**Distortion Analysis of CMOS Analog Integrated Circuits
Operating in the Moderate Inversion Region and
Implications for RF Applications**

by

William Toole, M.A.Sc., B. Eng.

A thesis submitted to the
Faculty of Graduate Studies and Research
in partial fulfillment of the requirements
for the Degree of

Doctor of Philosophy

in

Electrical Engineering

Ottawa-Carleton Institute for Electrical Engineering

Department of Electronics

Faculty of Engineering

Carleton University

Ottawa, Canada

August, 2004

© copyright

2004, William Toole

Library and
Archives Canada

Bibliothèque et
Archives Canada

Published Heritage
Branch

Direction du
Patrimoine de l'édition

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 0-612-97848-6

Our file Notre référence

ISBN: 0-612-97848-6

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract

Distortion analysis of MOSFET analog integrated circuits operating in the moderate inversion region and the implications for RF circuits are investigated in this thesis. With the advent of deep submicron MOSFET devices, the modern CMOS transistor has sufficient current gain, even in the moderate inversion region, for many RF applications. However, the linearity of a MOSFET operating in the moderate inversion region is largely unknown. As such, a “single-piece” expression for the drain current of a MOSFET is derived for distortion analysis. The model is valid in the saturation region over all regions of inversion and is used to evaluate the distortion components of various circuit configurations by means of Volterra Series representation. Analysis and measurements reveal a significant peaking, or “sweet-spot” of the third-order intercept point in the moderate inversion region. As a result, a significant increase in linearity with low power consumption is possible. Analysis and measurements are presented that show the dependence of distortion components on the frequency of operation, as well as device and circuit transistor parameters. These parameters are investigated and the implications for RF circuit applications are addressed.

Acknowledgements

The work described in this thesis could not have been accomplished without the help and support of many people. Foremost, I would like to thank my research advisor Professor Calvin Plett for his excellent advice and guidance during the past several years. I would not be where I am now without your help. Thanks.

I am grateful to Professor Sorin Voinigescu of the University of Toronto for his very useful comments on my thesis as well as the other members of my thesis examining board, Professor Emad Gad of the University of Ottawa, Professor Victor Aitken and Professor Tad Kwasniewski, both from Carleton University.

Of my friends and colleagues at Carleton, Professor Garry Tarr was very helpful on many occasions, especially with discussions on CMOS process and modelling issues. Thanks are also owed to Professor Tad Kwasniewski for his help and guidance.

A huge round of thanks is also owed to the guys at Skyworks Systems Ottawa Design Centre for their help and support. Most notably, Mark Cloutier provided a huge assistance with his guidance and insights into many of the issues I faced. You are an excellent mentor and I hope to be like you when I grow up. Many thanks also goes to Christian Cojocaru for being a driving force with this project and for the help with the design and layout of the various circuits. Then there is Tolga Pamir for his assistance and suggestions for the testing of the circuits, Chris Karabatsos for the help with the layout, Viorel Vaduva for assisting with the design and layout of the test boards, Oleg Kartsev for the building of the test boards, and Ardeshir Namdar for pushing me to get this thesis done. Yes Ardeshir, it is finally done. And finally, I owe Ed MacRobbie many thanks for allowing me access to all the excellent help at Skyworks and for supporting this project. Thanks Ed.

I am very grateful for financial assistance received in the form of scholarships, research and equipment grants from the National Science and Engineering Research Council Graduate Scholarship, MICRONET, a Canadian Centre of Excellence, Canadian Microelectronics Corporation, and from Skyworks Solutions Inc.

Finally, I owe a huge thanks to my wife, without whose support, my life would be a complete shambles. A life time of lawn mowing and vacuuming are owed.

Contents

Abstract	ii
Acknowledgements	iii
List of Figures	viii
List of Tables	xvi
List of Abbreviations and Symbols	xviii
1 Introduction	1
1.1 Introduction	1
1.2 Motivation	1
1.3 Objective	3
1.4 Contributions	4
1.5 Organization	4
2 Linearity Concepts	6
2.1 Introduction	6
2.2 Definition of a Linear System	6
2.3 Harmonic Distortion from Single-tone Inputs	7
2.3.1 1 dB Compression Point	10
2.4 Harmonic Distortion from Two-tone Inputs	11
2.4.1 Third Order Intercept Point	13
2.5 Power Series Expansions of Nonlinear Transconductance	15
2.5.1 One-dimensional Transconductance	16
2.5.2 Two-dimensional Transconductance	17
2.5.3 Three-dimensional Transconductance	19
2.6 Power Series Expansions of Nonlinear Capacitors	22

2.7	Summary	23
3	Volterra Series Analysis	25
3.1	Introduction	25
3.2	Basics of Volterra series	26
3.2.1	Time Domain Representation	27
3.2.2	Frequency Domain Representation	29
3.3	Nonlinear Response to Two-Tone Input	30
3.4	Cascade Connection of Nonlinear Systems	32
3.5	Linear Feedback	35
3.6	Calculation of Volterra Kernels	37
3.6.1	Direct Calculation of Nonlinear Responses	38
3.7	Summary	48
4	MOSFET I-V Models	49
4.1	Introduction	49
4.2	Overview of a MOSFET Transistor	50
4.2.1	MOSFET Transistor Operation	50
4.2.2	MOSFET Transistor Characteristics	52
4.3	MOSFET Modelling Issues	54
4.4	A Simple Moderate Inversion Nonlinear Model	56
4.4.1	Long Channel Model	56
4.4.2	Short Channel Effects	59
4.4.3	Source-Drain Resistance	66
4.4.4	Final Model for Saturation Region in Moderate Inversion	68
4.5	Capacitors in a MOSFET	76
4.5.1	Extrinsic Capacitors	76
4.5.2	Intrinsic Capacitances	78
4.6	Evaluation of Nonlinear Coefficients	79
4.6.1	Expressions for $c_{(1,0,0)}$, $c_{(2,0,0)}$, and $c_{(3,0,0)}$	81
4.6.2	Expressions for $c_{(0,1,0)}$, $c_{(0,2,0)}$, and $c_{(0,3,0)}$	96
4.6.3	Expressions for $c_{(0,0,1)}$, $c_{(0,0,2)}$, and $c_{(0,0,3)}$	98
4.6.4	Expressions for the cross-modulation terms	102
4.7	Summary	104

5	Distortion Analysis and Measurement of Weakly Nonlinear Amplifiers in Moderate Inversion	107
5.1	Introduction	107
5.2	Common Source Amplifier	108
5.2.1	Influence of g_m Only	108
5.2.2	Influence of Output Conductance	110
5.2.3	Influence of Load Resistance	117
5.2.4	Influence of a Tuned Load	120
5.2.5	Influence of C_{gd}	123
5.2.6	Influence of a Nonlinear C_{gs} and C_{db}	131
5.3	Source Degenerated Amplifier	139
5.3.1	First-Order Response	140
5.3.2	Third-Order Response	144
5.4	Measurement and Simulation Results	147
5.4.1	Influence of Device Width	152
5.4.2	Influence of Technology Scaling	152
5.4.3	Influence of Load Resistance and Gain	154
5.4.4	Influence of Drain-Source Bias	154
5.4.5	Influence of Signal Level	155
5.4.6	Additional Results	156
5.5	Common-gate and Cascode Amplifier	162
5.6	An LNA Design for Bluetooth	163
5.6.1	Differential LNA Design	163
5.6.2	LNA Measurements	165
5.7	Summary	170
6	Conclusions	172
6.1	Summary	172
6.2	Contributions	174
6.2.1	Publications	175
6.3	Recommendations for Future Research	176
	Appendix A Taylor Series Expansions	178
A.1	Taylor Series Expansion of a Single Tone	178

A.2	Taylor Series Expansion of Two Tones	180
A.2.1	Unequal Amplitudes	180
A.2.2	Equal Amplitudes	181
Appendix B	Basics of Volterra Series	182
B.1	Sinusoidal Response of a Second-Order Volterra Series System	182
B.2	Sinusoidal Response of a Third-Order Volterra Series System	184
B.3	Synthesis of Volterra Operators	186
B.3.1	Basic Second-Order System	186
B.3.2	Basic Third-Order System	188
B.4	Cascade Connection of Nonlinear Systems	189
B.4.1	Fourier Transforms	192
B.5	Nonlinear Circuit Analysis with Volterra Series	196
B.5.1	First-order Kernels	197
B.5.2	Second-order Kernels	199
B.5.3	Third-order Kernels	201
Appendix C	Evaluation of Derivatives of I_{ds}	204
C.1	Derivatives for $c_{(1,0,0)}$, $c_{(2,0,0)}$, and $c_{(3,0,0)}$	204
C.2	Derivatives for $c_{(0,1,0)}$, $c_{(0,2,0)}$, and $c_{(0,3,0)}$	207
C.3	Derivatives for $c_{(0,0,1)}$, $c_{(0,0,2)}$, and $c_{(0,0,3)}$	208
References		217

List of Figures

1.1	Measured V_{IIP3} using two tones spaced at 1 MHz centred at 10 MHz and 2 GHz versus I_{DS} of a CMOS common-source amplifier with $W/L = 127.5\mu\text{m}/0.18\mu\text{m}$	2
2.1	Plot of output voltage (in dBV) of the fundamental tone versus input voltage and 1 dB compression point.	10
2.2	The output spectrum of a weakly nonlinear system excited by two sinusoidal inputs at ω_1 and ω_2	12
2.3	Plot of input and output amplitudes of the fundamental and $IM3$ versus input voltage and extrapolated $IIP3$ point.	14
3.1	Block level system characterized by a Volterra Series.	27
3.2	The system $\mathbf{G}$ formed by the cascade connection of $\mathbf{H}$ and $\mathbf{F}$	33
3.3	The system $\mathbf{G}$ formed by the cascade connection of $\mathbf{H}$ and $\mathbf{F}$. The arrows indicate the two contributions to the second-order kernel of the overall system.	34
3.4	A general linear feedback system.	35
3.5	(a) A common-source amplifier. (b) AC-equivalent circuit.	39
3.6	Linearized circuit for calculating first-order kernels.	40
3.7	Linearized circuit for calculating second-order kernels.	41
4.1	An n-channel MOSFET transistor in the inversion region with negative bias between the body and the source.	51
4.2	$\text{Log}(I_{DS})$ - V_{GS} characteristics of an n-channel MOSFET for a fixed V_{DS}	52
4.3	I_{DS} - V_{DS} characteristics for various V_{GS} . (a) Linear y axis; (b) Logarithmic y axis	53
4.4	Nonlinear model of a MOSFET.	56
4.5	Surface mobility vs. transverse field.	60

4.6	Typical measured electron drift velocity versus longitudinal electric field.	62
4.7	Channel at (a) pinch off; (b) above pinch-off.	64
4.8	First-order modeling of the saturation region analogous to Early voltage.	65
4.9	(a) Components of the series resistance in a MOSFET, and (b) equivalent lumped model.	66
4.10	A lightly doped drain structure.	67
4.11	A MOS transistor including parasitic source and drain resistances. . .	67
4.12	Comparison of measured and calculated I_{DS} for a typical $0.18\mu\text{m}$ transistor using (4.38) with $V_{DS} = 1\text{V}$, $V_{SB} = 0\text{V}$ and with the parameters listed in Table 4.1.	70
4.13	Comparison of measured, simulated, and calculated $\log I_{DS}$ and derivatives versus V_{GS} for a typical $0.18\mu\text{m}$ transistor using (4.38) with the parameters listed in Table 4.1.	71
4.14	Comparison of measured and analytical solution of I_{DS} versus V_{DS} using (4.38) with $V_{GS} = 0, 4, 0.5, 0.6$ and 0.7 V , $V_{SB}=0\text{V}$ and with the parameters listed in Table 4.1 on (a) linear scale and (b) on logarithmic scale.	72
4.15	Comparison of $long$ versus V_{GS} for $V_{DS} = -0.3, 0$ and $+0.3\text{V}$ using the parameters listed in Table 4.1.	73
4.16	Comparison of $mobhot$ versus V_{GS} for $\alpha = 0$ (no short channel effects), $0.5, 1$, and 2 V^{-1} using the parameters listed in Table 4.1.	74
4.17	Comparison of $mobhot$ versus V_{GS} for $V_{SB} = -0.3, 0$, and 0.3 V with $\alpha = 2.0\text{V}^{-1}$ and using the parameters listed in Table 4.1.	75
4.18	Comparison of V_A versus V_{GS} for different V_{Asat} , V_{ACLM} , and P_{VAG} with $V_{DS} = 600\text{mV}$ and using the parameters listed in Table 4.1. . . .	76
4.19	Comparison of clm versus V_{GS} for different V_{Asat} and V_{ACLM} with $V_{DS} = 600\text{mV}$ and using the parameters listed in Table 4.1.	77
4.20	Intrinsic capacitances $C_{gs,i}$, $C_{gb,i}$, and $C_{sb,i}$ versus V_{GS} with $V_{DS} = 600\text{mV}$, $V_{SB} = 0\text{V}$ and using the parameter listed in Table 4.1.	80
4.21	An evaluation of the three terms in (4.61) versus V_{GS} with $V_{DS} = 600\text{mV}$ and $V_{SB} = 0$ for parameters listed in Table 4.1	82

