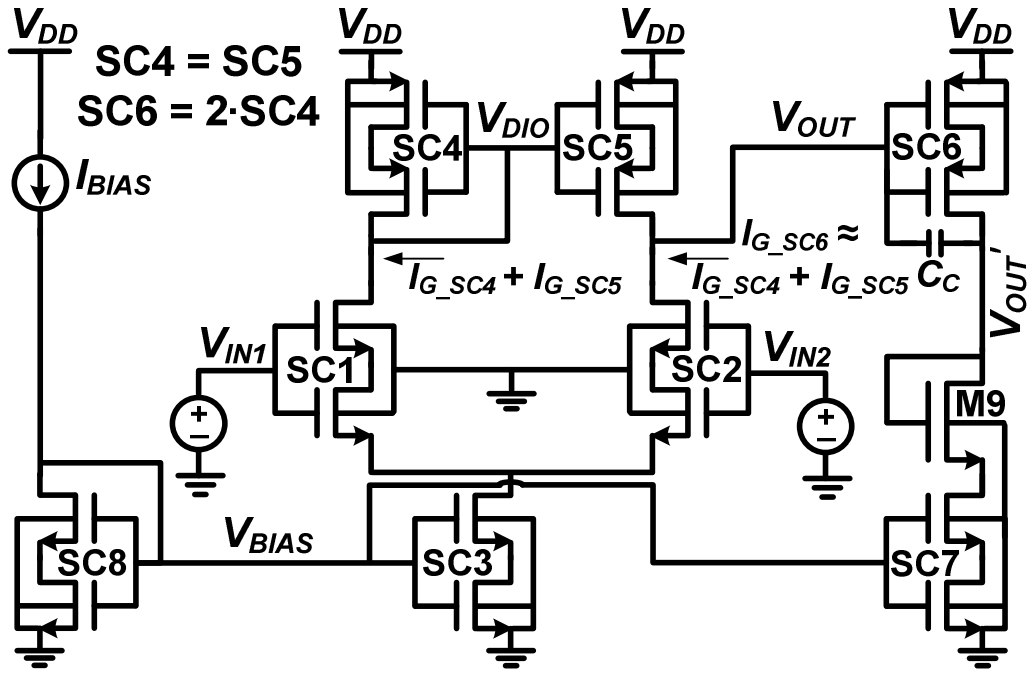

Figure 4.9: Two-stage self-cascode operational amplifier. SC1 and SC2 form the input pair. SC4 and SC5 form the active load. SC6 forms the second stage. SC3, SC7, SC8, and I_{BIAS} form the bias network. V_{DD} is the supply voltage. V_{IN1} and V_{IN2} are the common-mode input voltages. M9 is a diode-connected transistor used to force similar drain voltages between SC4, SC5, and SC6. V_{OUT} and V_{OUT}' are the output voltages of the first and second stages. C_C is the compensation capacitor.

Applying the gate-balance technique in combination with self-cascode structures is advantageous because it minimizes the effects of drain voltage differences while also increasing the amplifier's voltage gain. For example, consider Figure 4.9, which shows a gate-balanced self-cascode two-stage amplifier. If the cascoding devices are chosen to have relatively short channel lengths, their gate current will be minimal and thus the effects of drain voltage differences between them will be minimal. Also, because of the shielding provided by these devices, the gate and drain voltages of the cascoded devices will be similar. Therefore, the cascoded devices will ideally have equal voltages on all terminals and thus draw equal gate currents. This is important because these devices have longer channel lengths and therefore draw more gate current than the cascoding devices. The shielding provided by the cascoding devices allows the amplifier's balance to be set by the gate currents of the cascoded devices. This can be achieved by designing with a large S_F. Considering that gate current is generally undesirable, it may not be a good strategy to intentionally increase the gate current of the device being cascoded as a means of dwarfing the gate current through the cascoding device. Instead, the gate current through both devices should be minimized in such a way that their total contribution can be made as small as possible. However, if the impact of drain voltage differences between cascoding devices cannot be made negligible by sizing and biasing techniques, a diode-connected transistor can be used to minimize the voltage differences. For example, in Figure 4.9, M9 can be used to force the drain voltages of SC4-SC6 to be similar. This ensures that the gate currents of the cascoding devices of SC4-SC6 are similar and helps maintain gate current balance between all of these devices.

4.5.3 Input Current Cancellation

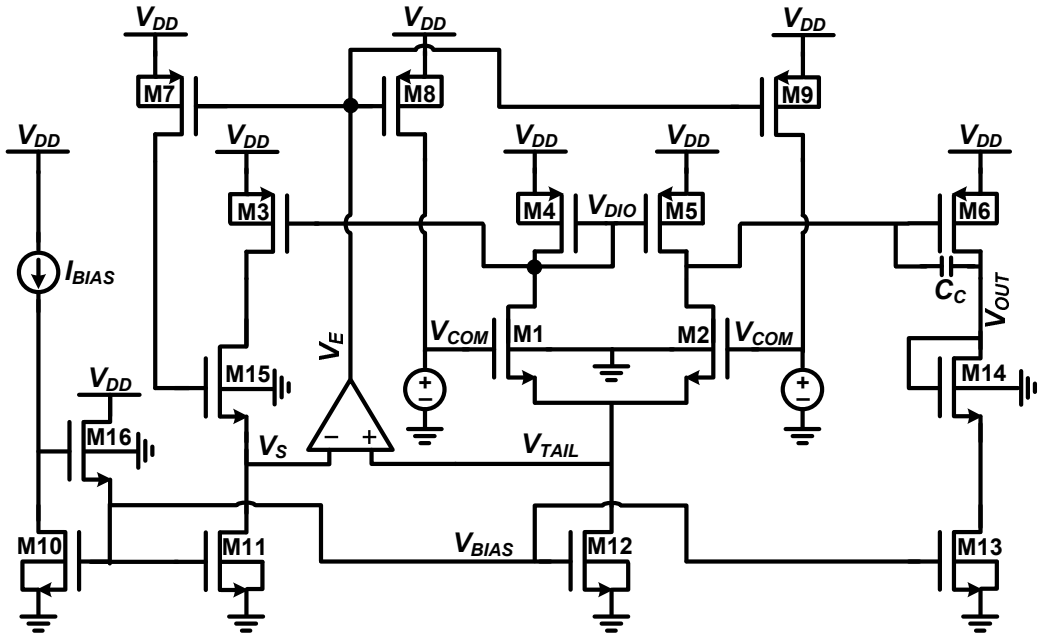

Figure 4.10: Differential amplifier with input current cancellation. M1 and M2 form the input pair. M12 is the tail current source. M4 and M5 form an active load. M16 is a helper transistor. V_{DD} is the supply voltage, V_{COM} is the common-mode input voltage, V_{DIO} is the diode-connected voltage of M4 and M5, V_{BIAS} is the gate-bias voltage of M10-M13, I_{BIAS} is the bias current, and V_{OUT} is the output voltage. C_C is the compensation capacitor. The input current cancellation network is formed by the error amplifier, M3, M7-M9, M11, and M15. V_S is the source voltage of M15 and V_E is the output voltage of the error amplifier.

One technique that could be used to cancel the effects of amplifier input current is shown in Figure 4.10. M1-M2, M4-M6, M10, and M12-M14 form a two-stage differential amplifier similar to that of Figure 4.8. M16 is a helper transistor used to block gate current from flowing into M10-M13. The input current cancellation network is formed by the error amplifier, M3, M7-M9, M11, and M15. The network attempts to minimize the input current provided by the input common-mode voltage sources, V_{COM}, to M1 and M2. This effectively increases the amplifier's low-frequency input resistance. The technique works as follows. M15 is sized equal to M1-M2. The error amplifier forces the tail voltage of M1 and M2, V_{TAIL}, to be equal to the source voltage of M15, V_S. M3 is used to bias the drain terminal of M15. It is equal in size to M4 and M5 and has the same gate bias voltage as M4 and M5. Therefore, M15 ideally supplies the same

amount of drain current as M4 and M5. M11 is used to bias the source terminal of M15. Its width is equal to half the width of M12 and its channel length is the same as that of M12. Therefore, M11 sources half the current of M12, which is ideally equal to the source current flowing through either M1 or M2. These bias conditions force M15 to have the same drain current and the same source current as M1 and M2. If this is true, M15 must have the same terminal voltages as M1 and M2. Specifically, $V_{GS15} = V_{GS1} = V_{GS2}$ and $V_{DS15} = V_{DS1} = V_{DS2}$. This implies that all three of these transistors draw the same amount of gate current. The gate current of M15 is supplied by M7, which is regulated by the error amplifier. The error amplifier also regulates M8 and M9. This implies that the gate currents of M1 and M2 are supplied by M8 and M9. If M8 and M9 supply I_{G1} and I_{G2}, then the V_{COM} voltage sources are not supplying gate current, effectively increasing the input resistance of the amplifier. Note that the gate balancing technique can be applied between M3, M4, M5, and M6. A similar technique can be applied using BJTs [201].

The error amplifier of Figure 4.10 allows the input resistance of the amplifier to remain high with changes in V_{COM}. For example, as V_{COM} increases, V_{TAIL} increases such that the drain currents of M1 and M2 do not change. The output voltage of the error amplifier, V_E, is adjusted such that $V_S = V_{TAIL}$ and $I_{D7} = I_{D8} = I_{D9} = I_{G15} = I_{G1} = I_{G2}$. An example of a transistor-level implementation of the error amplifier is shown in Figure 4.11. The amplifier has a PMOS differential input stage. This type of input stage was chosen because V_{TAIL} of Figure 4.10 has a relatively small absolute voltage, making it easier to bias with a PMOS input pair than an NMOS input pair. M9 is the second stage

of the amplifier. A second stage was used to increase the output voltage swing and to balance the gate currents between M7, M8 and M9.

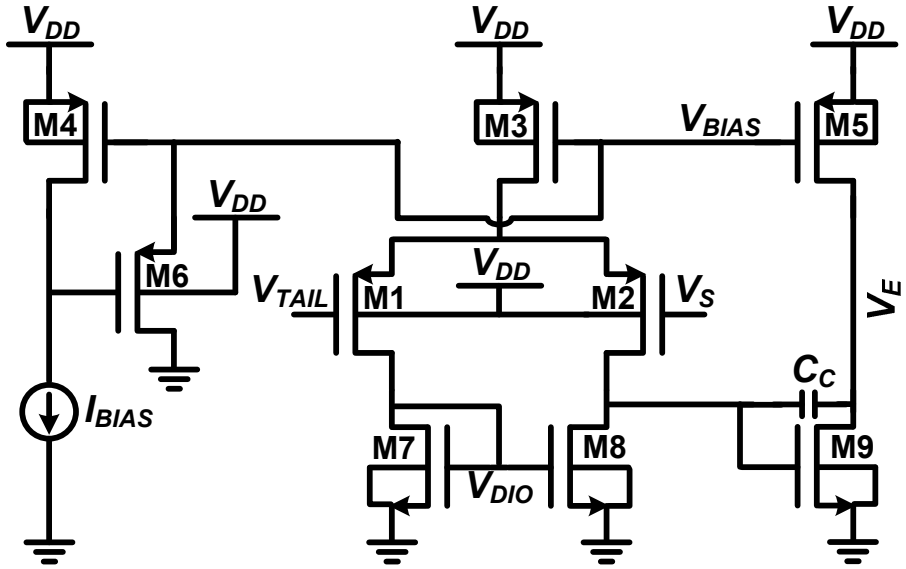

Figure 4.11: Transistor-level schematic of the error amplifier in Figure 4.10. M1 and M2 form the input pair. M3, M4, M5, and I_{BIAS} form the bias network. M7 and M8 form an active load. V_{DD} is the supply voltage, V_{TAIL} is connected to the tail voltage of M1 and M2 in Figure 4.10, V_S is connected to the source voltage of M15 in Figure 4.10, V_{BIAS} is the gate-bias voltage for M3, V_{DIO} is the diode-connected voltage of M7 and M8, V_E is the output voltage and is connected to the gate terminals of M7, M8, and M9 in Figure 4.10. M9 is the second stage of the amplifier. It is used to restore balance to the amplifier. C_C is the compensation capacitor.

4.5.4 Simulation Strategy

The circuit in Figure 4.7 was simulated to show that gate current disrupts the balance of differential amplifiers. The two-stage self-cascode operational amplifier shown in Figure 4.9 was simulated to show that amplifier balance can be restored using the gate balancing technique. Also, the voltage gain, $A_V = v_{out}'/v_{in}$, where v_{in} is the small-signal input voltage and v_{out}' is the small-signal output voltage was simulated for the two-stage self-cascode operational amplifier shown in Figure 4.9. The results were compared to the voltage gain of the simple operational amplifier shown in Figure 4.8. This was done to show the voltage gain enhancement that can be achieved using self-cascode structures. A self-cascode version of the differential amplifier of Figure

4.10 was simulated to show that amplifier input resistance can be increased by applying the input current cancellation technique. The results were compared to the two-stage self-cascode amplifier of Figure 4.9.

4.6 The AC Simulation of Ultra-Thin Oxide CMOS Amplifiers

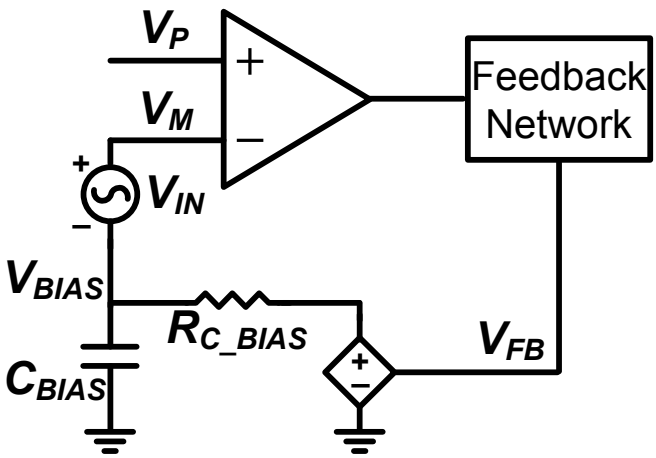

Figure 4.12: Circuit technique used to maintain the DC bias point when performing amplifier AC simulations [48]. V_P and V_M represent the amplifier's non-inverting and inverting input voltages, V_{FB} is the feedback voltage, V_{IN} is the small-signal input voltage, R_{C_BIAS} and C_{BIAS} create a low-pass filter, and the VCVS is used to copy the DC component of V_{FB} to V_{BIAS}.

Gate current also impacts the simulation of amplifiers. For example, when performing an AC simulation on an amplifier, a DC bias point must be chosen. This bias point is important because it plays a role in determining small-signal transistor parameters like g_m, r_O, and C_{gs} [44]. These parameters are used by simulators to calculate an amplifier's open-loop AC response. Typically, when performing an AC simulation, the correct DC bias point is the one found in the closed-loop configuration [48]. However, when the feedback loop is broken, this bias point is lost. To break the feedback loop but maintain the bias point, a simple circuit technique is employed. This technique, which is shown in Figure 4.12, uses a VCVS, a resistor (R_{C_BIAS}), and a capacitor (C_{BIAS}) [48]. V_P and V_M represent the amplifier's non-inverting and inverting input terminal voltages, V_{FB} is the feedback voltage, V_{IN} is the small-signal input voltage,

and V_{BIAS} is the DC bias voltage to be copied from the output (V_{FB}) to the input (V_M). Assuming that V_P is set by external circuitry (for example, a bandgap voltage reference), this technique sizes R_{C_BIAS} and C_{BIAS} such that they form a low-pass filter that can transfer the DC value of V_{FB} through the VCVS to V_{BIAS}. Example values of R_{C_BIAS} and C_{BIAS} are 100 MΩ and 500 μF. Their exact values are not important; they just need to be sized large enough to force V_{BIAS} to be the DC value of V_{FB}. The VCVS is used to prevent any current flow through R_{C_BIAS}. This is important, because in closed-loop operation V_{FB} is connected to V_M, which is typically the gate of a MOSFET that ideally draws no DC current. The VCVS also prevents R_{C_BIAS} and C_{BIAS} from loading down the feedback network. The other alternative to this approach is to use ideal voltages sources on the amplifier's input terminals. However, if this is done, the impact of process variations on the DC bias point cannot be simulated.

The technique shown in Figure 4.12 fails if non-negligible input current flows into the inverting or non-inverting input terminals of the amplifier. In ultra-thin oxide CMOS, this input current could be gate current due to direct tunneling. In the closed-loop configuration, the input current through the amplifier's inverting input terminal is provided by V_{FB}. If the loop is broken and the technique in Figure 4.12 applied, V_{FB} no longer supplies this current, which changes its DC bias point. The circuit shown in Figure 4.13 represents a potential solution to this problem. This figure is similar to Figure 4.12 except for the addition of the amplifier input current (I_{IN_A}), feedback output current (I_L), two current-controlled current sources (CCCSs), an inductor (L_{BIAS}), and a resistor (R_{L_BIAS}).

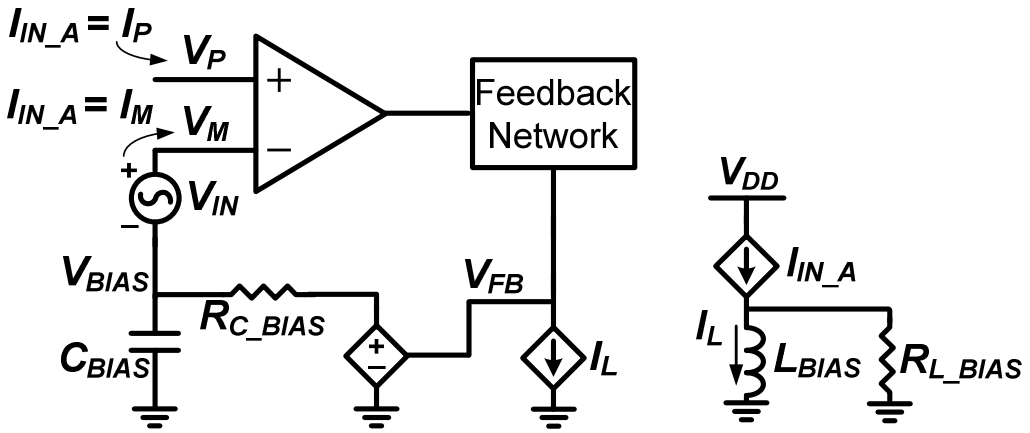

Figure 4.13: Circuit technique used to maintain the DC bias point when performing amplifier AC simulations in the presence of non-negligible amplifier input current. V_P and V_M represent the amplifier's non-inverting and inverting input voltages, V_{FB} is the feedback voltage, V_{IN} is the small-signal input voltage, R_{C_BIAS} and C_{BIAS} create a low-pass filter, and the VCVS is used to copy the DC component of V_{FB} to V_{BIAS}. I_{IN_A} and I_{IN_B} represent the amplifier input current. R_{L_BIAS} and L_{BIAS} from a low-pass filter that transfers the DC current component of I_{IN_A} via two CCCSs to V_{FB}.

The DC voltage transfer of V_{FB} to V_{BIAS} works exactly the same as in Figure 4.12. The DC current transfer works as follows. The input current through the inverting terminal, I_{IN_A}, is copied to the CCCS connected to R_{L_BIAS} and L_{BIAS}. I_{IN_A} contains a DC component and an AC component. The DC component comes from V_{BIAS} and the AC component comes from the small-signal input voltage, V_{IN}. R_{L_BIAS} and L_{BIAS} are used to filter the AC component. This is done by making L_{BIAS} relatively large and R_{L_BIAS} relatively small. Example values of R_{L_BIAS} and L_{BIAS} are 1 kΩ and 1 mH. At DC, L_{BIAS} acts like a short and thus the DC component of I_{IN_A} flows through it. At frequencies greater than DC, the impedance of L_{BIAS} increases while the impedance of R_{L_BIAS} remains the same. Therefore, if the impedance of R_{L_BIAS} is much smaller than L_{BIAS} at these frequencies, the high frequency components of I_{IN_A} will flow through R_{L_BIAS}. This implies I_L is equal to the DC component of I_{IN_A}. This current is what is needed at the output node to maintain the amplifier's DC bias point. The CCCS connected to V_{FB} is used to copy I_L to the output node. This forces $V_{BIAS} = V_{FB}$ and $I_{IN\_A} = I_L$, which restores

the correct DC bias point. This technique can be used to maintain DC bias point stability when performing AC simulations of closed-loop amplifiers.

4.6.1 Simulation Strategy

The buffer amplifier shown in Figure 4.16 (see Section 4.8.1) was simulated using the techniques shown in Figure 4.12 and Figure 4.13. The open-loop DC bias point of each technique was recorded and compared to the closed-loop DC bias point to determine which technique provided better accuracy. The amplifier output resistance of each technique was also recorded and compared.

4.7 Impact of Gate Current on Sub-1 V Bandgap Voltage References

This section describes the approach that was taken to minimize the negative effects of gate current on sub-1 V bandgap voltage references. A mathematical analysis was performed on the voltage reference shown in Figure 3.22 (see Appendix B.1). Assuming no gate current, an equation for the output voltage, V_{REF}, can be written as:

$$V_{REF} = \frac{MBV_t \ln(N) + 3MV_{EB1}}{3M + B} \tag{4.6}$$

where $N = A_{E2}/A_{E1}$, $B = R_2/R_1$, $M = R_4/R_1$, $I_1 = I_2 = I_3$, and $R_2 = R_3$. Assuming that the temperature slope of the resistors is negligible, this equation contains a PTAT component dependent upon the difference in V_{EB} voltages of two forward-biased PNP BJTs (see (3.17)) and a CTAT component dependent upon the V_{EB} voltage of a PNP BJT. Therefore, it can be differentiated with respect to temperature, set equal to zero, and solved for B to determine the R_2/R_1 ratio that forces V_{REF} to remain constant with temperature. If this is done, an equation for B can be written as:

$$B = -3\left(\frac{\partial V_{EB1}}{\partial T}\Big/\frac{\partial V_t \ln(N)}{\partial T}\right). \tag{4.7}$$

This equation is of the same form as (3.15). Therefore, it fulfills the requirements of a bandgap voltage reference. Given that (4.7) can be used to solve for B, M can be solved for by rewriting (4.6):

$$M = \frac{BV_{REF}}{3V_{EB1} + BV_t \ln(N) - 3\mathrm{V_{REF}}}. \tag{4.8}$$

Given that N is known, B is obtained from (4.7), and V_{EB1} is obtained from simulation, the only unknown in this equation is V_{REF}. As noted in [116] and [192], if V_{REF} is set equal to V_{EB1} at a desired temperature, the contributions of R_2 and R_3 are effectively nulled. Applying this to (4.8) yields:

$$M = \frac{V_{EB1}}{V_t \ln(N)}. \tag{4.9}$$

This equation shows M is ideally independent of R_2 and R_3 and mathematically proves that allowing V_{REF} to equal V_{EB1} effectively nulls their contributions at a desired temperature.

To account for amplifier non-idealities, the circuit shown in Figure 4.14 can be analyzed. This circuit is a schematic representation of the voltage reference in [116] (see Figure 3.22) that includes input offset voltage (V_{OS}), input bias current (I_{IN_B}), and input offset current (I_{OS}). The input offset voltage is modeled using a voltage source between the inverting terminal of the amplifier and the node connecting I_1 and Q1. The input offset voltage represents the amount of voltage needed to balance the common-mode

response. The input bias current and input offset current represent the gate current flowing into the input terminals of the error amplifier.

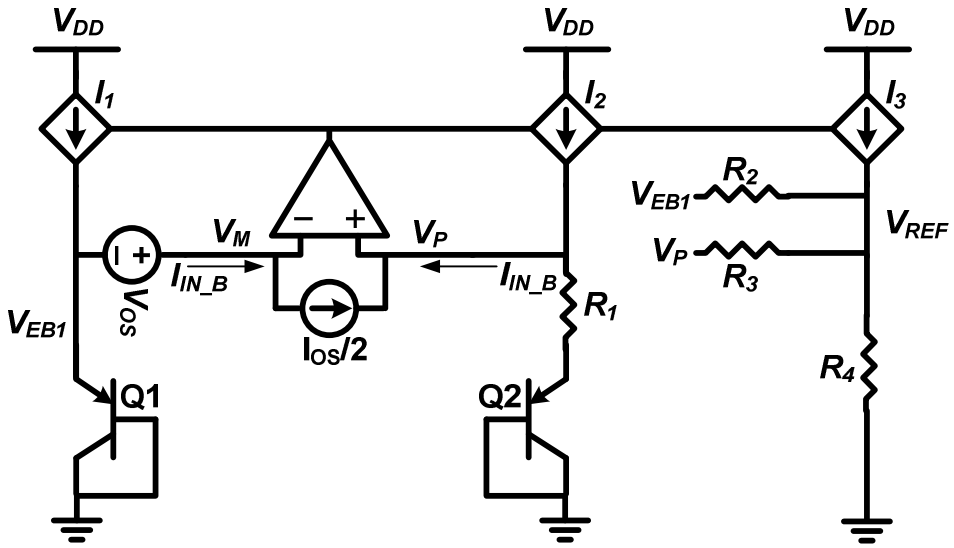

Figure 4.14: Sub-1 V bandgap voltage reference including amplifier input offset voltage and amplifier input current. Q1 and Q2 are diode-connected PNP BJTs. I_1, I_2, and I_3 are voltage-controlled current sources. The error amplifier ensures $V_{EB1} + V_{OS} = V_{EB2} + V_{R1}$. V_P and V_M represent the non-inverting and inverting input voltages of the amplifier. R_1, R_2, R_3, and R_4 are resistors used to zero the temperature slope and set the output voltage, V_{REF}. I_{IN_B} and I_{OS} represent the input bias current and the input offset current of the amplifier.

The model for I_{IN_B} and I_{OS} is explained in Appendix B.2. Using Figure 4.14, an equation for V_{REF} can be written as (see Appendix B.2):

$$V_{REF} = \frac{MBV_t \ln(N) + 3MV_{EB1}}{3M + B} + \frac{V_{OS}M(B+2)}{3M + B} + \frac{I_{IN\_B}R_1MB}{3M + B}. \qquad (4.10)$$

This equation shows that V_{REF} is a function of V_t, V_{EB1}, V_{OS} and I_{IN_B}. V_{OS} and I_{IN_B} are undesirable and introduce non-idealities that degrade performance. In CMOS technologies with $t_{ox} > 3$ nm, I_{IN_B} is negligible and can be ignored. Therefore, in these technologies, the main source of non-ideality is V_{OS}. To reduce its impact on performance, transistor area is increased [35], [37]. However, in CMOS technologies with $t_{ox} < 3$ nm, I_{IN_B}, which is proportional to device area, is not negligible. Therefore, increasing area to improve performance is a difficult strategy to employ because the

impact of I_{IN_B} is increased. A circuit technique is needed to reduce the impact of I_{IN_B} while allowing device area to be increased such that the effects of V_{OS} are reduced.

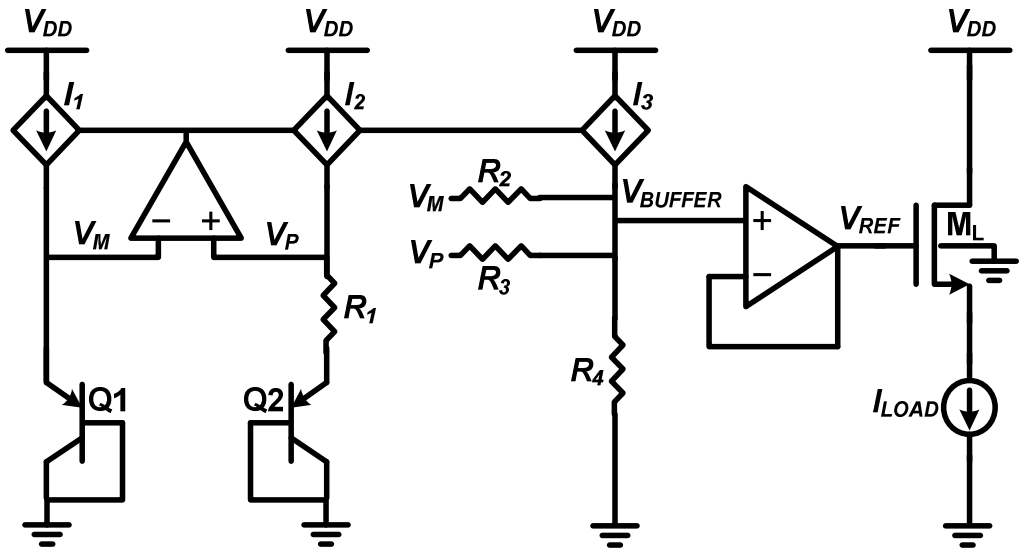

Figure 4.15: Sub-1 V bandgap voltage reference that minimizes the effects of amplifier input current. Q1 and Q2 are diode-connected PNP BJTs. I_1, I_2, and I_3 are voltage-controlled current sources. The error amplifier ensures $V_{EB1} = V_{EB2} + V_{R1}$. V_P and V_M represent the non-inverting and inverting input voltages of the amplifier. R_1, R_2, R_3, and R_4 are resistors used to zero the temperature slope and set the buffer voltage, V_{BUFFER}. V_{BUFFER} is the voltage transferred by the buffer to output of the reference, V_{REF}. The buffer is added to drain the input current of the error amplifier out of I_3. M_L and I_{LOAD} represent the load transistor and load current.

The circuit shown in Figure 4.15 attempts to reduce the impact of gate current with the addition of a buffer amplifier. The non-inverting input terminal of the buffer is used to drain I_{IN_B} from I_3. Note that I_3 contains I_{IN_B} because I_1 and I_2, which both supply I_{IN_B} to the error amplifier, are mirrors and designed to be equal to I_3. If the non-inverting input terminal of the buffer drains all of I_{IN_B} from I_3, no amplifier input current flows into R_4 and transistor area can be increased to minimize the effects of V_{OS}. This implies (4.6) can be used to approximate V_{BUFFER}, which is forced to equal to V_{REF} by the action of the buffer because no amplifier input current flows into R_4. In technologies with significant gate current, this technique can be employed when designing a bandgap voltage reference of the forms presented in [116] and [193]. If not

used, the input current from the error amplifier, which has a nonlinear temperature coefficient, degrades performance by flowing into R_4. This causes the absolute voltage of the reference to change and it also creates a non-zero temperature slope.

4.8 The Design of an Ultra-Thin Oxide Sub-1 V Bandgap Voltage Reference

This section describes the transistor-level design of the voltage reference shown in Figure 4.15. It is broken into six subsections. The first subsection describes how self-cascode structures and the gate-balancing technique were applied in the design of the voltage reference. The second subsection describes a novel startup circuit that accounts for the presence of gate current. The third subsection describes the impact of amplifier input current on the performance of the reference. The fourth subsection describes the design tradeoff between power and area. The fifth subsection discusses amplifier compensation. The last subsection presents the simulation strategy.

4.8.1 Self-Cascoding and Gate-Balancing

Figure 4.16 shows a transistor-level schematic of the voltage reference in Figure 4.15. The transistor pairs labeled *SCX* represent self-cascode structures. SC1-SC5 form the error amplifier, SC6-SC8 form I_1-I_3, SC9-SC10 form the bias network for the error amplifier, and SC11-SC19 form the buffer amplifier. M_L and I_{LOAD} form the load.

The gate-balancing technique presented in Figure 4.8 was applied in Figure 4.16 between nodes V_A and V_B. V_B drives the gates of SC4, SC5, and SC16. SC4 and SC5 are equal in area. SC16 is twice the area of SC4 and SC5. Therefore, V_B drives the equivalent of four equally sized self-cascode structures. V_A drives the gates of SC6-SC9, which are equal in area to SC4. Therefore, V_A and V_B drive an equal amount of gate area.

If each of these self-cascode structures leak an equal amount of gate current, the error amplifier should remain balanced.