4.22	$c_{(1,0,0)}$ versus V_{GS} using the parameters listed in Table 4.1 for (a) $V_{SB} = -300, 0,$ and $+300$ mV with $V_{DS} = 600$ mV and (b) $V_{DS} = 400, 600,$ and 800 mV with $V_{SB} = 0$	83
4.23	$c_{(1,0,0)}$ versus V_{GS} for different L_{eff}/W_{eff} with $V_{DS} = 600$ mV, $V_{SB} = 0$ and using the parameters listed in Table 4.1.	84
4.24	$c_{(1,0,0)}$ versus V_{GS} for different values of V_{th} with $V_{DS} = 600$ mV and $V_{SB} = 0$ using the parameters listed in Table 4.1.	84
4.25	$c_{(1,0,0)}$ versus V_{GS} for different values of n with $V_{DS} = 600$ mV and $V_{SB} = 0$ using the parameters listed in Table 4.1.	85
4.26	$c_{(1,0,0)}$ versus V_{GS} for different values of α with $V_{DS} = 600$ mV and $V_{SB} = 0$ using the parameters listed in Table 4.1.	85
4.27	An evaluation of the six terms in (4.64) versus V_{GS} with $V_{DS} = 600$ mV and $V_{SB} = 0$ for parameters listed in Table 4.1	87
4.28	$c_{(2,0,0)}$ versus V_{GS} for different values of V_{th} with $V_{DS} = 600$ mV and $V_{SB} = 0$ using the parameters listed in Table 4.1.	88
4.29	$c_{(2,0,0)}$ versus V_{GS} for different values of n with $V_{DS} = 600$ mV and $V_{SB} = 0$ using the parameters listed in Table 4.1.	88
4.30	$c_{(2,0,0)}$ versus V_{GS} for different values of α with $V_{DS} = 600$ mV and $V_{SB} = 0$ using the parameters listed in Table 4.1.	89
4.31	An evaluation of the terms in (4.67) versus V_{GS} with $V_{DS} = 600$ mV and $V_{SB} = 0$ for parameters listed in Table 4.1	90
4.32	A comparison of (4.68) to the full expression (4.67) versus V_{GS} with $V_{DS} = 600$ mV and $V_{SB} = 0$ for parameters listed in Table 4.1	91
4.33	$\partial long/\partial V_{GS}$ versus V_{GS} using the parameters listed in Table 4.1.	93
4.34	$c_{(3,0,0)}$ versus V_{GS} for different values of α with $V_{DS} = 600$ mV and $V_{SB} = 0$ using the parameters listed in Table 4.1.	94
4.35	$c_{(3,0,0)}$ versus V_{GS} for different values of V_{th} with $V_{DS} = 600$ mV and $V_{SB} = 0$ using the parameters listed in Table 4.1.	94
4.36	$c_{(3,0,0)}$ versus V_{GS} for different values of n with $V_{DS} = 600$ mV and $V_{SB} = 0$ using the parameters listed in Table 4.1.	95
4.37	$c_{(3,0,0)}$ versus V_{GS} for different values of P_{VAG} with $V_{DS} = 600$ mV, $V_{SB} = 0$ and using the parameters listed in Table 4.1.	95

4.38	A comparison of the measured and calculated output conductance versus V_{DS} with $V_{SB} = 0$ for parameters listed in Table 4.1	97
4.39	Measured $c_{(0,2,0)}$ versus V_{DS} with $V_{SB} = 0$ using a Savitzky Golay smoothing filter of fourth order with 20 coefficients.	98
4.40	g_{mb} versus V_{GS} for $V_{SB} = 0, 0.3$, and 0.6 V using the parameters listed in Table 4.1.	100
4.41	$c_{(0,0,2)}$ versus V_{GS} for $V_{SB} = 0, 0.3$, and 0.6 V using the parameters listed in Table 4.1.	101
4.42	$c_{(0,0,3)}$ versus V_{GS} for $V_{SB} = 0, 0.3$, and 0.6 V using the parameters listed in Table 4.1.	103
4.43	(a) $c_{(1,1,0)}$ and (b) $c_{(2,1,0)}$ versus V_{GS} with $V_{DS} = 600\text{mV}$, $V_{SB}=0\text{V}$, 300mV , and 600mV and with the parameters listed in Table 4.1. . . .	105
5.1	(a) A common-source amplifier. (b) AC-equivalent circuit.	109
5.2	V_{IIP3}^2 , given by (5.2), versus V_{GS} when considering only the influence of g_m . The parameters used for (5.2) are listed in Table 4.1 and $V_{SB} = 0\text{V}$	110
5.3	Equivalent circuits of the common source amplifier of Figure 5.1 including the influence of g_o for determining (a) the fundamental response, (b) the second-order response, and (c) the third-order response. . . .	111
5.4	Normalized third-order intermodulation product of the output voltage of a common source amplifier versus R_L and the corresponding A_v with $f = 0.01f_T$, $V_{GS} = 600\text{mV}$, $V_{DS} = 600\text{mV}$, $V_{SB} = 0\text{V}$, and using the parameters listed in Table 4.1.	117
5.5	(a) Amplitude and (b) phase response of the normalized $IM3$ versus f/f_T with $V_{GS} = 600\text{mV}$, $V_{DS} = 600\text{mV}$, $V_{SB} = 0\text{V}$, $R_L = 500\Omega$ and with the parameters listed in Table 4.1.	118
5.6	Normalized third-order intermodulation product of the output voltage of a common source amplifier versus V_{GS} with $R_L = 100\Omega$, $V_{DS} = 600\text{mV}$, $V_{SB} = 0\text{V}$, and $v_s = 1\text{V}$. The parameters listed in Table 4.1 are used for determining the nonlinear coefficients.	119
5.7	$IIP3$ of a common source amplifier versus V_{GS} for different R_L with $V_{DS} = 600\text{mV}$, $V_{SB} = 0\text{V}$, and using the parameters listed in Table 4.1 .	120
5.8	A common source amplifier with a tuned load.	121

5.9	Normalized third-order intermodulation product of the output voltage of a common source amplifier with a tuned load centred at 1 GHz versus R_L with $V_{GS} = 600\text{mV}$, $V_{DS} = 600\text{mV}$, $V_{SB} = 0\text{V}$, and using the parameters listed in Table 4.1.	122
5.10	Amplitude response of the normalized $IM3$ versus frequency with $V_{GS} = 600\text{mV}$, $V_{DS} = 600\text{mV}$, $V_{SB} = 0\text{V}$, $R_L = 100$ and 500Ω and with the parameters listed in Table 4.1.	122
5.11	Equivalent circuits of the common source amplifier of Figure 5.1 including the influence of C_{gd} for determining (a) the fundamental response, (b) the second-order response, and (c) the third-order response. . . .	123
5.12	Shunt-shunt feedback configuration using y-parameter representation.	125
5.13	Final block diagram representation of the overall feedback system. . .	126
5.14	Frequency response of $T(\omega)$ for different values of C_{gd} and R_L with g_o omitted, $V_{GS} = 600\text{mV}$, $V_{DS} = 600\text{mV}$, and with the parameters listed in Table 4.1.	127
5.15	$IIP3$ of a common source amplifier versus V_{GS} for different C_{gd} with $f_{in} = 1$ GHz, $V_{DS} = 600\text{mV}$, $V_{SB} = 0\text{V}$, $R_L = 100\Omega$ and using the parameters listed in Table 4.1	131
5.16	(a) Equivalent AC circuit of the common source amplifier of Figure 5.1 including the influence of a nonlinear C_{gs} and the equivalent circuits for determining (b) the fundamental response, (c) the second-order response, and (d) the third-order response.	133
5.17	The normalized $IM3$ and the dominant terms of (5.82) versus f/f_T with $V_{GS} = 600\text{mV}$, $V_{DS} = 600\text{mV}$, $V_{SB} = 0\text{V}$, $mobhot = 1$, $clm = 1$ and using the parameters listed in Table 4.1.	135
5.18	The normalized $IM3$ and the dominant terms of (5.82) versus f/f_T when biased close to the sweet spot with $V_{GS} = 700\text{mV}$, $V_{SB} = 0\text{V}$, $mobhot = 1$, $clm = 1$ and using the parameters listed in Table 4.1. . .	137
5.19	$IIP3$ of a common source amplifier with nonlinear C_{gs} versus V_{GS} for different f/f_T with $V_{SB} = 0\text{V}$, $mobhot = 1$, $clm = 1$ and using the parameters listed in Table 4.1.	137

5.20	The normalized $IM3$ and the dominant terms of (5.82) versus R_L and the corresponding A_v with $f = 0.01f_T$, $R_S = 50\Omega$, $C_L = 75\text{fF}$, $V_{GS} = 700\text{mV}$, $V_{SB} = 0\text{V}$, $mobhot = 1$, $clm = 1$ and using the parameters listed in Table 4.1.	138
5.21	$IIP3$ of a common source amplifier with nonlinear C_{gs} versus V_{GS} for different R_L with $f = 0.01f_T$, $R_S = 50\Omega$, $C_L = 75\text{fF}$, $V_{SB} = 0\text{V}$, $mobhot = 1$, $clm = 1$ and using the parameters listed in Table 4.1. . .	138
5.22	A common-source amplifier with inductive source degeneration. . . .	139
5.23	Equivalent AC circuit of a common-source amplifier with inductive source degeneration used for solving the first-order responses.	140
5.24	Series-series feedback configuration using z -parameter representation. . . .	141
5.25	Loop gain, $T(\omega)$, given by (5.103), versus f/f_T for different values of L_S and R_S	143
5.26	Equivalent AC circuit of a common-source amplifier with inductive source degeneration used for solving the third-order responses.	144
5.27	The frequency response of (5.105) versus f/f_T with $V_{GS} = 700\text{mV}$, $V_{DS} = 600\text{mV}$, $V_{SB} = 0\text{V}$, $R_S = 50\Omega$, $R_L = 100\Omega$, and using the parameters listed in Table 4.1.	146
5.28	$IIP3$ of a source degenerated common source amplifier versus V_{GS} for different L_S with $f_{in} = 1\text{GHz}$, $V_{DS} = 600\text{mV}$, $V_{SB} = 0\text{V}$, $R_L = 100\Omega$ and using the parameters listed in Table 4.1	146
5.29	$IIP3$ of a source degenerated common source amplifier versus V_{GS} for different frequencies with $L_S = 1\text{nH}$, $V_{DS} = 600\text{mV}$, $V_{SB} = 0\text{V}$, $R_L = 100\Omega$ and using the parameters listed in Table 4.1	147
5.30	Measurement setup for measuring $IIP3$ of a single transistor common source amplifier.	148
5.31	Input versus output voltage levels of the fundamental tone at 10 MHz and $IM3$ at 9MHz of a single transistor common source amplifier with $W/L = 127.5\mu/0.18\mu$, $V_{GS} = 600\text{mV}$, $V_{DS} = 1\text{V}$ and $R_L = 50\Omega$	149
5.32	Measured, simulated, and calculated $IIP3$ versus V_{GS} for a common source amplifier with $V_{DS} = 1\text{V}$, $R_L = 50\Omega$ and using the parameters listed in Table 4.1.	150

5.33	Measured, simulated, and calculated $IIP3$ versus I_{DS} for a common source amplifier with $V_{DS} = 1\text{ V}$, $R_L = 50\Omega$ and using the parameters listed in Table 4.1.	151
5.34	Comparison of measured $IIP3$ versus V_{GS} for different widths for $0.18\mu\text{m}$ technology with $V_{DS} = 1\text{ V}$ and $R_L = 50\Omega$	152
5.35	Comparison of measured $IIP3$ versus I_{DS} per μm of width for 0.25 and $0.18\mu\text{m}$ technology.	153
5.36	Comparison of measured $IIP3$ versus V_{GS} for $0.25\mu\text{m}$ technology with $W/L = 220\mu\text{m}/0.25\mu\text{m}$, and $V_{DS} = 1\text{ V}$	154
5.37	Measured $IIP3$ versus V_{DS} and different V_{GS} for $0.25\mu\text{m}$ technology with $W/L = 220\mu\text{m}/0.25\mu\text{m}$, and $V_{DS} = 1\text{ V}$	155
5.38	Measured V_{GS} and corresponding I_{DS} versus v_{in} for $0.25\mu\text{m}$ technology with $W/L = 220\mu\text{m}/0.25\mu\text{m}$, and $V_{DS} = 1\text{ V}$	156
5.39	Comparison of simulated $IIP3$ of a common source amplifier with a tuned load at 2.4 GHz to a resistive load versus V_{GS} for $R_L = 50\Omega$ and 500Ω	158
5.40	Simulated amplitude response of the normalized $IM3$ versus frequency with $V_{GS} = 600\text{mV}$, $V_{DS} = 600\text{mV}$, $V_{SB} = 0\text{ V}$, $R_L = 500\Omega$	158
5.41	Simulation results for $IIP3$ of a common source amplifier with inductive source degeneration versus V_{GS} operating at 2.4 GHz	159
5.42	Simulated effect of process variation on sweet spot for $0.18\mu\text{m}$ single transistor amplifier versus V_{GS} over fast and slow corners	160
5.43	Simulated effect of process variation on sweet spot for $0.18\mu\text{m}$ single transistor amplifier versus I_{DS} over fast and slow corners	160
5.44	Simulated effect of process variation over fast and slow corners on sweet spot for $0.18\mu\text{m}$ single transistor amplifier versus (a) V_{GS} and (b) I_{DS}	161
5.45	(a) Common source amplifier with source degeneration, (b) common gate amplifier, and (c) cascode amplifier with biasing details not shown.	162
5.46	Basic differential LNA structure.	164
5.47	A cross section of a deep N-well.	165
5.48	Microphotograph of standard version LNA and deep Nwell version LNA.	166
5.49	Measured input-referred third order intercept point for the standard version LNA with $V_{dd} = 1.8\text{ V}$ and $I_{core} = 2.15\text{mA}$	167

5.50	Comparison of S21 for standard version and deep Nwell version LNAs.	167
5.51	Measured gain and noise figure versus V_{dd} for standard version and deep Nwell version LNAs.	168
5.52	Measured $IIP3$ versus the core bias current for standard version and deep Nwell version LNAs.	169
B.1	Basic second-order system.	187
B.2	Basic second-order memoryless system.	187
B.3	Basic third-order system.	188
B.4	Basic third-order memoryless system.	189
B.5	The system $\mathbf{G}$ formed by the cascade connection of $\mathbf{H}$ and $\mathbf{F}$	189
B.6	(a) A common-source amplifier. (b) AC-equivalent circuit.	197
B.7	Linearized circuit for calculating first-order kernels.	198
B.8	Linearized circuit for calculating second-order kernels.	199

List of Tables

2.1	First-order, second-order, and third-order nonlinear coefficients for AC drain current model given by (2.54).	19
2.2	First-order, second-order, and third-order nonlinear coefficients for AC drain current model given by (2.54) and (2.64).	22
3.1	Output response of a nonlinear system described by a Volterra series up to third order when excited by a two-tone sinusoidal input.	31
3.2	Nonlinear second-order current sources to directly compute second-order intermodulation products at $ \omega_1 \pm \omega_2 $ and at $2\omega_1$ (or $2\omega_2$) where $V_{i,m,n}$ is the controlling voltage for the transconductance and $V_{j,m,n}$ is the controlling voltage for the output conductance.	42
3.3	Nonlinear third-order current sources to directly compute third-order intermodulation products at $2\omega_1 \pm \omega_2$ where $V_{i,m,n}$ is the controlling voltage for the transconductance, $V_{j,m,n}$ is the controlling voltage for the output conductance, and $V_{k,m,n}$ is the controlling voltage for the bulk conductance.	45
3.4	Nonlinear third-order current sources to directly compute third-order harmonics at $3\omega_1$ where $V_{i,m,n}$ is the controlling voltage for the transconductance, $V_{j,m,n}$ is the controlling voltage for the output conductance, and $V_{k,m,n}$ is the controlling voltage for the bulk conductance.	46
4.1	Parameters used for fitting (4.38) to measured data and calculated parameters from foundry supplied BSIM3v3 model files for a $0.18 \mu\text{m}$ process.	71
5.1	Summary of measurement results for version with standard transistors and comparison to other reported designs.	169

B.1	Nonlinear second-order current sources for different basic nonlinearities for the calculation of second order Volterra kernels. $H_{1_k}(\cdot)$ is the first-order transfer function of the controlling voltage of the transconductance and $H_{1_m}(\cdot)$ is the first-order transfer function of the controlling voltage of the output conductance.	200
B.2	Nonlinear third-order current sources for different basic nonlinearities for the calculation of second order Volterra kernels. $H_{1_k}(\cdot)$ and $H_{2_k}(\cdot)$ are the first-order and second-order transfer function of the controlling voltage of the transconductance. $H_{1_m}(\cdot)$ and $H_{2_m}(\cdot)$ are the first-order and second-order transfer function of the controlling voltage of the output conductance. $H_{1_n}(\cdot)$ and $H_{2_n}(\cdot)$ are the first-order and second-order transfer function of the controlling voltage of the bulk conductance.	202

List of Abbreviations and Symbols

α	combined mobility degradation, series drain-source resistance and velocity saturation coefficient
β_R	series drain-source resistance mobility degradation coefficient
$\mathcal{E}_x$	electric field intensity
$\mathcal{E}_{sat}$	critical electric field intensity at which the carrier velocity become saturated
λ	channel modulation factor
ϵ_{ox}	dielectric permittivity of SiO ₂
ϵ_s	dielectric permittivity of Si
γ	body effect coefficient
μ_0	low electric field electron surface mobility
μ_{eff}	effective electron surface mobility
ϕ	surface inversion potential
ϕ_F	Fermi level potential
ϕ_t	Thermal voltage
θ	mobility degradation coefficient
θ_B	bulk mobility degradation coefficient
A	peak amplitude (in volts) of a sinusoidal input
A_v	Voltage gain
c_m	first-order nonlinearity coefficient in the one-dimensional power series expansion of the function that describes the nonlinear current and the controlling voltage m
$c_{(m,n)}$	second-order nonlinearity coefficient in the two-dimensional power series expansion of the function that describes the nonlinear current and the controlling voltages m , and n

$c_{(m,n,p)}$	third-order nonlinearity coefficient in the three-dimensional power series expansion of the function that describes the nonlinear current and the controlling voltages m , n , and p
$C_{x,e}$	extrinsic capacitance
$C_{x,i}$	intrinsic capacitance
C_{db}	drain-bulk capacitance of a MOSFET
C_{fr}	fringing capacitance
C_{gb}	gate-bulk capacitance of a MOSFET
C_{gd}	gate-drain capacitance of a MOSFET
C_{gs}	gate-source capacitance of a MOSFET
C_j	junction capacitance
C_{sb}	source-bulk capacitance of a MOSFET
C'_{ox}	gate oxide capacitance per unit area
C_{ox}	total gate capacitance
C_{ov}	overlap capacitance
C/I	carrier to interferer ratio
CLM	channel length modulation
dB	decibels $10 \log(x)$
dBm	$10 \log_{10}(\text{Power in Watts}/1 \text{ mW})$
dBV	$20 \log_{10}(\text{Voltage in } V_p)$
dB μ V	$20 \log_{10}(\text{Voltage in } \mu V_p)$
dBV $_{rms}$	$20 \log_{10}(\text{Voltage in } V_{rms})$
dB μ V $_{rms}$	$20 \log_{10}(\text{Voltage in } \mu V_{rms})$
DIBL	drain induced barrier lowering
$f[\cdot]$	function of $[\cdot]$
f	frequency (Hz)
f_T	cut-off frequency
G	conductance
g_m	transconductance
g_{mb}	bulk conductance
g_o	output conductance
HD_2, HD_3	second and third-order harmonic distortion

$\mathbf{H}_n[\cdot]$	n th-order Volterra operator
$H_n(j\omega_1, j\omega_2, \dots, j\omega_n)$	n -dimensional Fourier transform of the n th-order Volterra kernel also referred to as n th-order transfer function
$h_n(\tau_1, \tau_2, \dots, \tau_n)$	n -th order Volterra kernel (time domain representation)
i_{NL2x}, i_{NL3x}	second and third-order nonlinear current source due to x
I_o	total current through a component where $I_o = I_O + i_o$
I_O	DC portion of the current through a component
i_o	AC portion of the current through a component
I_{DS}, i_{ds}, I_{ds}	drain-source current
IIP_3	input referred third-order intercept point
IM_2, IM_3	second and third-order intermodulation distortion
j	$\sqrt{-1}$
k	Boltzmann constant ($1.3809 \times 10^{-8} fC \cdot V/K$)
K	transconductance parameter of a MOSFET
K_m	Amplitude response at frequency $m\omega$ of a nonlinear system to a single-tone sinusoidal input
$K_{m,n}$	Amplitude response at frequency $m\omega + n\omega$ of a nonlinear system to a two-tone sinusoidal input
L	inductance
LAN	local area network
L, L_{eff}	effective gate length
LDD	lightly doped drain
LNA	low noise amplifier
MOSFET	metal oxide semiconductor field effect transistor
MESFET	metal semiconductor field effect transistor
n	subthreshold slope factor
N_A	acceptor concentration
SCBE	substrate current induced body effect
SNR	signal to noise ratio
P_{VAG}	gate bias Early voltage correction factor
q	magnitude of an electron charge ($1.602 \times 10^{-4} fC$)

Q_o	total charge of a capacitor where $Q_o = Q_O + q_o$
Q_O	DC portion of the charge of capacitor
q_o	AC portion of the charge of capacitor
R	resistance
r_o	transistor output resistance
RF	radio frequency
T	temperature in Kelvin
$T(\omega)$	closed loop gain at ω
t_{ox}	gate oxide thickness
t_{poly}	gate poly thickness
v	electron velocity
v_{sat}	electron saturation velocity
V_A	Early voltage
V_{ACLM}	Early voltage due to channel length modulation
V_{ADIBL}	Early voltage due to drain induced barrier lowering
V_{Asat}	Early voltage at $V_{DS} = V_{DS(sat)}$
V_c	total value of a voltage that controls a nonlinear circuit element where $V_c = V_C + v_c$
V_C	DC portion of a voltage that controls a nonlinear circuit element
v_c	AC portion of a voltage that controls a nonlinear circuit element
$V_{c,m,n}$	steady state voltage that controls a nonlinear circuit element due to two sinusoid excitations at the frequency $m\omega_1$ and $n\omega_2$
V_{DS}, v_{ds}, V_{ds}	drain-source voltage
$V_{DS(sat)}$	drain-source saturation voltage
V_{FB}	flat band voltage
V_{GS}, v_{gs}, V_{gs}	gate-source voltage
V_{IIP3}	Input referred third-order intercept point (in V_p)
V_{IM3}	third-order intermodulation product (in V_p)
V_{IP1}	Input referred one dB compression point (in V_p)
V_{OD}, v_{od}, V_{od}	overdrive voltage ($V_{od} = V_{gs} - V_{th}$)
V_{OIP3}	Output referred third-order intercept point (in V_p)

V_p	peak voltage
V_{SB}, v_{sb}, V_{sb}	source-bulk voltage
V_{th}	threshold voltage
V_{T0}	threshold voltage of a long channel device with $V_{SB} = 0V$
W, W_{eff}	effective gate width
$\omega, \omega_1, \omega_2$	2π frequency (rad/sec)
$Y(\omega)$	admittance at ω
$Z(\omega)$	impedance at ω

Chapter 1

Introduction

1.1 Introduction

Radio frequency design in CMOS technology has garnered much attention over the past decade and is increasingly becoming a serious option for many RF wireless systems. CMOS appears to be particularly well suited for low-tier, short range, wireless systems, such as cordless phones, wireless LAN systems and emerging short range standards like Bluetooth. While such systems generally do not have overly stringent sensitivity requirements, low power consumption is an absolute must, necessitating operation of the transistors further into the moderate inversion region than would be necessary in higher power designs. In addition, these systems must operate in the presence of large blocking signals, and, as a result, the linearity requirements of the system can become the limiting specification.

1.2 Motivation

Due to the exponential dependence of the drain current on the gate-drain voltage, the weak and moderate inversion regions are generally dismissed for low-distortion applications [1]. While the cut-off frequency, f_T , of a transistor biased in weak inversion is too low for RF applications, there is sufficient current gain for most RF applications while biased in the moderate inversion region, especially with modern sub-micron devices. In addition, two-tone intermodulation measurements, such as shown in Figure 1.1 and in [2], show a significant peaking, or “sweet spot”, in the third-order intercept point, $IP3$, in the moderate inversion region. Moreover, the

Figure 1.1: Measured V_{IIP3} using two tones spaced at 1 MHz centred at 10 MHz and 2 GHz versus I_{DS} of a CMOS common-source amplifier with $W/L = 127.5\mu\text{m}/0.18\mu\text{m}$.

peak shifts systematically to higher drain current per unit width as CMOS technology scales to smaller gate lengths. This peaking occurs as the mechanism for the drain current flow when biased in the weak inversion region changes from diffusion, a largely exponential behaviour, to drift, an approximately linear behaviour when biased in the strong inversion region (especially with minimum length modern sub-micron technology). Over a very narrow bias range in the moderate inversion region, the current flow follows a nearly ideal square-law behaviour with minimized third-order distortion. For $0.18\mu\text{m}$ CMOS technology, the peak occurs at approximately 1 - 2 mA for transistors with a width of $100\mu\text{m}$, a typical size for RF applications. This peak falls directly into the region of interest for low voltage and low power RF designs. However, the issue of linearity of a MOSFET transistor in the moderate inversion region is not well studied. Although known to MESFET designers [3], [4], the existence of a linearity sweet spot in MOSFETs, is largely unknown to RF CMOS circuit designers. As CMOS technology continues to scale to smaller device sizes, and with the quest for lower power consumption, more RF circuits will be designed for operation in the moderate inversion region. Operation in this region has not been

fully explored as of yet and designers can be easily misled by results if they are not aware of the sweet spot. As a result, more study and analysis of linearity in the moderate inversion region is needed.

The emphasis in previous work was on analysis of linearity of the MOSFET in the strong inversion region. Very little consideration was given to the moderate inversion region. Methods of characterization or model verification are addressed in [2], [5], [6], and while the existence of the sweet spot in the moderate inversion region is shown, no analysis is provided. A detailed and excellent treatment of distortion analysis of MOSFETs is provided in [7], but only the weak and strong inversion regions are considered. Similarly, only the strong inversion region is considered in [1], [8]-[10]. Although the moderate inversion region is considered in [11], the analysis focuses on the effect of the linearity of a CMOS switching pair and the existence of a sweet spot is missed. An enhanced linear region close to the sweet spot has been demonstrated using circuit linearization techniques in [12]-[15] by combining the nonlinearity characteristics of two or more transistors, however, the inherent linearity of the individual transistors received little attention. Some analysis is given in [14] and [15] but only g_m nonlinearities are considered and other practical considerations such as the effects of source degeneration and process variation are not addressed. Finally, analysis was presented recently in [16], however, the authors use only Taylor series analysis and do not employ Volterra series analysis even though they include memory elements in their analysis. They also dismiss the effects of the cross-modulation terms from the Taylor series expansion and do not provide a drain current model.

1.3 Objective

The objective of this thesis is primarily two-fold;

1. to provide insights into the mechanisms of the nonlinear distortion in a MOSFET, especially in the moderate inversion region; and,
2. to investigate the implications of the linearity sweet spot in RF applications.

To achieve these goals a MOSFET model is adapted from a “single-piece” expression [17] that is valid over all regions of inversion from which the power series coefficients

are extracted for use with the Volterra series analysis. The complexity of the MOSFET model is limited somewhat in order to keep the analysis from becoming too intractable. Volterra series analysis is used for an in-depth study of the nonlinear distortion of a common source amplifier, the most commonly used configuration for RF amplifiers. The analysis is then extended to include other configurations such as tuned loads, inductive source degeneration, cascode amplifiers and common gate amplifiers.

1.4 Contributions

This thesis contributes new analysis not presented elsewhere of the linearity of the moderate inversion region of a MOSFET. The analysis is confirmed with measurements and shows that a region of significantly reduced third-order distortion, referred to as the “linearity sweet spot” exists in the moderate inversion region. Implications of the linearity sweet spot on typical RF circuits is discussed in detail. In addition, new analysis of the effects of the nonlinear capacitors of a MOSFET’s linearity is presented. The design and implementation of a low power RF CMOS LNA operating in the moderate inversion region is also presented that demonstrates operation in the moderate inversion region is possible for certain applications such as a Bluetooth receiver. By operating in the moderate inversion region the design was able to achieve significant power reduction compared to other recent designs and yet still achieve comparable performance.

1.5 Organization

Chapter 2 defines several basic parameters that are often used to describe the nonlinear behaviour of analog integrated circuits. These parameters often make use of the power series coefficients of a nonlinear device. However, the drain current of a MOSFET depends on three controlling voltages and so a multi-dimensional power series expansion of a transconductor is described in this chapter. Examples of one, two and three-dimensional power series expansion of a transconductor are given.

Chapter 3 provides background material for using Volterra series analysis. Time domain and frequency domain representations of the Volterra series are presented. The analysis of cascade connections as well as the analysis of a nonlinear system with linear feedback is also presented. The final section of Chapter 3 discusses the “Direct Calculation Method” [7] for determining the harmonic and intermodulation products of a nonlinear circuit. A short example is provided.