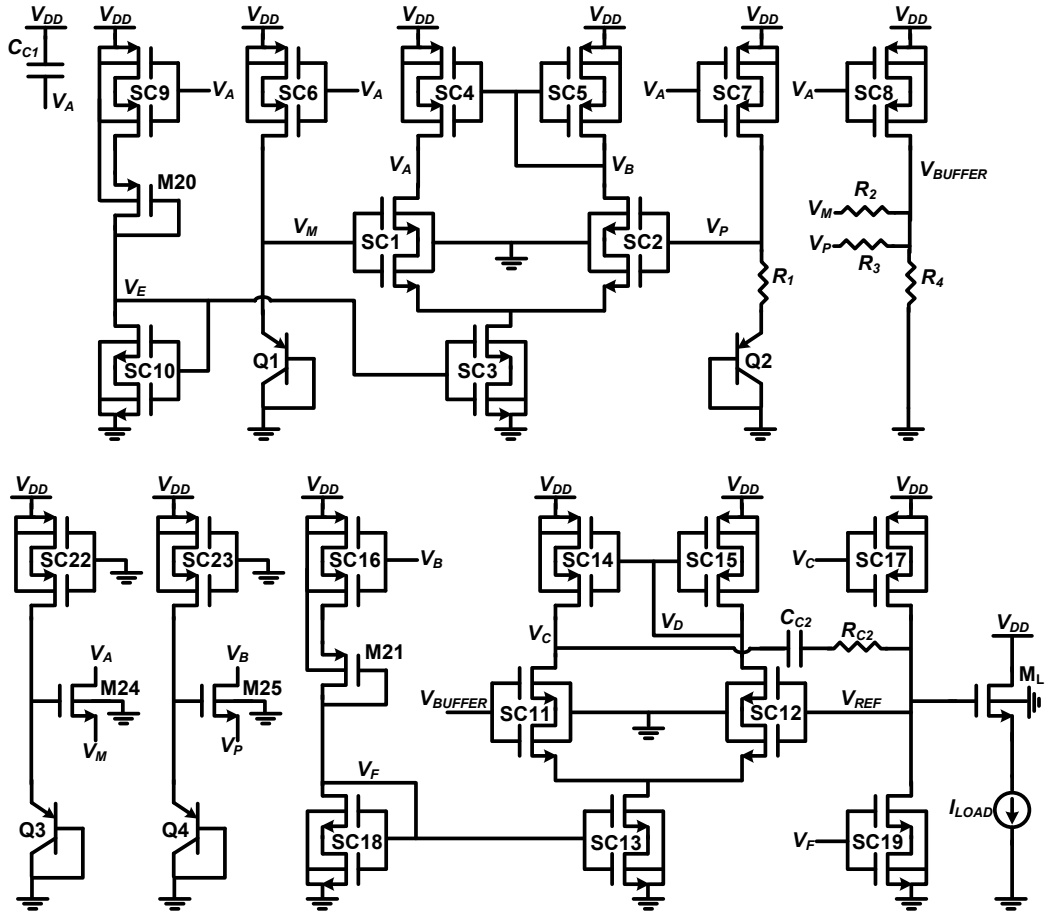

Figure 4.16: Transistor level schematic of Figure 4.15. SC1-SC5, SC9, SC10, and M20 form the error amplifier. SC6-SC8 form I_1-I_3. SC13-S19 form the buffer amplifier. SC22, SC23, M24, M25, Q3 and Q4 form the startup circuit. M_L and I_{LOAD} form the load transistor and load current. R_1, R_2, R_3, and R_4 are resistors used to zero the temperature slope and set the buffer voltage, V_{BUFFER}. C_{C1}, C_{C2}, and R_{C2} form the compensation networks for the error amplifier and the buffer amplifier.

This technique was also applied between nodes V_C and V_D. V_D drives the gates of SC14 and SC15, which have equal areas. V_C drives the gate of SC17, which is twice the area of SC14. Therefore, V_C and V_D both drive the equivalent of two equally sized self-cascode structures, which allows the buffer to remain balanced. M20 is a diode-connected transistor used to minimize the drain voltage differences between SC9 and SC4-SC8. M21 is a diode-connected transistor used to minimize the drain voltage

differences between SC16 and SC4-SC5. Note that a PMOS device was used to minimize the drain voltage differences in lieu of an NMOS device (see M9 in Figure 4.8) to maximize the voltage drop for a given device area and drain current. The voltage drop increased because the threshold voltage of the PMOS transistor was greater than the threshold voltage of the NMOS transistor. Also, the carrier mobility of the PMOS transistor was less than the carrier mobility of the NMOS transistor.

4.8.2 Startup

Figure 4.16 contains a startup circuit specifically designed to minimize the impact of gate current on voltage reference performance. The startup circuit is made up of SC22, SC23, M24, M25, Q3, and Q4. The startup circuit works as follows: SC22, SC23, Q3, and Q4 are used to bias M24 and M25. If the reference fails to start, negligible current flows through Q1 and Q2. Therefore, the gate voltages of M24 and M25 will be larger than their source voltages. This will cause them to begin conducting. The current out of their source terminals will be fed directly into the emitter terminals of Q1 and Q2. This causes the emitter voltages of Q1 and Q2 to rise, forcing SC1 and SC2 to conduct. The conduction of these self-cascode structures forces the feedback loop of the amplifier to place the reference in the desired operating condition. Once in this condition, the startup circuit turns off because V_{GS24} and V_{GS25} are extremely small.

The negative effects of gate current are balanced and minimized because the gate and source terminals of M24 and M25 are designed to change similarly with temperature. This occurs because these terminals are all connected to an emitter terminal of a diode-connected PNP BJT. If M24 and M25 are sized equally, they leak the same

amount of gate current because they have equal voltages on their terminals. This balances their gate current contribution. Minimization occurs by sizing SC22, SC23, Q3, and Q4 such that V_{GS24} and V_{GS25} are as small as possible over the temperature range of the voltage reference.

4.8.3 Impact of Amplifier Input Current

Equation (4.10) showed that I_{IN_B} factored directly into the output voltage of the reference. In Figure 4.16, this current is represented by I_{G1} and I_{G2}. Although gate current is ideally independent of temperature under constant terminal voltage conditions (see [144]–[145]), I_{G1} and I_{G2} change with temperature via the terminal voltages of SC1 and SC2. These currents are CTAT because V_{GS1} and V_{GS2} are CTAT. V_{GS1} and V_{GS2} are CTAT because V_{G1} and V_{G2} are ideally equal to the emitter voltage of Q1, which is a forward-biased diode with a temperature slope of ≈ −1.8 mV/°C [157]. To minimize the impact of I_{IN_B} on performance, I_{R1} can be increased. For example, (B.20) can be solved for R_1 and the result can be substituted into (4.10) to obtain:

$$V_{REF} = \frac{MBV_t \ln(N) + 3MV_{EB1}}{3M + B} + \frac{V_{OS}M(B + 2)}{3M + B} + \frac{I_{IN\_B}(\Delta V_{EB1} + V_{OS})MB}{I_{R1}(3M + B)}. \quad (4.11)$$

The third term of this equation is dependent upon the ratio of I_{IN_B} to I_{R1}. As I_{R1} increases, the relative impact of this term decreases, thus reducing the impact of I_{IN_B}. This assumes that I_{IN_B} does not increase at the same rate as I_{R1}. Referring to Figure 4.16, this can be understood by assuming $I_{R1} = I_{D2}$. Therefore, as I_{R1} increases, I_{D2} increases. Assuming that β_{F_MOS} increases with increases in drain current (see Section 5.1.3), the

relative impact of I_{IN_B} will decrease, which implies that increasing I_{R1} reduces the impact of I_{IN_B} on performance.

The buffer is used to minimize the impact of amplifier input current on performance. For example, $V_{GS1} = V_{GS2} = V_{GS11} = V_{GS12}$ at a specific temperature because V_{REF} is designed to equal V_{EB1} at that specific temperature,. This ensures that the gate current mirrored by SC1 and SC2 into SC8 is equal to the gate current drawn by SC11 and SC12 at the temperature where $V_{REF} = V_{EB1}$. Therefore, at this specific temperature, SC11 prevents this current from flowing into R_4 and impacting the ideal performance of the reference. As temperature changes, V_{EB1} no longer equals V_{REF}, resulting in V_{GS1} and V_{GS2} not equaling V_{GS11} and V_{GS12}. Therefore, the gate current of SC11 is slightly different than what is mirrored into SC8 by SC1 and SC2. This is undesired and suggests a small amount of gate current will flow into R_4. Because I_{IN_B} is CTAT, more CTAT current than expected is flowing. To account for this extra CTAT current, R_2 and R_3 can be slightly increased. By increasing R_2 and R_3, the CTAT currents I_{R2} and I_{R3} are reduced, which forces the total CTAT current flowing into R_4 to be closer to what it would be if no gate current were flowing into R_4. The net effect of this technique is an increase in B.

4.8.4 Power and Area Tradeoffs

The amount of current flowing in each of the current mirrors of Figure 4.16 has a significant impact on total power consumption and area. Equation (B.4) shows that this current is directly dependent upon R_1. Therefore, to reduce power, R_1 should be large. This results in larger R_2, R_3, and R_4 values, which increases the total area of the reference.

For example, if $N = 8$ and $I_{R1} = 2$ μA, $R_1 = 26.9$ kΩ. In the obtained technology, a precision poly resistor of this value can be made using an area of 69.5 μm^2 ($W = 1$ μm, $L = 69.5$ μm). On the other hand, if $N = 8$ and $I_{R1} = 12$ μA, $R_1 = 4.5$ kΩ. In the obtained technology, a precision poly resistor of this value can be made using an area of 11.5 μm^2 ($W = 1$ μm, $L = 11.5$ μm). This example demonstrates the tradeoff between resistor area and power. I_{R1} should be selected based on the application in which the reference is going to be used. For example, in a low-power application, I_{R1} would be small and resistor area would be large. In area-sensitive applications, I_{R1} could be increased, which would result in less overall area consumed by the reference. Note that the operating temperature range of the reference and the voltage headroom needed across SC6 and SC7 may limit increases in I_{R1}. Specifically, as I_{R1} increases, the voltage across Q1 and Q2 increases, which implies the voltage headroom of I_1 and I_2 decreases. As temperature decreases, the voltage headroom across I_1 and I_2 further decreases because of the CTAT nature of Q1 and Q2. This decrease in voltage headroom may cause I_1, I_2, and I_3 to stop acting like current mirrors, thus degrading reference performance. Therefore, increases in I_{R1} are limited by the voltage headroom requirements of SC6-SC8 in Figure 4.16.

Another concern of the voltage reference in Figure 4.16 is the total number of resistors. It is desirable to minimize the number of resistors to reduce area. To achieve the highest degree of matching between the resistors, they should be composed of series and parallel strings of a unit resistor, R_U [116]. As shown in Figure 4.17, this can lead to a seemingly excessive number of resistors. For example, if $R_1 = 5$ kΩ, $B = 30$, and M = 15: $R_2 = R_3 = 150$ kΩ and $R_4 = 75$ kΩ. If $R_1 = R_U$, R_2 and R_3 would each be made

using 30 unit resistors while R_4 would be made using 15 resistors. Therefore, 76 total unit resistors would be needed for R_1-R_4.

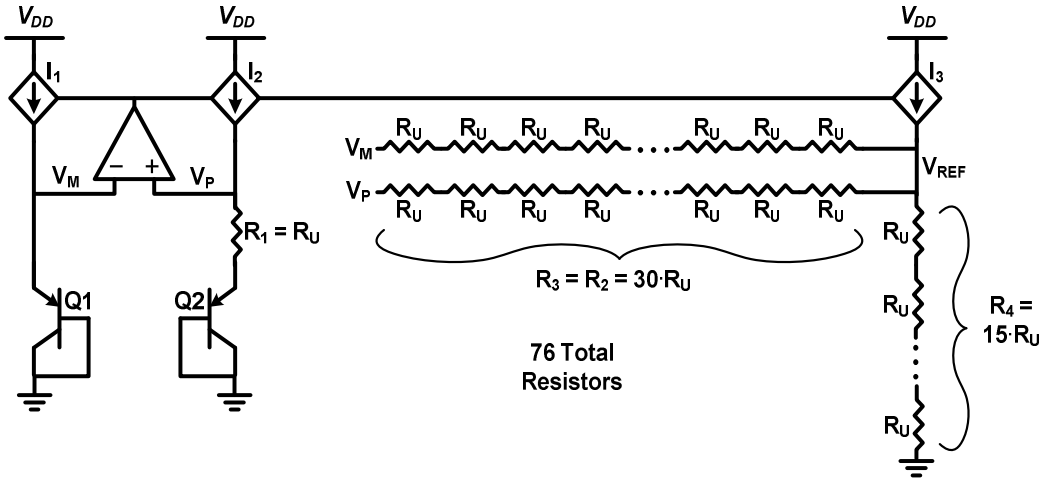

Figure 4.17: High-level schematic of [116] with excessive resistors. V_{DD} is the supply voltage, Q1 and Q2 are diode-connected PNP BJTs. I_1, I_2, and I_3 are voltage-controlled current sources. The error amplifier ensures $V_{EB1} = V_{EB2} + V_{R1}$. V_P and V_M represent the voltages on the non-inverting and inverting terminals of the amplifier. R_1, R_2, R_3, and R_4 are represented by series or parallel combinations a unit resistor (R_U). V_{REF} is the output voltage.

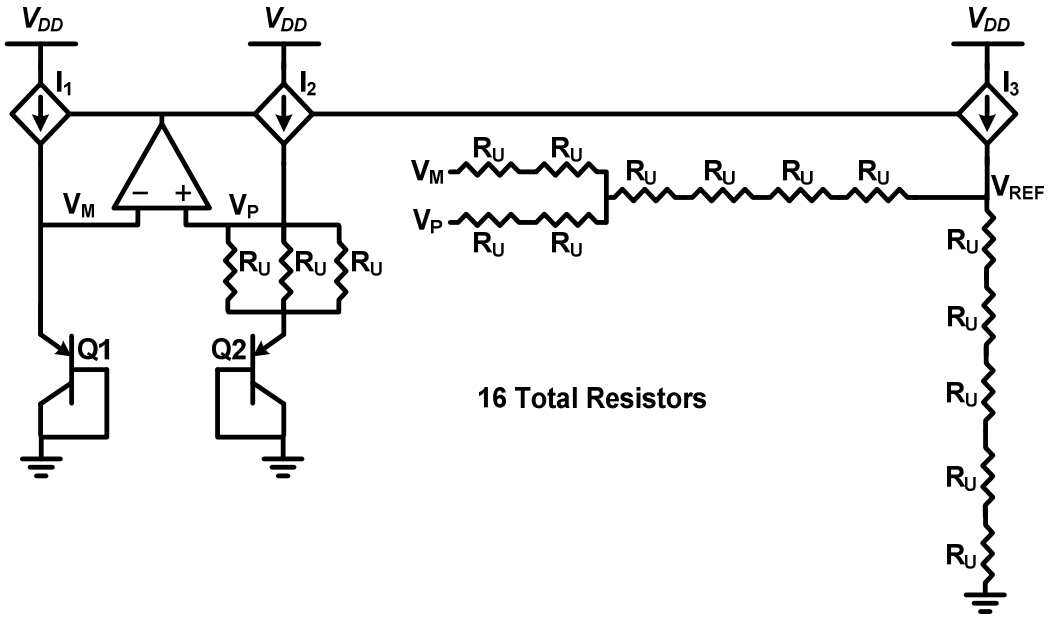

Figure 4.18: High-level schematic of [116] with combined resistors. V_{DD} is the supply voltage, Q1 and Q2 are diode-connected PNP BJTs. I_1, I_2, and I_3 are voltage-controlled current sources. The error amplifier ensures $V_{EB1} = V_{EB2} + V_{R1}$. V_P and V_M represent the voltages on the non-inverting and inverting terminals of the amplifier. R_1, R_2, R_3, and R_4 are represented by series or parallel combinations a unit resistor (R_U). V_{REF} is the output voltage.

To decrease the total number of resistors, the technique presented in [116] can be used. First, R_1 is made using parallel combinations of R_U. Next, the resistors making up R_2 and R_3 are combined. This can be done by recognizing that ideally $V_P = V_M$. Therefore, one end of R_2 and one end of R_3 are both ideally connected to $V_P = V_M$. The other end of R_2 and the other end of R_3 are both physically connected to V_{REF}. This implies that R_2 and R_3 can be analyzed as if they are in parallel and suggests they can be combined [116]. The limit of combination occurs when the combined portions of R_2 and R_3 degrade performance. This can be observed by performing Monte Carlo and process corners analyses. If this technique is applied in the previous example, the number of resistors can be reduced from 76 to 16 (see Figure 4.18).

4.8.5 Amplifier Compensation

The error amplifier and buffer amplifier in Figure 4.16 must be compensated. The compensation of the error amplifier is achieved using a capacitor, C_{C1}, with one end tied to V_A and the other end tied to V_{DD}. The compensation of the buffer is achieved using a series combination of a resistor (R_{C2}) and capacitor (C_{C2}) between V_C and V_{REF}. These compensation techniques are heavily covered in textbooks [36], [40], [44], [48]. Their effectiveness is determined by performing an AC simulation and calculating the phase and gain margins. To ensure stability, the phase margin should be $\geq 45°$ and the gain margin should be ≥ -10 db [44]. The capacitors used in these techniques cannot be made using ultra-thin oxide MOSFETs (see Section 3.2.6). To avoid the effects of gate current they can be made by using reverse-biased diode capacitance, metal-insulator-metal capacitance, or metal-oxide-metal capacitance.

4.8.6 Simulation Strategy

An ultra-thin oxide version of the voltage reference shown in Figure 3.22 was compared to thick and ultra-thin oxide versions of the reference presented in [116]. The thick-oxide reference was designed to show that a sub-1 V bandgap voltage reference can achieve a high level of performance in a nanoscale CMOS technology. Monte Carlo analyses were used to evaluate its results. All of the transistors in the thick-oxide reference were then switched to ultra-thin oxide and the reference was re-simulated. This was done to show the performance degradations caused by gate current. A ultra-thin oxide version of Figure 4.16 was then designed and simulated. A Monte Carlo analysis was performed and the results were compared to the previous two references. Five other analyses were used to characterize the ultra-thin oxide sub-1 V bandgap voltage reference. The first was a +/− 3-sigma process corners simulation of V_{REF} vs. T. The second was a +/− 3-sigma process corners simulation of V_{REF} vs. V_{DD}. The third was a transient startup corners analysis of V_{REF} vs. time (t). The fourth was a simulation to study the impact of loading (M_L and I_{LOAD} in Figure 4.16) on performance. The fifth was a sensitivity analysis, which was performed to determine which of BSIM4's direct tunneling parameters the reference was most sensitive too.

Large-area devices ($W \cdot L > 100\ \mu m^2$) were used in this work. The motivation for using device areas this large stems from the matching typically required in voltage references [35]. However, because gate current increases with area, matching and gate current trade off with each other. The impact of this tradeoff was determined by performing Monte Carlo analyses. For example, when designing the ultra-thin oxide bandgap voltage reference of Figure 4.16, a starting area was chosen for each device. A

Monte Carlo analysis was performed on the voltage reference. Device area was then increased and the Monte Carlo analysis was re-run. This process was repeated until the best possible performance was obtained. The optimum device area occurred when the combined negative effects of gate current and mismatch were at a minimum. When the device area was smaller than optimum, performance was constrained by mismatch. When device area was larger than optimum, performance was constrained by gate current. This approach represents a design methodology that can be employed when the combined negative effects of mismatch and gate current need to be minimized.

4.9 Topics Not Addressed in This Work

No attempts were made to model direct tunneling in this work. There were two major reasons for not modeling. First, accurate models already exist [13], [14], [31], [132]–[136]. Many of these models show excellent correlation with measurement across a wide range of device dimensions, terminal voltages, and temperature. Also, the physical basis of these models are similar in the sense that they all depend on the five components of direct tunneling described in Section 3.2.3 (I_{GCS}, I_{GCD}, I_{GS}, I_{GD}, I_{GB}). This implies that the academic community generally agrees on how direct tunneling should be modeled. Many of these models were developed over 10 years ago. This suggests they have been subjected to academic scrutiny, without failure, for this period of time. Also, the model presented in [136] is a part of BSIM4, which is widely used in industry. For example, IBM relies on BSIM4 to model its 65 nm 10SF technology [15]. This implies ultra-thin oxide CMOS circuits are being designed using the BSIM4 direct tunneling model, which validates its ability to accurately predict behavior. Given that models like this exist, any new attempt may be redundant and of little additional value.

The second reason for not attempting to model direct tunneling was the lack of published circuit techniques to deal with its negative effects on analog design. This lack of publications directly motivated this work and implied any headway that could be made in this area had potential value. Specifically, this work aimed to be the first to provide analog circuit solutions to direct tunneling. These solutions were not based on a simple direct tunneling equation. There is no "square-law" equivalent for direct tunneling. Most compact models rely on approximations, fitting parameters, and smoothing functions to correctly describe its behavior. This type of modeling is not exclusive to direct tunneling and is therefore not a concern [15]. Physical intuition was used to develop circuit solutions. Specifically, this work used the fact that direct tunneling is modeled as having five components which are strong functions of a particular set of voltages. Also, it made use of the approximation that β_{F_MOS} is roughly proportional to $1/L^2$ [18]. Therefore, even though a single self-contained equation was not used, the proposed circuit solutions are rooted in accepted theory and physically verified models.

CHAPTER 5
RESULTS

This chapter presents the results of this work. It has seven sections. The first six sections presents simulation results from the six simulation strategy subsections of the previous chapter. The first section presents simulation results of the gate current metrics described in Section 4.2. It also contains a subsection that presents a channel length selection methodology for ultra-thin oxide MOSFETs. The second section presents simulation results that characterize the impact of body biasing on gate current (Section 4.3). The third, fourth, and fifth sections present simulation results of the current mirror and amplifier techniques described in Sections 4.4, 4.5, and 4.6. The sixth section presents simulation results comparing the thick-oxide voltage reference presented in [116] to the ultra-thin oxide voltage reference described in Section 4.8. The last section presents the design of a chip that was awarded via the MOSIS Education Program [38].

5.1 Gate Current Performance Metrics

This section presents simulation results of the gate current metrics described in Section 4.2. It has three subsections. The simulation results from the first subsection characterize the impact of gate current on diode-connected transistors. The simulation results from the second subsection characterize the impact of V_{DS} on gate current. The last subsection presents a channel length selection methodology for ultra-thin oxide MOSFETs.

5.1.1 Impact of Gate Current on Diode-Connected Transistors

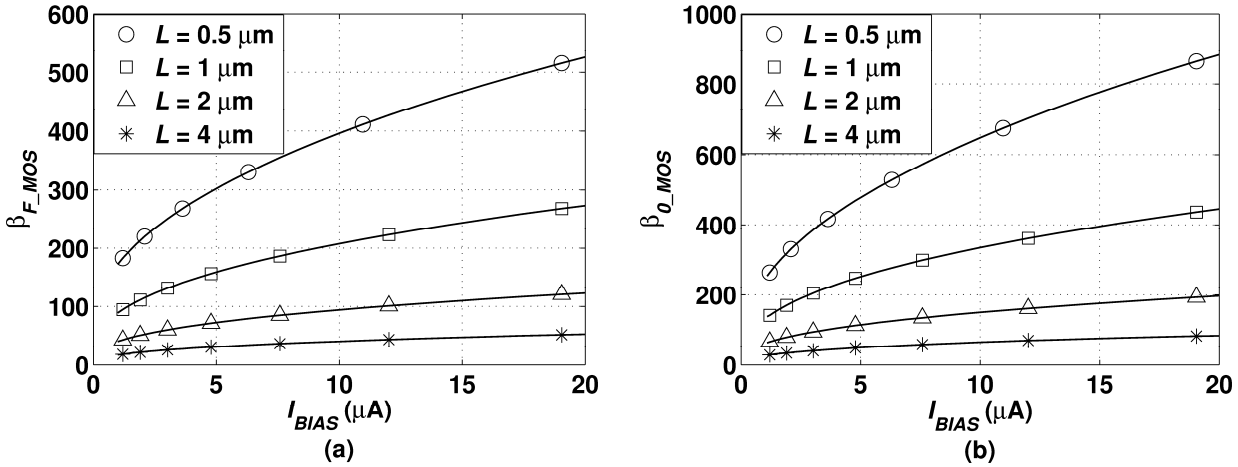

Figure 5.1: (a) β_{F_MOS} vs. I_{BIAS}. (b) β_{0_MOS} vs. I_{BIAS}. Both graphs refer to the circuit shown in Figure 4.1 (a). Transistor area was held constant at 100 μm^2. The legends specify L.

The circuit in Figure 4.1 (a) was simulated to determine the impact of gate current on transistors with $V_{GD} = 0$. Under this condition, the ultra-thin oxide MOSFET acts similar to a BJT because I_{GD} has negligible impact on the directionality of I_G. Two scenarios were simulated. The first scenario kept device area constant at 100 μm^2 while varying L and I_{BIAS}. This was done to determine the impact of L and I_{BIAS} on β_{F_MOS}, β_{0_MOS}, and r_{π_MOS}. The results for this scenario are shown in Figure 5.1 and Figure 5.2.

Figure 5.1 (a) plots β_{F_MOS} vs. I_{BIAS} and Figure 5.1 (b) plots β_{0_MOS} vs. I_{BIAS}. The results show that β_{F_MOS} and β_{0_MOS} increase significantly with reductions in L. For example, as L decreased from 4 μm to 500 nm (I_{BIAS} = 5 μA), β_{F_MOS} increased from 30 to 310 and β_{0_MOS} increased from 50 to 490. These results confirm what was presented in [18], which is that β_{F_MOS} and β_{0_MOS} both increase significantly with reductions in L.

Figure 5.1 also shows that β_{F_MOS} and β_{0_MOS} increase significantly with increases in I_{BIAS}. For example, as I_{BIAS} increased from 1 μA to 20 μA, β_{F_MOS} increased from 89 to 275 and β_{0_MOS} increased from 137 to 445. As I_{BIAS} increases, the transistor approaches saturation and the dominant current mechanism changes from diffusion to drift. This

causes the device to act more like a MOSFET and less like a BJT. The results from Figure 5.1 (b) suggest that β_{F_MOS} and β_{0_MOS} can be increased at the expense of power and voltage by increasing the bias current.

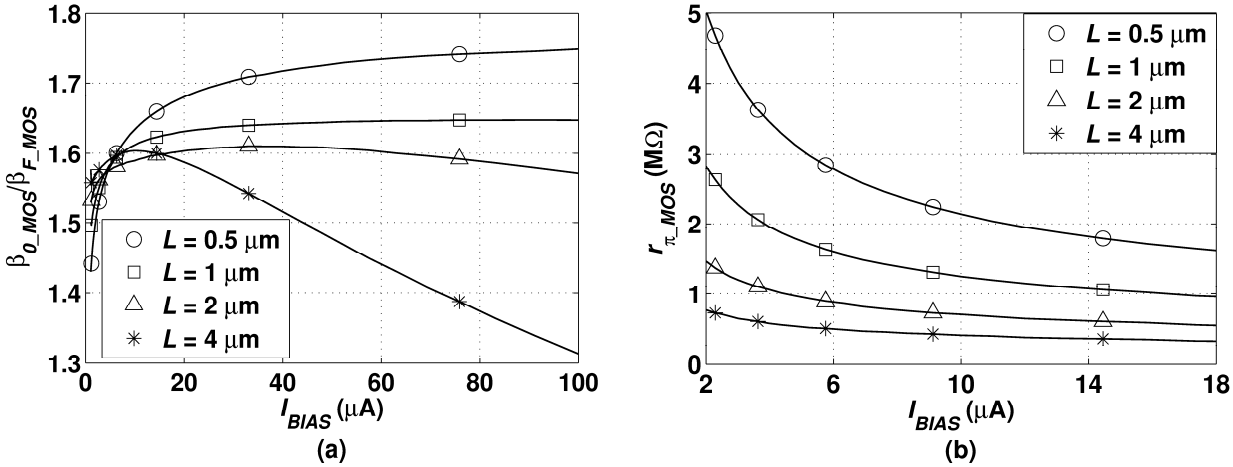

Figure 5.2: (a) $\beta_{0\_MOS}/\beta_{F\_MOS}$ vs. I_{BIAS}. (b) r_{π_MOS} vs. I_{BIAS}. Both graphs refer to the circuit shown in Figure 4.1 (a). Transistor area was held constant at 100 μm^2. The legends specify L.

Figure 5.2 (a) plots $\beta_{0\_MOS}/\beta_{F\_MOS}$ vs. I_{BIAS}. The results show that $\beta_{0\_MOS}/\beta_{F\_MOS}$ is greater than one over a wide range of bias currents and channel lengths. This implies that β_{0_MOS} and β_{F_MOS} are not equal, which demonstrates a difference between ultra-thin oxide MOSFETs and BJTs, where β_F ideally equals β_0. However, the plot shows that $\beta_{0\_MOS}/\beta_{F\_MOS}$ does not change significantly with changes in L and I_{BIAS}. For example, as I_{BIAS} increased from 5 μA to 80 μA (L = 1 μm), $\beta_{0\_MOS}/\beta_{F\_MOS}$ only changed 3.125% (1.6 to 1.65). This implies that β_{0_MOS} is typically greater than β_{F_MOS} and that their ratio remains relatively constant over a wide range of bias currents.

Figure 5.2 (b) plots r_{π_MOS} vs. I_{BIAS}. The results show that r_{π_MOS} is a strong function of L. For example, as L increased from 500 nm to 4 μm (I_{BIAS} = 4 μA), r_{π_MOS} decreased from 3.5 MΩ to 0.6 MΩ. This suggests that the effects of r_{π_MOS} may become important when designing with long-channel ultra-thin oxide MOSFETs. r_{π_MOS} is also a strong function of bias current. For example, as I_{BIAS} increased from 2 μA to 18 μA

(L = 500 nm), r_{π_MOS} decreased from 7.1 MΩ to 1.6 MΩ. Therefore, r_{π_MOS} generally increases with decreasing I_{BIAS} and decreasing L. This suggests that the effects of r_{π_MOS} can be minimized by using low-power short-channel devices.

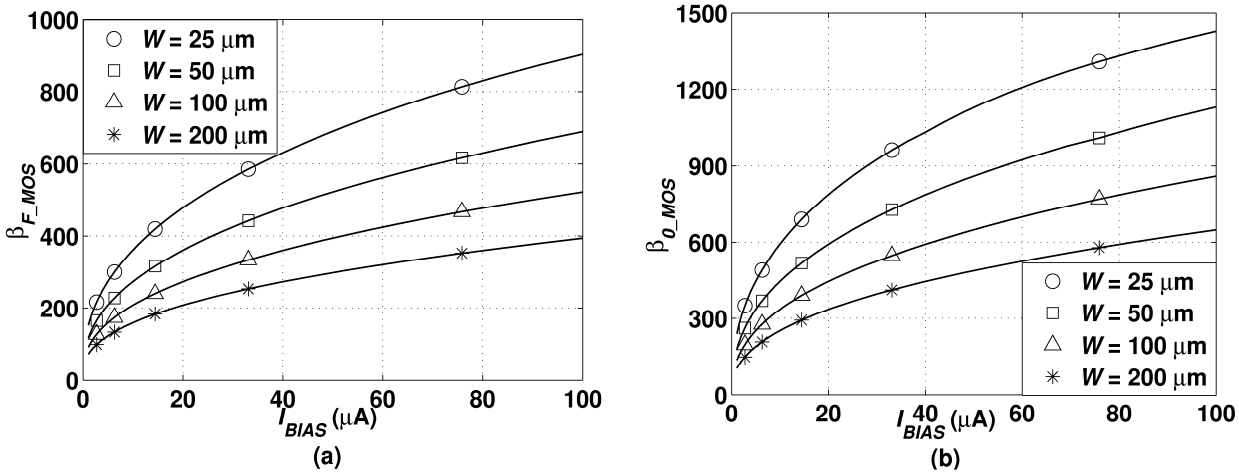

Figure 5.3: (a) β_{F_MOS} vs. I_{BIAS}. (b) β_{0_MOS} vs. I_{BIAS}. Both graphs refer to the circuit shown in Figure 4.1 (a). L = 1 µm in both graphs. The legends specify W.

The second scenario in which Figure 4.1 was simulated kept L constant at 1 µm while varying W and I_{BIAS}. This was done to determine the impact of W on β_{F_MOS} and β_{0_MOS}. The results are shown in Figure 5.3. The plots show that β_{F_MOS} and β_{0_MOS} generally increase with increasing I_{BIAS}. For example, in Figure 5.3 (a), β_{F_MOS} increased from 486 to 830 as I_{BIAS} increased from 20 µA to 80 µA (W = 25 µm). Likewise, in Figure 5.3 (b), β_{0_MOS} increased from 800 to 1330 as I_{BIAS} increased from 20 µA to 80 µA (W = 25 µm). These metrics increase with increases in I_{BIAS} because the device is approaching saturation and operating more like a MOSFET and less like a BJT.

Figure 5.3 also shows that β_{F_MOS} and β_{0_MOS} decrease with increasing W. For example, in Figure 5.3 (a), β_{F_MOS} decreased from 630 to 270 as W increased from 25 µm to 200 µm (I_{BIAS} = 40 µA). Likewise, in Figure 5.3 (b), β_{0_MOS} decreased from 1030 to 444 as W increased from 25 µm to 200 µm (I_{BIAS} = 40 µA). The reduction of β_{F_MOS} and

β_{0_MOS} with increases in W seems to disagree with what was claimed in [18], where β_{F_MOS} was shown to be relatively independent of W. This discrepancy may be due to the fact that constant current was used in Figure 5.3 whereas constant voltage was used in [18]. Increasing W with constant current results in a reduction of V_{BIAS} and smaller β_{F_MOS} values because the MOSFET (drift current) approaches the sub-V_{TH} region and begins to act like a BJT (diffusion current) [44]. Therefore, in current-mode circuits, β_{F_MOS} cannot be considered to be independent of W. However, increasing W with constant voltage, and maintaining saturation, results in increased power and relatively constant β_{F_MOS} and β_{0_MOS} values [18]. Therefore, the impact of W on β_{F_MOS} and β_{0_MOS} is a function of the type of design (current or voltage) being performed.