In Chapter 4, a semi-empirical single-piece expression for modeling the drain current is described from which the nonlinear coefficients are derived for later use of the direct calculation method for Volterra Series analysis. The single-piece expression also includes second-order effects such as channel-length modulation, mobility degradation and velocity saturation, however, simple expressions for second-order effects are used in order to keep the drain current model and the analysis from becoming too intractable. Generally low intermodulation distortion amplifiers are biased for class-A operation, and so the analysis is restricted to the saturation region.

In Chapter 5, Volterra series analysis using the direct calculation method is used in conjunction with the power series coefficients derived in Chapter 4 to determine the intermodulation components of a common-source amplifier. The effects of nonlinear intrinsic capacitances, tuned loads, and inductive source degeneration are investigated. Measurement results at low frequencies as well as RF frequencies are used to confirm the analysis. Finally, a design example of an LNA operating in the moderate inversion region at 2.4 GHz is presented and compared to several other designs demonstrating that operation on the moderate inversion region is feasible.

Chapter 6 draws conclusions and discusses possible future research subjects.

Chapter 2

Linearity Concepts

2.1 Introduction

The nonlinear behaviour of analog integrated circuits is often described in terms of parameters that are measured in the frequency domain [7]. These parameters are defined in this chapter.

Beginning with a definition of a linear system, the effects of harmonic distortion, gain compression, and intermodulation distortion are described by considering the output of a general nonlinear analog circuit. This circuit is excited by a single sinusoidal source and later, by a two-tone sinusoidal source.

Often, a power series expansion of a *weakly nonlinear* device is sufficiently accurate to evaluate the nonlinear behaviour of that device. The drain current of a MOSFET, however, is dependent on three voltages, the gate-source voltage, the drain-source voltage and the bulk-source voltage, and requires a multi-dimensional power series expansion to fully evaluate its nonlinear behaviour. Hence, Section 2.5 describes the multi-dimensional power series expansion of a transconductor and examples of one, two, and three-dimensional power series expansions of a transconductor are given. Furthermore, the effects of nonlinear parasitic capacitors must also be taken into account when considering the nonlinear behaviour of a MOSFET, and so the nonlinear coefficients of a capacitor are defined in Section 2.6.

2.2 Definition of a Linear System

A system is linear if its output can be expressed as the superposition of the responses to individual inputs [18]. In other words, if two inputs $x_1(t)$ and $x_2(t)$ produce outputs

$y_1(t)$ and $y_2(t)$ such that

$$y_1(t) = f[x_1(t)] \quad (2.1)$$

$$y_2(t) = f[x_2(t)] \quad (2.2)$$

then a system is linear if, and only if

$$f[ax_1(t) + bx_2(t)] = ay_1(t) + by_2(t) \quad (2.3)$$

where a and b are constants. Otherwise, the system is considered to be *nonlinear*.

One of the consequences of a linear system is that the output spectrum of a linear system can only have tones at the same frequencies as the input spectrum. A nonlinear system, however, can have additional spectral components at the output that were not present at the input. These spectral components are referred to as *harmonic distortion* components. Often the harmonic distortion components are undesired and can interfere and distort the desired components. Section 2.3 describes in more detail the effects of distortion resulting from a single tone sinusoidal input applied to a nonlinear system and Section 2.4 considers the effects from a two tone sinusoidal input.

2.3 Harmonic Distortion from Single-tone Inputs

If a sinusoid is applied to a nonlinear system, frequency components that are integer multiples of the input frequency, referred to as harmonics, are generated at the output of the system. These harmonic frequency components are known as harmonic distortion. The relationship between the input signal, $x(t)$, of a circuit and the output, $y(t)$, in a nonlinear system can often be described by the power series

$$y(t) = c_0 + c_1x(t) + c_2x(t)^2 + c_3x(t)^3 + \dots \quad (2.4)$$

where $x(t)$ and $y(t)$ can either be a voltage or a current signal. When $y(t)$ is explicitly given by the analytic relationship

$$y(t) = f[x(t)] \quad (2.5)$$

then the coefficients $c_1, c_2, c_3, \dots$ can be described by the coefficients of a Taylor series of f , where

$$c_1 = \frac{df}{dx} \quad (2.6)$$

$$c_2 = \frac{1}{2} \frac{d^2 f}{dx^2} \quad (2.7)$$

$$c_3 = \frac{1}{6} \frac{d^3 f}{dx^3} \quad (2.8)$$

assuming that the series of (2.4) converges to $f[x(t)]$.

The coefficient c_0 describes the DC behaviour of the circuit and c_1 describes the linear behaviour. The coefficients c_2 and c_3 are called the second-order and third-order nonlinear coefficients respectively and describe the second-order and third-order nonlinear behaviour of the circuit. The coefficients are independent of frequency in circuits where there is little to no delay between the output and input, or if there are no capacitors or inductors, or the frequency is so low that the capacitors and inductors do not play a significant role. Such circuits are often referred to as a “memoryless circuits” [7]. Conversely, circuits where the inductors and capacitors play a significant role are referred to as “circuits with memory”. For simplicity sake, the present analysis assumes memoryless circuits.

If a single-tone sinusoidal input, $x(t) = A \cos(\omega t + \theta)$, is applied, then equation (2.4) yields, to third-order,

$$y(t) = c_0 + c_1 A \cos(\omega t + \theta) + c_2 A^2 \cos^2(\omega t + \theta) + c_3 A^3 \cos^3(\omega t + \theta) + \dots \quad (2.9)$$

Using the trigonometric identities

$$\cos^2 x = \frac{1}{2} + \frac{1}{2} \cos 2x \quad (2.10)$$

$$\cos x \cos y = \frac{1}{2} \cos(x + y) + \frac{1}{2} \cos(x - y) \quad (2.11)$$

(2.9) can be rewritten as

$$y(t) = K_0 + K_1 \cos(\omega t + \theta) + K_2 \cos(2\omega t + 2\theta) + K_3 \cos(3\omega t + 3\theta) + \dots \quad (2.12)$$

where K_n is the output amplitude response at $n\omega$ and, to a maximum of third-order, is given by

$$K_0 = c_0 + \frac{1}{2}A^2c_2 + \dots \quad (2.13)$$

$$K_1 = Ac_1 + \frac{3}{4}A^3c_3 + \dots \quad (2.14)$$

$$K_2 = \frac{1}{2}A^2c_2 + \dots \quad (2.15)$$

$$K_3 = \frac{1}{4}A^3c_3 + \dots \quad (2.16)$$

It can be seen from (2.15) that the second-order nonlinear coefficient c_2 gives rise to a second-order harmonic at 2ω . c_2 also contributes to a second-order signal at DC, given by K_0 , which results in a DC shift from the nominal operating point at c_0 . In addition, from Appendix A (see (A.6)), higher even-order coefficients in the series expansion, c_4 and c_6 , also contribute to the signal at 2ω and DC in addition to tones at 4ω and 6ω . Similarly, from (2.16), the third-order coefficient c_3 gives rise to a third-order harmonic at 3ω that is proportional to c_3 and the cube of the input amplitude. From (2.14), the third-order distortion also contributes to a signal at ω in addition to the linear portion with c_1 . If c_1 and c_3 are the same sign, the amplitude of the output tone at ω increases nonlinearly, due to the $3c_3A^3/4$ term, as the amplitude of the input signal, A , increases. This is referred to as *gain expansion*. If c_1 and c_3 have opposite signs then the circuit experiences *gain compression* where the amplitude of the fundamental output tone at ω decreases nonlinearly with increasing amplitude at the input. From (A.12), the fifth-order coefficient, c_5 , also contributes to a response at ω and 3ω , in addition to 5ω . Similarly, higher odd-order coefficients will also contribute to the distortion of the fundamental signal and give rise to distortion at odd-order harmonics. However, the contribution to the distortion of the fundamental signal at ω by odd-order coefficients higher than and including c_5 are often negligible and so only third-order distortion effects are commonly considered. The analysis

presented in this thesis will only consider effects up to and including third-order effects.

2.3.1 1 dB Compression Point

The 1 dB compression point is defined as the input signal level that causes the small signal gain at the fundamental harmonic to drop by one dB [18]. This is illustrated in Figure 2.1 which plots the log of output amplitude of the fundamental tone as a function of the input signal level and shows the point where the output has decreased from the ideal linear gain by 1 dB. The 1 dB compression point can be expressed

Figure 2.1: Plot of output voltage (in dBV) of the fundamental tone versus input voltage and 1 dB compression point.

mathematically as

$$\frac{\text{Fundamental Output}}{\text{Linear Output}} = -1\text{dB} = 0.89125 \quad (2.17)$$

From (2.14) the output of the fundamental, to third-order, is $c_1A + \frac{3}{4}c_3A^3$. c_1A is the undistorted or linear output, and hence, (2.17) yields

$$1 + \frac{3}{4}\frac{c_3}{c_1}A^2 = 0.89125 \quad (2.18)$$

assuming c_1 and c_3 have opposite signs. Solving for A and setting $A = V_{IP1}$, where V_{IP1} is the input referred 1 dB compression point, yields

$$|V_{IP1}|^2 = 0.145 \left| \frac{c_1}{c_3} \right| \quad (2.19)$$

where it can be seen that the 1 dB compression point is proportional to the linear gain, c_1 , and inversely proportional to c_3 .

2.4 Harmonic Distortion from Two-tone Inputs

If a nonlinear system described by (2.4) is excited by two sinusoids, $x(t) = A \cos(\omega_1 t) + A \cos(\omega_2 t)$, distortion components are generated at the harmonics of ω_1 and ω_2 . However, additional distortion components, referred to as intermodulation products, are also generated at the output. If $x(t)$ is applied to the input, then equation (2.4) yields, to third-order,

$$\begin{aligned} y(t) = & K_{0,0} + K_{1,0} \cos(\omega_1 t) + K_{0,1} \cos(\omega_2 t) + K_{2,0} \cos(2\omega_1 t) + \\ & K_{0,2} \cos(2\omega_2 t) + K_{3,0} \cos(3\omega_1 t) + K_{0,3} \cos(3\omega_2 t) + \dots + \\ & K_{1,\pm 1} \cos((\omega_1 \pm \omega_2)t) + K_{\pm 1,1} \cos((\omega_2 \pm \omega_1)t) + \\ & K_{2,\pm 1} \cos((2\omega_1 \pm \omega_2)t) + K_{\pm 1,2} \cos((2\omega_2 \pm \omega_1)t) + \dots \end{aligned} \quad (2.20)$$

where $K_{n,m}$ is the amplitude response of a tone at $n\omega_1 + m\omega_2$. The responses at the different tones are, to third-order

$$K_{0,0} = c_0 + A^2 c_2 + \dots \quad (2.21)$$

$$K_{1,0} = K_{0,1} = A c_1 + \frac{9}{4} A^3 c_3 + \dots \quad (2.22)$$

$$K_{2,0} = K_{0,2} = \frac{1}{2} A^2 c_2 + \dots \quad (2.23)$$

$$K_{3,0} = K_{0,3} = \frac{1}{4} A^3 c_3 + \dots \quad (2.24)$$

$$K_{1,\pm 1} = K_{\pm 1,1} = A^2 c_2 + \dots \quad (2.25)$$

$$K_{2,\pm 1} = K_{\pm 1,2} = \frac{3}{4} A^3 c_3 + \dots \quad (2.26)$$

The output spectrum is shown in Figure 2.2. As in the case for a single sinusoidal input, c_2 gives rise to a DC shift and to second-order harmonics, given by (2.21) and (2.23), when a two tone input is applied. For a two tone input, the magnitude of the DC shift is twice that of a single tone input but the magnitude of each second harmonic is the same as the single tone case. In addition, c_2 contributes to even-order intermodulation products at $\omega_1 + \omega_2$ and $\omega_1 - \omega_2$, given by (2.25), whose magnitudes are twice that of the second-order harmonics. Similarly with a single tone input, c_3 also gives rise to a distortion component in the fundamentals tones at ω_1 and ω_2 , (2.22), for a two tone input. However, the magnitude of the third-order distortion of the fundamental tones is three times that of the distortion caused by a single tone. Hence, the fundamental signal can be expected to compress about 6 dB sooner than when a single tone input is applied. As in the case of the second-order harmonic tones, the magnitude of the third harmonic tones at $3\omega_1$ and $3\omega_2$, given by (2.24), are the same for a single tone input. In addition to third harmonics, c_3 contributes to third-order intermodulation products at $2\omega_1 \pm \omega_2$ and $2\omega_2 \pm 2\omega_1$, given by (2.26), and can be expected to be three times larger than the third-order harmonic terms.

Figure 2.2: The output spectrum of a weakly nonlinear system excited by two sinusoidal inputs at ω_1 and ω_2 .

2.4.1 Third Order Intercept Point

The third-order intermodulation products (*IM3*), given by $K_{2,\pm 1}$ and $K_{\pm 1,2}$, are of special importance in radio systems since they can reside in the input band and degrade the carrier to interferer ratio (C/I) and the signal to noise ratio (SNR). The third-order intercept point (*IP3*) is a measure of the third-order nonlinearity of a circuit and is the theoretical point where the amplitude of the third-order intermodulation tone is equal to the amplitude of the fundamental tone [19]. For small input signal levels, the fundamental tone increases linearly (1 dB/dB) with increasing input levels and the *IM3* products increase as the cube of the input signal amplitude (3 dB/dB). The theoretical point at which the fundamental and the *IM3* products are equal is given by

$$\frac{K_{2,\pm 1}}{c_1 A} = 1 \quad (2.27)$$

Solving for A in (2.27) with (2.26) and setting $A = V_{IIP3}$ yields the well known result

$$|V_{IIP3}|^2 = \frac{4}{3} \left| \frac{c_1}{c_3} \right| \quad (2.28)$$

where V_{IIP3} is the input referred third-order intercept point, *IIP3*. As in the case with the 1 dB compression point, *IIP3* is proportional to the linear gain, c_1 , and inversely proportional to the third-order nonlinear coefficient c_3 .

In reality, the fundamental and *IM3* product would not actually intersect since both tones will begin to compress at large signal levels, as shown in Figure 2.3. The point where the two extrapolated lines intersect is *IIP3*. As a result, *IIP3* can be determined by measuring the amplitude of fundamental and *IM3* output tones at signal levels well below the 1dB compression point and extrapolating from the measurement results. Since *IM3* rises with a slope of 3 and the fundamental rises with a slope of 1 on a log scale plot, *IIP3* can be solved from

$$\frac{V_{OIP3} - V_{out}}{V_{IIP3} - V_{in}} = 1 \quad (2.29)$$

$$\frac{V_{OIP3} - V_{IM3}}{V_{IIP3} - V_{in}} = 3 \quad (2.30)$$

and

$$G = V_{OIP3} - V_{IIP3} = V_{out} - V_{in} \quad (2.31)$$

to yield

$$V_{IIP3} = V_{in} + \frac{1}{2}(V_{out} - V_{IM3}) \quad (2.32)$$

where V_{out} is the output voltage of the fundamental, V_{in} is the input voltage of the fundamental, V_{IM3} is the output voltage of the $IM3$ product, V_{OIP3} is the output referred $IP3$ voltage, and V_{IIP3} is the input referred $IP3$ voltage. All the voltages are in dBV. Hence, to determine V_{IIP3} , only the difference between the fundamental and the $IM3$ product at the output for a given input level is required. The above analysis applies equally if the signals are expressed in terms of power levels in dBm rather than dBV.

Figure 2.3: Plot of input and output amplitudes of the fundamental and $IM3$ versus input voltage and extrapolated $IIP3$ point.

Relationship between 1 dB compression and IP3

Most often in weakly nonlinear circuits, measurements show a relationship between the 1 dB compression point and the *IIP3* point. Taking the ratio of (2.19) and (2.28) gives

$$\frac{|V_{IIP3}|^2}{|V_{IP1}|^2} = \frac{\frac{4}{3} \left| \frac{c_1}{c_3} \right|}{0.145 \left| \frac{c_1}{c_3} \right|} = 9.195 = 9.64\text{dB} \quad (2.33)$$

Hence, the *IIP3* point is approximately 10 dB above the *single tone* 1 dB compression point. Equation (2.33) is modified slightly for a two tone input since the peak amplitude of a two tone signal is 6dB larger than a single tone. The 1 dB compression point for a two tone input is given by

$$|V_{IP1}|^2 = 0.0483 \left| \frac{c_1}{c_3} \right| \quad (2.34)$$

resulting in and *IIP3*/1 dB compression ratio of

$$\frac{|V_{IIP3}|^2}{|V_{IP1}|^2} = \frac{\frac{4}{3} \left| \frac{c_1}{c_3} \right|}{0.0483 \left| \frac{c_1}{c_3} \right|} = 27.58 = 14.4\text{dB} \quad (2.35)$$

The “general rule” of a 10 dB difference for a single tone and 15 dB difference for two tones only applies if the third-order nonlinearities are dominant and, more importantly, if the linearity of the fundamental signal is not limited or clipped by some other means, such as a transistor cutting off.

2.5 Power Series Expansions of Nonlinear Transconductance

A nonlinear transconductor is a device in which the current is a nonlinear function of one or more controlling voltages. Typically, a power series is used for distortion analysis to approximate the behaviour of nonlinear devices where small excursions around an operating point are assumed. Most computations found in literature have been simplified such that only one voltage controls the transconductance, greatly

simplifying the distortion analysis [7]. In reality though, transconductors, such as the MOSFET transistor, are dependent on more than one controlling voltage and so multi-dimensional power series expansions must be used to fully analyze the nonlinear behaviour with respect to each controlling voltage. For this reason, descriptions of one, two, and three-dimensional power series expansions will be presented here along with brief examples using simple models for the MOSFET. A more complex and accurate model for the MOSFET will be presented later in Chapter 4 along with expressions for the nonlinear coefficients.

2.5.1 One-dimensional Transconductance

A one-dimensional nonlinear transconductor is a device in which the current, $I_o(t)$, is a nonlinear function of a single controlling voltage, $V_c(t)$, where $V_c(t)$ represents the sum of the DC voltage, V_C , and the AC voltage, $v_c(t)$. The function can be expanded into a power series around the quiescent point with

$$I_o(t) = f(V_c(t)) = f(V_C + v_c(t)) \quad (2.36)$$

$$= \sum_{n=0}^{\infty} \frac{1}{n!} \left. \frac{\partial^n I_o(V_c)}{\partial V_c^n} \right|_{V_c=V_C} v_c^n \quad (2.37)$$

$$= c_0 + c_1 \cdot v_c + c_2 \cdot v_c^2 + c_3 \cdot v_c^3 + c_4 \cdot v_c^4 + \dots \quad (2.38)$$

where $\partial^n I_o(V_c)/\partial V_c^n$ is the n th derivative of I_o with respect to V_c , evaluated at V_C , and

$$c_n = \frac{1}{n!} \left. \frac{\partial^n I_o(V_c)}{\partial V_c^n} \right|_{V_c=V_C} \quad n = 1, 2, 3, \dots \quad (2.39)$$

The time dependency has been omitted for simplicity, that is, $v_c = v_c(t)$. The coefficient c_1 is the small-signal transconductance, g_m , and the higher order coefficients, c_2 , c_3 , etc. are referred to as the second-order, and third-order nonlinear coefficients, and so on.

From (2.38), the current can be separated into a DC current, with

$$I_0 = c_0 = f(V_C) \quad (2.40)$$

and an AC current,

$$i_o = c_1 v_c + c_2 \cdot v_c^2 + c_3 \cdot v_c^3 + c_4 \cdot v_c^4 + \dots \quad (2.41)$$

An example of a one-dimensional transconductor can be found with the simplified square law model of a MOSFET operating in the saturation region, given by

$$I_{ds} = \mu_0 C'_{ox} \frac{W}{L} \frac{(V_{gs} - V_{th})^2}{2}, \quad V_{ds} > V_{DS(sat)} \quad (2.42)$$

where μ_0 is the mobility factor, C'_{ox} is the gate oxide capacitance per unit area, W is the effective width, L is the effective length, V_{th} is the threshold voltage and assumed to be constant, V_{ds} is the drain-source voltage, $V_{DS(sat)}$ is the drain-source saturation voltage, and V_{gs} is the controlling gate-source voltage. Other effects such as body effect, channel length modulation, mobility degradation, velocity saturation, etc., have been ignored. In this over simplified case of a nonlinear transconductor, the channel current is controlled only by the gate-source voltage. Using the definition of (2.39), yields the following nonlinear coefficients;

$$c_1 = g_m = \frac{K}{2} (V_{GS} - V_{th}) \quad (2.43)$$

$$c_2 = \frac{K}{2} \quad (2.44)$$

$$c_3 = 0 \quad (2.45)$$

where $K = \mu_0 C'_{ox} W/L$. The analysis reveals that third-order and higher order nonlinear terms do not exist for the ideal square law model. This is unrealistic for a real device and will be discussed in more detail in Chapter 3.

2.5.2 Two-dimensional Transconductance

A two-dimensional transconductor is a device in which the current is controlled by two different voltages, V_{c1} and V_{c2} , where $I_o = f(V_{c1}, V_{c2})$. Using a two-dimensional

power series expansion [7] around the quiescent point $I_O = f(V_{C1}, V_{C2})$ yields

$$I_o(t) = f(V_{c1}(t), V_{c2}(t)) = f(V_{C1} + v_{c1}(t), V_{C2} + v_{c2}(t)) \quad (2.46)$$

$$= \sum_{m=0}^{\infty} \sum_{n=0}^m \left[\frac{\partial^m I_o(V_{c1}, V_{c2})}{\partial V_{c1}^{m-n} \partial V_{c2}^n} \cdot \frac{v_{c1}^{m-n}}{(m-n)!} \cdot \frac{v_{c2}^n}{n!} \right] \quad (2.47)$$

$$= c_0 + c_{(1,0)}v_{c1} + c_{(2,0)}v_{c1}^2 + c_{(3,0)}v_{c1}^3 + c_{(4,0)}v_{c1}^4 + \dots \quad (2.48)$$

$$+ c_{(0,1)}v_{c2} + c_{(0,2)}v_{c2}^2 + c_{(0,3)}v_{c2}^3 + c_{(0,4)}v_{c2}^4 + \dots$$

$$+ c_{(1,1)}v_{c1}v_{c2} + c_{(2,1)}v_{c1}^2v_{c2} + c_{(1,2)}v_{c1}v_{c2}^2 + \dots$$

where

$$c_{(m,n)} = \frac{1}{m!} \frac{1}{n!} \frac{\partial^{m+n} I_o}{\partial V_{c1}^m \partial V_{c2}^n} \bigg|_{\substack{V_{c1}=V_{C1} \\ V_{c2}=V_{C2}}} \quad (2.49)$$

are the nonlinear coefficients. The coefficients $c_{(1,0)}$, and $c_{(0,1)}$, are nothing else than the linear small-signal parameters g_m , and g_o respectively, assuming V_{c1} is the input voltage and V_{c2} is the output voltage. $c_{(2,0)}$ and $c_{(3,0)}$ describe the transconductance variation with V_{C1} . $c_{(0,2)}$ and $c_{(0,3)}$ describe the variation of the output conductance with V_{C2} . The other nonlinear terms in (2.48) are referred to as cross-modulation terms. $c_{(1,1)}$ describes the dependence of g_m with V_{C2} or the dependence of g_o with V_{C1} . $c_{(2,1)}$ describes the second-order dependence of g_o on V_{C1} , $c_{(1,2)}$ represents second-order dependence of g_m with V_{C2} , and so on for the remaining terms[4].

From (2.48), the current can also be separated into a DC current, given by

$$I_O = c_o = f(V_{C1}, V_{C2}) \quad (2.50)$$

and three separate AC currents, i_1 , i_2 and i_3 , such that

$$i_1 = c_{1_{v1}} \cdot v_{c1} + c_{2_{v1}} \cdot v_{c1}^2 + c_{3_{v1}} \cdot v_{c1}^3 + \dots \quad (2.51)$$

$$i_2 = c_{1_{v2}} \cdot v_{c2} + c_{2_{v2}} \cdot v_{c2}^2 + c_{3_{v2}} \cdot v_{c2}^3 + \dots \quad (2.52)$$

$$i_3 = c_{2_{v1,v2}} \cdot v_{c1} \cdot v_{c2} + c_{3_{2v1,v2}} \cdot v_{c1}^2 \cdot v_{c2} + c_{3_{v1,2v2}} \cdot v_{c1} \cdot v_{c2}^2 + \dots \quad (2.53)$$

The current i_1 represents the AC current due to v_{c1} only and i_2 represents the current due to v_{c2} only, whereas i_3 represents the current due to the cross products of v_{c1} and v_{c2} .

An example of a two-dimensional transconductance can be illustrated with a slightly more complicated model than (2.42) where channel length modulation is included such that

$$I_{ds} = \frac{K}{2}(V_{gs} - V_{th})^2(1 + \lambda V_{ds}), \quad V_{ds} > V_{DS(sat)} \quad (2.54)$$

The parameter λ is the channel modulation factor. The nonlinear coefficients are given in Table 2.1. Coefficients that are derived from the second derivative of I_{ds} with respect to V_{ds} are zero due to the current's linear dependence on V_{ds} . Similarly, coefficients that are derived from the third derivative with respect to V_{gs} are zero due to the quadratic dependency. The transconductance, g_m , and the second-order coefficient, $c_{(2,0)}$, are both increased by a factor of $1 + \lambda V_{DS}$, compared to the one-dimensional analysis. The coefficient $c_{(0,1)}$ is the output conductance g_o and is proportional to λ and the square of V_{GS} .

Table 2.1: First-order, second-order, and third-order nonlinear coefficients for AC drain current model given by (2.54).

$c_{(1,0)} (g_m)$	$K(V_{GS} - V_{th})(1 + \lambda V_{DS})$	$c_{(1,1)}$	$K\lambda(V_{GS} - V_{th})$
$c_{(2,0)}$	$\frac{K}{2}(1 + \lambda V_{DS})$	$c_{(2,1)}$	$\frac{K\lambda}{2}$
$c_{(3,0)}$	0	$c_{(1,2)}$	0
$c_{(0,1)} (g_o)$	$\frac{K}{2}\lambda(V_{GS} - V_{th})^2$		
$c_{(0,2)}$	0		
$c_{(0,3)}$	0		