5.1.2 Impact of V_{DS} on Gate Current

The circuit in Figure 4.1 (b) was simulated to determine the impact of V_{DS} and I_{BIAS} on I_G, α_{F_MOS}, β_{F_MOS}, and r_{μ_MOS}. Transistor dimensions of $W = 100$ µm and $L = 1$ µm were chosen for this simulation. The results are shown in Figure 5.4 and Figure 5.5.

Figure 5.4 (a) plots I_G vs. V_{DS}. This plot shows that the directionality of I_G is a function of V_{DS} for small I_{BIAS} values. For example, as V_{DS} increased from 0.45 V to 1.0 V for $I_{BIAS} = 2$ µA, I_G decreased from 10 nA to −81 nA. This shows that the negative contributions of I_{GD} can be strong enough to change the direction of I_G. It also suggests that at a certain V_{DS} value, $I_G = 0$ and $\beta_{F\_MOS} \approx \infty$. However, to achieve this condition, a relatively large amount of voltage must be placed across the drain and source terminals of the device. In technologies with supply voltages of 1 V or less, increasing V_{DS} above

0.5 V to maximize β_{F_MOS} may not be practical. The plot also shows that the directionality of I_G remains constant (positive) as I_{BIAS} increases. For example, as V_{DS} increased from 0.15 V to 1.0 V for I_{BIAS} = 32 µA, I_G decreased from 98 nA to 30 nA. This suggests that I_G can be made unidirectional at the expense of power by designing with larger bias currents.

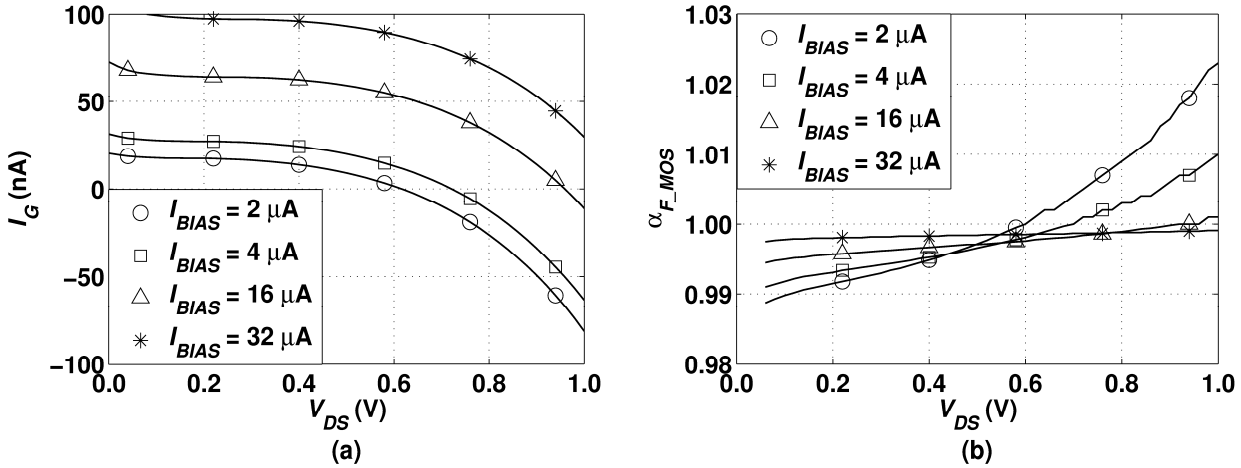

Figure 5.4: (a) I_G vs. V_{DS}. (b) α_{F_MOS} vs. V_{DS}. Both graphs refer to the circuit shown in Figure 4.1 (b). L = 1 µm and W = 100 µm for both graphs. The legends specify I_{BIAS}.

The impact of I_G's bi-directionality is shown in Figure 5.4 (b), which plots α_{F_MOS} vs. V_{DS}. In BJTs, α_F is typically less than one. However, as shown in Figure 5.4 (b), α_{F_MOS} can be greater than one. For example, as V_{DS} increased from 0.2 V to 1.0 V for I_{BIAS} = 4 µA, α_{F_MOS} increased from 0.99 to 1.01. This demonstrates a difference between ultra-thin oxide MOSFETs and BJTs. This difference only occurs at relatively small bias currents. Therefore, to avoid the bi-directionality of I_G, I_{BIAS} should be increased such that the positive contributions of I_{GCS}, I_{GCD}, and I_{GS} dominate the negative contribution of I_{GD}.

Figure 5.5 (a) plots β_{F_MOS} vs. V_{DS}. The plot shows that β_{F_MOS} increases with increasing V_{DS}. For example, as V_{DS} increased from 0.1 V to 0.4 V for I_{BIAS} = 2 µA,

β_{F_MOS} increased from 97 to 188. The increases in β_{F_MOS} with increasing V_{DS} can be explained by reduced r_O and the increasing negative contributions of I_{GD}. Specifically, as V_{DS} increases, MOSFET output resistance generally decreases (larger I_D) and I_G generally decreases due to the increasing negative contributions of I_{GD}. The plot also shows that β_{F_MOS} increases with increasing I_{BIAS}. For example, as I_{BIAS} increased from 2 µA to 16 µA for V_{DS} = 0.2 V, β_{F_MOS} increased from 116 to 233. The increases in β_{F_MOS} with increasing I_{BIAS} occur because of reduced r_O and because the device tends to operate more like a MOSFET and less like a BJT.

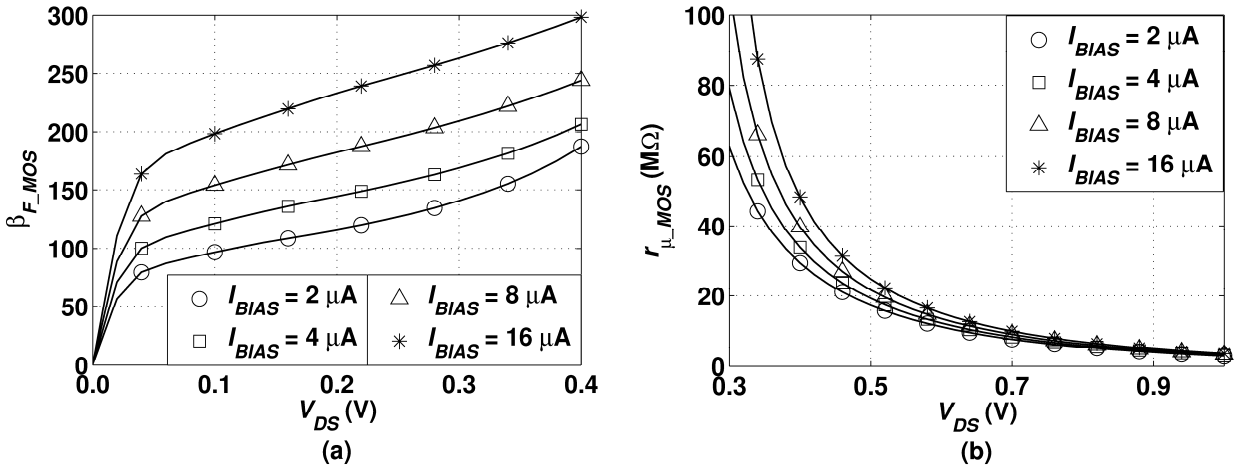

Figure 5.5: (a) β_{F_MOS} vs. V_{DS}. (b) r_{μ_MOS} vs. V_{DS}. Both graphs refer to the circuit shown in Figure 4.1 (b). L = 1 µm and W = 100 µm for both graphs. The legends specify I_{BIAS}.

Figure 5.5 (b) plots r_{μ_MOS} vs. V_{DS}. In general, the results show that r_{μ_MOS} is large enough to be considered negligible in most applications. For example, as V_{DS} increased from 0.4 V to 0.8 V for I_{BIAS} = 2 µA, r_{μ_MOS} decreased from 30 MΩ to 5 MΩ. These small-signal resistance values are generally much larger than anything they would be in parallel with. Therefore, the effects of r_{μ_MOS} can generally be assumed negligible in ultra-thin oxide analog CMOS design.

5.1.3 Channel Length Selection Methodology

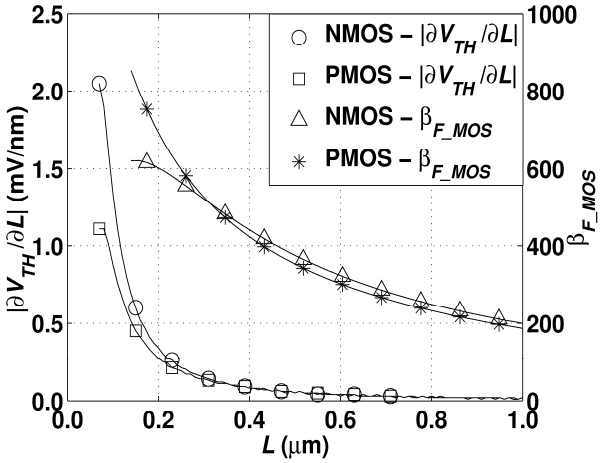

Figure 5.6: Simulated $|\partial V_{TH}/\partial L|$ vs. L and β_{F_MOS} vs. L for NMOS and PMOS transistors with $W \cdot L = 100\ \mu m^2$ and $I_D = 10\ \mu A$.

The preceding analysis has shown that gate current is a strong function of channel length. The use of long-channel ultra-thin oxide devices is generally restricted because β_{F_MOS} is roughly proportional to $1/L^2$. This proportionality suggests L should be set to the process minimum. However, this is not practical for several reasons. First, L is typically increased to improve r_o-degrading effects such as drain-induced barrier lowering and channel length modulation [44]. Second, due to the halo implant, the threshold voltage, V_{TH}, rapidly increases as L decreases [50], [52]. Therefore, for a given I_D and MOSFET aspect ratio ($A_R \equiv W/L$), operating at smaller channel lengths increases V_{TH} and the required gate-to-source voltage, V_{GS}, to supply the drain current. This limits voltage headroom, which is a major concern in technologies with $V_{DD} \leq 1$ V [18]. Third, the rapid increases in V_{TH} caused by the halo implant limits achievable matching [95]. For example, consider Figure 5.6 which plots $|\partial V_{TH}/\partial L|$ vs. L and β_{F_MOS} vs. L for NMOS and PMOS devices in the obtained 65 nm technology. As L approaches the process minimum, $|\partial V_{TH}/\partial L|$ becomes exponential-like and approaches a maximum value of 2 mV/nm in NMOS devices. Operating on the exponential-like portion of this curve

exacerbates mismatch because small differences in L result in significant differences in V_{TH} [95].

The previous paragraph suggests that a minimum analog channel length, L_{MIN_A}, is needed to balance gate current with r_o-degradations, reduced supply voltages, and mismatch. The ITRS defines L_{MIN_A} as $5 \cdot L_{MIN}$, where L_{MIN} is the process minimum [17]. This approach yields a value of $L_{MIN\_A} = 250$ nm in the obtained technology ($L_{MIN} = 50$ nm). Referring to Figure 5.6, this can be validated by observing that $|\partial V_{TH}/\partial L|$ is approximately 200 µV/nm for both devices. It can also be seen that β_{F_MOS} is relatively large, approximately 580 for both devices. Therefore, for traditional ultra-thin oxide CMOS technologies, an L_{MIN_A} value in the 200 nm to 300 nm range helps reduce the impact of r_o-degradations, reduced supply voltages, and matching limitations while still allowing for relatively large β_{F_MOS} values.

The restriction of long-channel devices stems from β_{F_MOS} being proportional to $1/L^2$. This proportionality suggests that a maximum analog channel length, L_{MAX_A}, is needed to prevent extremely small β_{F_MOS} values. One approach is to restrict β_{F_MOS} to a minimum value, $\beta_{F\_MOS\_MIN}$. For example, assuming that I_D and the device area are known from matching considerations, L can be increased until $\beta_{F\_MOS} = \beta_{F\_MOS\_MIN}$. The channel length at which this equality occurs represents L_{MAX_A}.

Figure 5.7 plots L_{MAX_A} vs. I_D for NMOS and PMOS devices with an area of 100 µm^2. A $\beta_{F\_MOS\_MIN}$ value of 100 was chosen. The results show that L_{MAX_A} increases as I_D increases for both devices. For example, the NMOS L_{MAX_A} changed from 1 µm to 3.3 µm as I_D changed from 2 µA to 64 µA. One possible explanation for this behavior is

as follows. For a relatively small drain current, the device operates in the weak inversion region. In this region, MOSFETs function similar to BJTs because they are dominated by diffusion current [44]. In traditional ultra-thin oxide CMOS technologies, this BJT-like behavior is more-pronounced because MOSFET gate current is somewhat similar to BJT base current [18]. Therefore, for a given L, β_{F_MOS} will be smaller for a MOSFET operated in the weak inversion region (small I_D) compared to a MOSFET operated in the strong inversion region (large I_D) because it acts more like a BJT in the weak inversion region. Of course, for either region, β_{F_MOS} decreases with increases in L. To increase L_{MAX_A} and generally avoid operating in the weak and moderate inversion regions, I_D can be increased or device area can be decreased. However, both of these approaches should be weighed against power limitations, voltage headroom, and matching requirements.

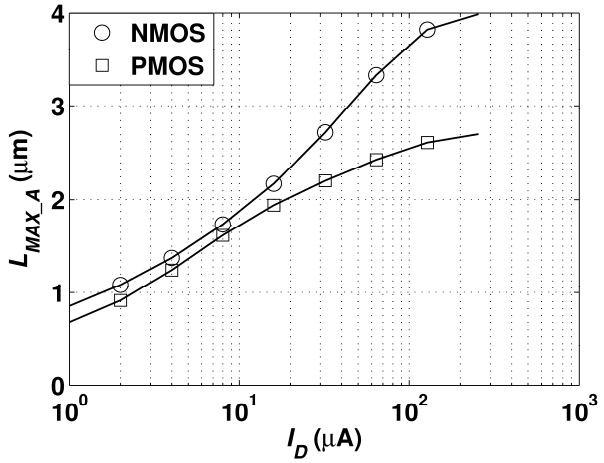

Figure 5.7: Simulated L_{MAX} vs. I_D for NMOS and PMOS transistors with $W \cdot L$ = 100 μm^2 for $\beta_{F\_MOS\_MIN}$ = 100.

Figure 5.7 also shows that the L_{MAX_A} of the PMOS device is consistently shorter than the L_{MAX_A} of the NMOS device. One possible explanation for this stems from differences in $|V_{GS}|$. For example, assuming a constant I_D and equal device dimensions, $|V_{GSP}|$ could be greater than V_{GSN} because of differences in threshold voltage ($|V_{THP}|$ >

V_{THN}) or channel mobility ($\mu_n > \mu_p$). Gate current is a strong function of $|V_{GS}|$ and a weak function of threshold voltage and channel mobility [13], [136]. Therefore, for a given I_D, $|I_{GP}|$ will be larger than I_{GN} ($\beta_{F\_MOS\_P} < \beta_{F\_MOS\_N}$) because $|V_{GSP}| > V_{GSN}$. As L increases, $\beta_{F\_MOS\_P}$ will approach $\beta_{F\_MOS\_MIN}$ quicker than $\beta_{F\_MOS\_N}$ because $|I_{GP}| > I_{GN}$. This results in the L_{MAX_A} of the PMOS device being shorter than the L_{MAX_A} of the NMOS device.

5.2 Impact of Body Biasing on Gate Current

This section presents simulation results that characterize the impact of body biasing on gate current. It is broken into two subsections. The first subsection presents the results for constant terminal voltages (Figure 4.2). The second subsection presents the results for constant drain current (Figure 4.3).

5.2.1 Constant Terminal Voltages

The circuit in Figure 4.2 was simulated to determine the impact of V_{BS} on I_G when a MOSFET is under constant terminal voltages. With respect to an NMOS transistor under constant terminal voltage conditions, increases in V_{BS} decrease V_{TH} and therefore increase I_D. I_G is not a strong function of V_{TH} [13], [136]. Therefore, increasing V_{BS} yields larger β_{F_MOS} values because I_D increases and I_G remains relatively constant. For example, consider Figure 5.8, which plots β_{F_MOS} vs. $|V_{BS}|$ and the percent reduction in I_G vs. $|V_{BS}|$ for an NMOS transistor and a PMOS transistor under constant terminal voltages. Both devices were designed with W = 100 µm and L = 1 µm. Note that V_{BS} of the NMOS device and V_{SB} of the PMOS device were both kept greater than 0 V. The results show that β_{F_MOS} increases significantly with increases in $|V_{BS}|$. For example, β_{F_MOS} increased from approximately 240 to 1200 for both devices as $|V_{BS}|$ was swept from 0 V to 0.5 V. Note that $|V_{BS}|$ was not swept above this voltage to avoid forward-biasing the

body-to-source diode. The increases in β_{F_MOS} were not caused by significant reductions in I_G. For example, Figure 5.8 (b) shows that I_G was reduced by a maximum of 10% for both devices across the entire voltage range. This small decrease can mostly likely be attributed to the dependence of the probability of direct tunneling on the gate-to-body voltage, V_{GB} [13]. Therefore, the improvements in β_{F_MOS} can be mostly attributed to significant increases in I_D.

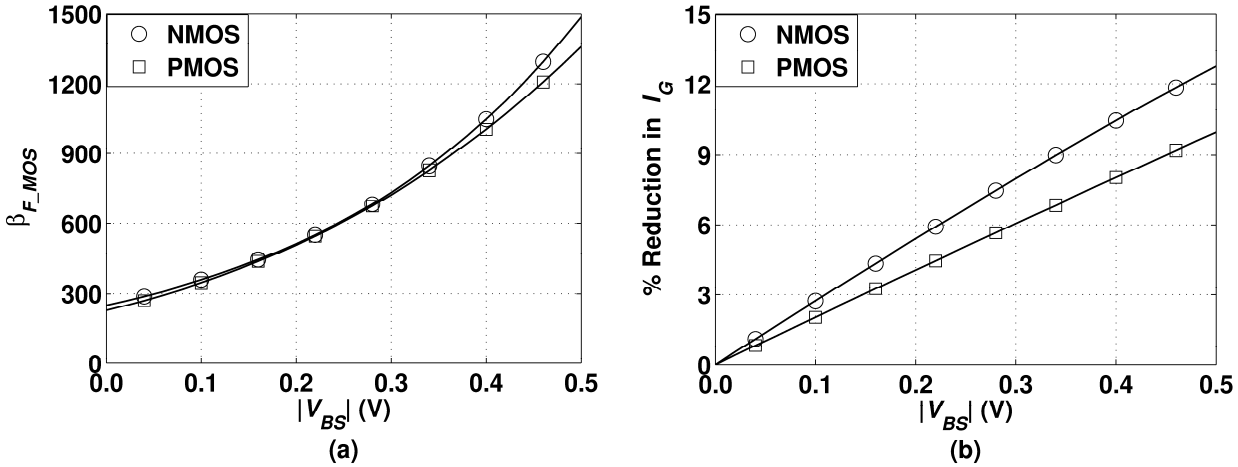

Figure 5.8: Simulated (a) β_{F_MOS} vs. $|V_{BS}|$ and (b) percent reduction in I_G vs. $|V_{BS}|$ for an NMOS transistor and a PMOS transistor under a constant voltage condition. Each transistor was sized with W = 100 µm and L = 1 µm and had an I_D of 16 µA at $|V_{BS}|$ = 0 V. V_{BS} of the NMOS device and V_{SB} of the PMOS device were both kept greater than 0 V.

One potential application of the constant terminal voltage condition is forward body-biased transistors. Forward body biasing is used in digital circuits to reduce critical path delay [88]. Along with reducing delay, Figure 5.8 suggests it also helps reduce the relative impact of gate current.

5.2.2 Constant Drain Current

The circuit in Figure 4.3 was simulated to determine the impact of V_{BS} on I_G for a MOSFET with constant drain current. With respect to an NMOS transistor under constant drain current conditions, increases in V_{BS} decrease V_{TH} and thus reduce the V_{GS}

value needed to supply I_D. I_G is a strong function of V_{GS} [13], [136]. Therefore, increasing V_{BS} yields larger β_{F_MOS} values because I_G decreases with reductions in V_{GS}. For example, Figure 5.9 plots β_{F_MOS} vs. $|V_{BS}|$ and the percent reduction in I_G vs. $|V_{BS}|$ for an NMOS transistor and a PMOS transistor under constant drain current. Both transistors were designed with W = 100 µm and L = 1 µm. The results show β_{F_MOS} values similar to those of the constant voltage condition of Figure 5.8. However, the increases in β_{F_MOS} are not caused by increases in I_D. Instead, they are caused by significant reductions in I_G. For example, as $|V_{BS}|$ was swept from 0 V to 0.5 V, I_G was reduced by approximately 80% for both devices. Therefore, the improvements in β_{F_MOS} can be attributed to significant reductions in I_G.

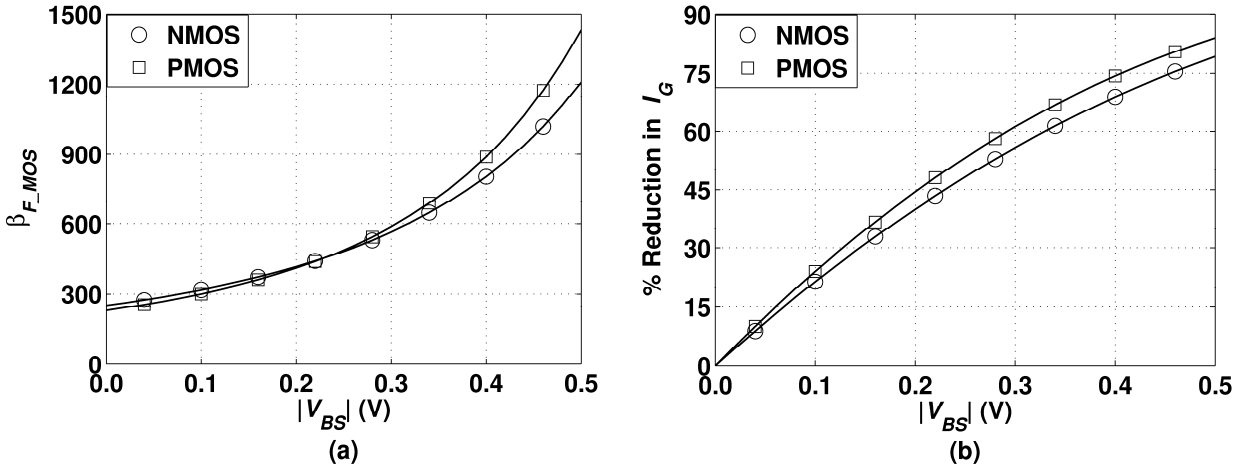

Figure 5.9: Simulated (a) β_{F_MOS} vs. $|V_{BS}|$ and (b) percent reduction in I_G vs. $|V_{BS}|$ for an NMOS transistor and a PMOS transistor under a constant current condition. Each transistor was sized with W = 100 µm and L = 1 µm and had an I_D of 16 µA at $|V_{BS}|$ = 0 V. V_{BS} of the NMOS device and V_{SB} of the PMOS device were both kept greater than 0 V.

One potential application of the constant drain current condition is the input differential pair of an amplifier. For example, MOSFET input pairs often have their body terminals tied to a power or ground and their source terminals tied to the output node of a current mirror. By tying their body and source terminals together, the total amount of

gate current flowing through an input pair can be significantly reduced, resulting in less amplifier input current. The only downside of this technique is that the input pair must be placed in a separate well.

5.3 The Design of Ultra-Thin Oxide CMOS Current Mirrors

This section presents simulation results of the current mirror techniques described in Section 4.4. It is broken into four subsections. The first subsection presents a current mirror comparison. The second subsection presents the results of self-cascode current mirrors. The third subsection presents the results of self-cascode current mirrors with a helper transistor. The fourth subsection presents the results of triple self-cascode current mirrors.

5.3.1 Current Mirror Comparison

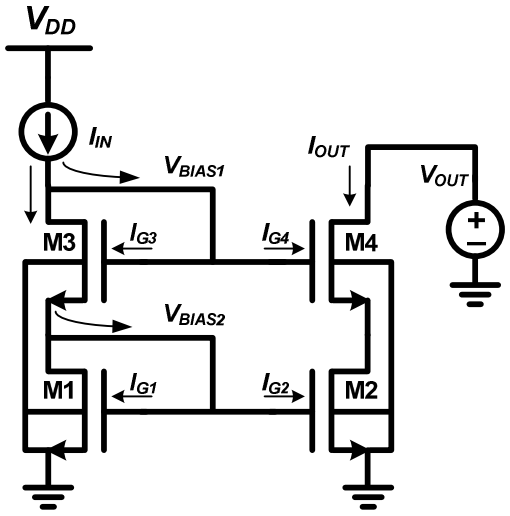

Figure 5.10: Basic Cascode Current Mirror. V_{DD} is the supply voltage, I_{IN} is the input current, V_{OUT} is the output voltage, I_{OUT} is the output current. V_{BIAS1} is the gate-bias voltage of M3 and M4. V_{BIAS2} is the gate-bias voltage of M1 and M2. M1-M4 form the basic cascode current mirror.

The impact of gate current on current mirrors was investigated by simulating a simple current mirror (Figure 3.17), a cascode current mirror (Figure 5.10), and a self-cascode current mirror (Figure 4.4). For all three mirrors, I_{IN} was set to 2 μA and the

desired current gain was $A_i = 1$. The transistor dimensions for the simple and cascode mirrors were $W = 10$ µm and $L = 10$ µm. The self-cascode current mirror was designed using self-cascode structures where the devices being cascoded had $W = 10$ µm and $L = 10$ µm and the cascoding devices had $W = 30$ µm and $L = 3.33$ µm. The results are shown in Figure 5.11. Figure 5.11 (a) plots A_i vs. V_{OUT} for all three mirrors. The results show that the desired current gain was not achieved by any of the mirrors. For example, the current gain of the simple current mirror went from 0.69 to 0.95 as V_{OUT} increased from 0.2 V to 1.0 V. This was expected considering the simple current mirror relies on single devices that exhibit poor output resistance.

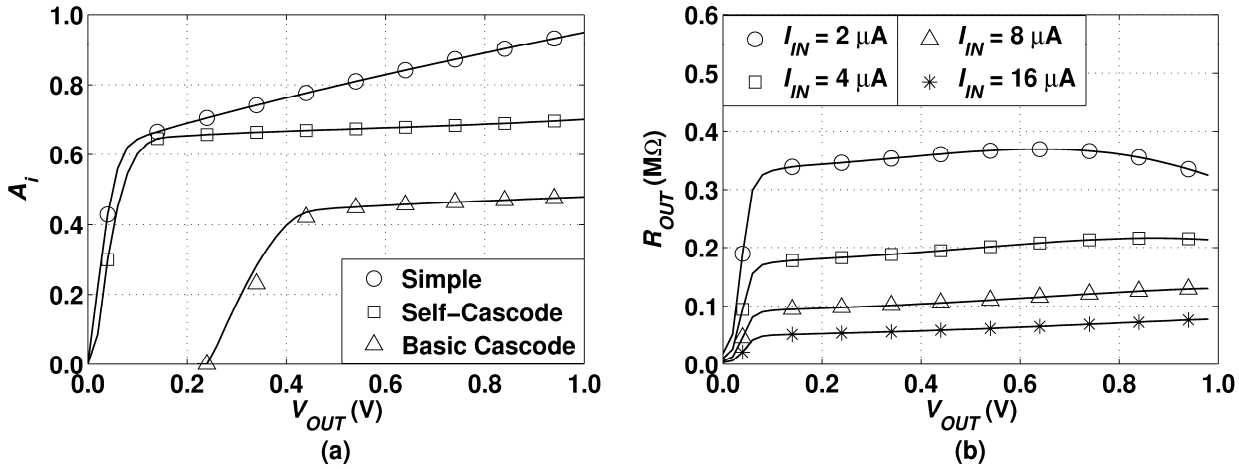

Figure 5.11: (a) A_i vs. V_{OUT} for the three types of current mirrors noted in the legend (I_{IN} = 2 µA). W = 10 µm and L = 10 µm for all devices in the simple and basic cascode current mirrors. The cascoded devices of the self-cascode current mirror were designed with W = 10 µm and L = 10 µm. The cascoding devices of the self-cascode current mirror were designed with W = 30 µm and L = 3.33 µm (b) R_{OUT} vs. V_{OUT} for a simple current mirror with W = 10 µm and L = 10 µm. The legend specifies I_{IN}.

Figure 5.11 (b) plots R_{OUT} vs. V_{OUT} for the simple current mirror for four different I_{IN} values. The results show that the output resistance of the simple current mirror was never greater than 400 kΩ for all simulated values of I_{IN}. These results quantify the poor output resistance of single transistors in nanoscale CMOS technologies.

Figure 5.11 (a) shows that the current gain of the basic cascode current mirror saturated at approximately 0.4. The current gain saturated at this value because I_{IN} supplied significant gate current to four relatively large transistors. It took approximately 400 mV across the current mirror to achieve this saturation. Considering that V_{DD} = 1 V, this may be too much voltage headroom to spend on a current mirror. These results explain why basic cascode structures are generally avoided in nanoscale CMOS technologies.

Figure 5.11 (a) shows that the current gain of the self-cascode current mirror saturated at approximately 0.6. However, unlike the cascode current mirror, it only took 150 mV across V_{OUT} to achieve this saturation. This was a significant improvement over the basic cascode current mirror and suggested that a reliable current mirror could be designed if the gain degradations caused by gate current could be overcome.

5.3.2 Self-Cascode Current Mirrors

To reduce the impact of gate current on the self-cascode current mirror of Figure 4.4, transistors M1-M4 should be sized such that their gate current is minimized. This can be accomplished using the channel length selection methodology outlined Section 5.1.3. For example, assuming I_{IN}, I_{OUT}, and the area needed for M1 and M2 to meet matching requirements are known, L_1-L_2 can be set equal to L_{MAX_A} and L_3-L_4 can be set equal to L_{MIN_A}. Setting L_3 and L_4 equal to L_{MIN_A} helps minimize the gate current of M3 and M4 and also increases S_{F3} and S_{F4}, which allows the current mirror to provide high output resistances at low output voltages. The only unknowns with this approach are W_3 and W_4, which can be used to set S_{F3} and S_{F4}.

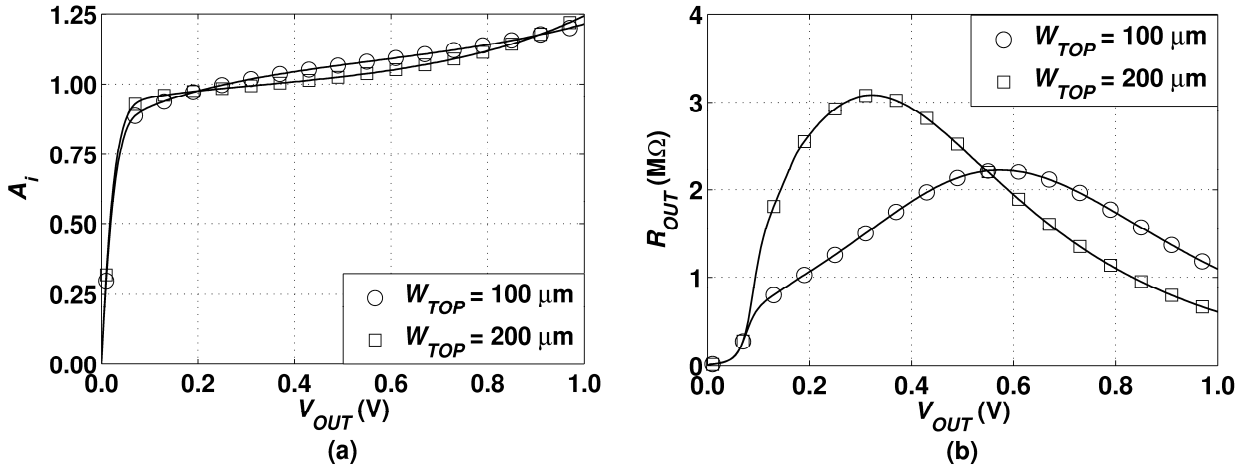

Figure 5.12: (a) A_i vs. V_{OUT} and (b) R_{OUT} vs. V_{OUT} for a self-cascode current mirror with I_{IN} = 2 µA. Both graphs refer to Figure 4.4. The cascoded devices were designed with W = 100 µm and L = 1 µm. The cascoding devices were designed with L = 0.25 µm. The legends specify the width of the cascoding transistors.