2.5.3 Three-dimensional Transconductance

A three-dimensional transconductor is a device in which the current is controlled by three different voltages, V_{c1} , V_{c2} and V_{c3} where $I_o = f(V_{c1}, V_{c2}, V_{c3})$. Using a three-dimensional power series expansion around the quiescent point $I_O = f(V_{C1}, V_{C2}, V_{C3})$

yields

$$I_o(t) = f(V_{c1}(t), V_{c2}(t), V_{c3}(t)) \quad (2.55)$$

$$\begin{aligned} &= f(V_{C1} + v_{c1}(t), V_{C2} + v_{c2}(t), V_{C3} + v_{c3}(t)) \\ &= \sum_{m=0}^{\infty} \sum_{n=0}^m \sum_{k=0}^{m-n} \left[\frac{\partial^m I_o(V_{c1}, V_{c2}, V_{c3})}{\partial V_{c1}^n \partial V_{c2}^n \partial V_{c3}^{m-n-k}} \cdot \frac{v_{c1}^m}{m!} \cdot \frac{v_{c2}^n}{n!} \cdot \frac{v_{c3}^{m-n-k}}{(m-n-k)!} \right] \end{aligned} \quad (2.56)$$

$$\begin{aligned} &= c_0 + c_{(1,0,0)} \cdot v_{c1} + c_{(2,0,0)} \cdot v_{c1}^2 + c_{(3,0,0)} \cdot v_{c1}^3 + c_{(4,0,0)} \cdot v_{c1}^4 + \dots \quad (2.57) \\ &\quad + c_{(0,1,0)} \cdot v_{c2} + c_{(0,2,0)} \cdot v_{c2}^2 + c_{(0,3,0)} \cdot v_{c2}^3 + c_{(0,4,0)} \cdot v_{c2}^4 + \dots \\ &\quad + c_{(0,0,1)} \cdot v_{c3} + c_{(0,2,0)} \cdot v_{c3}^2 + c_{(0,3,0)} \cdot v_{c3}^3 + c_{(0,4,0)} \cdot v_{c3}^4 + \dots \\ &\quad + c_{(1,1,0)} \cdot v_{c1} \cdot v_{c2} + c_{(2,1,0)} \cdot v_{c1}^2 \cdot v_{c2} + c_{(1,2,0)} \cdot v_{c1} \cdot v_{c2}^2 + \dots \\ &\quad + c_{(1,0,1)} \cdot v_{c1} \cdot v_{c3} + c_{(2,0,1)} \cdot v_{c1}^2 \cdot v_{c3} + c_{(1,0,2)} \cdot v_{c1} \cdot v_{c3}^2 + \dots \\ &\quad + c_{(0,1,1)} \cdot v_{c2} \cdot v_{c3} + c_{(0,2,1)} \cdot v_{c2}^2 \cdot v_{c3} + c_{(0,1,2)} \cdot v_{c2} \cdot v_{c3}^2 + \dots \\ &\quad + c_{(1,1,1)} \cdot v_{c1} \cdot v_{c2} \cdot v_{c3} + \dots \end{aligned}$$

where

$$c_{(m,n,p)} = \frac{1}{m!} \frac{1}{n!} \frac{1}{p!} \frac{\partial^{(m+n+p)} I_o}{\partial^m V_{c1} \partial^n V_{c2} \partial^p V_{c3}} \bigg|_{\substack{V_{c1}=V_{C1} \\ V_{c2}=V_{C2} \\ V_{c3}=V_{C3}}} \quad (2.58)$$

are the nonlinear coefficients. As in the case for the two-dimensional power series expansion, the coefficients $c_{(1,0,0)}$, and $c_{(0,1,0)}$, are nothing else than the linear small-signal parameters g_m , g_o . $c_{(0,0,1)}$ is the bulk conductance, g_{mb} , assuming V_{c3} is a bulk voltage. $c_{(2,0,0)}$ and $c_{(3,0,0)}$ describe the transconductance variation with V_{C1} . $c_{(0,2,0)}$ and $c_{(0,3,0)}$ describe the variation of the output conductance with V_{C2} and $c_{(0,0,2)}$ and $c_{(0,0,3)}$ represent the bulk conductance variation with V_{C3} . The other nonlinear terms in (2.57) are referred to as cross-modulation terms. $c_{(1,1,0)}$ describes the dependence of g_m with V_{C2} or the dependence of g_o with V_{C1} . $c_{(2,1,0)}$ describes the second-order dependence of g_o on V_{C1} , $c_{(1,2,0)}$ represents second-order dependence of g_m with V_{C2} , and so on for the remaining terms [4].

From (2.57), the current can also be separated into a DC current, given by

$$I_O = c_o = f(V_{C1}, V_{C2}, V_{C3}) \quad (2.59)$$

and four separate AC currents, i_1 , i_2 , i_3 , and i_4 , such that

$$i_1 = c_{(1,0,0)} \cdot v_{c1} + c_{(2,0,0)} v_{c1}^2 + c_{(3,0,0)} \cdot v_{c1}^3 + c_{(4,0,0)} \cdot v_{c1}^4 + \dots \quad (2.60)$$

$$i_2 = c_{(0,1,0)} \cdot v_{c2} + c_{(0,2,0)} \cdot v_{c2}^2 + c_{(0,3,0)} \cdot v_{c2}^3 + c_{(0,4,0)} \cdot v_{c2}^4 + \dots \quad (2.61)$$

$$i_3 = c_{(0,0,1)} \cdot v_{c3} + c_{(0,0,2)} \cdot v_{c3}^2 + c_{(0,0,3)} \cdot v_{c3}^3 + c_{(0,0,4)} \cdot v_{c3}^4 + \dots \quad (2.62)$$

$$\begin{aligned} i_4 = & c_{(1,1,0)} \cdot v_{c1} \cdot v_{c2} + c_{(2,1,0)} \cdot v_{c1}^2 \cdot v_{c2} + c_{(1,2,0)} \cdot v_{c1} \cdot v_{c2}^2 + \dots \quad (2.63) \\ & + c_{(1,0,1)} \cdot v_{c1} \cdot v_{c3} + c_{(2,0,1)} \cdot v_{c1}^2 \cdot v_{c3} + c_{(1,0,2)} \cdot v_{c1} \cdot v_{c3}^2 + \dots \\ & + c_{(0,1,1)} \cdot v_{c2} \cdot v_{c3} + c_{(0,2,1)} \cdot v_{c2}^2 \cdot v_{c3} + c_{(0,1,2)} \cdot v_{c2} \cdot v_{c3}^2 + \dots \\ & + c_{(1,1,1)} \cdot v_{c1} \cdot v_{c2} \cdot v_{c3} + \dots \end{aligned}$$

The current i_1 represents the AC current due to v_{c1} only, i_2 represents the current due to v_{c2} only, i_3 represents the current due to v_{c3} only, and i_4 represents the current due to the cross products of v_{c1} , v_{c2} and v_{c3} .

An example of three-dimensional transconductance can be obtained by considering the additional complexity of body effect to (2.54), where the threshold voltage is dependent on the source-bulk voltage V_{sb} and is given by

$$V_{th} = V_{T0} + \gamma \left(\sqrt{V_{sb} + \phi} - \sqrt{\phi} \right) \quad (2.64)$$

The parameter V_{T0} is the threshold voltage of a long channel device at $V_{SB} = 0V$, ϕ is the surface inversion potential, and γ is the body effect coefficient.

The resulting coefficients are listed in Table 2.2 and are essentially the same as those in Table 2.1 except for the addition of coefficients with respect to v_{sb} . The first-order coefficient, $c_{(0,0,1)}$, is also referred to as the bulk transconductance g_{mb} . As previously seen, coefficients that are derived from the second derivative of I_{ds} with respect to V_{ds} are zero due to the drain current's linear dependence on V_{ds} . Similarly, coefficients that are derived from the third derivative with respect to V_{gs} are zero due to the quadratic dependency.

Table 2.2: First-order, second-order, and third-order nonlinear coefficients for AC drain current model given by (2.54) and (2.64).

$c_{(1,0,0)} (g_m)$	$K(V_{GS} - V_{th})(1 + \lambda V_{DS})$	$c_{(1,1,0)}$	$K\lambda(V_{GS} - V_{th})$
$c_{(2,0,0)}$	$\frac{K}{2}(1 + \lambda V_{DS})$	$c_{(1,0,1)}$	$-\frac{K(1+\lambda V_{DS})\gamma}{2\sqrt{\phi+V_{SB}}}$
$c_{(3,0,0)}$	0	$c_{(0,1,1)}$	$-\frac{K\lambda\gamma(1+\lambda V_{DS})}{2\sqrt{\phi+V_{SB}}}$
$c_{(0,1,0)} (g_o)$	$\frac{K}{2}\lambda(V_{GS} - V_{th})^2$	$c_{(2,1,0)}$	$\frac{K\lambda}{2}$
$c_{(0,2,0)}$	0	$c_{(1,2,0)}$	0
$c_{(0,3,0)}$	0	$c_{(2,0,1)}$	0
$c_{(0,0,1)} (g_{mb})$	$\frac{K(V_{GS}-V_{th})(1+\lambda V_{DS})\gamma}{2\sqrt{\phi+V_{SB}}}$	$c_{(1,0,2)}$	$\frac{K(1+\lambda V_{DS})\gamma}{8(\phi+V_{SB})^{3/2}}$
$c_{(0,0,2)}$	$-\frac{K\gamma(V_{GS}-V_{T0}+\gamma\sqrt{\phi})(1+\lambda V_{DS})}{8(\phi+V_{SB})^{3/2}}$	$c_{(0,2,1)}$	0
$c_{(0,0,3)}$	$\frac{K\gamma(V_{GS}-V_{T0}+\gamma\sqrt{\phi})(1+\lambda V_{DS})}{16(\phi+V_{SB})^{5/2}}$	$c_{(0,1,2)}$	$\frac{K\gamma(V_{GS}-V_{T0}+\gamma\sqrt{\phi})}{8(\phi+V_{SB})^{3/2}}$
		$c_{(1,1,1)}$	$-\frac{K\lambda\gamma}{2\sqrt{\phi+V_{SB}}}$

2.6 Power Series Expansions of Nonlinear Capacitors

The nonlinear relationship between the charge, Q_o , of a capacitor and the controlling voltage, V_c , is given by

$$Q_o(t) = f(V_c(t)) = f(V_C + v_c(t)) \quad (2.65)$$

$$= \sum_{n=0}^{\infty} \frac{1}{n!} \frac{\partial^n Q_o(V_c)}{\partial V_c^n} \bigg|_{V_c=V_C} v_c^n \quad (2.66)$$

$$= Q_O + C_1 \cdot v_c + c_{2C1} \cdot v_c^2 + c_{3C1} \cdot v_c^3 + c_{4C1} \cdot v_c^4 + \dots \quad (2.67)$$

where Q_o represents the sum of the DC charge, Q_O , and the AC charge, q_o , and $\partial^n Q_o(V_c)/\partial V_c^n$ is the nth derivative of Q_o with respect to V_c , evaluated at V_C . C_1 is

the capacitance, and is given by

$$C_1 = \left. \frac{\partial Q_o(V_c)}{\partial V_c} \right|_{V_c=V_C} \quad (2.68)$$

c_{2C1} and c_{3C1} are the second-order and third-order nonlinear capacitance coefficients and are given by

$$c_{2C1} = \left. \frac{1}{2} \frac{\partial^2 Q_o(V_c)}{\partial V_c^2} \right|_{V_c=V_C} \quad (2.69)$$

$$c_{3C1} = \left. \frac{1}{6} \frac{\partial^3 Q_o(V_c)}{\partial V_c^3} \right|_{V_c=V_C} \quad (2.70)$$

They may be rewritten in the more convenient form as

$$c_{2C1} = \left. \frac{1}{2} \frac{\partial C_1(V_c)}{\partial V_c} \right|_{V_c=V_C} \quad (2.71)$$

$$c_{3C1} = \left. \frac{1}{6} \frac{\partial^2 C_1(V_c)}{\partial V_c^2} \right|_{V_c=V_C} \quad (2.72)$$

The AC part of the charge on the capacitor is expressed as

$$q_o(t) = C_1 v_c(t) + c_{2C1} v_c^2(t) + c_{3C1} v_c^3(t) + \dots \quad (2.73)$$

The AC current through a capacitor is found by taking the derivative of q_o with respect to time such that

$$i_o(t) = \frac{d}{dt} (C_1 v_c(t) + c_{2C1} v_c^2(t) + c_{3C1} v_c^3(t) + \dots) \quad (2.74)$$

2.7 Summary

In this chapter, parameters used to measure the nonlinear behaviour of a circuit have been described. The parameters and distortion products have been described in terms of the nonlinear coefficients of the circuit. Often, a power series expansion of a weakly

nonlinear device is sufficiently accurate to evaluate the nonlinear behaviour of that device but the drain current of a MOSFET is dependent on three voltages and requires a multi-dimension power series expansion to fully evaluate its nonlinear behaviour. As such, one, two, and three-dimensional power series expansions of a transconductor were described and examples were given. The power series expansion of a nonlinear capacitor was also given since the nonlinear behaviour of a MOSFET can also be influenced by its nonlinear capacitors. Expressions for the nonlinear coefficients of a more complex and accurate MOSFET model will be given in Chapter 4 which then will be used in conjunction with the Volterra Kernels as described in Chapter 3 and Chapter 5.

Chapter 3

Volterra Series Analysis

3.1 Introduction

In 1930 Vito Volterra, a Spanish mathematician, first introduced the concept of what is now known as the Volterra series in his “Theory of Functionals” [20]. Over the years, Volterra series analysis has been shown to be a powerful tool for the computation of the nonlinear behaviour of analog integrated circuits. Unlike numerical analysis methods, Volterra series analysis can provide valuable insight into the mechanisms of the nonlinear behaviour of analog circuits and the relative contribution of various circuit parameters and elements. This provides useful information for the designer as to which circuit parameters and elements to modify in order to obtain the required specifications. Since the first use of Volterra series for nonlinear circuit analysis in 1967 [21], it has been used extensively for distortion analysis in nonlinear electrical circuits, [3], [7], [8], [11], [22]-[29], to name a few.

The Volterra series description of a nonlinear system is equivalent to the Taylor series approximation of an analytic function. As with the Taylor series, a nonlinear system that is excited by a small amplitude signal can be accurately described by the first few terms of a Volterra series. More terms are needed for higher input amplitudes until the series diverges for very high amplitudes, just as the Taylor series can. The series often diverges due to strong nonlinearities caused by the high amplitude levels [7]. Hence, the Volterra series is only suitable for analysis of weakly nonlinear circuits.

The Volterra series approach is especially useful for high frequency analysis. Unlike the Taylor series, the Volterra series retains phase information and can describe nonlinear circuits with memory elements such as inductors and capacitors. At very high frequencies, these circuit elements often play a crucial role in the circuit's

behaviour. Furthermore, the Volterra series approach can provide insight into systems operations, such as cascaded systems and feedback systems.

This chapter aims to provide the background information for using Volterra series. The general form of the time domain and frequency domain representations of the Volterra series are first presented and are later used to describe the nonlinear response of a general system to a two tone input. This is then extended to a cascaded system and a feedback system. Several algorithms exist [7], [30], [31], for determining the frequency domain Volterra kernels in terms of circuit parameters. These kernels can then be used to determine the harmonic and intermodulation products of a nonlinear circuit. However, an algorithm, referred to as the “Direct Calculation Method” [7] can be used to determine the harmonic and intermodulation distortion products directly without having to use the intermediate step of determining the Volterra kernels. Both approaches produce the same answer. The Direct Calculation Method algorithm is described in the last section through the use of an example. An example of the more traditional approach is provided in Appendix B. It should be pointed out that it is not intended to provide the detailed proofs of the final results in this Chapter. Some of the basic proofs are provided in Appendix B and a more thorough study can be found in several texts such as [7], [27], and [28].

3.2 Basics of Volterra series

Volterra series describe the output of a nonlinear system as the sum of the response of n th-order operators [7]. A block level diagram is shown in Figure 3.1, where $\mathbf{H}_1$ is a first-order *operator* and represents the linearized system, $\mathbf{H}_2$ is a second-order operator, $\mathbf{H}_3$ is a third-order operator, and so on. Each operator can be described in either the time domain or frequency domain as a transfer function and is referred to as a *Volterra kernel*.

The Volterra series is in actuality a power series. Moreover, the Volterra series is a power series with memory since, as will be shown, the integrals for $\mathbf{H}_n$ are convolutions. This means that the Volterra series can describe weakly nonlinear circuits with memory elements such as capacitors and inductors, whereas a Taylor series cannot. Furthermore, if the system described by a Volterra series has no memory elements, then the Volterra series can be shown to reduce to a Taylor series [7].

Figure 3.1: Block level system characterized by a Volterra Series.

As with the Taylor series representation of an analytical function, the Volterra series of a nonlinear system may exhibit convergence difficulties. When the input amplitude increases, the higher-order terms in the Volterra series increase relatively more than the lower-order terms and at a certain amplitude, divergence may occur. Hence, a radius of convergence can be defined, however, the computation is complicated [7] and will not be considered here. It is assumed that the signals applied to the circuits under consideration are small enough such that the Volterra series converges.

3.2.1 Time Domain Representation

The Volterra series is represented in the time domain as

$$\begin{aligned}
 y(t) = & \int_{-\infty}^{\infty} h_1(\tau_1) x(t - \tau_1) d\tau_1 + \\
 & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_2(\tau_1, \tau_2) x(t - \tau_1) x(t - \tau_2) d\tau_1 d\tau_2 + \\
 & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_3(\tau_1, \tau_2, \tau_3) x(t - \tau_1) x(t - \tau_2) x(t - \tau_3) d\tau_1 d\tau_2 d\tau_3 + \dots \\
 & \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h_n(\tau_1, \dots, \tau_n) x(t - \tau_1) \dots x(t - \tau_n) d\tau_1 \dots d\tau_n + \dots
 \end{aligned} \tag{3.1}$$

where $y(t)$ is the output, $x(t)$ is the input, $h_n(\tau_1, \dots, \tau_n)$ are called *Volterra kernels*, and τ_n are time variables. $h_n(\tau_1, \dots, \tau_n)$ is also equivalent to an n -dimensional impulse

response. The Volterra kernels can be expressed in terms of circuit parameters and elements but are left in the general form.

In terms of the Volterra series operators, the Volterra series can be represented as

$$y(t) = \mathbf{H}_1[x(t)] + \mathbf{H}_2[x(t)] + \mathbf{H}_3[x(t)] + \cdots + \mathbf{H}_n[x(t)] + \cdots \quad (3.2)$$

$$= \sum_{n=1}^{\infty} \mathbf{H}_n[x(t)] \quad (3.3)$$

The first-order Volterra operator is the convolution integral and is given by

$$\mathbf{H}_1[x(t)] = \int_{-\infty}^{\infty} h_1(\tau_1)x(t - \tau_1)d\tau_1 \quad (3.4)$$

The function $h_1(\tau_1)$ is called the *first-order Volterra kernel* of the system and is also commonly known as the unit impulse response to a linear time invariant (LTI) system.

The second-order Volterra operator is known as a two-dimensional convolution integral [7] and is given by

$$\mathbf{H}_2[x(t)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_2(\tau_1, \tau_2)x(t - \tau_1)x(t - \tau_2)d\tau_1d\tau_2 \quad (3.5)$$

The function $h_2(\tau_1, \tau_2)$ is denoted as the *second-order Volterra kernel* and is equivalent to a two-dimensional impulse response [28]. Since the Volterra kernel is independent of the input signal, the output of a second-order system can be computed to any kind of input, such as a sinusoid, a sum of sinusoids, a Gaussian input and so on.

The third-order Volterra operator is three-dimensional convolution integral and is given by

$$\mathbf{H}_3[x(t)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_3(\tau_1, \tau_2, \tau_3)x(t - \tau_1)x(t - \tau_2)x(t - \tau_3)d\tau_1d\tau_2d\tau_3 \quad (3.6)$$

The function $h_3(\tau_1, \tau_2, \tau_3)$ is denoted as the *third-order Volterra kernel* and can be considered as a three-dimensional impulse response.

Extending the concept further yields the general n th order Volterra operator

$$\mathbf{H}_n[x(t)] = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h_n(\tau_1, \dots, \tau_n) x(t - \tau_1) \dots x(t - \tau_n) d\tau_1 \dots d\tau_n \quad (3.7)$$

which is seen to be the n th-order convolution integral. In [7] and [27] it is shown that the kernels h_n can always be made symmetric, which means that $h_n(\tau_1, \dots, \tau_n)$ has the same value regardless of the permutation of $\tau_1, \dots, \tau_n$. The rest of this work assumes the kernels are symmetric.

3.2.2 Frequency Domain Representation

The frequency domain response of a linear system can be obtained by multiplying the Fourier transform of the impulse response with the Fourier transform of the input signal. Similarly, since an n th-order Volterra kernel can be considered as an n th-order impulse response, this approach can be extended to n th-order operators.

A multidimensional Fourier transform, $H_n(j\omega_1, \dots, j\omega_n)$ of the n th-order kernel, $h(\tau_1, \dots, \tau_n)$, is given by

$$H_n(j\omega_1, \dots, j\omega_n) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h_n(\tau_1, \dots, \tau_n) e^{(j\omega_1\tau_1 + \dots + j\omega_n\tau_n)} d\tau_1 \dots d\tau_n \quad (3.8)$$

The n th-order transfer function is symmetric when its time-domain equivalent is symmetric. The complex conjugate of a symmetric transfer function is

$$H_n(-j\omega_1, -j\omega_2, \dots, -j\omega_n) = H_n^*(j\omega_1, j\omega_2, \dots, j\omega_n) \quad (3.9)$$

The response of a linear system to a sinusoid, $A \cos \omega t$, is well known and is given by

$$\begin{aligned} y_1(t) &= A \operatorname{Re} [H_1(j\omega) e^{j\omega t}] \\ &= A |H_1(j\omega)| \cos(\omega t + \arg[H_1(j\omega)]) \end{aligned} \quad (3.10)$$

The response of a second-order system excited by a sinusoidal, $A \cos \omega t$, is derived in Appendix B.1 and can be written in terms of the second-order nonlinear transfer

function, $H_2(j\omega_1, j\omega_2)$, where

$$\begin{aligned} y_2(t) &= \frac{A^2}{2} \text{Re} [H_2(j\omega, j\omega)e^{j2\omega t}] + \frac{A^2}{2} \text{Re} [H_2(j\omega, -j\omega)] \\ &= \frac{A^2}{2} |H_2(j\omega, j\omega)| \cos(2\omega t + \arg [H_2(j\omega, j\omega)]) + \frac{A^2}{2} H_2(j\omega, -j\omega) \end{aligned} \quad (3.11)$$

The output is a constant and a sinusoid at 2ω . Similarly, the output of a third-order system, derived in Appendix B.2, is a sinusoid at ω and 3ω , and is given by

$$\begin{aligned} y_3(t) &= \frac{A^3}{4} \text{Re} [H_3(j\omega, j\omega, j\omega)e^{j3\omega t}] + \frac{3A^3}{4} \text{Re} [H_3(j\omega, j\omega, -j\omega)e^{j\omega t}] \\ &= \frac{A^3}{4} |H_3(j\omega, j\omega, j\omega)| \cos(3\omega t + \arg [H_3(j\omega, j\omega, j\omega)]) + \\ &\quad + \frac{3A^3}{4} |H_3(j\omega, j\omega, -j\omega)| \cos(\omega t + \arg [H_3(j\omega, j\omega, -j\omega)]) \end{aligned} \quad (3.12)$$

It should be noted though, that the results given here are general and that the Fourier transforms, $H_1(j\omega_1)$, $H_2(j\omega_1, j\omega_2)$, and $H_3(j\omega_1, j\omega_2, j\omega_3)$ are complex numbers that model phase shifts on the harmonics due to the capacitors and inductors in the nonlinear network.

3.3 Nonlinear Response to Two-Tone Input

When a nonlinear system that can be described by a third-order Volterra series is excited by a two tone sinusoidal input, $x(t) = A_1 \cos \omega_1 t + A_2 \cos \omega_2 t$, the output is given by the sum of the response listed in Table 3.1. The responses in Table 3.1, taken from [7], are classified as linear responses, harmonics, intermodulation products, desensitization, compression, or expansion responses. The responses to a memoryless system can also be found from the table. In Appendix B.3, it is shown that the Volterra kernel transforms of a memoryless system are equal to the nonlinear coefficients, where $H_1(j\omega) = c_1$, $H_2(j\omega_1, j\omega_2) = c_2$, and $H_3(j\omega_1, j\omega_2, j\omega_3) = c_3$, and so Table 3.1 can also be used for memoryless systems. For example, with $A_1 = A_2 = A$, the third-order harmonic becomes $1/4A^3c_3$ as previously found in (2.24).

The normalized harmonic distortion components and normalized intermodulation products, in terms of Volterra kernel transforms can easily be found from the

Table 3.1: Output response of a nonlinear system described by a Volterra series up to third order when excited by a two-tone sinusoidal input. Taken from [7]

Order	Frequency of Response	Amplitude of Response	Type of Response
1 1	ω_1 ω_2	$A_1 H_1(j\omega_1) $ $A_2 H_1(j\omega_2) $	linear
2 2	0 0	$\frac{1}{2}A_1^2 H_2(j\omega_1, -j\omega_1) $ $\frac{1}{2}A_2^2 H_2(j\omega_2, -j\omega_2) $	DC shift
2 2	$2\omega_1$ $2\omega_2$	$\frac{1}{2}A_1^2 H_2(j\omega_1, j\omega_1) $ $\frac{1}{2}A_2^2 H_2(j\omega_2, j\omega_2) $	2nd harmonics
2 2	$\omega_1 + \omega_2$ $ \omega_1 - \omega_2 $	$A_1A_2 H_2(j\omega_1, j\omega_2) $ $A_1A_2 H_2(j\omega_1, -j\omega_2) $	2nd-order intermodulation products
3 3	$3\omega_1$ $3\omega_2$	$\frac{1}{4}A_1^3 H_3(j\omega_1, j\omega_1, j\omega_1) $ $\frac{1}{4}A_2^3 H_3(j\omega_2, j\omega_2, j\omega_2) $	3rd harmonics
3 3	$\omega_1 + \omega_2 - \omega_2 = \omega_1$ $\omega_1 - \omega_1 + \omega_2 = \omega_2$	$\frac{2}{3}A_1A_2^2 H_3(j\omega_1, j\omega_2, -j\omega_2) $ $\frac{2}{3}A_1^2A_2 H_3(j\omega_1, -j\omega_1, j\omega_2) $	3rd-order desensitization
3 3	$2\omega_1 - \omega_1 = \omega_1$ $2\omega_2 - \omega_2 = \omega_2$	$\frac{3}{4}A_1^3 H_3(j\omega_1, j\omega_1, -j\omega_1) $ $\frac{3}{4}A_2^3 H_3(j\omega_2, j\omega_2, -j\omega_2) $	3rd-order compression
3 3 3 3	$2\omega_1 + \omega_2$ $ 2\omega_1 - \omega_2 $ $\omega_1 + 2\omega_2$ $ \omega_1 - 2\omega_2 $	$\frac{3}{4}A_1^2A_2 H_3(j\omega_1, j\omega_1, j\omega_2) $ $\frac{3}{4}A_1^2A_2 H_3(j\omega_1, j\omega_1, -j\omega_2) $ $\frac{3}{4}A_1A_2^2 H_3(j\omega_1, j\omega_2, j\omega_2) $ $\frac{3}{4}A_1A_2^2 H_3(j\omega_1, -j\omega_2, -j\omega_2) $	3rd-order intermodulation products

table as well. The second-order and third-order harmonic distortion components, normalized with respect to the fundamental component are

$$HD_2 = \frac{A_1}{2} \left| \frac{H_2(j\omega_1, j\omega_1)}{H_1(j\omega_1)} \right| \quad (3.13)$$

$$HD_3 = \frac{A_1^2}{4} \left| \frac{H_3(j\omega_1, j\omega_1, j\omega_1)}{H_1(j\omega_1)} \right| \quad (3.14)$$

Similarly, the normalized second-order intermodulation products are

$$IM_2 = A_1 \left| \frac{H_2(j\omega_1, j\omega_2)}{H_1(j\omega_1)} \right| \quad (3.15)$$

for the sum frequency $\omega_1 + \omega_2$ and

$$IM_2 = A_1 \left| \frac{H_2(j\omega_1, -j\omega_2)}{H_1(j\omega_1)} \right| \quad (3.16)$$

for the difference frequency $\omega_1 - \omega_2$. The third-order intermodulation product can be given by

$$IM_3 = \frac{3}{4} A_1^2 \left| \frac{H_3(j\omega_1, j\omega_1, -j\omega_2)}{H_1(j\omega_1)} \right| \quad (3.17)$$

from which the third-order intercept point is given by

$$IIP3^2 = \frac{4}{3} \left| \frac{H_1(j\omega_1)}{H_3(j\omega_1, j\omega_1, -j\omega_2)} \right| \quad (3.18)$$

3.4 Cascade Connection of Nonlinear Systems

Figure 3.2 depicts a nonlinear system $\mathbf{H}$, connected in cascade with the nonlinear system $\mathbf{F}$, resulting in the overall system $\mathbf{G}$. The Volterra kernels for the overall system is computed assuming that blocks $\mathbf{H}$ and $\mathbf{F}$ do not load one another. This is often not a valid assumption but the idealized results reveal some valuable insights [7]. The Volterra operators of the system $\mathbf{G}$, up to third-order, are given by

$$\mathbf{G}_1 = \mathbf{H}_1 \mathbf{F}_1 \quad (3.19)$$

$$\mathbf{G}_2 = \mathbf{H}_2 \mathbf{F}_1 + \mathbf{H}_1 \mathbf{F}_2 \quad (3.20)$$

$$\mathbf{G}_3 = \mathbf{H}_3 \mathbf{F}_1 + \mathbf{F}_2 [\mathbf{H}_1 + \mathbf{H}_2] - \mathbf{H}_1 \mathbf{F}_2 - \mathbf{H}_2 \mathbf{F}_2 + \mathbf{H}_1 \mathbf{F}_3 \quad (3.21)$$

The details of the derivations can be found in Appendix B.4. It should be noted, in the expression for $\mathbf{G}_3$, given by (3.21), $\mathbf{F}_2 [\mathbf{H}_1 + \mathbf{H}_2] \neq \mathbf{F}_2 \mathbf{H}_1 + \mathbf{F}_2 \mathbf{H}_2$.