Figure 5.12 plots A_i vs. V_{OUT} and the output resistance, R_{OUT}, vs. V_{OUT} for an NMOS self-cascode current mirror with a desired unity current gain and an I_{IN} of 2 µA. The cascoded transistors of the mirror had an area of 100 µm^2 and a channel length of L_{MAX_A} ($L_{MAX\_A\_2\mu A}$ = 1 µm). The cascoding devices of the mirror were sized with $L = L_{MIN\_A}$ = 0.25 µm. The width of the cascoding device, W_{TOP}, was a variable. W_{TOP} values of 100 µm and 200 µm were simulated. The mirror with a W_{TOP} value of 200 µm had an S_F value of 8 while the mirror with a W_{TOP} value of 100 µm had an S_F value of 4. The results show that strategically sized self-cascode current mirrors are capable of minimizing the impact of I_G on the current gain under relatively small currents while still producing high output resistances at low output voltages. For example, the output resistance of the mirror with W_{TOP} = 100 µm reached a value of 1 MΩ at V_{OUT} = 0.1 V. Its current gain was within 5% of the desired value for 0.1 V ≤ V_{OUT} ≤ 0.6 V. The mirror achieved a peak output resistance of 3.1 MΩ at V_{OUT} = 0.33 V. W_{TOP} had a noticeable impact on the output resistance. For example, the difference in output resistance between the two W_{TOP} values was 1.6 MΩ at V_{OUT} = 0.2 V. This suggests that at relatively small

input currents, large output resistances with minimal voltage overhead can be obtained by increasing S_F via the width of the cascoding transistor.

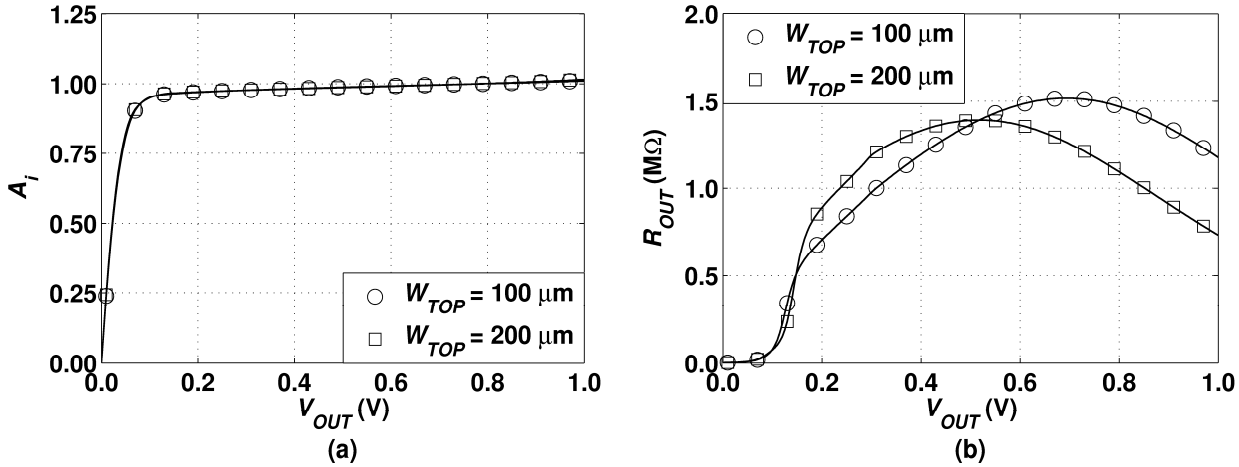

Figure 5.13: (a) A_i vs. V_{OUT} and (b) R_{OUT} vs. V_{OUT} for a self-cascode current mirror with I_{IN} = 16 µA. Both graphs refer to Figure 4.4. The cascoded devices were designed with W = 100 µm and L = 1 µm. The cascoding devices were designed with L = 0.25 µm. The legends specify the width of the cascoding transistors.

Figure 5.13 plots A_i vs. V_{OUT} and the output resistance, R_{OUT}, vs. V_{OUT} for an NMOS self-cascode current mirror with a desired unity current gain and an I_{IN} of 16 µA. The cascoded devices had an area of 100 µm^2 and a channel length of L_{MAX_A} ($L_{MAX\_A\_16\mu A}$ = 2 µm). The cascoding devices of were sized with $L = L_{MIN\_A}$ = 0.25 µm. The width of the cascoding device, W_{TOP}, was a variable. W_{TOP} values of 100 µm and 200 µm were simulated. The mirror with a W_{TOP} value of 200 µm had an S_F value of 32 while the mirror with a W_{TOP} value of 100 µm had an S_F value of 16. The results show that strategically sized self-cascode current mirrors are capable of minimizing the impact of I_G on the current gain under relatively large current conditions while producing high output resistances at low output voltages. For example, the output resistance of the mirror with W_{TOP} = 200 µm reached a value of 1.39 MΩ at V_{OUT} = 0.5 V. Its current gain was within 5% of the desired value for 0.1 V ≤ V_{OUT} ≤ 1 V. The impact of W_{TOP} on performance was not as noticeable in Figure 5.13. For example, the difference in output

resistance between the two W_{TOP} values was only 180 kΩ at V_{OUT} = 0.2 V. This suggests that at relatively large input currents, S_F can be reduced by decreasing the width of the cascoding transistor without a significant impact on current mirror performance.

One concern with the architecture of Figure 4.4 is the bi-directionality of I_{G4}. Ideally, I_{G4} flows into the gate of M4 and is supplied by I_{IN}. However, if the gate-to-drain voltage of M4, V_{GD4}, is large and negative, I_{G4} could flow out of the gate of M4 [17]. This is caused by the gate-to-drain overlap current, I_{GD4}, which is a strong function of V_{GD4} and it suggests I_{OUT} is directly supplying I_{G4} and indirectly supplying some of I_{G1}-I_{G3} [13], [136]. This could potentially degrade R_{OUT} as V_{OUT} increases because V_{OUT} would be supplying an undesired current. For example, consider the 2 µA self-cascode current mirror of Figure 5.12, where A_i increased by 0.19 and R_{OUT} decreased by 1.3 MΩ as V_{OUT} increased from 0.6 V to 1.0 V. These degradations were caused by I_{OUT} directly supplying I_{G4} and indirectly supplying some of I_{G1}-I_{G3}. To avoid this problem, I_{IN} can be chosen large enough such that I_{G4} is always supplied by I_{IN} or V_{OUT} can be restricted to a voltage range where I_{G4} is always supplied by I_{IN}.

A self-cascode current mirror (Figure 4.4) was compared to a simple current mirror (Figure 3.17) to illustrate the output resistance enhancements that can be obtained by following the channel length selection methodology of Section 5.1.3. The simple current mirror was designed using W = 100 µm and L = 1 µm. The self-cascode current mirror was designed using self-cascode structures where the devices being cascoded had W = 100 µm and L = 1 µm and the cascoding devices had W = 100 µm and $L = L_{MIN\_A}$ = 0.25 µm. The desired current gain was A_i = 1. The results are shown in

Figure 5.14 for I_{IN} values of 2 µA and 16 µA. Figure 5.14 (a) plots R_{OUT} vs. V_{OUT} for the self-cascode current mirror. The plot shows that the self-cascode current mirror achieves relatively high output resistances across a wide voltage range. For example, the 2 µA self-cascode current mirror had an output resistance greater than 1 MΩ for $0.2\ \text{V} \leq V_{OUT} \leq 1.0\ \text{V}$. The 16 µA self-cascode current mirror had an output resistance greater than 1 MΩ for $0.31\ \text{V} \leq V_{OUT} \leq 1.0\ \text{V}$.

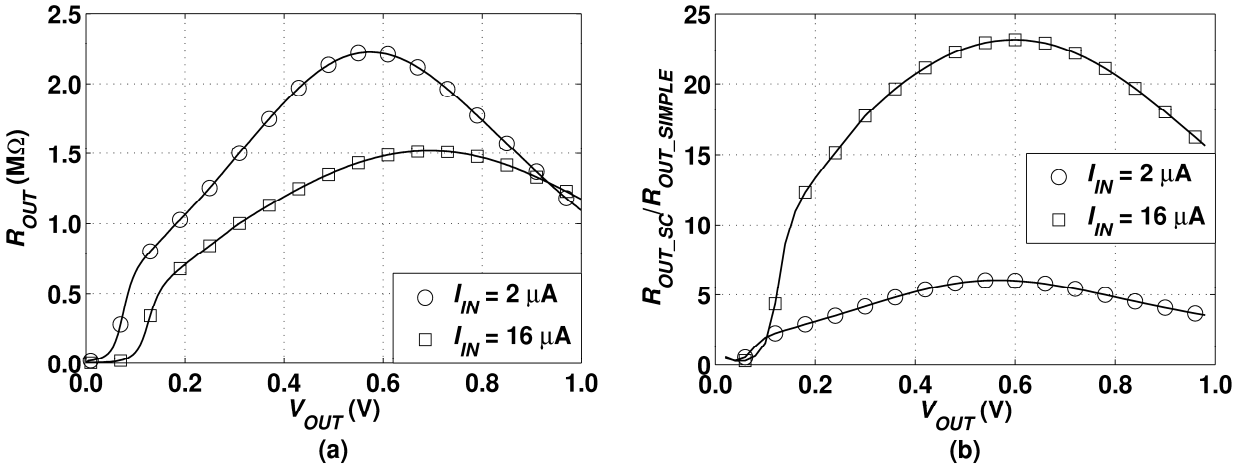

Figure 5.14: (a) R_{OUT} vs. V_{OUT} for the self-cascode current mirror of Figure 4.4. The cascoded devices were designed with W = 100 µm and L = 1 µm. The cascoding devices were designed with W = 100 µm and L = 0.25 µm. The legend specifies I_{IN}. (b) $R_{OUT\_SC}/R_{OUT\_SIMPLE}$ vs. V_{OUT}. The simple current mirror was designed with W = 100 µm and L = 1 µm. The legend specifies I_{IN}.

Figure 5.14 (b) plots the ratio of output resistances between the two mirrors vs. V_{OUT}. The plot shows that the self-cascode current mirror is capable of consistently providing 5-to-10 times the output resistance of a simple current mirror across a wide voltage range. For example, the 2 µA self-cascode current mirror had an output resistance at least five times that of the simple current mirror for $0.38\ \text{V} \leq V_{OUT} \leq 0.77\ \text{V}$. The 16 µA self-cascode current mirror had an output resistance at least ten times that of the simple current mirror for $0.15\ \text{V} \leq V_{OUT} \leq 1.0\ \text{V}$. These results suggest that

self-cascode current mirrors represent a desirable low-voltage alternative to simple current mirrors in ultra-thin oxide technologies.

5.3.3 Self-Cascode Current Mirrors with a Helper Transistor

Proper sizing and biasing may not always be enough to overcome the current gain degradations of (4.4). For example, the channel length selection methodology described in Section 5.1.3 may fail if the desired current gain is greater than one or if channel lengths longer than L_{MAX_A} are used for M1 and M2 of Figure 4.4.

As the desired current gain increases, the widths of M2 and M4 are scaled to be A_i times larger than M1 and M3 ($W_4 = A_i \cdot W_3$, $W_2 = A_i \cdot W_1$). Therefore, as A_i increases, I_{G2} and I_{G4} will increase because of the increases in area of M2 and M4 ($I_G \propto W \cdot L$). This will cause more of I_{IN} to flow into the gates of M2 and M4, thus further degrading the current gain.

Assuming constant area, β_{F_MOS1} and β_{F_MOS2} will decrease if channel lengths longer than L_{MAX_A} are used for M1 and M2 in Figure 4.4 ($\beta_{F\_MOS} \propto 1/L^2$) [19]–[20]. This will cause I_{G1} and I_{G2} to increase and thus degrade A_i. One possible solution to these problems is shown in Figure 4.5. This figure is similar to Figure 4.4 except for the addition of a helper transistor, M5. This additional transistor is used to supply gate current to M1-M4. A similar technique has been applied using BJTs [44]. Assuming that M5 is relatively small, its gate current is negligible. This forces all of I_{IN} into the drain of M3 and implies that I_{OUT} will mirror I_{IN} because of the high output resistance provided by the self-cascode structures.

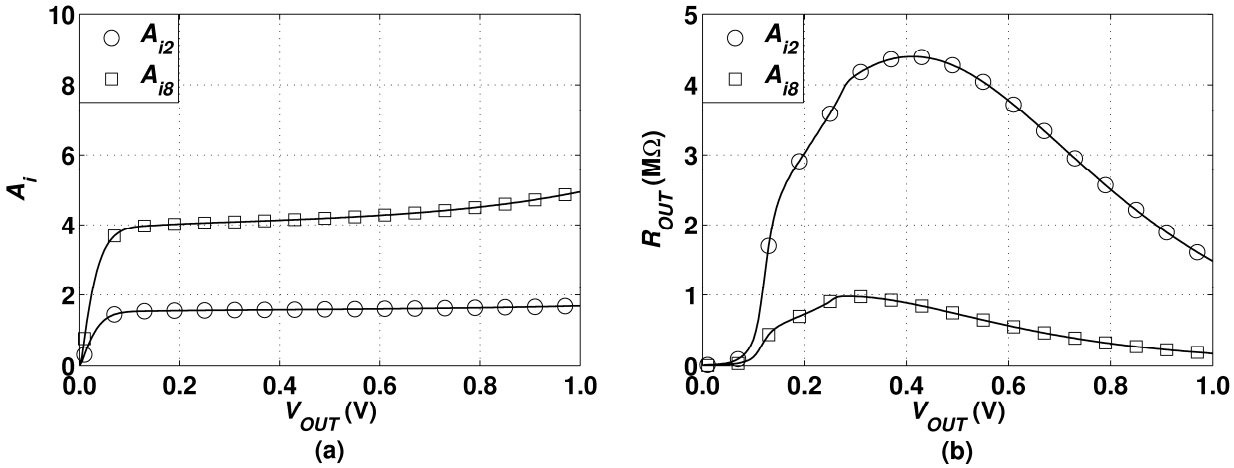

Figure 5.15: (a) A_i vs. V_{OUT} and (b) R_{OUT} vs. V_{OUT} for the self-cascode current mirror of Figure 4.4. The cascoded devices were designed with W = 20 µm and L = 5 µm. The cascoding devices were designed with W = 40 µm and L = 1.25 µm. I_{IN} was 2 µA. The legends specify the desired current gain.

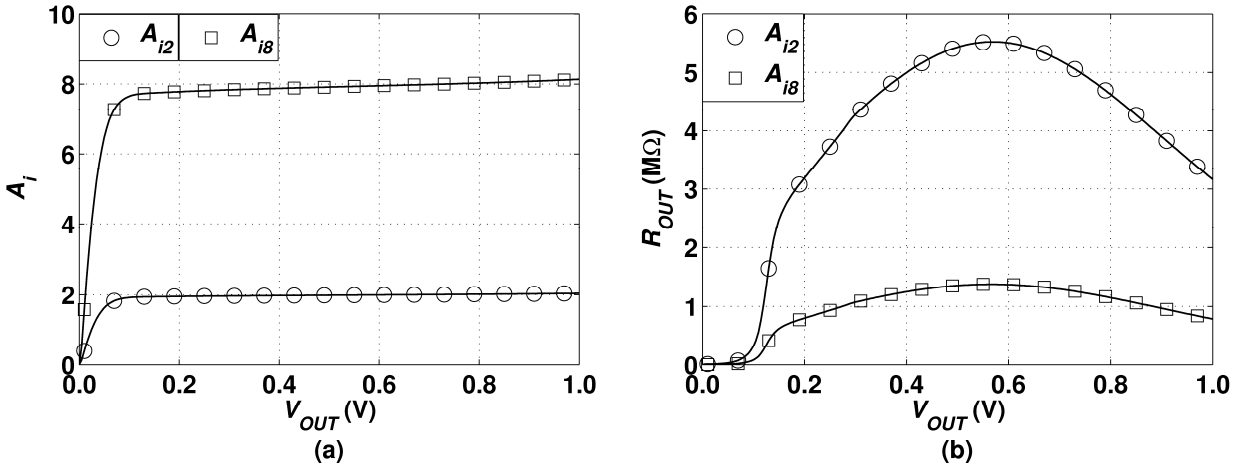

Figure 5.16: (a) A_i vs. V_{OUT} and (b) R_{OUT} vs. V_{OUT} for the self-cascode current mirror of Figure 4.5. The cascoded devices were designed with W = 20 µm and L = 5 µm. The cascoding devices were designed with W = 40 µm and L = 1.25 µm. I_{IN} was 2 µA. The helper transistor was designed with W = 5 µm, L = 0.5 µm. The legends specify the desired current gain.

Figure 5.15 and Figure 5.16 plot A_i vs. V_{OUT} and R_{OUT} vs. V_{OUT} for four self-cascode current mirrors: two without a helper transistor (Figure 4.4, Figure 5.15) and two with a helper transistor (Figure 4.5, Figure 5.16). For all four mirrors, I_{IN} was 2 µA, and the MOSFETs were sized as follows: $L_1 = L_2 = 5$ µm, $W_1 = 20$ µm, $W_2 = A_i \cdot 20$ µm, $L_3 = L_4 = 1.25$ µm, $W_3 = 40$ µm, $W_4 = A_i \cdot 40$ um, $L_5 = 0.5$ µm, and $W_5 = 5$ µm. Target A_i values of 2 and 8 were chosen. Figure 5.15 (a) shows that the current gain was significantly lower than its desired value for both mirrors without a helper transistor. For

example, the helper-less mirror with a desired current gain of 2 achieved a maximum gain of 1.69 and the mirror with a desired current gain of 8 achieved a maximum gain of 4.95. Figure 5.16 (a) shows that both mirrors with a helper transistor were within 5% of their target gain value for 0.1 V $\leq V_{OUT} \leq$ 1 V. With respect to output resistance, Figure 5.16 (b) shows that R_{OUT} of the mirrors with a helper transistor was larger than those without a helper transistor. For example, the mirror with a helper transistor and desired current gain of 2 had an output resistance at least 0.5 MΩ greater than that of the helper-less mirror for 0.39 V $\leq V_{OUT} \leq$ 1 V. For the mirrors with a desired current gain of 8, the mirror with a helper transistor had an R_{OUT} greater than 1 MΩ for 0.28 V $\leq V_{OUT} \leq$ 0.88 V, while the helper-less mirror never achieved an R_{OUT} of 1 MΩ.

5.3.4 Triple Self-Cascode Current Mirrors

The triple self-cascode current mirror of Figure 4.6 was simulated to determine the impact of an extra self-cascode on mirror performance. The triple self-cascode current mirror was designed with W = 100 μm for all transistors. The bottom transistors had channel lengths of 1 μm, the middle transistors had channel lengths of 0.5 μm, and the top transistors had channel lengths of 0.25 μm. The helper transistor was designed with W = 5 μm and L = 0.5 μm.

The results are shown in Figure 5.17 for input currents of 2 μA, 4 μA, 8 μA, and 16 μA. Figure 5.17 (a) plots A_i vs. V_{OUT}. The plot shows that the current gain of the triple self-cascode current mirror was within 5% of its target value across the four different input currents for 0.28 V $\leq V_{OUT} \leq$ 0.60 V. This suggests that the triple

self-cascode current mirror is capable of providing the desired current gain across a wide range of output voltages and input currents.

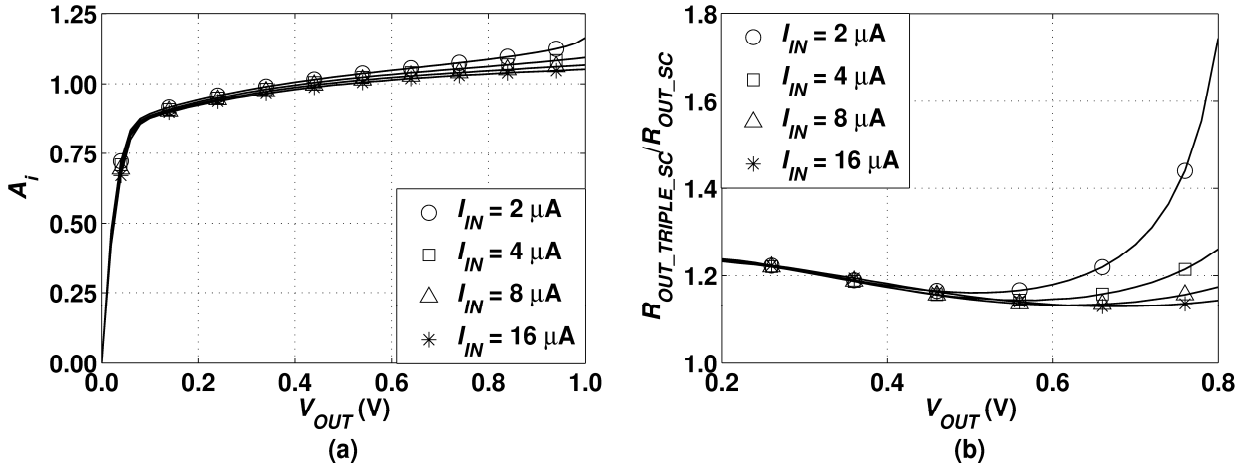

Figure 5.17: (a) A_i vs. V_{OUT} for the triple self-cascode current mirror of Figure 4.6 (b). The cascoded devices of the triple self-cascode current mirror were designed with W = 100 μm and L = 1 μm. The middle cascoding devices of the triple self-cascode current mirror were designed with W = 100 μm and L = 0.5 μm. The top cascoding devices of the triple self-cascode current mirror were designed with W = 100 μm and L = 0.25 μm. The legend specifies I_{IN}. (b) $R_{OUT\_TRIPLE\_SC}/R_{OUT\_SC}$ vs. V_{OUT}. The cascoded devices of the self-cascode current mirror were designed with W = 100 μm and L = 1 μm. The cascoding devices of the self-cascode current mirror were designed with W = 100 μm and L = 0.25 μm. The legend specifies I_{IN}.

Figure 5.17 (b) plots the ratio of output resistances between the triple self-cascode current mirror and a self-cascode current mirror (Figure 4.5) vs. V_{OUT} for the same input currents as Figure 5.17 (a). The self-cascode current mirror was designed with W = 100 μm for all transistors. The cascoded transistors had channel lengths of 1 μm and the cascoding transistors had channel lengths of 0.25 μm. The helper transistor was designed with W = 5 μm and L = 0.5 μm. Figure 5.17 (b) shows that the triple self-cascode current mirror achieves a greater output resistance than the self-cascode current mirror over a wide range of output voltages and input currents. For example, the triple self-cascode current mirror had an output resistance at least 1.1 times greater than the self-cascode current mirror for 0.2 V $\leq V_{OUT} \leq$ 0.8 V. The increase in output resistance is due to the r_O of the added device. However, this increase in output

resistance may not be significant enough to warrant the use of the third area-consuming device in most applications.

5.4 The Design of Ultra-Thin Oxide CMOS Differential Amplifiers

This section presents simulation results of the amplifier techniques described in Section 4.5. It is broken into three subsections. The first subsection characterizes the gate-balancing technique. The second subsection presents results comparing the voltage gain of a self-cascode amplifier to a simple amplifier The third subsection presents results characterizing the input current cancellation technique of Figure 4.10.

5.4.1 Gate Balancing

The simple differential amplifier of Figure 4.7 was simulated to show the imbalance created by gate current. Figure 5.18 (a) plots $I_{D1} - I_{D2}$ vs. I_{BIAS} and $V_{DIO} - V_{OUT}$ vs. I_{BIAS} for the differential amplifier of Figure 4.7. M1, M2, and M3 were sized with $W = 20$ µm and $L = 5$ µm. M4 and M5 were sized with $W = 40$ µm and $L = 5$ µm. The results show that gate current can cause extreme imbalance. For example, $V_{DIO} - V_{OUT}$ reached a peak value of 200 mV at $I_{BIAS} = 2$ µA and was greater than 30 mV for 2 µA $\leq I_{BIAS} \leq$ 256 µA. $I_{D1} - I_{D2}$ reached a peak value of 4.2 µA at $I_{BIAS} = 256$ µA and was greater than 260 nA for 2 µA $\leq I_{BIAS} \leq$ 256 µA.

To rectify this problem, the gate-balance technique described in Section 4.5.2 was implemented using the self-cascode amplifier shown in Figure 4.9. For example, Figure 5.18 (b) plots $I_{D1} - I_{D2}$ vs. I_{BIAS} and $V_{DIO} - V_{OUT}$ vs. I_{BIAS} for the self-cascode amplifier of Figure 4.9. SC1 and SC2 were sized with $W = 100$ µm, $L = 1$ µm, and $S_F = 8$. SC4 and SC5 were sized with $W = 200$ µm, $L = 1$ µm, and $S_F = 8$. SC3 was sized with

$W = 50\ \mu m$, $L = 2\ \mu m$, and $S_F = 16$. All cascoding transistors were sized with $L = L_{MIN\_A} = 0.25\ \mu m$. Note that a helper transistor could be added between the gate of SC3 and the drain of SC8 to improve the current gain. The results show a significant improvement compared to Figure 5.18 (a). For example, $V_{DIO} - V_{OUT}$ reached a peak value of 2 mV at $I_{BIAS} = 2\ \mu A$ and $I_{D1} - I_{D2}$ reached a peak value of 11 nA at $I_{BIAS} = 256\ \mu A$.

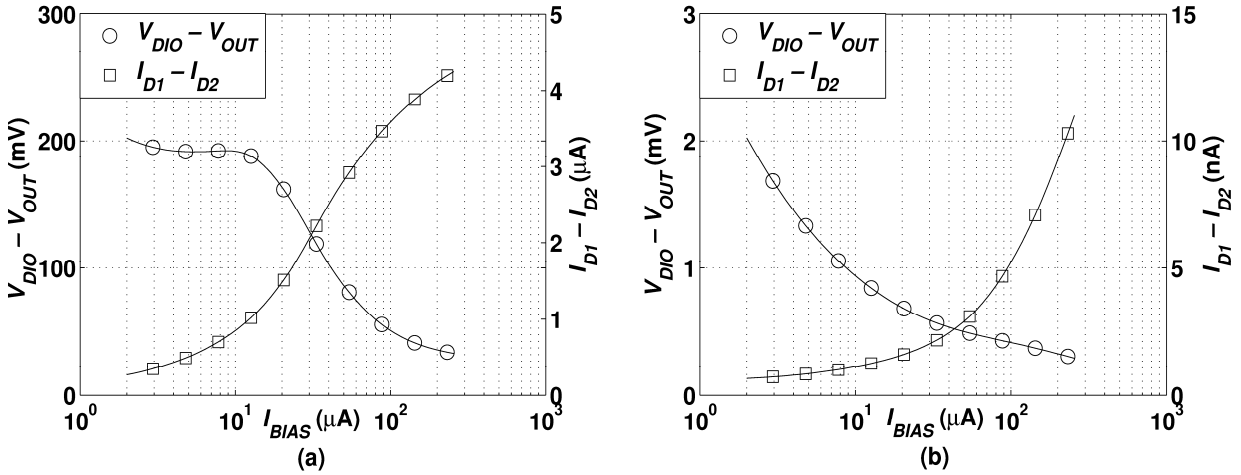

Figure 5.18: (a) $I_{D1} - I_{D2}$ vs. I_{BIAS} and $V_{DIO} - V_{OUT}$ vs. I_{BIAS} for the unbalanced amplifier of Figure 4.7. (b) $I_{D1} - I_{D2}$ vs. I_{BIAS} and $V_{DIO} - V_{OUT}$ vs. I_{BIAS} for the balanced self-cascode amplifier of Figure 4.9. V_{IN1} and V_{IN2} of both amplifier's were biased at 650 mV.

5.4.2 Amplifier Gain Comparison

Figure 5.19 compares the voltage gain of a balanced self-cascode amplifier (Figure 4.9) with a balanced simple amplifier (Figure 4.8). The transistors of the simple amplifier were sized equally to the cascoded transistors of the self-cascode amplifier. The results show that the self-cascode amplifier is able to produce a relatively large voltage gain (72.98 dB) compared to the simple amplifier (51.68 dB). This suggests that the combined use of the gate-balance technique with cautiously sized self-cascode structures can minimize the impact of gate current and r_o-degradations while allowing for the design of relatively high-gain amplifiers.

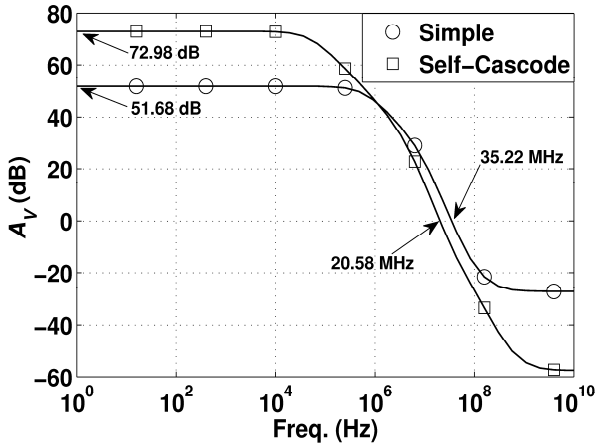

Figure 5.19: A_V vs. Frequency for the balanced simple amplifier (Figure 4.8) and the balanced self-cascode amplifier (Figure 4.9). I_{BIAS} = 16 µA. The load capacitance was 1 pF. V_{IN1} and V_{IN2} of both amplifier's were biased at 650 mV. The intrinsic gain of M1 in Figure 4.8 was 27.87 dB.

5.4.3 Input Current Cancellation

A self-cascode version of the differential amplifier of Figure 4.10 was simulated to show that amplifier input resistance can be increased by applying the input current cancellation technique described in Section 4.5.3. M1, M2, and M15 were sized with W = 20 µm and L = 5 µm. The bias current of the differential amplifier and the error amplifier was set equal to 1 µA. All current mirrors were made using self-cascode structures with W_{BOT} = 10 µm, L_{BOT} = 1 µm, W_{TOP} = 10 µm, and L_{TOP} = 0.25 µm. M7, M8, and M9 of Figure 4.10 were sized with W = 1 µm and L = 1 µm. M16 of Figure 4.10 and M6 of Figure 4.11were also sized with W = 1 µm and L = 1 µm. The results are shown in Figure 5.20, which plots gate current vs. V_{COM} for the amplifier with the input current cancellation technique applied and an amplifier without the input current cancellation technique. The amplifier without the input current cancellation technique was the same as the amplifier with the technique except that it did not have M3, M7-M9, M11, M15, and the error amplifier.

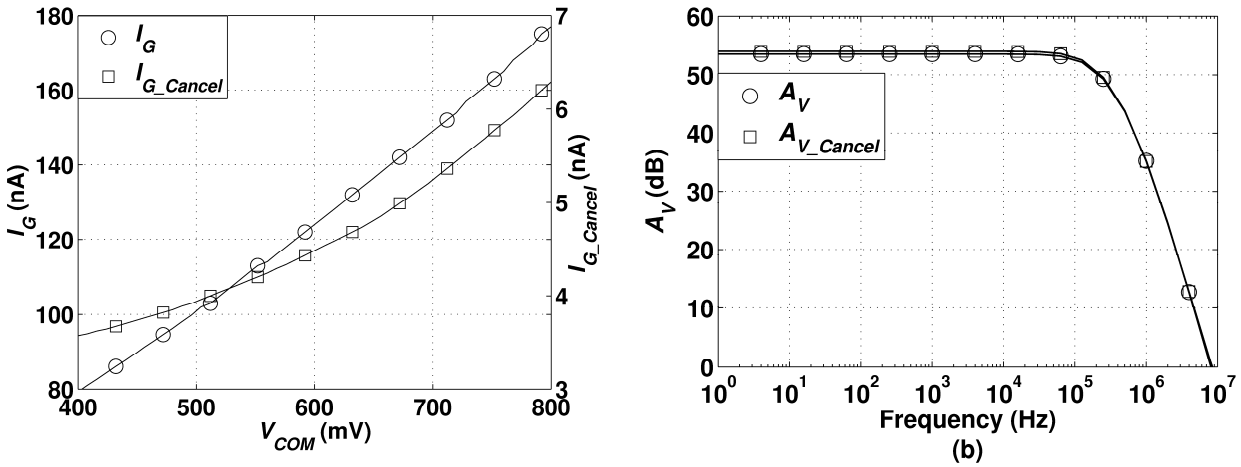

Figure 5.20: (a) I_G vs. V_{COM} and (b) A_V vs. Frequency for two self-cascode differential amplifiers. I_{G_Cancel} and A_{V_Cancel} refer to an amplifier with input current cancelation (Figure 4.10). I_G refers to an amplifier without input current cancellation. The amplifier without input current cancellations was the same as the amplifier with input current cancellation except that it did not have M3, M7-M9, M11, M15, and the error amplifier of Figure 4.10.

Figure 5.20 (a) shows that the gate current supplied by V_{COM} for the amplifier with input current cancellation, I_{G_Cancel}, was significantly less than I_G, the gate current supplied by V_{COM} for the amplifier without cancellation. For example I_{G_Cancel} had a minimum value of approximately 3 nA and a maximum value of approximately 7 nA for 400 mV $\leq V_{COM} \leq$ 800 mV. I_G had a minimum value of approximately 80 nA and a maximum value of approximately 180 nA across the same common-mode input range. Figure 5.20 (b) plots the voltage gain, A_V, vs. frequency for each amplifier. The results show that the voltage gain of the amplifier with cancellation, A_{V_Cancel}, is approximately equal to the voltage gain without cancellation, A_V. This suggests that the cancellation technique does not modify the nominal voltage gain of the amplifier. These results suggest that the input current cancellation technique can be used to significantly increase amplifier input resistance. Also, it allows for longer channel lengths to be used in input differential pairs.

5.5 The AC Simulation of Ultra-Thin Oxide CMOS Amplifiers

The impact of gate current on the AC simulation of ultra-thin oxide amplifiers was investigated using the voltage reference shown in Figure 4.16. The feedback loop of the buffer was broken and the traditional technique described in Section 4.6 was applied [48]. An AC simulation was performed and R_{OUT} of the buffer along with the DC bias point of V_{REF} were recorded. The new technique described in Section 4.6, which attempts to account for non-negligible amplifier input current, was then applied and the simulation was re-run.

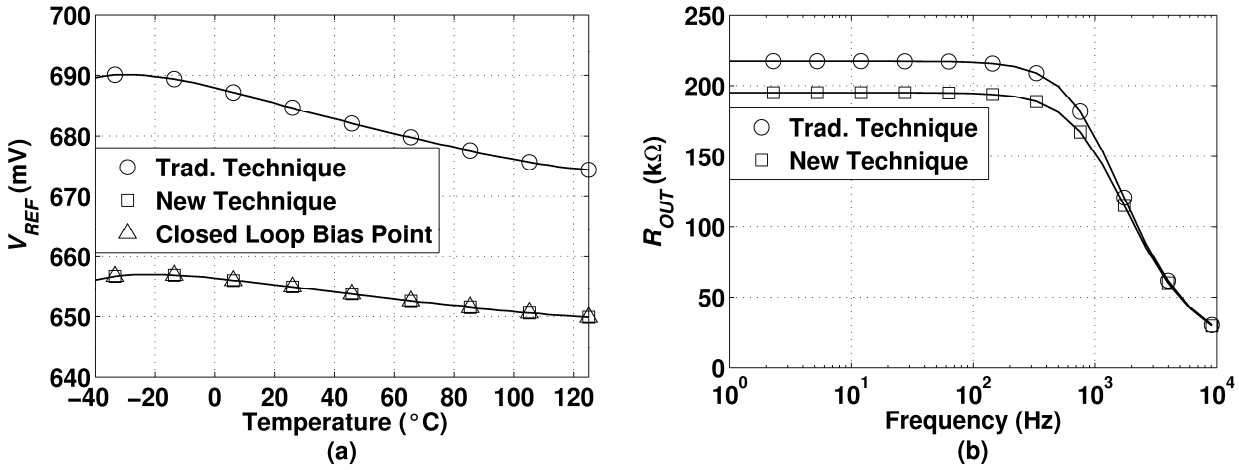

Figure 5.21: (a) V_{REF} vs. T and (b) R_{OUT} vs. frequency for the AC Simulation techniques described in Section 4.6.

The results of these two simulations are shown in Figure 5.21. Figure 5.21 (a) plots V_{REF} vs. temperature (°C). The correct DC bias point for V_{REF} was the value simulated when the amplifier was in the closed-loop configuration. The results show that the traditional technique led to differences in V_{REF} of up to 34 mV across the operating temperature range, while the new technique was able to maintain the correct DC bias point.

Figure 5.21 (b) plots R_{OUT} vs. frequency for both techniques. The plot shows significant differences in output resistance between the traditional technique and the new technique. For example, the traditional technique simulated a DC output resistance of 217 kΩ while the new technique simulated a DC output resistance of 195 kΩ. These results suggest that the new technique described in Section 4.6 should be applied when performing AC simulations on feedback amplifiers in technologies with non-negligible gate current.

5.6 The Design of an Ultra-Thin Oxide Sub-1 V Bandgap Voltage Reference

This section presents simulation results comparing the thick-oxide sub-1 V bandgap voltage reference presented in [116] to the ultra-thin oxide sub-1 V bandgap voltage reference described in Section 4.8. It contains three subsections. The first subsection presents the results of the thick-oxide voltage reference. The second subsection presents the results of the thick-to-ultra-thin voltage reference. The third subsection presents the results of the ultra-thin oxide voltage reference.

5.6.1 Thick-Oxide Sub-1 V Bandgap Voltage Reference

A thick-oxide version of the reference presented in [116] and [192] (see Figure 3.22) was designed and simulated in IBM's 10SF technology. The basis for this design came from a previous design that was fabricated in a 0.13 μm CMOS technology. Thick-oxide transistors were used to minimize the effect of gate current on performance. Self-cascode structures were used for all current mirrors, which were designed with drain currents of 2.5 μA at $T = 25$ °C. The thick-oxide voltage reference consumed approximately 15 μW of total power at $T = 25$ °C. The cascoding transistors of the

PMOS mirrors were sized with $W = 400$ µm and $L = 0.25$ µm. The cascoded transistors of the PMOS mirrors were sized with $W = 160$ µm and $L = 2.5$ µm. The cascoding transistors of the NMOS mirrors were sized with $W = 400$ µm and $L = 0.25$ µm. The cascoded transistors of the NMOS mirrors were sized with $W = 80$ µm and $L = 2.5$ µm. The NMOS input pair of the error amplifier was sized with $W = 800$ µm and $L = 0.5$ µm. A self-cascode structure was not used for the input pair so that the voltage headroom could be increased.

Seventy-two 3.2 µm x 3.2 µm diode-connected PNP transistors were used for Q2 of Figure 3.22. Nine 3.2 µm x 3.2 µm diode-connected PNP transistors were used for Q1 of Figure 3.22. The ratio of emitter areas between Q2 and Q1 was 8:1. V_{EB1} was found in simulation to be 653 mV at $T = 25$ °C, $\partial V_{EB1}/\partial T$ was found to be approximately −1.8 mV/°C, and $\partial \Delta V_{EB}/\partial T$ (see (3.17)) was found to be approximately 181 µV/°C.

From these numbers, (4.7) was used to calculate R_2/R_1 and R_3/R_1 values of 30. Equation (4.8) was used to calculate an R_4/R_1 ratio of approximately 12. R_1 was designed using a combination of three parallel precision poly-silicon unit resistors, with the unit resistance being 64.74 kΩ (L = 80 µm, W = 0.5 µm). R_2 and R_3 were combined into 11 unit resistors (see Section 4.8.4). R_4 was designed using 4 unit resistors.

The error amplifier was compensated using two vertical natural capacitors, which were both connected from the amplifier's output to V_{DD} [202]. Two capacitors were used to simplify the layout of the reference. The first capacitor was 9.98 pF and was sized with W = 150.18 µm and L = 40.215 µm. The second capacitor was 6.09 pF and was sized with W = 49.25 µm and L = 75.92 µm. The worst-case phase margin of the error amplifier across process corners was 51°. The worst-case gain margin of the error

amplifier across process corners was −20 dB. Dummy transistors were included on the PMOS mirrors, NMOS mirrors, and NMOS input pair. Dummy resistors were included in the resistor array.

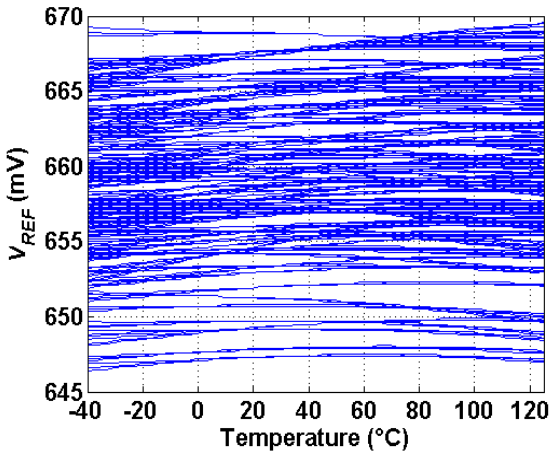

Figure 5.22: Monte Carlo analysis of V_{REF} vs. T for the thick-oxide sub-1 V bandgap voltage reference presented in [116]. The graph shows 300 Monte Carlo runs across three different supply voltages (0.9 V, 1.0 V, and 1.1 V). Each supply voltage simulated 100 runs.

Once the design was complete, a Monte Carlo analysis was performed at V_{DD} values of 0.9 V, 1.0 V, and 1.1 V. The analysis had 300 total runs, with each V_{DD} value simulating 100 runs. Each single run simulated V_{REF} vs. temperature. The temperature range was −40 °C to 125 °C. The results are shown in Figure 5.22, which plots V_{REF} vs. temperature. The results show that the minimum output voltage, V_{REF_MIN}, was 646.4 mV and the maximum output voltage, V_{REF_MAX}, was 669.6 mV. Averaging these two together gives an average output voltage, V_{REF_AVG}, of 658.0 mV. This implies that V_{REF} changed by ± 1.8% $\left(\frac{V_{REF\_MAX}-V_{REF\_MIN}}{2 \cdot V_{REF\_AVG}} \cdot 100\right)$ over a temperature range of 165 °C. The temperature coefficient was calculated as:

$$T_C = \frac{V_{REF\_MAX} - V_{REF\_MIN}}{V_{REF\_AVG} \cdot (T_{MAX} - T_{MIN})} 10^6 \tag{5.1}$$

where T_{MAX} = 125 °C is the maximum temperature and T_{MIN} = −40 °C is the minimum temperature. The temperature coefficient of the thick-oxide bandgap voltage reference was calculated to be 213.7 ppm/°C.

5.6.2 Thick-to-Ultra-Thin Oxide Sub-1 V Bandgap Voltage Reference

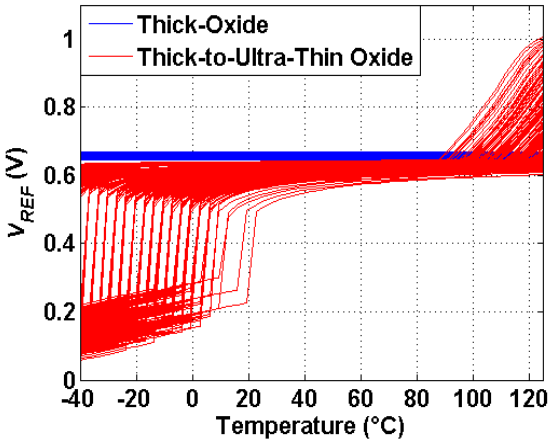

Figure 5.23: Comparison of the Monte Carlo analyses of the thick-oxide sub-1 V bandgap voltage reference presented in [116] and the thick-to-ultra-thin sub-1 V bandgap voltage reference shown in [116]. The graph shows 300 Monte Carlo runs across three different supply voltages (0.9 V, 1.0 V, and 1.1 V). Each supply voltage simulated 100 runs.

To show the effects of gate current on voltage reference performance, all of the devices in the thick-oxide reference were changed to ultra-thin oxide and the Monte Carlo analysis re-run. The results are shown in Figure 5.23, which plots the Monte Carlo results of the thick-oxide reference and the thick-to-ultra-thin reference on the same axes. The graph shows that the effects of gate current are devastating. For example, the performance metrics of the thick-to-ultra-thin oxide reference were: V_{REF_MIN} = 57.4 mV, V_{REF_MAX} = 1.006 V, V_{REF_AVG} = 531.7 mV, a percent change of ± 89.2%, and T_C = 10,821.4 ppm/°C. These performance metrics were so poor that the thick-to-ultra-thin reference could not be considered a voltage reference. The dominant

cause of this degradation was gate current, which demonstrates the necessity of a circuit methodology that can account for its presence.

5.6.3 Ultra-Thin Oxide Sub-1 V Bandgap Voltage Reference

This subsection contains eight subsections. The first subsection presents the general design strategy of the ultra-thin oxide sub-1 V bandgap voltage reference. The second subsection presents the impact of the error amplifier's PMOS active load on performance. The third subsection presents the impact of the error amplifier's input pair on performance. The fourth subsection presents the impact of gate current flowing into the output node on performance. The fifth subsection presents the results of Monte Carlo and process corners analyses that were performed on the reference. The sixth subsection presents results of startup analyses that were performed on the reference. The seventh subsection presents results of transistor loading analyses that were performed on the reference. The last subsection presents results of a sensitivity analysis that was performed on the reference.

5.6.3.1 General Design Strategy

The ultra-thin oxide voltage reference of Figure 4.16 was designed to investigate if the developed methodology could overcome the problems observed in Figure 5.23. The techniques described in Sections 4.2-4.8 were used in this design. Specifically, the gate-balancing technique was applied to both the error amplifier and buffer amplifier (see Section 4.5.2). Diode-connected transistors were used to minimize I_{GD} differences between SC9 and SC16 (see Section 4.8.1). Self-cascode structures were used to maximize output resistance while still allowing for low-voltage operation. They were

sized using the channel length selection methodology described in Section 5.1.3. This was done to minimize the total amount of gate current while still allowing for large-area devices to achieve a high degree of matching.

The self-cascode current mirrors were designed to have nominal drain currents of 3.3 μA at $T = 25$ °C. The reference consumed approximately 37 μW of total power at $T = 25$ °C. Note that the nominal drain current was made larger than the nominal drain current of the thick-oxide reference (2.5 μA). This was done because the relative effects of gate current decrease with increasing bias current (see Section 5.1.3). However, increases in the nominal drain current beyond 3.3 μA were limited by the minimum voltage headroom needed across the PMOS current mirrors, which was found to be approximately 100 mV (see Section 4.8.1). Specifically, V_{EB1}, which is a CTAT voltage, limited the current mirror's voltage headroom at cold temperatures. The voltage headroom was further limited by reductions in the supply voltage and the slow process corner. Therefore, the nominal drain current was found by setting the temperature to the process minimum (–40 °C), supply voltage to the process minimum (0.9 V), the process corner to slow, and verifying that the PMOS mirrors had at least 100 mV of headroom.

The cascoding transistors of the PMOS mirrors were designed with $W = 408.0$ μm and $L = 0.25$ μm. The cascoded transistors of the PMOS mirrors were designed with $W = 204.0$ μm and $L = 1.0$ μm. The cascoding transistors of the NMOS mirrors were designed with $W = 204.0$ μm and $L = 0.25$ μm. The cascoded transistors of the NMOS mirrors were designed with $W = 102.0$ μm and $L = 1.0$ μm. The area of the self-cascode current mirrors in the ultra-thin oxide reference (204 μm^2 for the PMOS mirrors, 102 μm^2

for the NMOS mirrors) was significantly less than the area of the self-cascode current mirrors in the thick-oxide reference (400 μm^2 for the PMOS mirrors, 200 μm^2 for the NMOS mirrors). This implies that the ultra-thin oxide reference may not match as well as the thick-oxide reference and illustrates a tradeoff between matching and gate current in ultra-thin oxide technologies.

Relatively large aspect ratios were used on the cascoded devices in the ultra-thin oxide reference (204/1 for the PMOS mirrors, 102/1 for the NMOS mirrors) compared to the thick-oxide reference (64/1 for the PMOS mirrors, 32/1 for the NMOS mirrors). This was done to minimize the relative impact of gate current ($\beta_{F\_MOS} \propto 1/L^2$) on the ultra-thin oxide voltage reference. It also placed the ultra-thin oxide current mirrors into the sub-threshold region of operation, where it was shown in Section 3.1.4.2 that high device output resistance could be obtained. One possible downside to this approach is degraded drain current matching. For example, when designing in saturation, I_D is roughly proportional to $(V_{GS} - V_{TH})$, which suggests that using small aspect ratios and thus large V_{GS} bias voltages helps wash out V_{TH} mismatch. However, in this design, device area was relatively large and V_{TH} mismatch was not a major concern.

The area of the NMOS current mirrors was less than the area of the PMOS current mirrors. This was done because the current matching of the PMOS mirrors was more important than the current matching in the NMOS current mirrors. For example, referring to Figure 4.15, the critical currents to be matched are I_1, I_2, and I_3, which are made up of PMOS self-cascode current mirrors in the transistor implementation. Also,

by making the NMOS current mirrors smaller, the impact of I_{GD} differences on SC3, SC10, SC13, and SC18 due to different output voltages is less of a concern.

The channel lengths of all mirror cascoding transistors were set to L_{MIN_A} (0.25 μm) to minimize the impact of I_{GD} on current mirror performance. The width of each mirror cascoding transistor was chosen to be equal to the width of the transistor it was cascoding. This approach helped increase S_F of each self-cascode structure while keeping the area of the cascoding transistors relatively small. The channel lengths of the transistors being cascoded were chosen to be 1 μm because that was the maximum analog channel length for the given temperature range, device area, and bias current (see Figure 5.7). Note that if the bias current were to be increased, the channel lengths of the cascoded transistors could potentially be increased.

Seventy-two 3.2 μm x 3.2 μm diode-connected PNP transistors were used for Q2 of Figure 4.16. Nine 3.2 μm x 3.2 μm diode-connected PNP transistors were used for Q1 of Figure 4.16. The ratio of emitter areas between Q2 and Q1 was 8:1. Note that the area of Q1 and Q2 could have been increased to decrease V_{EB1} and V_{EB2} such that a larger nominal drain current could have been used. However, it was found via simulation that further increasing the area of Q1 and Q2 had minimal impact on the nominal drain current. V_{EB1} was found in simulation to be 653 mV at $T = 25$ °C, $\partial V_{EB1}/\partial T$ was found to be approximately −1.8 mV/°C, and $\partial \Delta V_{EB}/\partial T$ (see (3.17)) was found to be approximately 181 μV/°C. These values are identical to the thick-oxide voltage reference because both references were designed using the same PNP BJTs. From these numbers, R_2/R_1 and R_3/R_1 were calculated to be 30 and the R_4/R_1 ratio was calculated to be 12. R_1 was

designed using a combination of three parallel precision poly-silicon unit resistors, with the unit resistance being 48.6 kΩ (L = 60.0 μm, W = 0.5 μm). R_2 and R_3 were combined using 14 unit resistors. The actual ratio used for R_2/R_1 and R_3/R_1 was 31 (not 30) because of the CTAT gate current mirrored from the input of the error amplifier into R_4 (see Section 4.8.3). R_4 was designed using 4 unit resistors.

The error amplifier was compensated using a 16 pF (W = 49.465 μm, L = 195.645 μm) vertical natural capacitor connected from its output to V_{DD} [202]. The worst-case phase margin of the error amplifier across process corners was 47°. The worst-case gain margin of the error amplifier across process corners was −12 dB. The buffer amplifier was compensated using an 8 pF (W = 72.965 μm, L = 67.66 μm) vertical natural capacitor in series with four unit resistors in parallel (W = 60 μm, L = 0.5 μm, $R_{PARALLEL}$ = 48.6 kΩ/4 = 12.15 kΩ). The resistor and capacitor compensation network was connected between V_{REF} and V_C of Figure 4.16. The worst-case phase margin of the buffer amplifier across process corners was 50°. The worst-case gain margin of the buffer amplifier across process corners was −10 dB.

5.6.3.2 Impact of Error Amplifier's PMOS Active Load

The channel lengths of the cascoded transistors in the PMOS active load of the error amplifier had a significant impact on reference performance. For example, to ideally avoid the effects of gate current, the channel lengths of these devices would be made as small as possible. However, if the channel length is made too short, the source-to-gate voltage across SC4 and SC5 drops below 100 mV under hot temperatures at the fast NMOS process corner and the fast PMOS process corner. For example,

consider Figure 5.24, which plots V_{REF} vs. T and V_{SG5} vs. T for V_{DD} = 0.9 V at the fast NMOS process corner and the fast PMOS process corner. The cascoded transistors of the PMOS mirrors were sized with W = 400 µm and L = 0.25 µm. The cascoding transistors of the PMOS mirrors were sized with W = 800 µm and L = 0.25 µm. The plot shows that at temperatures greater than 100 °C, V_{SG5} dropped below 100 mV. The self-cascode structures needed approximately 100 mV of voltage headroom to function as adequate current mirrors. If V_{SG5} is less than 100 mV when T > 100 °C the active load of the error amplifier no longer functions as a current mirror.

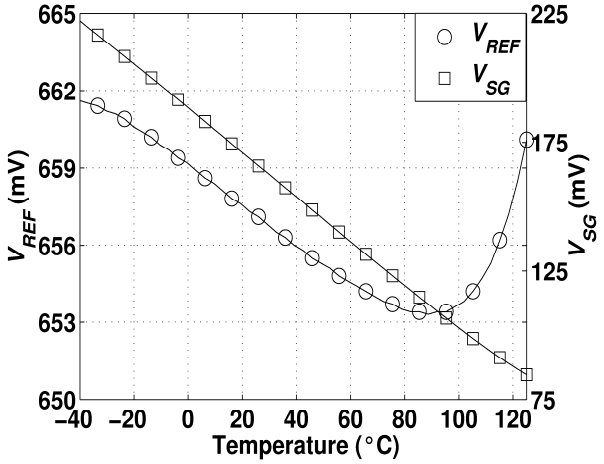

Figure 5.24: V_{REF} vs. T and V_{SG} of SC5 vs. T for V_{DD} = 0.9 V at the fast NMOS process corner and the fast PMOS process corner for the voltage reference of Figure 4.16. The cascoded transistors of the PMOS mirrors were sized with W = 400 µm and L = 0.25 µm. The cascoding transistors of the PMOS were sized with W = 800 µm and L = 0.25 µm.

The desired mirroring action of the reference was further degraded because SC6-SC9 had V_{SD} voltages much larger than 100 mV at temperatures greater than 100 °C. This implies that the currents in SC6-SC9 were not similar to the currents in SC4 and SC5 at temperatures above 100 °C because of significant differences in V_{SD}. This resulted in V_{REF} having a large temperature slope at hot temperatures. For example, V_{REF} only changed 7.3 mV as T increased from −40 °C to 100 °C, but it changed 6.4 mV as T

increased from 100 °C to 125 °C. This large change at hot temperatures was due to decreased V_{SG} voltages across the active load of the error amplifier. This problem was solved by increasing the channel length of all the PMOS cascoded transistors to 1 μm and decreasing the width to 204 μm. This approach allowed current mirror area to increase from 100 μm^2 to 204 μm^2 (improved matching) and also reduced the aspect ratio of the PMOS current mirrors from 1600 to 200. The reduction in aspect ratio forced the V_{SG} voltage of the active load to increase because the drain current remained constant. This increase in V_{SG} voltage improved the relative performance of the PMOS current mirrors such that they had more than 100 mV of headroom across the entire temperature range.

5.6.3.3 Impact of Error Amplifier's Input Pair

The input pair of both the error amplifier and the buffer amplifier had dimensions of W = 100 μm and L = 1 μm. A self-cascode structure was not used for either input pair so that the voltage headroom could be increased. This approach also limited the amount of gate current that was mirrored into R_4. The area of the thick-oxide reference's input pair (400 μm^2) was significantly larger than that of ultra-thin oxide reference (100 μm^2). This difference was due to the input current of the error amplifier. Specifically, the input current of the error amplifier in the thick-oxide reference was negligible. However, it was not negligible in the ultra-thin oxide reference. The buffer was used to limit the amount of error amplifier input current that got mirrored into R4 (see Section 4.7). However, the presence of the buffer does not imply that the input pair can be made arbitrarily large. The amount of input current drained by the buffer is a function of temperature and supply voltage. Therefore, it was necessary to size the input pair of the error amplifier such that the buffer would do an adequate job of draining the input current

across changes in temperature and supply voltage, while still being able to obtain a high degree of matching.

Dummy transistors were added to the input pair of both the error amplifier and the buffer amplifier. The gate, drain, and source terminals of the dummy transistors were tied to the source terminals of the transistors for which they were acting as dummies. Their body terminals were tied to the substrate. This suggests that these transistors would have non-zero gate-to-bulk current, I_{GB}. However, it was found via simulation that the I_{GB} component of these transistors was largely negligible and thus it did not impact the performance of the voltage reference. Note that the body biasing technique described in Section 4.3 could have been used to further reduce the impact of gate current on performance. However, this technique was not applied because the reference was designed to be used in a standard CMOS process that does not provide a separate well for the body terminal.

The channel length of the input pair of the error amplifier had a significant impact on performance. For example, to ideally avoid the effects of gate current, this channel length would be made as small as possible. However, if the channel length is made too short, the V_{DS} voltage across the input pair approaches zero under cold temperatures at the fast NMOS process corner and slow PMOS process corner. For example, consider Figure 5.25, which plots V_{REF} vs. T and V_{DS} of the input pair vs. T for V_{DD} = 0.9 V at the fast NMOS process corner and the slow PMOS process corner. The input pair was sized with W = 400 μm and L = 0.25 μm. The plot shows that at temperatures less than 0 °C, V_{DS} of the input pair dropped below 60 mV. This resulted in V_{REF} having a large

temperature slope. For example, V_{REF} only changed 0.6 mV as T decreased from 125 °C to 0 °C, but it changed 35.7 mV as T decreased from 0 °C to –40 °C. This large voltage change at cold temperatures was due to the low V_{DS} voltages on the input pair of the error amplifier.

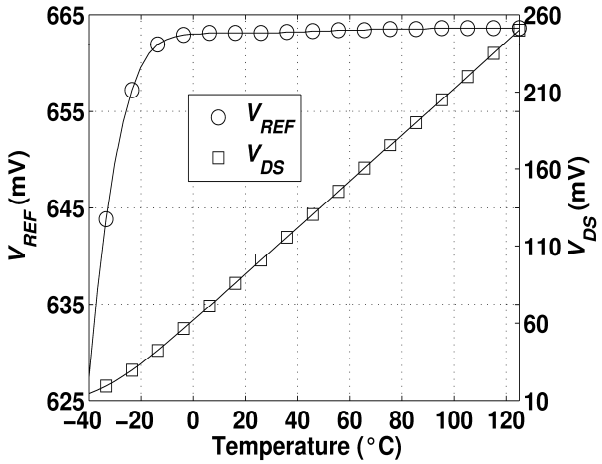

Figure 5.25: V_{REF} vs. T and V_{DS} of the error amplifier's input pair vs. T for V_{DD} = 0.9 V at the fast NMOS process corner and the slow PMOS process corner. The input pair was sized with W = 400 µm and L = 0.25 µm.

This problem was solved by increasing the channel length of the input pair to 1 µm and decreasing the width to 100 µm. This approach allowed the area to remain constant at 100 µm^2 and also reduced the aspect ratio of the input pair from 1600 to 100. This forced the V_{GS} voltage of the input pair to increase because the drain current remained constant. This increase in V_{GS} voltage was mostly due to a reduction in the source voltage, not an increase in gate voltage. The gate voltage remained constant because the gate terminal is connected to a diode-connected PNP, which provides the same voltage regardless of the size or aspect ratio of the input pair. Therefore, V_{GS} increased because of reductions in the source voltage. The drain voltage of the input pair, which was set by the PMOS active load, remained roughly constant. Therefore, V_{DS} of the input pair increased as L increased because the drain voltage remained constant and

the source voltage decreased. One observed advantage of increasing the channel length of the input pair was a decrease in the difference in drain voltages between SC3 and SC10. This resulted in improved NMOS current mirror performance.

5.6.3.4 Impact of Gate Current Flowing into the Output

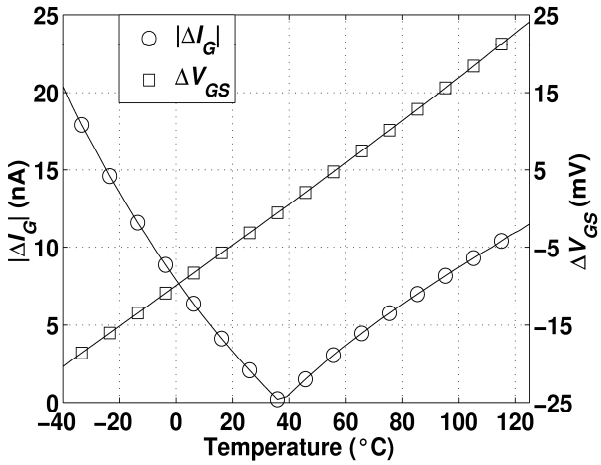

Figure 5.26: $I_{G2} - I_{G12}$ vs. T and $V_{GS2} - V_{GS12}$ (ΔV_{GS}) vs. T for V_{DD} = 1.1 V at the slow NMOS process corner and the slow PMOS process corner.

In Section 4.8.3, it was noted that at a specific temperature, the gate current mirrored by SC1 and SC2 into SC8 is equal to the gate current drawn by SC11 and SC12. Therefore, at this specific temperature, SC11 prevents this current from flowing into R_4 and impacting the performance of the reference. As temperature changes, V_{EB1} no longer equals V_{REF}, resulting in V_{GS1} and V_{GS2} not equaling V_{GS11} and V_{GS12}. Therefore, the gate current of SC11 is slightly different than what is mirrored into SC8 by SC1 and SC2. This is undesired and suggests that some gate current will flow into R_4 (see Section 4.8.3). For example, consider Figure 5.26, which plots $|I_{G2} - I_{G12}|$ vs. T and $V_{GS2} - V_{GS12}$ vs. T for V_{DD} = 1.1 V at the slow NMOS process corner and the slow PMOS process corner. The plot shows that $|I_{G2} - I_{G12}|$ and $V_{GS2} - V_{GS12}$ are relatively minimized around room temperature. This occurred because R_2 and R_3 were sized such that

$V_{EB1} = V_{REF}$ at this temperature. As the temperature changed, V_{EB1} no longer equaled V_{REF}. However, $|I_{G2} - I_{G12}|$ and $V_{GS2} - V_{GS12}$ were still both relatively minimized. Specifically, as the temperature increased from 27 °C to 125 °C, $I_{G2} - I_{G12}$ changed from 2.1 nA to 11.5 nA and $V_{GS2} - V_{GS12}$ changed from −3.1 mV to 24.1 mV. As the temperature decreased from 27 °C to −40 °C, $|I_{G2} - I_{G12}|$ changed from 2.1 nA to 20.4 nA and $V_{GS2} - V_{GS12}$ changed from −3.1 mV to −20.4 mV. $|I_{G2} - I_{G12}|$ and $V_{GS2} - V_{GS12}$ were minimized by applying the channel length selection methodology developed in Section 5.1.3.

5.6.3.5 Monte Carlo and Process Corners Analyses

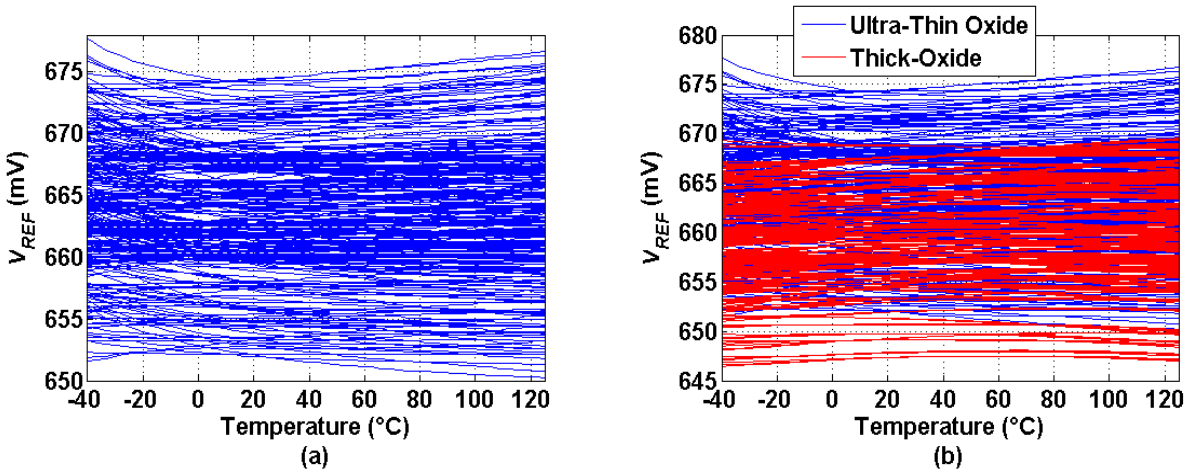

Figure 5.27: (a) Monte Carlo analysis of V_{REF} vs. T for the ultra-thin-oxide sub-1 V bandgap voltage reference shown in Figure 4.16. The graph shows 300 Monte Carlo runs across three different supply voltages (0.9 V, 1.0 V, and 1.1 V). Each supply voltage simulated 100 runs. (b) Comparison of the Monte Carlo analyses of the ultra-thin-oxide sub-1 V bandgap voltage reference shown of Figure 4.16 and the thick-oxide bandgap voltage reference presented in [116].

Once the design was complete, the Monte Carlo analysis performed on the previous two references was performed on the ultra-thin oxide reference. The results are shown in Figure 5.27. Figure 5.27 (a) shows that $V_{REF\_MIN} = 650.0$ mV, $V_{REF\_MAX} = 677.7$ mV, $V_{REF\_AVG} = 664.0$ mV, the percent change was ± 2.1%, and $T_C = 251.0$ ppm/°C. Table 5.1 compares the statistics of all three references. Figure

5.27 (b) plots the Monte Carlo results of the thick-oxide reference and the ultra-thin oxide reference on the same axes. The results show that the ultra-thin oxide reference of Figure 4.16 compares favorably to the thick-oxide reference and provides significant improvements over the ultra-thin oxide version of [116]. V_{REF_AVG} and T_C of the ultra-thin oxide voltage reference in Figure 4.16 are similar to the thick-oxide version of [116]. For example, the difference in average voltages between these two references is only 6.0 mV and the difference in temperature coefficients is only 37.3 ppm/°C.

Voltage Ref.	V_{REF_MIN} **(mV)**	V_{REF_MAX} **(mV)**	V_{REF_AVG} **(mV)**	**% change**	T_C **(ppm/°C)**
Thick - [116]	646.4	669.6	658.0	1.8	213.7
Ultra-Thin - [116]	57.4	1,006.0	531.7	89.2	10,821.4
Ultra-Thin - Fig. 4.14	650.0	677.7	664.0	2.1	251.0

Table 5.1: Comparison of the simulated voltage references.

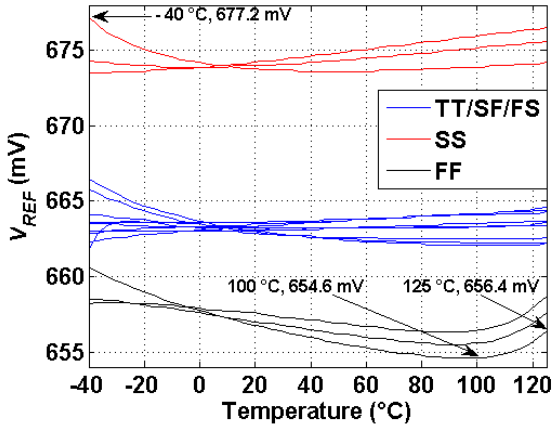

Figure 5.28: Process Corners analysis of V_{REF} vs. T for the ultra-thin oxide sub-1 V bandgap voltage reference shown in Figure 4.16.

A +/– 3-sigma process corners simulation of V_{REF} vs. T was performed on the ultra-thin oxide voltage reference. The temperature range was swept from –40 °C to 125 °C. The following MOSFET process corners were simulated: SS, SF, TT, FS, and FF. The following V_{DD} process corners were simulated: 0.9 V, 1.0 V, and 1.1 V. The process corners for the passive elements (resistors and capacitors) were set equal to the MOSFET process corner if the MOSFET process corner was equal to SS or FF.

Otherwise the passive element process corner was set equal to its typical value. There were 15 total corners in this simulation. The results are shown in Figure 5.28. The figure shows that V_{REF_MIN} = 654.6 mV, V_{REF_MAX} = 677.2 mV, V_{REF_AVG} = 666.8 mV, the percent change was ± 1.6%, and T_C = 189.0 ppm/°C. These results show that the voltage dispersion across process corners was similar to the Monte Carlo voltage dispersion. This suggests that the ultra-thin oxide voltage reference will function as desired in the presence of systematic and random process variations.

5.6.3.6 Startup Analyses

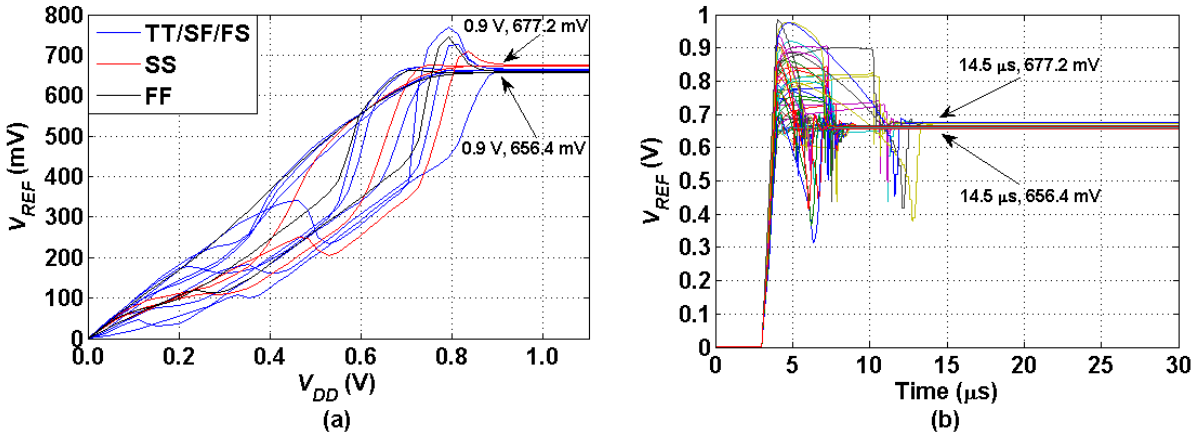

Figure 5.29: (a) Process Corners analysis of V_{REF} vs. V_{DD} for the ultra-thin oxide sub-1 V bandgap voltage reference shown in Figure 4.16. (b) V_{REF} vs. t for a V_{DD} rise time of 1 µs for the ultra-thin oxide sub-1 V bandgap voltage reference shown in Figure 4.16.

A +/– 3-sigma DC startup process corners analysis of V_{REF} vs. V_{DD} was performed on the ultra-thin oxide reference. The following MOSFET process corners were simulated: SS, SF, TT, FS, and FF. The following T process corners were simulated: –40 °C, 25 °C, and 125 °C. The process corners for the passive elements (resistors and capacitors) were set equal to the MOSFET process corner if the MOSFET process corner was equal to SS or FF. Otherwise the passive element process corner was set equal to its typical value. There were 15 total process corners for this simulation. The results are

shown in Figure 5.29 (a). The results show that for all corners V_{REF} was relatively settled for 0.9 V $\geq$ V_{DD} $\geq$ 1.1 V. Under this supply voltage range the minimum V_{REF} was 656.4 mV and the maximum V_{REF} was 677.2 mV. These results suggest that the voltage reference is capable of starting correctly when V_{DD} is ramped from zero to a final voltage between 0.9 V and 1.1 V.

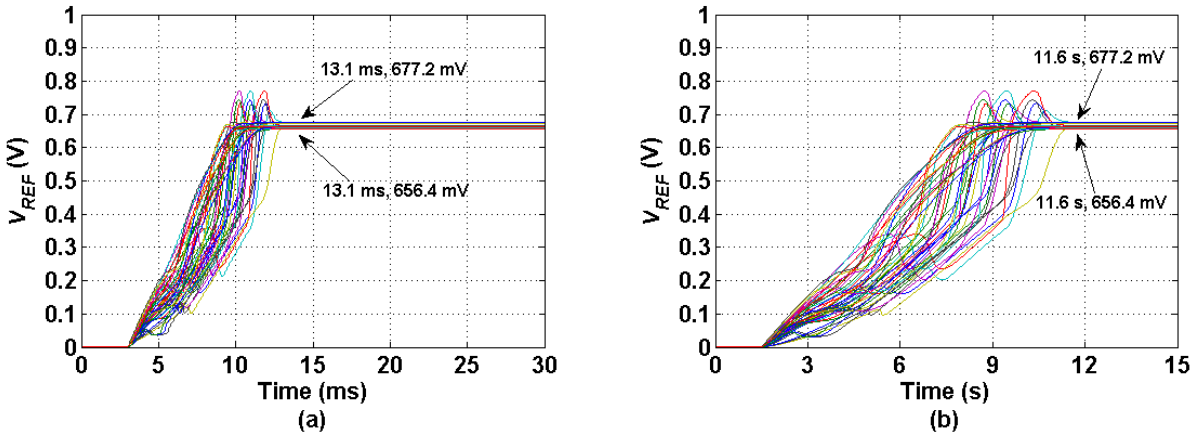

Figure 5.30: (a) V_{REF} vs. t for a V_{DD} rise time of 10 ms for the ultra-thin oxide sub-1 V bandgap voltage reference shown in Figure 4.16. (b) V_{REF} vs. t for a V_{DD} rise time of 10 s for the ultra-thin oxide sub-1 V bandgap voltage reference shown in Figure 4.16.

A transient startup process corners analysis of V_{REF} vs. time (t) was performed on the ultra-thin oxide voltage reference. After an initial delay, V_{DD} was stepped to its final value at a variable rise time. The simulated rise times were 1 µs, 10 ms, and 1 s. The following MOSFET process corners were simulated: SS, SF, TT, FS, and FF. The following T process corners were simulated: –40 °C, 25 °C, and 125 °C. The following V_{DD} process corners were simulated: 0.9 V, 1.0 V, and 1.1 V. The corners for the passive elements (resistors and capacitors) were set equal to the MOSFET process corner if the MOSFET process corner was equal to SS or FF. Otherwise the passive element process corner was set equal to its typical value. There were 45 total corners for this simulation. The results are shown in Figure 5.29 (b) and Figure 5.30. Figure 5.29 (b) plots V_{REF} vs. t

for a V_{DD} rise time of 1 µs and an initial delay of 3 µs. The results show that V_{REF} was settled to within 1% of its final value across all 45 corners within 11.5 µs of the supply ramp. Figure 5.30 (a) plots V_{REF} vs. t for a V_{DD} rise time of 10 ms and an initial delay of 3 ms. The results show that V_{REF} was settled to within 1% of its final value across all 45 corners within 8.1 ms of the supply ramp. Figure 5.30 (b) plots V_{REF} vs. t for a V_{DD} rise time of 10 s and an initial delay of 1.5 s. The results show that V_{REF} was settled to within 1% of its final value across all 45 corners within 8.6 s of the supply ramp. These results suggest that the voltage reference starts properly under transient power supply ramps across process, voltage, and temperature corners.

5.6.3.7 Transistor Loading

The impact of transistor loading on the voltage reference was also simulated. This simulation was performed because ultra-thin oxide MOSFETs draw gate current, which suggests that the voltage reference must be able to supply gate current to a loading transistor without changing its voltage or temperature characteristics. V_{REF} was loaded down with the gate of an NMOS transistor that had a PTAT current source connected to its source terminal (see M_L and I_{LOAD} in Figure 4.16). The current source had a temperature slope of 170 nA/°C and a room temperature value of 50 µA. The width of the loading transistor was set equal to 100 µm. A +/– 3-sigma process corners simulation of V_{REF} vs. T was then performed. The temperature range was swept from –40 °C to 125 °C. The following MOSFET process corners were simulated: SS, SF, TT, FS, and FF. The following V_{DD} process corners were simulated: 0.9 V, 1.0 V, and 1.1 V. The process corners for the passive elements (resistors and capacitors) were set equal to the MOSFET process corner if the MOSFET process corner was equal to SS or FF.

Otherwise the passive element process corner was set equal to its typical value. Three loading transistor channel lengths were simulated: 0.5 µm, 1 µm, and 2 µm. There were 45 total process corners in this simulation.

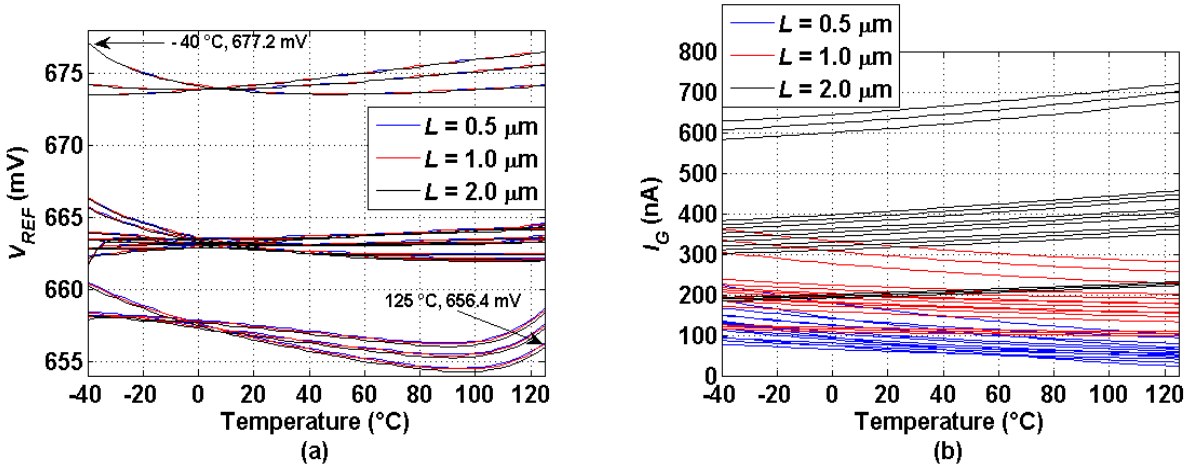

Figure 5.31: (a) Process Corners analysis of V_{REF} vs. T for the ultra-thin oxide sub-1 V bandgap voltage reference shown in Figure 4.16. (b) Process Corners analysis of I_G of the loading transistor vs. T for the ultra-thin oxide sub-1 V bandgap voltage reference shown in Figure 4.16. V_{REF} was loaded down with the gate of an NMOS transistor that had a PTAT current source connected to its source terminal (see M_L and I_{LOAD} in Figure 4.16). The current source had a temperature slope of 170 nA/°C and a value of 50 µA at T = 25 °C. Three loading transistor channel lengths were simulated: 0.5 µm, 1 µm, and 2 µm. The width of the loading transistor was set equal to 100 µm.

The results of the loading analysis are shown in Figure 5.31. Figure 5.31 (a) plots V_{REF} vs. T across all 45 corners. The results show that V_{REF} is relatively independent of the loading transistor. Specifically, there was no noticeable difference between Figure 5.31 (a) and Figure 5.28, which was unloaded. Figure 5.31 (b) plots I_G of the loading transistor vs. T. The plots show that the buffer was able to provide up to 720 nA of gate current to the loading transistor. These results suggest that the voltage reference is capable of providing load current while maintaining its voltage and temperature characteristics.

5.6.3.8 Sensitivity Analysis

A sensitivity analysis was performed on the ultra-thin oxide voltage reference. This analysis was done to determine which of BSIM4's direct tunneling parameters the reference was most sensitive too. This analysis could be used to potentially help explain why measured results do not match those obtained in simulation. For example, if the reference showed extreme sensitivity to a single direct tunneling model parameter and the measured results did not match those obtained in simulation, this model parameter may need to be adjusted such that future measured results match those obtained in simulation. BSIM4 has 21 total direct tunneling parameters [136]. Each of these 21 parameters is separately populated for the NMOS transistor and the PMOS transistor. In the sensitivity analysis, each of these parameters was varied ±100%, in 10% increments, in the same direction, for the both types of devices. The V_{REF} vs. T curve at the TT process corner with V_{DD} = 1.0 V was used to get the signature of each parameter.

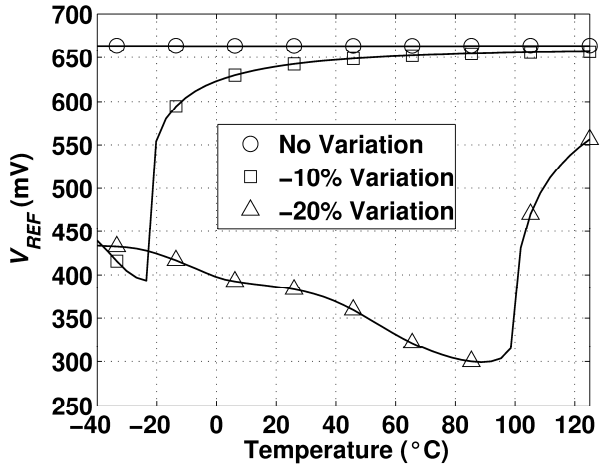

Figure 5.32: Sensitivity analysis of V_{REF} vs. T for the BSIM4 direct tunneling model parameter *aigc*. V_{DD} = 1.0 V. The process corner was TT.

The reference showed a high degree of sensitivity to the BSIM4 direct tunneling model parameter *aigc*, which is the major fitting parameter for I_{GCS} and I_{GCD}. For

example, Figure 5.32 plots V_{REF} vs. T for *aigc* under three conditions: not-varied, varied −10%, and varied −20%. Under the not varied condition, V_{REF_MAX} was 663.4 mV and V_{REF_MIN} was 663.1 mV. When *aigc* was varied −10%, V_{REF_MAX} decreased to 656.5 mV and V_{REF_MIN} decreased to 393.7 mV. When *aigc* was varied −20%, V_{REF_MAX} further decreased to 556.3 mV and V_{REF_MIN} further decreased to 259.3 mV. These results suggest that *aigc* must be characterized correctly if simulation results are going to match measurements.

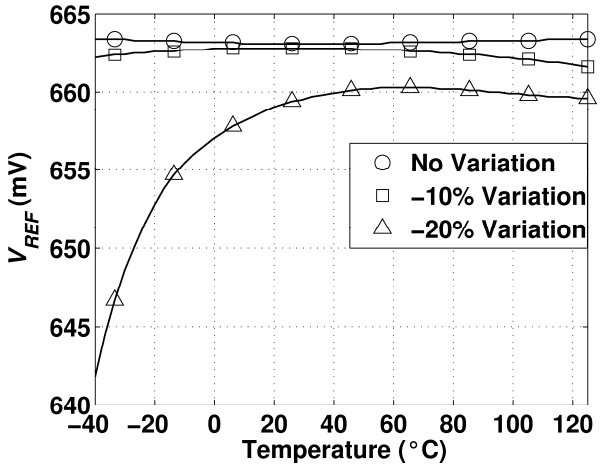

Figure 5.33: Sensitivity analysis of V_{REF} vs. T for the BSIM4 direct tunneling model parameter *poxedge*. V_{DD} = 1.0 V. The process corner was TT.

The reference showed a moderate degree of sensitivity to the BSIM4 direct tunneling model parameter *poxedge*, which is the major fitting factor for the oxide thickness. For example, Figure 5.33 plots V_{REF} vs. T for *poxedge* under three conditions: not varied, varied −10%, and varied −20%. Under the not-varied condition, V_{REF_MAX} was 663.4 mV and V_{REF_MIN} was 663.1 mV. When *poxedge* was varied −10%, V_{REF_MAX} decreased to 662.8 mV and V_{REF_MIN} decreased to 661.6 mV. When *poxedge* was varied −20%, V_{REF_MAX} further decreased to 660.3 mV and V_{REF_MIN} further decreased to 641.8 mV. These results suggest that moderate variations in *poxedge* could significantly

impact performance and that *poxedge* must be characterized relatively well if simulation results are expected to match measurements.

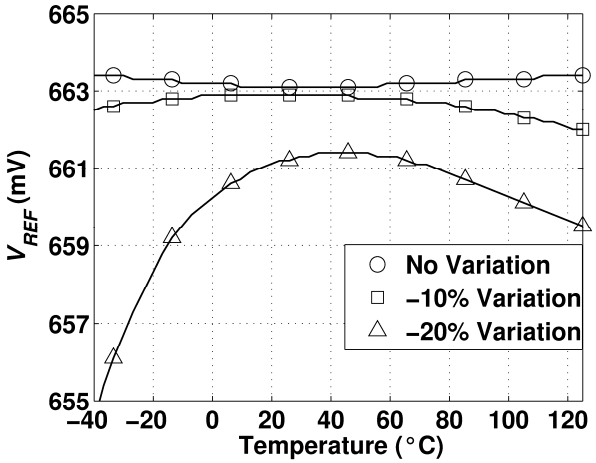

Figure 5.34: Sensitivity analysis of V_{REF} vs. T for the BSIM4 direct tunneling model parameter *aigsd*. V_{DD} = 1.0 V. The process corner was TT.

The reference also showed a moderate degree of sensitivity to the BSIM4 direct tunneling model parameter *aigsd*, which is the major fitting parameter for I_{GS} and I_{GD}. For example, Figure 5.34 plots V_{REF} vs. T for *aigsd* under three conditions: not varied, varied −10%, and varied −20%. Under the not-varied condition, V_{REF_MAX} was 663.4 mV and V_{REF_MIN} was 663.1 mV. When *aigsd* was varied −10%, V_{REF_MAX} decreased to 662.7 mV and V_{REF_MIN} decreased to 662 mV. When *aigsd* was varied −20%, V_{REF_MAX} further decreased to 661.4 mV and V_{REF_MIN} further decreased to 654.5 mV. These results suggest that moderate variations in *aigsd* could significantly impact performance and that *aigsd* must be characterized relatively well if simulation results are expected to match measurements.

The reference showed a low degree of sensitivity to the BSIM4 direct tunneling model parameter *toxref*, which is the nominal gate oxide thickness for direct tunneling. For example, Figure 5.35 plots V_{REF} vs. T for *toxref* under three conditions: not varied,

varied −50%, and varied +50%. Under the not-varied condition, V_{REF_MAX} was 663.4 mV and V_{REF_MIN} was 663.1 mV. When *toxref* was varied −50%, V_{REF_MAX} increased to 665.4 mV and V_{REF_MIN} increased to 664.2 mV. When *toxref* was varied +50%, V_{REF_MAX} decreased to 662.4 mV and V_{REF_MIN} decreased to 659.4 mV. These results suggest that large variations in *toxref* will not significantly impact performance. Considering that direct tunneling exponentially increases with decreasing oxide thickness, this implies that the developed voltage reference is able to maintain performance with changes in oxide thickness.

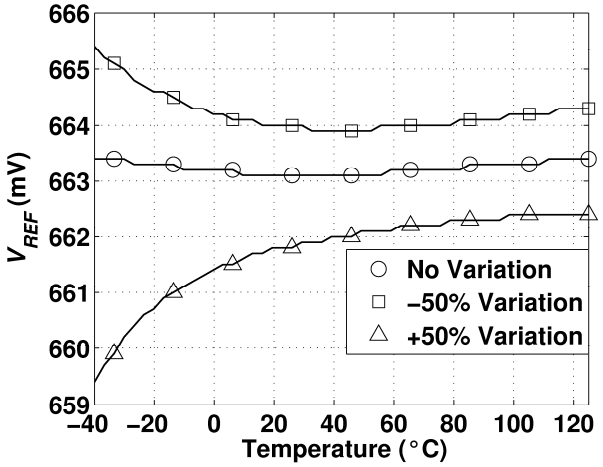

Figure 5.35: Sensitivity analysis of V_{REF} vs. T for the BSIM4 direct tunneling model parameter *toxref*. V_{DD} = 1.0 V. The process corner was TT.

5.7 Sponsored Fabrication

A sponsored fabrication of this work was awarded based on technical merit via the MOSIS Educational Program [38]. The target technology was IBM's 10SF technology. The design, simulation, and layout of a 2 mm x 2 mm chip was completed and sent to MOSIS. The design had 44 input/output pads. Forty die were to be shipped for testing. Twenty of these die were to be unpackaged and were to be tested on a

thermal chuck. The remaining twenty die were to be sealed in a moisture-insensitive conformally coated QFP44a package and tested in a thermal chamber [203].

The chip contained the design of seven different sub-1 V bandgap voltage references. Six of these references were variations on the ultra-thin oxide reference of Figure 4.16. The first reference was the standard ultra-thin oxide reference described in the previous subsection. Note that R_2/R_1 of this reference was 31. The second reference was the same as the standard ultra-thin oxide reference except that the body terminals of all transistors were tied to their source terminals. Specifically, the body terminals of the input pairs of the error amplifier and the buffer amplifier of Figure 4.16 were tied to their source terminals. The body terminals of all cascoding transistors were also tied to their source terminals. This was done to minimize the amount of gate current flowing through each transistor (see Section 4.3). R_2/R_1 of this reference was 30, which shows a slight decrease compared to the standard reference. This decrease in R_2/R_1 occurred because less error amplifier input current was mirrored into the output node (see Section 4.8.3).

The third reference was a standard ultra-thin oxide reference with no metal fill. The metal fill was to going to be placed by IBM's automatic metal filling process. This reference was then going to be compared to the standard reference to determine if the manually metal filled reference performed better than the automatically metal filled reference. The fourth reference was a rotated version of the standard reference. The reference was rotated 90°. It was going to be compared to the standard reference to determine if rotation had any effect on angled implants such that the rotated reference performed differently than the non-rotated reference. The fifth reference was a CTAT

version of the standard reference. This reference was given a CTAT output slope as a precaution to direct tunneling model imperfections. R_2/R_1 of this reference was 27. If the measured results of the standard reference showed a PTAT slope, this reference should have shown less of a CTAT slope than was seen in simulation. The sixth reference was a PTAT version of the standard reference. This reference was given a PTAT output slope as a precaution to direct tunneling model imperfections. R_2/R_1 of this reference was 33. If the measured results of the standard reference showed a CTAT slope, this reference should have shown less of a PTAT slope than was seen in simulation. The seventh reference was the thick-oxide reference of Figure 3.22. The design of this reference was the same as described in Section 5.6.1.

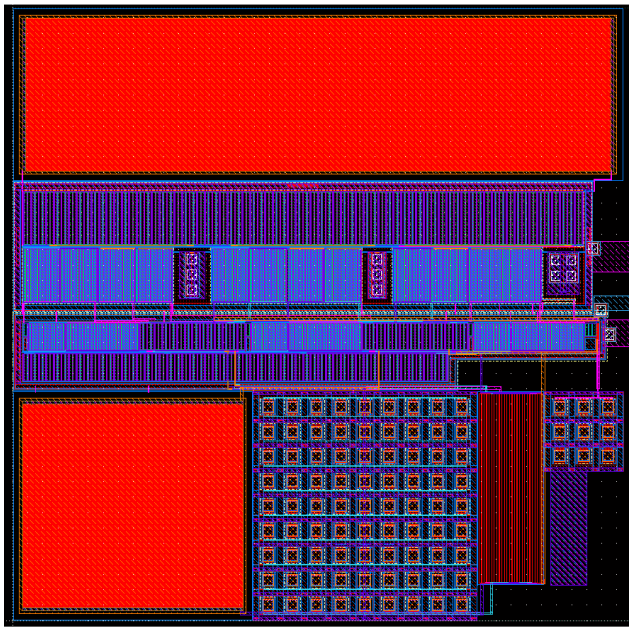

Figure 5.36: Layout of the standard ultra-thin oxide sub-1 V bandgap voltage reference of Figure 4.16 (202.165 µm by 198.1 µm).

All seven references were designed using interdigitation and common centroid layout techniques [91]. Guard rings were used to isolate resistors and different types of transistors. Figure 5.36 shows the layout of the standard ultra-thin oxide bandgap voltage

reference. It occupied an area of 202.165 µm by 198.1 µm. Figure 5.37 shows the layout of the thick-oxide reference. It occupied an area of 183.17 µm by 187.69 µm. Figure 5.38 shows the layout of the body-biased version of the standard ultra-thin oxide reference. It occupied an area of 248.265 µm by 209.43 µm.

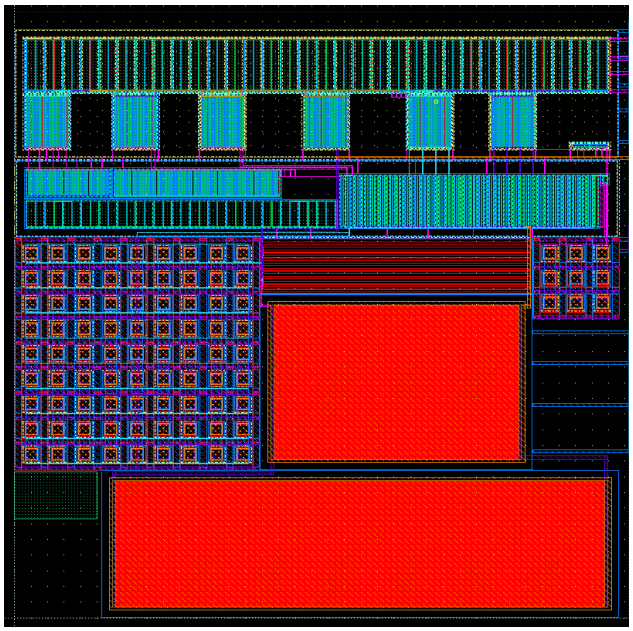

Figure 5.37: Layout of the thick-oxide bandgap voltage reference of Figure 3.22 (183.17 µm by 187.69 µm).

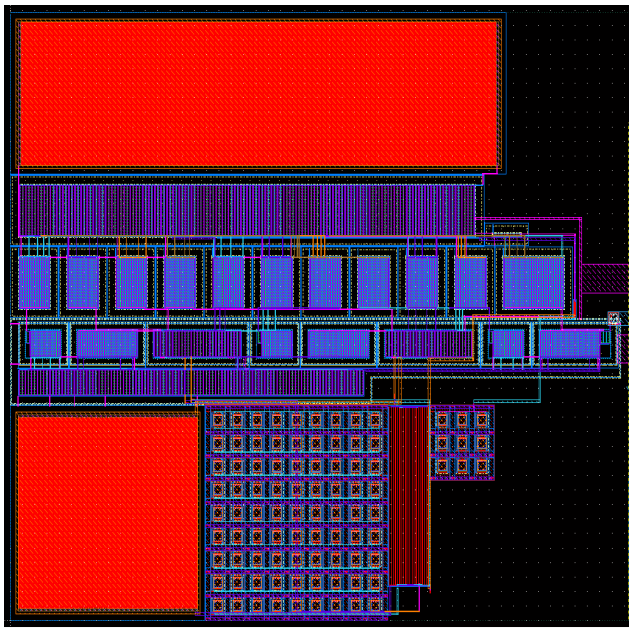

Figure 5.38: Layout of the body-biased version of the standard ultra-thin oxide bandgap voltage reference of Figure 4.16 (248.265 µm by 209.43 µm).

The chip contained the layout of three isolated transistors: a triple-well NFET, a dual-well NFET, and a dual-well PFET. Each transistor was designed with $W = 100$ μm and $L = 1$ μm. These transistors were going to be used to validate the BJT-like metrics of Section 4.2 and the sizing strategies of Section 5.1.3. They could have also been used to validate the direct tunneling model of BSIM4.

The chip contained the design of three NMOS self-cascode current mirrors (see Figure 4.4 and Figure 4.5). The first mirror was designed with a desired unity current gain. The cascoded transistors had $W = 102$ μm and $L = 2$ μm. The cascoding transistors had $W = 204$ μm and $L = 0.25$ μm. The second mirror was also designed with a desired unity current gain. The cascoded transistors had $W = 204$ μm and $L = 1$ μm. The cascoding transistors had $W = 204$ μm and $L = 0.25$ μm. The third current mirror was designed with a desired current gain of eight. The cascoded transistors had $W = 40$ μm and $L = 5$ μm. The cascoding transistors had $W = 40$ μm and $L = 1.25$ μm. All three current mirrors were designed using interdigitation and common centroid techniques. Guard rings were used to isolate different types of transistors. It was desired that these mirrors be used to validate the current mirror design strategies of Section 4.4.

The chip also contained the design of an ultra-thin oxide operational amplifier. The amplifier was sized equally to the buffer amplifier of the standard ultra-thin oxide sub-1 V bandgap voltage reference described in the previous section. It was desired that this amplifier be used to validate the amplifier design strategies of Section 4.5.

Figure 5.39 shows the complete layout of the chip. Note that ESD protection was included. Specifically, RC clamps were used for power supply pads. Double diodes and

SCRs were used for signal pads [204]. Note that all simulations were performed with the ESD protection present. Nine metal layers were available and were used for routing.

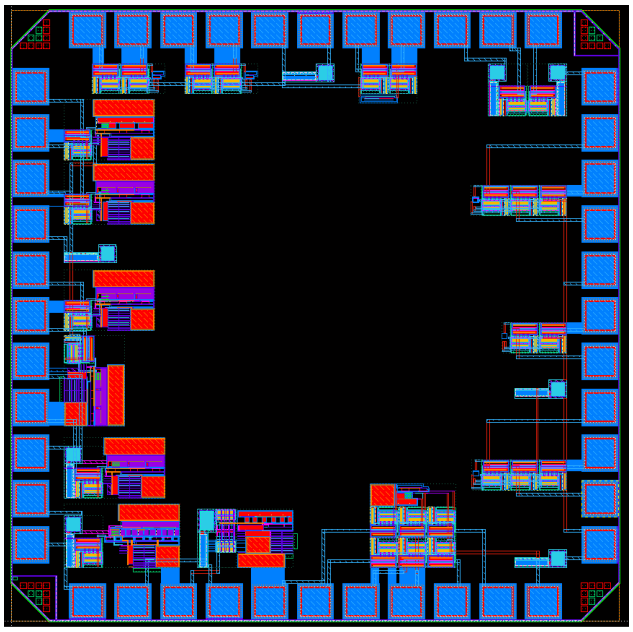

Figure 5.39: Complete layout of the designed chip.

For reasons beyond the author's control, the fabrication of the designed chip was delayed over 2 years. Therefore, fabrication results were unable to be included in this document. However, if fabrication does eventually occur after the publishing of this document, the results will be made available via a scholarly journal.

CHAPTER 6
CONCLUSION

This work developed a methodology that allows the design of analog systems with ultra-thin oxide MOSFETs. This methodology focused on transistor sizing, DC biasing, and the design of current mirrors and differential amplifiers. It attempted to minimize, balance, and cancel the negative effects of direct tunneling on analog design in traditional (non-high-κ/metal gate) ultra-thin oxide CMOS technologies. It showed that the tradeoff between gate current and mismatch can be minimized via informed device sizing. The methodology required only ultra-thin oxide devices and was investigated in IBM's 10SF 65 nm CMOS technology, which has a nominal V_{DD} of 1 V and a physical oxide thickness of 1.25 nm. Theoretical analysis and simulation were used to develop the methodology. The methodology attempted to not aggravate existing analog nanoscale CMOS problems such as reduced voltage headroom, decreased intrinsic gain, and reduced SNR. It focused on low-frequency performance because the effects of direct tunneling are negligible at higher frequencies. The results suggest that the methodology is effective and can be utilized to design useful analog circuits with traditional ultra-thin oxide MOSFETs.

A sub-1 V bandgap voltage reference was designed and implemented using the developed methodology in IBM's 10SF 65 nm process. It required only ultra-thin oxide MOSFETs and its performance was used to illustrate that the negative effects of direct tunneling can be suppressed by following the developed methodology. The voltage reference was used as a vehicle to prove that analog systems can be constructed with

ultra-thin oxide MOSFETs. Its performance (T_C = 251.0 ppm/°C) was compared to a thick-oxide voltage reference (T_C = 213.7 ppm/°C) as a means of demonstrating that ultra-thin oxide MOSFETs can achieve performance similar to that of thick(er) oxide MOSFETs. The results suggest that the developed methodology can be used to design analog systems with ultra-thin oxide MOSFETs.

A sponsored fabrication of this work was awarded based on technical merit via the MOSIS Educational Program. The target technology was IBM's 10SF technology. The design, simulation, and layout of a 2 mm x 2 mm chip was completed and sent to MOSIS. However, for reasons beyond the author's control, this fabrication was delayed over 2 years. Therefore, fabrication results were unable to be included in this document. If fabrication were to occur after the publishing of this document, it would be recommended that the measurements outlined in Section 5.7 be taken and that the results be made available via a scholarly journal.

APPENDIX A
Low-Frequency Small-Signal Analysis of the Self-Cascode Amplifier

A.1. Derivation of G_M, R_{OUT}, and A_V of the Self-Cascode Amplifier

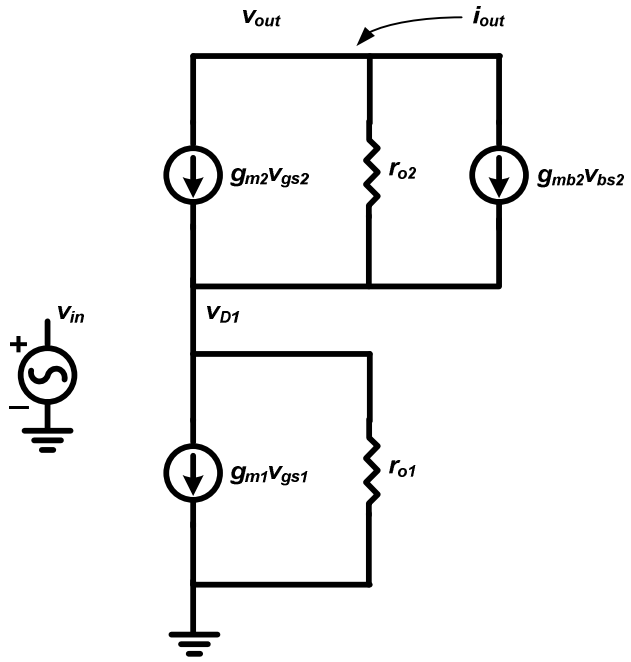

Figure A.1: Low-frequency small-signal equivalent of a self-cascode amplifier.

From [44], the ideal small-signal voltage gain of an amplifier, A_V, is defined as $-G_M \cdot R_{OUT}$, where G_M is the short-circuit transconductance and R_{OUT} is the output resistance. Specifically, G_M and R_{OUT} are defined as [44]:

$$G_M = \left.\frac{i_{out}}{v_{in}}\right|_{v_{out}=0} \tag{A.1}$$

$$R_{OUT} = \left.\frac{v_{out}}{i_{out}}\right|_{v_{in}=0} \tag{A.2}$$

where i_{out} is the small-signal output current, v_{in} is the small-signal input voltage, and v_{out} is the small-signal output voltage. Referring to Figure A.1, R_{OUT} of the self-cascode structure can be solved by using (A.2) to write two expressions for i_{out}:

$$i_{out} = \frac{v_{out} - v_{D1}}{r_{o2}} - g_{m2}v_{D1} - g_{mb2}v_{D1}. \quad \text{(A.3)}$$

$$i_{out} = \frac{v_{D1}}{r_{o1}}. \quad \text{(A.4)}$$

Setting these two equations equal to another and solving for v_{D1} yields:

$$v_{D1} = \frac{v_{out}r_{o1}}{r_{o2} + r_{01}r_{o2}(g_{m2} + g_{mb2}) + r_{o1}}. \quad \text{(A.5)}$$

Plugging (A.5) into (A.4) and solving for v_{out}/i_{out} gives an expression for R_{OUT}:

$$R_{OUT} = \frac{v_{out}}{i_{out}} = r_{o2} + r_{01}r_{o2}(g_{m2} + g_{mb2}) + r_{o1}. \quad \text{(A.6)}$$

To solve for G_M, Figure (A.1) is used to write two expressions for i_{out}:

$$i_{out} = g_{m2}(v_{in} - v_{D1}) - \frac{v_{D1}}{r_{o2}} - g_{mb2}v_{D1}. \quad \text{(A.7)}$$

$$i_{out} = g_{m1}v_{in} + \frac{v_{D1}}{r_{o1}}. \quad \text{(A.8)}$$

Setting these two equations equal to one another and solving for v_{D1} yields:

$$v_{D1} = \frac{v_{in}(g_{m2} - g_{m1})r_{o1}r_{o2}}{r_{o2} + r_{o1}r_{o2}(g_{m2} + g_{mb2}) + r_{o1}}. \quad \text{(A.9)}$$

Plugging (A.9) into (A.8) and solving for i_{out}/v_{in} gives an expression for G_M:

$$G_M = \frac{i_{out}}{v_{in}} = \frac{g_{m1}r_{o1}r_{o2}(g_{m2} + g_{mb2}) + g_{m1}r_{o1} + g_{m2}r_{o2}}{r_{o2} + r_{o1}r_{o2}(g_{m2} + g_{mb2}) + r_{o1}}. \quad \text{(A.10)}$$

Using (A.6) and (A.10) an expression for A_V can be written as [44]:

$$A_V = -G_M R_{OUT} = -[g_{m1}r_{o1}r_{o2}(g_{m2} + g_{mb2}) + g_{m1}r_{o1} + g_{m2}r_{o2}]. \quad \text{(A.11)}$$

APPENDIX B
Sub-1 V Voltage Reference Analyses

B.1. Analysis of Ideal Sub-1 V Bandgap Voltage Reference

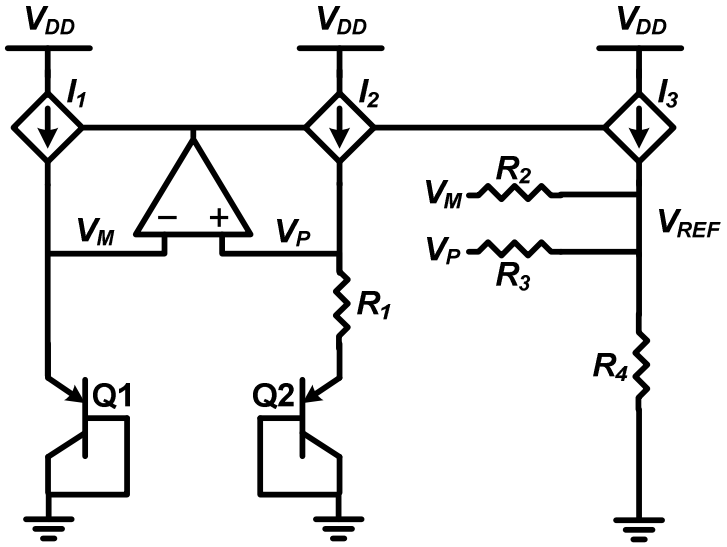

Figure B.1: Simplified representation of the sub-1 V bandgap voltage reference presented in [116].

The current through the emitter of a diode-connected PNP BJT can be approximated as [44]:

$$I_E \approx J_S A_E e^{\frac{V_{EB}}{V_t}} \tag{B.1}$$

where J_S is the saturation current density, A_E is the emitter area, V_{EB} is the emitter-base voltage, and $V_t = kT/q$ is thermal voltage (k = 8.602e−5 eV/K is Boltzmann's constant, T is temperature, and q = 1.602x10^{-19} C is the electronic charge). Referring to Figure B.1, $V_M = V_P = V_{EB1}$. Therefore, the current through R_1 can be written as:

$$I_{R1} = I_{E2} = (V_{EB1} - V_{EB2})/R_1 = \Delta V_{EB}/R_1. \tag{B.2}$$

Given $A_{E2} = N \cdot A_{E1}$ and letting $I_{E2} = I_{E1}$, ΔV_{EB} can be written as [44]:

$$\Delta V_{EB} = V_t \ln(N). \tag{B.3}$$

Therefore, I_{R1} can be expressed as:

$$I_{R1} = V_t \ln(N)/R_1. \tag{B.4}$$

I_{R1} is a component of I_2, which can be written as:

$$I_2 = I_{R1} + I_{R3}. \tag{B.5}$$

I_{R3} can be expressed as:

$$I_{R3} = (V_{EB1} - V_{REF})/R_3. \tag{B.6}$$

Plugging (B.4) and (B.6) into (B.5) yields:

$$I_2 = V_t \ln(N)/R_1 + (V_{EB1} - V_{REF})/R_3. \tag{B.7}$$

Letting $R_2 = R_3$, which implies $I_{R2} = I_{R3}$, and assuming $I_1 = I_2 = I_3$, an equation for I_{R4} can be written as:

$$I_{R4} = I_{R1} + 3I_{R2}. \tag{B.8}$$

Substituting (B.4) and (B.6) into (B.8) and writing an expression for V_{REF} yields:

$$V_{REF} = R_4 \cdot I_{R4} = R_4 \left(\frac{V_t \ln(N)}{R_1} + \frac{3(V_{EB1} - V_{REF})}{R_2} \right). \tag{B.9}$$

Rearranging and solving for V_{REF} gives:

$$V_{REF} = \frac{V_t \ln(N) R_2 R_4 + 3V_{EB1} R_1 R_4}{R_1 R_2 + 3R_1 R_4}. \tag{B.10}$$

Letting $R_4 = \mathrm{M} \cdot R_1$ and $R_2 = B \cdot R_1$, this equation can be simplified to:

$$V_{REF} = \frac{MBV_t \ln(N) + 3MV_{EB1}}{3M + B}. \tag{B.11}$$

Differentiating this equation with respect to temperature and setting the result equal to zero yields:

$$B = -3\left(\frac{\partial V_{EB1}}{\partial T}\Big/\frac{\partial V_t \ln(N)}{\partial T}\right). \tag{B.12}$$

This equation shows *B* is used to zero the temperature slope. To set a desired output voltage, (B.11) can be written in terms of M as:

$$M = \frac{BV_{REF}}{3V_{EB1} + BV_t \ln(N) - 3\mathrm{V_{REF}}}. \tag{B.13}$$

If $V_{EB1} = V_{REF}$, this equation can be written as:

$$M = \frac{V_{EB1}}{V_t \ln(N)}. \tag{B.14}$$

For a given *N* and a desired V_{EB1}, this equation can be used to solve for M.

B.2. Analysis of a Sub-1 V Bandgap Voltage Reference Including Offset Voltage, Input Bias Current, and Input Offset Current

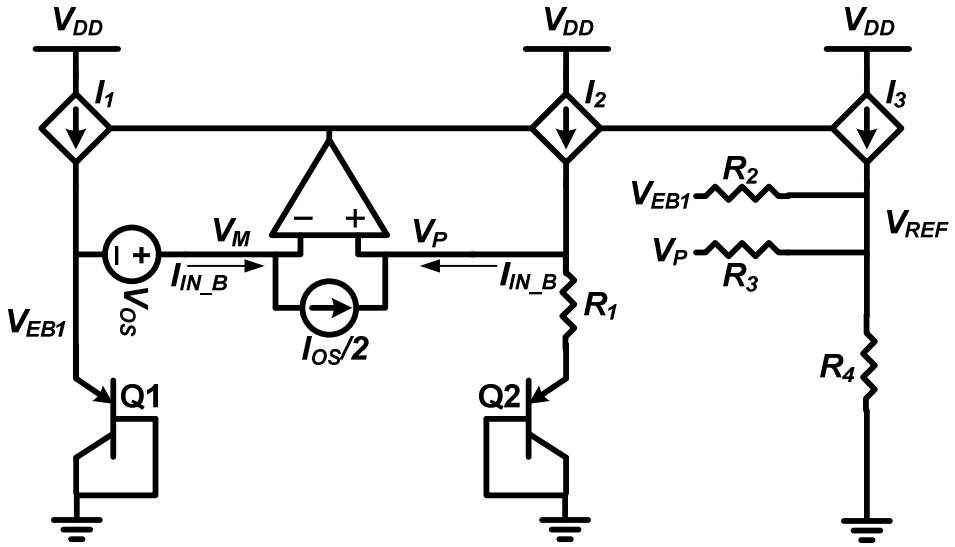

Figure B.2: Simplified representation of the sub-1 V bandgap voltage reference presented in [116]. The schematic includes input offset voltage, input bias current, and input offset current.

Referring to Figure B.2, the input bias current, I_{IN_B}, is defined as:

$$I_{IN\_B} = \frac{I_P + I_N}{2} \tag{B.15}$$

where I_P and I_N are defined as the currents flowing into the non-inverting and inverting input terminals of the amplifier. The input offset current is defined as:

$$I_{OS} = I_P - I_N. \tag{B.16}$$

Using (B.15) and (B.16), expressions for I_P and I_N can be written as:

$$I_P = I_{IN\_B} + \frac{I_{OS}}{2}. \tag{B.17}$$

$$I_N = I_{IN\_B} - \frac{I_{OS}}{2}. \tag{B.18}$$

These currents are taken into account by placing a current source with a value of $I_{OS}/2$ between the non-inverting and inverting terminals of the amplifier. This current source allows one to assume that I_{IN_B} flows out of I_1 and I_2. This allows I_1, I_2, and I_3 to be treated as if they are equal, which simplifies analysis [44]. Therefore, an equation for I_{R4} can be written as:

$$I_{R4} = 2I_{R3} + I_{R2} + I_{R1} + I_{IN\_B}. \tag{B.19}$$

Equations for I_{R1}, I_{R2}, and I_{R3} can be written as:

$$I_{R1} = \frac{V_{EB1} + V_{OS} - V_{EB2}}{R_1} = \frac{\Delta V_{EB} + V_{OS}}{R_1}. \tag{B.20}$$

$$I_{R2} = \frac{V_{EB1} - V_{REF}}{R_2}. \tag{B.21}$$

$$I_{R3} = \frac{V_P - V_{REF}}{R_3} = \frac{V_{EB1} + V_{OS} - V_{REF}}{R_3}. \tag{B.22}$$

Substituting (B.20), (B.21), and (B.22) into (B.19) yields:

$$\frac{V_{REF}}{R_4} = I_{R4} = 2\left(\frac{V_{EB1} + V_{OS} - V_{REF}}{R_3}\right) + \frac{V_{EB1} - V_{REF}}{R_2} + \frac{\Delta V_{EB} + V_{OS}}{R_1} + I_{IN\_B}. \tag{B.23}$$

Rearranging this equation and solving for V_{REF} gives:

$$V_{REF} = \frac{R_4[R_2R_3\Delta V_{EB} + R_1V_{EB1}(R_3 + 2R_2)]}{R_1(R_3R_4 + 2R_2R_4 + R_2R_3)} + \frac{R_2R_4(R_3 + 2R_1)V_{OS}}{R_1(R_3R_4 + 2R_2R_4 + R_2R_3)} + \frac{R_2R_3R_4I_{IN\_B}}{R_3R_4 + 2R_2R_4 + R_2R_3}. \tag{B.24}$$

Letting $R_2 = R_3$ and using (B.3) for ΔV_{EB}, V_{REF} can be expressed as:

$$V_{REF} = \frac{V_t \ln(N)R_2R_4 + 3V_{EB1}R_1R_4}{R_1R_2 + 3R_1R_4} + \frac{V_{OS}R_4(R_2 + 2R_1)}{R_1R_2 + 3R_1R_4} + \frac{I_{IN\_B}R_2R_4}{3R_4 + R_2}. \tag{B.25}$$

Letting $R_2 = B{\cdot}R_1$ and $R_4 = \mathrm{M}{\cdot}R_1$, this equation can be simplified to:

$$V_{REF} = \frac{MBV_t \ln(N) + 3MV_{EB1}}{3M + B} + \frac{V_{OS}M(B + 2)}{3M + B} + \frac{I_{IN\_B}R_1MB}{3M + B}. \tag{B.26}$$

where the first term of this equation is equal to (B.11). The second and third terms represent non-idealities caused by amplifier input offset voltage and amplifier input bias current.

REFERENCES

[1] D. Buss, B. L. Evans, J. Bellay, W. Krenik, B. Haroun, D. Leipold, K. Maggio, J.-Y. Yang, and T. Moise, "SOC CMOS Technology for Persoanl Internet Products," *IEEE Trans. Electron Devices*, vol. 50, no. 3, pp. 546-556, March 2003.

[2] S. Acharya. (2008, September 25) International Telecommunication Union Web Site. [Online]. http://www.itu.int/newsroom/press_releases/2008/29.html

[3] J. Rebello. (2010, September 17) isuppli.com. [Online]. http://www.isuppli.com/Mobile-and-Wireless-Communications/News/Pages/Global-Wireless-Subscriptions-Reach-5-Billion.aspx

[4] W. Krenik, D. D. Buss, and P. Rickert, "Cellular Handset Integration—SIP Versus SOC," *IEEE J. Solid-State Circuits*, vol. 40, no. 9, pp. 1839-1846, September 2005.

[5] G. E. Moore, "Cramming More Components onto Integrated Circuits," *Proc. IEEE*, vol. 86, no. 1, pp. 82-85, January 1998.

[6] Q. Huang, F. Piazza, P. Orsatti, and T. Ohguro, "The Impact of Scaling Down to Deep Submicron on CMOS RF Circuits," *IEEE J. Solid-State Circuits*, vol. 33, no. 7, pp. 1023-1036, July 1998.

[7] T. A. C. M. Classen, "An Industry Perspective on Current and Future State of the Art in System-on-Chip (SoC) Technology," *Proc. IEEE*, vol. 94, no. 6, pp. 1121-1137, June 2006.

[8] H.-S. P. Wong, D. J. Frank, P. M. Solomon, C. H. J. Wannn, and J. J. Welser, "Nanoscale CMOS," *Proc. IEEE*, vol. 87, no. 4, pp. 537-570, April 1999.

[9] B. Murmann, P. Nikaeen, D. J. Connelly, and R. W. Dutton, "Impact of Scaling on Analog Performance and Associated Modeling Needs," *IEEE Trans. Electron Devices*, vol. 53, no. 9, pp. 2160-2167, September 2006.

[10] B. Gilbert, "Analog at Milepost 2000: A Personal Perspective," *Proc. IEEE*, vol. 89, no. 3, pp. 289-304, March 2001.

[11] S. S. Rajput and S. S. Jamuar, "Low Voltage Analog Circuit Design Techniques," *IEEE Circuits and Sytstems Magazine*, vol. 2, no. 1, pp. 24-42, 2002.

[12] J. D. Plummer and P. B. Griffin, "Material and Process Limits in Silicon VLSI Technology," *Proc. IEEE*, vol. 89, no. 3, pp. 240-258, March 2001.

[13] R. van Langevelde, A. J. Scolten, R. Duffy, F. N. Cubaynes, M. J. Knitel, and D. B. M. Klaassen, "Gate current: Modeling, ΔL extraction and impact on RF performance," in *IEDM Tech. Dig.*, 2001, pp. 289-292.

[14] R. van Langevelde, A. J. Scholten, and D. B. M. Klaassen. (2003) Physical Background of MOS Model 11. [Online]. http://www.nxp.com/acrobat_download/other/models/nl_tn2003_00239.pdf

[15] W. Yang, M. V. Dunga, X. Xi, J. He, W. Liu, K. Cao, X. Jin, J. J. Ou, M. Chan, A. M. Niknejad, and C. Hu, "BSIM 4.6.2 MOSFET Model -User's Manual," UC Berkeley, Berkeley, 2008.

[16] K. Roy, S. Mukhopadhyay, and H. Mahmoodi-Meimand, "Leakage Current

Mechanisms and Leakage Reduction Techniques in Deep-Submicrometer CMOS Circuits," *Proc. IEEE*, vol. 91, no. 2, pp. 305-327, February 2003.

[17] (2009) International Technology Roadmap for Semiconductors. [Online]. http://www.itrs.net/

[18] A.-J. Annema, B. Nauta, R. van Langevelde, and H. Tuinhout, "Analog Circuits in Ultra-Deep-Submicron CMOS," *IEEE J. Solid-State Circuits*, vol. 40, no. 1, pp. 132-143, January 2005.

[19] E. Bohannon, C. Washburn, and P.R. Mukund, "Investigating the BJT-like Behavior of MOSFETs in Ultra-Deep-Submicron CMOS Technologies with Significant Gate Current," in *International Semiconductor Device Research Symposium*, College Park, Maryland, 2009, pp. 1-2.

[20] E. Bohannon, C. Washburn, and P. R. Mukund, "Analog IC Design in Ultra-Thin Oxide CMOS Technologies With Significant Direct Tunneling-Induced Gate Current," *IEEE Trans. Circuits Syst. I, Reg. Papers*, To Be Published.

[21] L.L. Lewyn, T. Ytterdal, C. Wulff, and K. Martin, "Analog Circuit Design in Nanoscale CMOS Technologies," *Proc. IEEE*, vol. 97, no. 10, pp. 1687-1714, October 2009.

[22] G. D. Wilk, R. M. Wallace, and J. M. Anthony, "High-κ gate dielectrics: Current Status and Materials Properties Considerations," *Journal of Applied Physics*, vol. 89, no. 10, pp. 5243-5275, May 2001.

[23] D. Lammers. (2008, September 29) Semiconductor.net. [Online]. http://www.semiconductor.net/article/203006-Citing_High_k_Costs_TSMC_Plans_Dual_Track_28_nm_Solutions_in_2010.php

[24] R. Goering. (2009, April 7) ChipeEstimate.com. [Online]. http://www.chipestimate.com/techtalk.php?d=2009-04-07

[25] L. Jelinek. (2009, June 16) isuppli.com. [Online]. http://www.isuppli.com/News/Pages/Is-Moore-s-Law-Becoming-Academic.aspx?

[26] D. Ponton, P. Palestri, D. Esseni, L. Selmi, M. Tiebout, B. Parvais, D. Siprak, and G. Knoblinger, "Design of Ultra-Wideband Low-Noise Amplifiers in 45-nm CMOS Technology: Comparison Between Planar Bulk and SOI FinFET Devices," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 56, no. 5, pp. 920-932, May 2009.

[27] S. Zafar, A. Kumar, E. Gusev, and E. Cartier, "Threshold voltage instabilities in high-κ gate dielectric stacks," *IEEE Transactions on Device and Materials Reliability*, vol. 5, no. 1, pp. 45-64, March 2005.

[28] K. Mistry, C. Allen, B. Beattie, D. Bergstrom, M. Bost, M. Brazier, M. Buehler, A. Cappellanni, R. Chau., and C.-H. Choi, "A 45nm Logic Technology with High-k+Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193nm Dry Patterning, and 100% Pb-free Packaging," in *IEDM Tech. Dig.*, Washington, D.C., 2007, pp. 247-250.

[29] M. LaPedus. (2010, September 27) eetimes.com. [Online]. http://www.eetimes.com/electronics-news/4208861/IBM--fab-club--denies-problems-with-high-k-semiconductor

[30] H. Wu, Y. Zhao, and M. H. White, "Quantum mechanical modeling of MOSFET

gate leakage for high-κ gate electrics," *Solid-State Electronics*, vol. 50, no. 6, pp. 1164-1169, June 2006.

[31] Y.-C. Yeo, T.-J. King, and C. Hu, "MOSFET Gate Leakage Modeling and Selection Guide for Alternative Gate Dielectrics Based on Leakage Considerations," *IEEE Trans. Electron Devices*, vol. 50, no. 4, pp. 1027-1035, April 2003.

[32] J. P. Halter and F. N. Najim, "A gate-level leakage power reduction method for ultra-low-power CMOS circuits," in *IEEE Custom Integrated Circuits Conference*, Santa Clara, 1997, pp. 475-478.

[33] D. Lee, W. Kwong, D. Blaauw, and D. Sylvester, "Analysis and minimization techniques for total leakage considering gate oxide leakage," in *Design Automation Conference*, 2003, pp. 175-180.

[34] F. Hamzaoglu and M. R. Stan, "Circuit-level techniques to control gate leakage for sub-100nm CMOS," in *International Symposium on Low Power Electronics and Design*, 2002, pp. 60-63.

[35] G. A. Rincón-Mora, *Voltage References: From Diodes to Precision High-Order Bandgap Circuits*. Piscataway, United States of America: IEEE Press, 2002.

[36] B. Razavi, "Bandgap References," in *Design of Analog CMOS Integrated Circuits*. New York, United States of America: McGraw Hill, 2001, ch. 11, pp. 377-397.

[37] M.J.M. Pelgrom, A.C.J. Duinmaijer, and A.P.G. Welbers, "Matching properties of MOS transistors," *IEEE J. Solid-State Circuits*, vol. 24, no. 5, pp. 1433-1439, October 1989.

[38] MOSIS Integrated Circuit Fabrication Service. [Online]. www.mosis.com

[39] E. Bohannon, "A Stable Voltage Reference Allowing Supply Voltages Approaching the Forward Voltage of a Silicon Diode with Compensation for Non-Negligible Input Current," Provisional Application, October 23, 2009.

[40] R. J. Baker, *CMOS Circuit Design, Layout, and Simulation*, 2nd ed. Piscataway, United States of America: IEEE Press, 2005.

[41] D. A. Johns and K W. Martin, *Analog Integrated Circuit Design*. United States of America: John Wiley & Sons, 1997.

[42] T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*, 2nd ed. Cambridge, United States of America: Cambridge University Press, 2004.

[43] Y. Tsividis, *Operation and Modeling of The MOS Transistor*, 2nd ed. New York, United States of America: Oxford University Press, 1999.

[44] P. R. Gray, P. J. Hurst, S. H. Lewis, and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, 4th ed. New York, United States of America: John Wiley and Sons, 2001.

[45] R. C. Jaeger and T. N. Blalock, *Microelectronic Circuit Design*, 2nd ed. New York, United States of America: McGraw-Hill, 2004.

[46] R. S. Muller and T. I. Kamins, *Device Electronics for Integrated Circuits*, 3rd ed. New York, United States of America: John Wiley & Sons, 2003.

[47] J. P. Uyemura, *Introduction to VLSI Circuits and Systems*. New York, United

States of America: John Wiley & Sons, 2002.

[48] P. E. Allen and D. R. Holberg, *CMOS Analog Circuit Design*, 2nd ed., A. S. Sedra, Ed. New York, United States of America: Oxford University Press, 2004.

[49] S. M. Sze and K. K. Ng, *Physics of Semiconductor Devices*, 3rd ed. Hoboken, United States of America: John Wiley & Sons, 2007.

[50] K. M. Cao, W. Liu, X. Jin, K. Vasanth, K. Green, J. Krick, T. Vrotsos, and C. Hu, "Modeling of Pocket Implanted MOSFETs for Anomalous Analog Behavior," in *IEDM Tech. Dig.*, Washington DC, 1999, pp. 171-174.

[51] K. Cao, "Advanced Compact Modeling of MOSFETs," University of California, Berkeley, Ph.D. Thesis 2002.

[52] A. Chatterjee, K. Vasanth, D. T. Grider, M. Nandakumar, G. Pollack, R. Aggarwal, M. Rodder, and H. Shichijo, "Transistor Design Issues in Integrating Analog Functions with High Performance Digital CMOS," in *Proc. VLSI Symposium*, 1999, pp. 147-148.

[53] A. P. Chandrakasan and R. W. Broderson, "Minimizing Power Consumption in Digital CMOS Circuits," *Proc. IEEE*, vol. 83, no. 4, pp. 498-523, April 1995.

[54] H. V. Deshpande, B. Cheng, and J. C. S. Woo, "Deep Sub-Micron CMOS device design for Low Power Analog Applications," in *Symposium on VLSI Technology Digest of Technical Papers*, 2001, pp. 87-88.

[55] H. V. Deshpande, B. Cheng, and J. C. S. Woo, "Channel Engineering for Analog Device Design in Deep Submicron CMOS Technology for System on Chip Applications," *IEEE Trans. Electron Devices*, vol. 49, no. 9, pp. 1558-1565, September 2002.

[56] H. V. Deshpande, B. Cheng, and J. C. S. Woo, "Analog Device Design for Low Power Mixed Mode Applications in Deep Submicron CMOS Technology," *IEEE Electron Device Lett.*, vol. 22, no. 12, pp. 588-590, December 2001.

[57] D. D. Buss, "Technology in the Internet Age," in *IEEE ISSCC 2002 Dig. Tech. Papers*, San Francisco, 2002, pp. 18-21.

[58] A.-J. Annema, "Analog Circuit Performance and Process Scaling," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 46, no. 6, pp. 711-725, June 1999.

[59] J. M. Rabaey, F. De Bernardinis, A. M. Niknejad, B. Nikolić, and A. Sangiovanni-Vincentelli, "Embedding Mixed-Signal Design in Systems-on-Chip," *Proc. IEEE*, vol. 94, no. 6, pp. 1070-1088, June 2006.

[60] B. Nauta and A.-J. Annema, "Analog/RF Circuit Design Techniques for Nanometerscale IC Technologies," in *Proc. ESSCIRC*, Grenoble, 2005, pp. 45-53.

[61] M. J. M. Pelgrom and M. Vertregt, "CMOS Technology for Mixed Signal ICs," *Solid-State Electronics*, vol. 41, no. 7, pp. 967-974, July 1997.

[62] A. Mercha, W. Jeamsaksiri, J. Ramos, D. Linten, S. Jenei, P. Wambacq, and S. Decoutere, "Impact of Scaling on Analog/RF CMOS Performance," in *International Conference on Solid-State and Integrated Circuits Technology*, 2004, pp. 147-152.

[63] S. Wong, C. Andre, and T. Salama, "Impact of Scaling on MOS Analog

Performance," *IEEE J. Solid-State Circuits*, vol. 18, no. 1, pp. 106-114, February 1983.

[64] D. J. Frank, R. H. Dennard, E. Nowak, P. M. Solomon, Y. Taur, and H.-S. P. Wong, "Device Scaling Limits of Si MOSFETs and Their Application Dependencies," *Proc. IEEE*, vol. 89, no. 3, pp. 259-288, March 2001.

[65] R. H. Dennard, F. H. Gaensslen, H.-N. Yu, V. L. Rideout, E. Bassous, and A. R. Leblanc, "Design of Ion-Implanted MOSFET's with Very Small Physical Dimensions," *Proc. IEEE*, vol. 87, no. 4, pp. 668-678, April 1999.

[66] Y. Taur, D. A. Buchanan, W. Chen, D. J. Frank, K. E. Ismail, S.-H. Lo, A. Sai-Halasz, R. G. Viswanathan, H.-J. C. Wann, S. J. Wind, and H.-S. Wong, "CMOS Scaling into the Nanometer Regime," *Proc. IEEE*, vol. 85, no. 4, pp. 486-504, April 1997.

[67] F. Fallah and M. Pedram, "Standby and Active Leakage Current Control and Minimization in CMOS VLSI Circuits," *IEICE Transactions on Electronics* , vol. 88, no. 4, pp. 505-519, 2005.

[68] S. Borkar, "Design Challenges of Technology Scaling," *IEEE Micro*, vol. 19, no. 4, pp. 23-29, July-August 1999.

[69] C. Hu, "Future CMOS Scaling and Reliability," *Proc. IEEE*, vol. 81, no. 5, pp. 682-689, May 1993.

[70] M. D. Cave, "Scalable Voltage Reference for Ultra Deep Submicron Technologies ," The University of Texas, Austin, Ph.D. Thesis 2005.

[71] Y. Chiu, B. Nikolić, and P. R. Gray, "Scaling of Analog-to-Digital Converters into Ultra-Deep-Submicron CMOS," in *IEEE Custom Integrated Circuits Conference*, 2005, pp. 375-382.

[72] H.-S. Lee and C. G. Sodini, "Analog-to-Digital Converters: Digitizing the Analog World," *Proc. IEEE*, vol. 96, no. 2, pp. 323-334, February 2007.

[73] L. W. Nagel, "Spice 2: A computer program to stimulate semiconductor circuits," UC Berkeley, Bekeley, 1975.

[74] B P. Wong, A. Mittal, Y. Cao, and G. Starr, *Nano-CMOS Circuit and Physical Design*. Hoboken, United States of America: John Wiley & Sons, 2005.

[75] J. E. Moon, T. Garfinkel, J. Chung, M. Wong, P. K. Ko, and C. Hu, "A new LDD structure: Total overlap with polysilicon spacer (TOPS)," *IEEE Electron Device Letters*, vol. 11, no. 5, pp. 221-223, May 1990.

[76] R. F. M. Roes, A. C. M. C. van Brandenburg, A. H. Montree, and P. H. Woerlee, "Implications of pocket optimisation on analog performance in deep sub-micron CMOS," in *Proc. SSDRC*, 1999, pp. 176-179.

[77] J.-C. Guo, "Halo and LDD Engineering for Multiple VTH High Performance Analog CMOS Devices," *IEEE Transactions on Semiconductor Manufacturing*, vol. 20, no. 3, pp. 313-322, August 2007.

[78] S. Chakraborty, A. Mallik, C. K. Sarkar, and V. R. Rao, "Impact of Halo Doping on the Subthreshold Performance of Deep-Submicrometer CMOS Devices and Circuits for Ultralow Power Analog/Mixed-Signal Applications," *IEEE Trans.*

Electron Devices, vol. 54, no. 2, pp. 241-248, February 2007.

[79] W. Liu, X. Jin, Y. King, and C. Hu, "An Efficient and Accurate Compact Model for Thin-Oxide-MOSFET Intrinsic Capacitance Considering the Finite Charge Layer Thickness," *IEEE Trans. Electron Devices*, vol. 46, no. 5, pp. 1070-1072, May 1999.

[80] S.-H. Lo, D. A. Buchanan, and Y. Taur, "Modeling and Characterization of Quantization, Polysilicon, Depletion, and Direct Tunneling Effects in MOSFETs with Ultrathin Oxides," *IBM J. Res. & Develp.*, vol. 43, no. 3, pp. 327-337, May 1999.

[81] Y-C. King, H. Fujioka, S. Kamohara, and C. Hu, "Dc electrical oxide thickness model for quantization of the inversion layer in MOSFETs," *Semiconductor Science and Technology*, vol. 13, no. 8, pp. 963-966, 1998.

[82] B. Hoeneisen and C. A. Mead, "Fundamental Limitations in Microelectronics–I. MOS Technology," *Solid-State Electronics*, vol. 15, pp. 819-829, 1972.

[83] P. A. Packan, "Pushing the Limits," *Science*, vol. 285, no. 5436, pp. 2079-2081, September 1999.

[84] J. D. Meindl, "Low Power Microelectronics: Retrospect and Prospect," *Proc. IEEE*, vol. 83, no. 4, pp. 619-635, April 1995.

[85] E. J. Nowak, "Maintaining the benefits of CMOS scaling when scaling bogs down," *IBM J. Res. & Dev.*, vol. 46, no. 2/3, pp. 169-180, March/May 2002.

[86] L. Wang, "Quantum Mechanical Effects on MOSFET Scaling Limit," Georgia Institute of Technology, Atlanta, Ph.D. Thesis 2006.

[87] A. P. Chandrakasan, S. Sheng, and R. W. Brodersen, "Low-Power CMOS Digital Design," *IEEE J. Solid-State Circuits*, vol. 27, no. 4, pp. 473-484, April 1992.

[88] J. W. Tschanz, S. G. Narendra, Y. Ye, B. A. Bloechel, S. Borkar, and V. De, "Dynamic sleep transistor and body bias for active leakage power control of microprocessors," *IEEE J. Solid-State Circuits*, vol. 38, no. 11, pp. 1838-1845, November 2003.

[89] R. H. Havemann and J. A. Hutchby, "High-Performance Interconnects: An Integration Overview," *Proc. IEEE*, vol. 89, no. 5, pp. 586-601, May 2001.

[90] R. Ho, K. W. Mai, and M. A. Horowitz, "The Future of Wires," *Proc. IEEE*, vol. 89, no. 4, pp. 490-504, April 2001.

[91] A. Hastings, *The Art of Analog Layout*. Upper Saddle River, United States of America: Prentice-Hall, 2001.

[92] P. G. Drennan, M. L. Kniffin, and D. R. Locascio, "Implications of Proximity Effects for Analog Design," in *IEEE Custom Integrated Circuits Conference*, 2006, pp. 169-176.

[93] J. Croon, W. Sansen, and H. Maes, *Matching properties of deep sub-micron MOS transistors*. Dordecht, The Netherlands: Springer International Series, 2005.

[94] A. Asenov, S. Kaya, and J. H. Davies, "Intrinsic Threshold Voltage Fluctuations in Decanano MOSFETs Due to Local Oxide Thickness Fluctuations," *IEEE Trans. Electron Devices*, vol. 49, no. 1, pp. 112-119, January 2002.

[95] J. B. Johnson, T. B. Hook, and Y.-M. Lee, "Analysis and Modeling of Threshold Voltage Mismatch for CMOS at 65 nm and Beyond," *IEEE Electron Device Letters*, vol. 29, no. 7, pp. 802-804, July 2008.

[96] P. G. Drennan and C. C. McAndrew, "Understanding MOSFET Mismatch for Analog Design," in *IEEE Custom Integrated Circuits Conference*, 2002, pp. 449-452.

[97] P. G. Drennan and C. C. McAndrew, "A Comprehensive MOSFET Mismatch Model," in *IEDM Tech. Dig.*, 1999, pp. 741-744.

[98] C. C. Enz and G. C. Temes, "Circuit Techniques for Reducing the Effects of Op-Amp Imperfections: Autozeroing, Correlated Double Sampling, and Chopper Stabilization," *Proc. IEEE*, vol. 84, no. 11, pp. 1584-1614, November 1996.

[99] K. Agarwal and S. Nassif, "Characterizing Process Variation in Nanometer CMOS," in *Proc. DAC*, San Diego, United States of America, 2007, pp. 396-399.

[100] S. Borkar, "Designing Reliable Systems from Unreliable Components: The Challenges of Transistor Variability and Degradation," *IEEE Micro*, vol. 25, no. 6, pp. 10-16, November 2005.

[101] S. Asai and Y. Wada, "Technology Challenges for Integration Near and Below 0.1 μm," *Proc. IEEE*, vol. 85, no. 4, pp. 505-520, April 1997.

[102] K. Bernstein, D. J. Frank, A. E. Gattiker, W. Haensch, B. L. Ji, S. R. Nassif, E. J. Nowak, D. J. Pearson, and N. J. Rohrer, "High-Performance CMOS Variability in the 65-nm Regime and Beyond," *IBM J. Res. & Dev.*, vol. 50, no. 4/5, pp. 433-449, July/September 2006.

[103] A. Balasubramanian, P. R. Fleming, B. L. Bhuva, O. A. Amusan, and L. W. Massengill, "Effects of Random Dopant Fluctuations (RDF) on the Single Event Vulnerability of 90 and 65 nm CMOS Technologies," *IEEE Transactions on Nuclear Science*, vol. 54, no. 6, pp. 2400-2406, December 2007.

[104] J. Dubois, J. Knol, M. Bolt, H. Tuinhout, J. Schmitz, and P. Stolk, "Impact of source/drain implants on threshold voltage matching in deep sub-micron CMOS technologies," in *Proc. ESSDERC*, 2002, pp. 115-118.

[105] J. A. Croon, E. Augendre, S. Decoutere, W. Sansen, and H. E. Maes, "Influence of Doping Profile and Halo Implantation on the Threshold Voltage Mismatch of a 0.13 μm CMOS Technology," in *Proc. ESSDERC*, 2002, pp. 579-582.

[106] B. H. Calhoun, Y. Cao, X. Li, K. Mai, L. T. Pileggi, R. A. Rutenbar, and K. L. Shepard, "Digital Circuit Design Challenges and Opportunities in the Era of Nanoscale CMOS," *Proc. IEEE*, vol. 96, no. 2, pp. 343-365, February 2008.

[107] T. C. Chen, "Where CMOS is Going: Trendy Hype vs. Real Technology," in *IEEE ISSCC 2006 Dig. Tech. Papers*, 2006, pp. 1-18.

[108] S. Hanson, B. Zhai, K. Bernstein, D. Blaauw, A. Bryant, L. Chang, K. K. Das, W. Haensch, E. J. Nowak, and D. M. Sylvester, "Ultralow-voltage, minimum-engery CMOS," *IBM J. Res. & Dev.*, vol. 50, no. 4/5, pp. 469-490, July/September 2006.

[109] M. Steyaert, V. Peluso, J. Bastots, P. Kinget, and W. Sansen, "Custom Analog Low Power Design: The problem of low voltage and mismatch," in *IEEE Custom Integrated Circuits Conference*, 1997, pp. 285-292.

[110] B. P. Wong, F. Zach, V. Moroz, A. Mittal, G. W. Starr, and A. Kahng, *Nano-CMOS Design for Manufacturability*. Hoboken, United States of America: John Wiley & Sons, 2009.

[111] P. Hasler and T. S. Lande, "Overview of floating-gate devices, circuits, and systems," *IEEE Tran. on Circuits and Systems—II: Analog and Digital Signal Processing*, vol. 48, no. 1, pp. 1-3, January 2001.

[112] B. J. Blalock and P. E. Allen, "A low-voltage, bulk-driven MOSFET current mirror for CMOS technology," in *Proc. ISCAS'95*, 1995, pp. 1972-1975.

[113] S. Yan and E. Sanchez-Sinencio, "Low Voltage Analog Circuit Design Techniques: A Tutorial," *IEICE Trans. Analog Integrated Circuits and Systems*, vol. E00-A, no. 2, pp. 1-17, February 2000.

[114] S. Chatterjee, Y. Tsividis, and P. Kinget, "0.5-V analog circuit techniques and their application in OTA and filter design," *IEEE J. Solid-State Circuits*, vol. 40, no. 12, pp. 2373-2387, December 2005.

[115] C. Urban, J. E. Moon, and P. R. Mukund, "Scaling the Bulk-Driven MOSFET," in *International Conference on Microelectronics*, Marrakech, 2009, pp. 42-45.

[116] C. Washburn. (2005, April 6) A Planet Analog exclusive: A bandgap reference for 90nm and beyond. [Online]. http://www.planetanalog.com/features/power/showArticle.jhtml?articleID=160501575

[117] D. J. Comer, D. T. Comer, and C. S. Petrie, "The utility of the composite cascode in analog CMOS design," *Int. J. Electronics*, vol. 91, no. 8, pp. 491-502, August 2004.

[118] C. G. Montoro, M. C. Schneider, and I. J. B. Loss, "Series-Parallel Association of FET's for High Gain and High Frequency Application," *IEEE J. Solid-State Circuits*, vol. 29, no. 9, pp. 1094-1101, September 1994.

[119] H. S. Momose, M. Ono, T. Yoshitomi, T. Ohguro, S.-i. Nakamure, M. Saito, and H. Iwai, "1.5 nm Direct-Tunneling Gate Oxide Si MOSFET's," *IEEE Trans. Electron Devices*, vol. 43, no. 8, pp. 1233-1242, August 1996.

[120] A. Ghetti, C.-T. Liu, M. Mastrapasqua, and E. Sangiorgi, "Characterization of tunneling current in ultra-thin gate oxide," *Solid-State Electonics*, vol. 44, no. 9, pp. 1523-1531, September 2000.

[121] E. Takeda, C. Y. Yang, and A. Miura-Hamada, *Hot-Carrier Effects in MOS Devices*. San Diego, United States of America: Academic Press, 1995.

[122] E. Takeda, "Hot-carrier effects in submicrometre MOS VLSIs," *IEE Proc. Solid-State and Electronic Devices I*, vol. 131, no. 5, pp. 153-162, October 1984.

[123] J. E. Chung, M.-C. Jeng, J. E. Moon, P. K. Ko, and C. Hu, "Low-voltage hot-electron currents and degradation in deep-submicrometer MOSFETs," *IEEE Trans. Electron Devices*, vol. 37, no. 7, pp. 1651-1657, July 1990.

[124] J. P. McKelvey, *Solid State Physics for Engineering and Materials Science*. Malabar, United States of America: Krieger Publishing Company, 1993.

[125] S. Gasiorowicz, *Quantum Physics*, 3rd ed.: John Wiley & Sons, 2007.

[126] K. F. Schuegraf and C. Hu, "Hole Injection SiO2 Breakdown Model for Very Low

Voltage Lifetime Extrapolation," *IEEE Trans. Electron Devices*, vol. 41, no. 5, pp. 761-767, May 1994.

[127] Z. A. Weinberg, "On tunneling in metal-oxide-silicon structures," *Journal of Applied Physics*, vol. 53, no. 7, pp. 5052-5056, July 1982.

[128] S. Aritome, R. Shirota, G. Hemink, T. Endoh, and F. Masuoka, "Reliability issues of flash memory cells," *Proc. IEEE*, vol. 81, no. 5, pp. 776-788, May 1993.

[129] N. Yang, "A Comparative Study of Gate Direct Tunneling and Drain Leakage Currents in N-MOSFET's with Sub-2-nm Gate Oxides," *IEEE Trans. Electron Devices*, vol. 47, no. 8, pp. 1636-1644, August 2000.

[130] W. K. Henson, N. Yang, S. Kubicek, E. M. Vogel, J. J. Wortman, K. De Meyer, and A. Naem, "Analysis of Leakage Currents and Impact of Off-State Power Consumption for CMOS Technology in the 100-nm regime," *IEEE Trans. Electron Devices*, vol. 47, no. 7, pp. 1393-1400, July 2000.

[131] H.-S. P. Wong, "Beyond the Conventional Transistor," *IBM J. Res. & Dev.*, vol. 46, no. 2/3, pp. 133-168, March/May 2002.

[132] W.-C. Lee and C. Hu, "Modeling CMOS Tunneling Currents Through Ultrathin Gate Oxide Due to Conduction- and Valence-Band Electron and Hole Tunneling," *IEEE Trans. Electron Devices*, vol. 48, no. 7, pp. 1366-1373, July 2001.

[133] N. Yang, W. K. Henson, J. R. Hauser, and J. J. Wortman, "Modeling Study of Ultrathin Gate Oxides Using Direct Tunneling Current and Capacitance-Voltage Measurements in MOS Devices," *IEEE Trans. Electron Devices*, vol. 46, no. 7, pp. 1464-1471, July 1999.

[134] S. H. Lo, D. A. Buchanan, Y. Taur, and W. Wang, "Quantum Mechanical Modeling of Electron Tunneling Current from the Inversion Layer of Ultra-Thin-Oxide nMOSFET's," *IEEE Electron Device Lett.*, vol. 18, no. 5, pp. 209-211, May 1997.

[135] C.-H. Choi, K.-H. Oh, J.-S. Goo, Z. Yu, and R. W. Dutton, "Direct Tunneling Current Model for Circuit Simulation," in *IEDM Tech. Dig.*, 1999, pp. 735-738.

[136] K. M. Cao, W.-C. Lee, W. Liu, X. Jin, P. Su, S. K. H. Fung, J. X. An, B. Yu, and C. Hu, "BSIM4 Gate Leakage Model Including Source-Drain Partition," in *IEDM Tech. Dig.*, 2000, pp. 815-818.

[137] W. Liu, *MOSFET Models for SPICE Simulation including BSIM3v3 and BSIM4*. New York, United States of America: John Wiley & Sons, 2001.

[138] A. Mercha, W. Jeamsaksiri, J. Ramos, D. Linten, S. Jenei, P. Wambacq, and S. Decoutere, "Impact of Scaling on Analog/RF CMOS Performance," in *International Solid-State and Integrated Circuits Technology*, 2004, pp. 147- 152.

[139] D. M. Binkley, *Tradeoffs and Optimization in Analog CMOS Design*. West Sussex, England: John Wiley & Sons, 2008.

[140] E. Kougianos and S. P. Mohanty, "Impact of gate-oxide tunneling on mixed-signal design and simulation of a nano-CMOS VCO," *Microelectronics Journal*, vol. 40, no. 1, pp. 95-103, January 2009.

[141] F. Crupi, P. Magnone, A. Pugliese, and G. Cappuccino, "Performance of current

mirror with high-κ gate dielectrics," *Microelectronic Engineering*, vol. 85, no. 2, pp. 284-288, February 2008.

[142] A. J. Scholten, L. F. Tiemeijer, R. van Langevelde, R. J. Havens, A. T. A. Z.-v. Duijnhoven, and V. C. Venezia, "Noise Modeling for RF CMOS Circuit Simulations," *IEEE Trans. Electron Devices*, vol. 50, no. 3, pp. 618-632, March 2003.

[143] E. Bohannon, C. Urban, M. Pude, Y. Nishi, A. Gopalan, and P. R. Mukund, "Passive and Active Reduction Techniques for On-Chip High-Frequency Digital Power Supply Noise," *IEEE Transactions on VLSI Systems*, vol. 18, no. 1, pp. 157-161, January 2010.

[144] K. F. Schuegraf and C. Hu, "Effects of Temperature and Defects on Breakdown Lifetime of Thin SiO2 at Very Low Voltages," in *Reliability Physics Symposium*, San Jose, 1994, pp. 126-135.

[145] A. Yassine and R. Hijab, "Temperature dependence of gate current in ultra thin SiO2 in direct-tunneling regime," in *1997 IEEE International Integrated Reliability Workshop Final Report*, Lake Tahoe, 1997, pp. 56-61.

[146] P. Häfliger and H. K. O. Berge, "Exploiting Gate Leakage in Deep-Submicrometer CMOS for Input Offset Adaptation," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 54, no. 2, pp. 127-130, February 2007.

[147] H. K. O. Berge and P. Häfliger, "A Gate Leakage Feedback Element in an Adaptive Amplifier Application," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 55, no. 2, pp. 101-105, February 2008.

[148] D. Lee, D. Blaauw, and D. Sylvester, "Gate Oxide Leakage for Current Analysis and Reduction for VLSI Circuits," *IEEE Transactions on VLSI Systems*, vol. 12, no. 2, pp. 155-166, February 2004.

[149] C.-H. Choi, K.-Y. Nam, Z. Yu, and R. W. Dutton, "Impact of Gate Direct Tunneling Current on Circuit Performance: A Simulation Study," *IEEE Trans. Electron Devices*, vol. 48, no. 12, pp. 2823-2829, December 2001.

[150] Y. C. Yeo, Q. Lu, W. C. Lee, T.-J. King, C. Hu, X. Wang, X. Gui, and T. P. Ma, "Direct Tunneling Gate Leakage Current in Transistors with Ultrathin Silicon Nitride Gate Dielectrics," *IEEE Electron Device Lett.*, vol. 21, no. 11, pp. 540-542, November 2000.

[151] B. Cheng, M. Cao, R. Rao, A. Inani, P. V. Voorde, W. M. Greene, J. M. C. Stork, Z. Yu, P. M. Zeitzoff, and J. C. S. Woo, "The Impact of High-κ Gate Dielectrics and Metal Gate Electrodes on Sub-100 nm MOSFET's," *IEEE Trans. Electron Devices*, vol. 46, no. 7, pp. 1537-1544, July 1999.

[152] E. P. Gusev, E. Cartier, D. A. Buchanan, M. Gribelyuk, M. Copel, H. Okorn-Schmidt, and C. D'Emic, "Ultrathin high-K metal oxides on silicon: processing, characterization and integration issues," *Microelectronic Engineering*, vol. 59, no. 1-4, pp. 341-349, November 2001.

[153] E. M. Vogel, K. Z. Ahmed, B. Hornung, W. K. Henson, P. K. McLarty, G. Lucovsky, J. R. Hauser, and J. J. Wortman, "Modeled Tunnel Currents for High Dielectric Constant Dielectrics," *IEEE Trans. Electron Devices*, vol. 45, no. 6,

pp. 1350-1355, June 1998.

[154] Y.-C. Yeo, T.-J. King, and C. Hu, "Direct tunneling leakage current and scalability of alternative gate dielectrics," *Applied Physics Letters*, vol. 81, no. 11, pp. 2091-2093, September 2002.

[155] K. Kuhn, C. Kenyon, A. Kornfeld, M. Liu, A. Maheshwari, W.-k. Shih, S. Sivakumar, G. Taylor, P. VanDerVoorn, and K. Zawadzki, "Managing Process Variation in Intel's 45nm CMOS Technology," *Intel Technology Journal*, vol. 12, no. 2, pp. 93-110, February 2008.

[156] R. J. Widlar, "New Developments in IC Voltage Regulators," *IEEE J. Solid-State Circuits*, vol. 6, no. 1, pp. 2-7, February 1971.

[157] I. M. Filanovsky and A. Allam, "Mutual Compensation of Mobility and Threshold Voltage Temperature Effects with Applications in CMOS Circuits," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 48, no. 7, pp. 876-884, July 2001.

[158] Y. P. Tsividis, "Accurate Analysis of Temperature Effects in IC-VBE Characteristics with Application to Bandgap Reference Sources," *IEEE J. Solid-State Circuits*, vol. SC-15, no. 6, pp. 1076-1084, December 1980.

[159] B.-S. Song and P. R. Gray, "A Precision Curvature-Compensated CMOS Bandgap Reference," *IEEE J. Solid-State Circuits*, vol. SC-18, no. 6, pp. 634-643, December 1983.

[160] P. Malcovati, F. Maloberti, C. Fiocchi, and M. Pruzzi, "Curvature-Compensated BiCMOS Bandgap with 1-V Supply Voltage," *IEEE J. Solid-State Circuits*, vol. 36, no. 7, pp. 1076-1081, July 2001.

[161] J. Michejda and S. K. Kim, "A Precision CMOS Bandgap Reference," *IEEE J. Solid-State Circuits*, vol. SC-19, no. 6, pp. 1014-1021, December 1984.

[162] A. P. Brokaw, "A simple three-terminal IC bandgap reference," *IEEE J. Solid-State Circuits*, vol. 9, no. 6, pp. 388-393, December 1974.

[163] Y. P. Tsividis, "A CMOS Voltage Reference," *IEEE J. Solid-State Circuits*, vol. SC-13, no. 6, pp. 774-778, December 1978.

[164] K. E. Kuijk, "A Precision Voltage Source," *IEEE J. Solid-State Circuits*, vol. SC-8, no. 3, pp. 222-226, June 1973.

[165] E. A. Vittoz, "MOS Transistors Operated in the Lateral Bipolar Mode and Their Application in CMOS Technology," *IEEE J. Solid-State Circuits*, vol. SC-18, no. 3, pp. 273-279, June 1983.

[166] A. Boni, "Op-Amps and Startup Circuits for CMOS Bandgap References With Near 1-V Supply," *IEEE J. Solid-State Circuits*, vol. 37, no. 10, pp. 1339-1343, October 2002.

[167] V. Gupta and G. A. Rincón-Mora, "Predicting The Effects of Error Sources in Bandgap Reference Circuits and Evaluating Their Design Implications," in *Proc. MWSCAS*, 2002, pp. III-575- III-578.

[168] V. Gupta and G. A. Rincón-Mora, "Predicting and Designing for the Impact of Process Variations and Mismatch on the Trim Range and Yield of Bandgap References," in *Proc. ISQED*, 2005, pp. 503- 508.

[169] S. Sengupta, L. Carastro, and P. E. Allen, "Design Considerations in Bandgap References Over Process Variations," in *Proc. ISCAS'05*, 2005, pp. 3869-3872.

[170] G. V. Ceekala, L. D. Lewicki, J. B. Wieser, D. Varadarajan, and J. Mohan, "A Method for Reducing the Effects of Random Mismatches in CMOS Bandgap References," in *IEEE ISSCC 2002 Dig. Tech. Papers*, 2002, pp. 392-393.

[171] J. P. M. Brito, H. Klimach, and S. Bampi, "A Design Methodology for Matching Improvement in Bandgap References," in *Proc. ISQED*, 2007, pp. 586-594.

[172] V. Gupta and G. A. Rincón-Mora. (2006, March 24) Reduce transistor mismatch errors without costly trimming and noisy chopping schemes. [Online]. http://www.planetanalog.com/showArticle.jhtml?articleID=184400051&queryText=son

[173] P. R. Kinget, "Device Mismatch: An Analog Design Perspective," in *Proc. ISCAS'07*, 2007, pp. 1245-1248.

[174] A.-J. Annema, "Low-Power Bandgap References Featuring DTMOST's," *IEEE J. Solid-State Circuits*, vol. 34, no. 7, pp. 949-955, July 1999.

[175] K. N. Leung and P. T. Mok, "A CMOS Voltage Reference Based on Weighted ΔVGS for CMOS Low-Dropout Linear Regulators," *IEEE J. Solid-State Circuits*, vol. 38, no. 1, pp. 146-150, January 2003.

[176] K. N. Leung, P. K. T. Mok, and C. Y. Leung, "A 2-V 23-μA 5.3-ppm/°C Curvature-Compensated CMOS Bandgap Voltage Reference," *IEEE J. Solid-State Circuits*, vol. 38, no. 3, pp. 561-564, March 2003.

[177] R. T. Perry, S. H. Lewis, A. P. Brokaw, and T. R. Viswanathan, "A 1.4 V Supply CMOS Fractional Bandgap Reference," *IEEE J. Solid-State Circuits*, vol. 42, no. 10, pp. 2180-2186, October 2007.

[178] A. Becker-Gomez, T. L. Viswanathan, and T. R. Viswanathan, "A Low-Supply-Voltage CMOS Sub-Bandgap Reference," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 55, no. 7, pp. 609-613, July 2008.

[179] Y. Jiang and E. K. F. Lee, "Design of Low-Voltage Bandgap Reference Using Transimpedance Amplifier," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 47, no. 6, pp. 552-555, June 2000.

[180] V. Srinivasan, G. Serrano, C. M. Twigg, and P. Hasler, "A Compact Programmable CMOS Reference With ±40μV Accuracy," in *IEEE Custom Integrated Circuits Conference*, 2008, pp. 611-614.

[181] T. Borejko and W. A. Pleskacz, "A Resistorless Voltage Reference Source for 90 nm CMOS Technology with Low Sensitivity to Process and Temperature Variations," in *Proc. DDECS*, 2008, pp. 1-6.

[182] S. Ying, L. Wengao, C. Zhongjian, G. Jun, T. Ju, and J. Lijiu, "A Precise Compensated Bandgap Reference without Resistors," in *International Conference on Solid-State and Integrated Circuits Technology*, 2004, pp. 1583-1586.

[183] X. Xing, Z. Wang, and D. Li, "A Low Voltage High Precision CMOS Bandgap Reference," in *Proc. 25th Norchip Conference*, Aalborg, 2007, pp. 1-4.

[184] S. F. Ashrafi, S. M. Atarodi, and M. Chahardori, "A New Low Voltage, High PSRR, CMOS Bandgap Voltage Reference," in *Proc. IEEE SOCC*, 2008,

pp. 345-348.

[185] G. Giustolisi, G. Palumbo, M. Criscione, and F. Cutri, "A Low-Voltage Low-Power Voltage Reference Based on Subthreshold MOSFETs," *IEEE J. Solid-State Circuits*, vol. 38, no. 1, pp. 151-154, January 2003.

[186] G. De Vita and G. Iannaccone, "A Sub-1-V, 10 ppm/°C, Nanopower Voltage Reference Generator," *IEEE J. Solid-State Circuits*, vol. 42, no. 7, pp. 1536-1542, July 2007.

[187] F. Bedeschi, E. Bonizzoni, A. Fantini, C. Resta, and G. Torelli, "A Low-Power Low-Voltage Mosfet-Only Voltage Reference," in *Proc. ISCAS'04*, 2004, pp. I-57-I-60.

[188] G. Di Naro, G. Lombardo, C. Paolino, and G. Lullo, "A Low-Power Fully-MOSFET Voltage Reference Generator for 90 nm CMOS Technology," in *Proc. ICICDT*, 2006, pp. 1-4.

[189] H. C. Lai and Z. M. Lin, "An Ultra-Low Temperature-Coefficient CMOS Voltage Reference," in *Proc. EDSSC*, 2007, pp. 369-372.

[190] J. Ma, Y. Li, C. Zhang, and Z. Wang, "A 1V Ultra-Low Power High Precision CMOS Voltage Reference," in *Proc. EDSSC*, 2007, pp. 847-850.

[191] L. Testa, H. Lapuyade, M. Cimino, Y. Deval, J. L. Carbonero, and J. B. Begueret, "A Bulk-Controlled Temperature and Power Supply Independent CMOS Voltage Reference," in *Proc. TAISA*, 2008, pp. 109-112.

[192] C. Washburn, "Low-Voltage Bandgap Voltage Reference Circuit," U.S. Patent 7,113,025, September 26, 2006.

[193] H. Banba, H. Shiga, A. Umezawa, T. Miyaba, T. Tanzawa, S. Atsumi, and K. Sakui, "A CMOS Bandgap Reference Circuit with Sub-1-V Operation," *IEEE J. Solid-State Circuits*, vol. 34, no. 5, pp. 670-674, May 1999.

[194] K. N. Leung and P. K. T. Mok, "A Sub-1-V 15-ppm/°C CMOS Bandgap Voltage Reference Without Requiring Low Threshold Voltage Device," *IEEE J. Solid-State Circuits*, vol. 37, no. 4, pp. 526-530, April 2002.

[195] J. Doyle, Y. J. Lee, Y.-B. Kim, H. Wilsch, and F. Lombardi, "A CMOS Subbandgap Reference Circuit with 1-V Power Supply Voltage," *IEEE J. Solid-State Circuits*, vol. 39, no. 1, pp. 252-255, January 2004.

[196] M.-D. Ker and J.-S. Chen, "New Curvature-Compensation Technique for CMOS Bandgap Reference With Sub-1-V Operation," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 53, no. 8, pp. 667-671, August 2006.

[197] K. Lasanen, V. Korkala, E. Räisänen-Ruotsalainen, and J. Kostamovaara, "Design of A 1-V Low Power CMOS Bandgap Reference Based on Resistive Subdivision," in *Proc. IEEE MWSCAS Symposium*, 2002, pp. III-564-III-567.

[198] K. Sanborn, D. Ma, and V. Ivanov, "A Sub-1-V Low-Noise Bandgap Voltage Reference," *IEEE J. Solid-State Circuits*, vol. 42, no. 11, pp. 2466-2481, November 2007.

[199] K. Pan, J. Wu, and P. Wang, "A High Precision CMOS Bandgap Reference," in *International Conference on ASIC*, 2007, pp. 692-695.

[200] S. Mukhopadhyay and K. Roy, "Accurate Modeling of Transistor Stacks to Effectively Reduce Total Standby Leakage in Nano-Scale CMOS Circuits," in *Symposium on VLSl Circuits Digest of Technical Papers*, 2003, pp. 53-56.

[201] H. Camenzind, *Designing Analog Chips*, 2nd ed. College Station, United States of America: Virtualbookworm.com Publishing, 2005.

[202] F. Chen, F. Ungar, A. H. Fischer, J. Gil, A. Chinthakindi, T. Goebel, M. Shinosky, D. Coolbaugh, V. Ramachandran, Y. K. Siew, E. Kaltalioglu, S. O. Kim, and K. Park, "Reliability Characterization of BEOL Vertical Natural Capacitor using Copper and Low-k SiCOH Dielectric for 65nm RF and Mixed-Signal Applications," in *IEEE International Reliability Physics Symposium*, San Jose, CA, 2006, pp. 490-495.

[203] (2008) www.rohm.com. [Online]. http://www.rohm.com/products/databook/pack/pdf/qfp44.pdf

[204] O. Semenov, H. Sarbishaei, and M. Sachdev, *ESD protection device and circuit design for advanced CMOS technologies*.: Springer, 2008.