Figure 3.2: The system $\mathbf{G}$ formed by the cascade connection of $\mathbf{H}$ and $\mathbf{F}$.

The Fourier transforms, derived in Appendix B.4.1, are found to be

$$G_1(j\omega_1) = H_1(j\omega_1)F_1(j\omega_1) \quad (3.22)$$

$$\begin{aligned} G_2(j\omega_1, j\omega_2) &= H_2(j\omega_1, j\omega_2)F_1(j\omega_1 + j\omega_2) \\ &\quad + H_1(j\omega_1)H_2(j\omega_2)F_2(j\omega_1, j\omega_2) \end{aligned} \quad (3.23)$$

$$\begin{aligned} G_3(j\omega_1, j\omega_2, j\omega_3) &= H_1(j\omega_1)H_1(j\omega_2)H_1(j\omega_3)F_3(j\omega_1, j\omega_2, j\omega_3) \\ &\quad + H_3(j\omega_1, j\omega_2, j\omega_3)F_1(j\omega_1 + j\omega_2 + j\omega_3) \\ &\quad + \frac{2}{3} \left[H_1(j\omega_1)H_2(j\omega_2, j\omega_3)F_2(j\omega_1, j\omega_2 + j\omega_3) \right. \\ &\quad + H_1(j\omega_2)H_2(j\omega_1, j\omega_3)F_2(j\omega_2, j\omega_1 + j\omega_3) \\ &\quad \left. + H_1(j\omega_3)H_2(j\omega_1, j\omega_2)F_2(j\omega_3, j\omega_1 + j\omega_2) \right] \end{aligned} \quad (3.24)$$

From (3.22), it can be seen that the cascaded linear transform simply consists of the linear transform of both stages. Furthermore, the overall input-output relationship of G_1 is not affected by the order of the connections. However, this is not the case for higher order transforms. Figure 3.3 depicts the possible paths for some of the higher order transforms. In the case of the second-order system, the second-order transform consists of two contributions, as given in (3.23). The first term comes from the second-order nonlinearity of $\mathbf{H}$ that is then fed through the linear portion of $\mathbf{F}$. The higher-order nonlinearities of $\mathbf{F}$ do not need to be considered since they will generate nonlinearities higher than second-order. The second term in (3.23) is caused by the second-order nonlinearity of $\mathbf{F}$ that is fed from the linear portion of $\mathbf{H}$. For the third-order cascaded transform, the first two terms in (3.23) are similar to the second-order case. The last term in (3.23) between the square brackets is caused by the second-order kernel of $\mathbf{F}$ acting on the sum of the linearly amplified signal of $\mathbf{H}$

and the second-order nonlinearity stage of $\mathbf{H}$. The term consists of three terms so as to make $G_3(j\omega_1, j\omega_2, j\omega_3)$ symmetric [7] and hence the factor of 3 in the denominator of the common factor $2/3$. The factor of 2 is necessary since there are two ways to combine a first-order and second-order signal.

Figure 3.3: The system $\mathbf{G}$ formed by the cascade connection of $\mathbf{H}$ and $\mathbf{F}$. The arrows indicate the two contributions to the second-order kernel of the overall system.

If $\mathbf{H}$ and $\mathbf{F}$ are memoryless systems, then the transfer functions, (3.22), (3.23), and (3.23), are independent of the frequency variables and they simplify to

$$G_1 = H_1 F_1 \quad (3.25)$$

$$G_2 = H_2 F_1 + H_1^2 F_2 \quad (3.26)$$

$$G_3 = H_1^3 F_3 + H_3 F_1 + 2H_1 H_2 F_2 \quad (3.27)$$

If the two cascaded systems are nearly identical such that $H_1 \approx F_1$, $H_2 \approx F_2$ and $H_3 \approx F_3$, then it can be seen for G_2 that the second stage contributes a factor of H_1 more than the first stage's second-order nonlinearity. Similarly for G_3 , the second stage contributes a factor of approximately H_1^2 more. Hence, generally speaking, the final stage of a cascaded system tends to contribute the largest to the nonlinear distortion of the overall system. As a result, each stage of a cascaded system generally needs to be more linear than the preceding stage.

3.5 Linear Feedback

Linear feedback is used extensively in analog circuits. Feedback has many benefits such as stabilization of circuit performance against parameter changes, increase of bandwidth, input impedance transformation and, generally, reduction of nonlinear distortion. However, it can be shown that feedback may not significantly reduce the open-loop third-order distortion [22], and may, in fact, increase it. In addition to linear feedback, nonlinear feedback can also be used but applications with *deliberate* use of nonlinear feedback are seldom found in analog design and so will not be considered here. A detailed treatment of nonlinear feedback can be found in [7] and [28].

Figure 3.4 shows a general feedback system, where $\mathbf{H}$ represents the nonlinear element and $\mathbf{F}$ corresponds to the linear feedback network. In this case $F_2(j\omega_1, j\omega_2) = 0$ and $F_3(j\omega_1, j\omega_2, j\omega_3) = 0$. The overall closed-loop system is represented by $\mathbf{G}$.

Figure 3.4: A general linear feedback system.

The nonlinear transfer functions from order one to three for the closed loop system are derived using the same approach for a cascade system. The details can be found in [22] or [7] and are summarized below with

$$G_1(j\omega_1) = R(j\omega_1)H_1(j\omega_1) \quad (3.28)$$

$$G_2(j\omega_1, j\omega_2) = R(j\omega_1)R(j\omega_2)R(j\omega_1 + j\omega_2)H_2(j\omega_1, j\omega_2) \quad (3.29)$$

$$\begin{aligned} G_3(j\omega_1, j\omega_2, j\omega_3) = & R(j\omega_1)R(j\omega_2)R(j\omega_3)R(j\omega_1 + j\omega_2 + j\omega_3) \\ & \cdot [H_3(j\omega_1, j\omega_2, j\omega_3) - 2H_2(j\omega_1, j\omega_2) \\ & \cdot (F_1(j\omega_1 + j\omega_2)R(j\omega_1 + j\omega_2)H_2(j\omega_3, j\omega_1 + j\omega_2))] \end{aligned} \quad (3.30)$$

where

$$R(j\omega) = \frac{1}{1 + H_1(j\omega)F_1(j\omega)} = \frac{1}{1 + T(j\omega)} \quad (3.31)$$

is approximately the inverse of the loop gain $T(j\omega)$.

The expression for the linear gain $G_1(j\omega)$ is the well known linear gain for a feedback system where the gain is reduced by a factor of $R(j\omega_1)$. Similarly, the second-order transfer function is reduced by a factor of $R(j\omega_1)R(j\omega_2)R(j\omega_1 + j\omega_2)$ and the third-order transfer function is suppressed by a factor of $R(j\omega_1)R(j\omega_2)R(j\omega_3)R(j\omega_1 + j\omega_2 + j\omega_3)$. The second-order nonlinearity component also contributes to the third-order transfer function of $\mathbf{G}$ and can be used in conjunction with the loop gain to zero the third-order nonlinearity since the sign of the second-order term is opposite to the third-order term. Conversely, as will be shown later, if the system $\mathbf{H}$ has no third-order nonlinearities, or the second-order nonlinearity is dominant, the overall feedback system $\mathbf{G}$ will exhibit higher third-order distortion than the open loop system $\mathbf{H}$ [22].

For a memoryless system, the above transforms simplify to

$$G_1 = H_1 R \quad (3.32)$$

$$G_2 = H_2 R^3 \quad (3.33)$$

$$G_3 = [H_3 - 2H_2^2 F_1 R] R^4 \quad (3.34)$$

If the loop gain, $F_1 H_1 = T$, is large the following approximations can be made

$$G_1 \approx \frac{1}{F_1} \quad (3.35)$$

$$G_2 \approx \frac{H_2}{T^3} \quad (3.36)$$

$$G_3 \approx \left[H_3 - \frac{2H_2^2}{H_1} \right] \frac{1}{T^4} \quad (3.37)$$

When comparing the harmonic distortion of the memoryless feedback system to an open loop system, the reduction in distortion is clearly evident. For the open

loop configuration

$$HD_{2 \text{ open loop}} = \frac{A_1}{2} \frac{H_2}{H_1} \quad (3.38)$$

$$HD_{3 \text{ open loop}} = \frac{A_1^2}{4} \frac{H_3}{H_1} \quad (3.39)$$

whereas, for the closed loop

$$HD_{2 \text{ closed loop}} = \frac{A_1}{2} \frac{1}{T} \frac{H_2}{H_1} \quad (3.40)$$

$$HD_{3 \text{ closed loop}} = \frac{A_1^2}{4} \frac{1}{T} \left| \frac{H_3}{H_1} - \frac{2H_2^2}{H_1^2} \right| \quad (3.41)$$

The second harmonic distortion is reduced by a factor of T. The third harmonic is also reduced by a factor of T but there is also an additional reduction due to the negative contribution of the second-order nonlinearity, unless H_2 is dominant as previously discussed.

3.6 Calculation of Volterra Kernels

The calculation of harmonics and intermodulation products based on the Volterra series can be approached by several methods. The three main methods are;

1. the Volterra series method;
2. the direct calculation method; and
3. the perturbation method.

The Volterra series method calculates the Volterra kernels of the node voltages and currents in a circuit, from which the harmonic and intermodulation products are determined by using Table 3.1. The method computes the Volterra kernels in increasing order by each time solving the same linearized equivalent network of the original circuit. This linear circuit is used to compute the first-order kernels. The second-order kernels are solved using the same linearized network but with different inputs that depend on the first-order kernels. The third-order kernels are computed similarly, but with inputs that depend on the first and second-order kernels. Higher

order kernels are computed in the same manner. An example of the Volterra series method can be found in Appendix B.5

The Volterra series method, however, has two main disadvantages.

1. It requires redundant computations. If, for instance, only the third harmonic distortion is required, then from Table 3.1, only the Volterra Kernel $H_3(j\omega_1, j\omega_1, j\omega_1)$ is required. If, however, the third-order intermodulation is required then the Kernel must be re-evaluated for $H_3(j\omega_1, j\omega_1, -j\omega_2)$.
2. The Volterra series method requires the use of tensors for circuits with more than one input port, such as mixers. However, the use of tensors can be quite cumbersome.

The direct calculation method is a simpler method that avoids the use of tensors and the redundancy of the Volterra series approach. The analysis is limited to circuits with at most two input ports or with two input frequencies applied at one input port. The method leads to the same results as the Volterra series approach [8] but is simpler to use. For this reason, this method will be used for this work.

The perturbation method [24], leads to the same results as the Volterra series and the direct calculation method. However, as with the Volterra series method, the perturbation method is only applicable for single input circuits and requires redundant calculations.

3.6.1 Direct Calculation of Nonlinear Responses

This method is explained using an example common-source amplifier circuit, depicted in Figure 3.5(a). The source and bulk are grounded and the input to the amplifier is a voltage source with a source resistance R_S . The output is loaded with a resistance R_L and a capacitance C_L . A simplified AC-equivalent circuit is shown in Figure 3.5(b). When in strong inversion, the gate-source capacitance, C_{gs} , is nearly constant with respect to V_{gs} when the transistor is operating in the saturation region and so C_{gs} is assumed to be constant for now. The gate-drain and drain-bulk capacitances, C_{gd} and C_{db} , are neglected for simplicity. The drain current is assumed to be a function of only V_{gs} and so the nonlinearities due to the output conductance are not considered for this simple example. The case with V_{ds} will be addressed later. It should also be

noted that since only single input circuits will be considered for this work, the example given here is not exactly applicable to a *two-port input* circuit but is applicable for a *two-tone input*. A detailed example of a two-port input can be found in [7].

Figure 3.5: (a) A common-source amplifier. (b) AC-equivalent circuit.

First-order Responses

The first-order kernels are computed with the linearized circuit first with the matrix equation

$$\mathbf{Y}(\omega_1) \cdot \mathbf{V}_{1,0} = \mathbf{IN}_{1,0} \quad (3.42)$$

where $\mathbf{Y}(\omega_1)$ is the admittance matrix of the circuit at frequency ω_1 , $\mathbf{V}_{1,0}$ is the vector of the node voltages $V_{x,1,0}$ at frequency ω_1 , and $\mathbf{IN}_{1,0}$ is the vector of the input excitations. The notation $V_{x,m,n}$ represents the phasor of the voltage response at node x at the frequency $m\omega_1 + n\omega_2$.

Figure 3.6: Linearized circuit for calculating first-order kernels.

The linearized circuit is shown in Figure 3.6. By applying Kirchhoff's current law at nodes **1** and **2**, the equation (3.42) yields

$$\begin{bmatrix} Y_i(\omega_1) & 0 \\ g_m & Y_o(\omega_1) \end{bmatrix} \cdot \begin{bmatrix} V_{1,1,0} \\ V_{2,1,0} \end{bmatrix} = \begin{bmatrix} G_s V_{in} \\ 0 \end{bmatrix} \quad (3.43)$$

where

$$Y_i(\omega) = G_s + j\omega C_{gs} \quad (3.44)$$

is the input admittance,

$$Y_o(\omega) = G_L + j\omega C_L \quad (3.45)$$

is the output admittance, $G_s = 1/R_s$, and $G_L = 1/R_L$.

Using Cramer's rule, the response at frequency ω_1 at the output node 1 is given by

$$V_1(\omega_1) = V_{1,1,0} = -\frac{G_s}{Y_i(\omega_1)} V_{in} \quad (3.46)$$

and, the response at frequency $\pm\omega_2$ at the output node 1 is given by

$$V_1(\pm\omega_2) = V_{1,0,\pm 1} = -\frac{G_s}{Y_i(\pm\omega_2)} V_{in} \quad (3.47)$$

Similarly, the response at frequency ω_1 at the output node 2 is given by

$$V_2(\omega_1) = V_{2,1,0} = -\frac{g_m G_s}{Y_i(\omega_1) Y_o(\omega_1)} V_{in} \quad (3.48)$$

and, the response at frequency $\pm\omega_2$ at the output node 2 is given by

$$V_2(\pm\omega_2) = V_{2,0,\pm 1} = -\frac{g_m G_s}{Y_i(\pm\omega_2)Y_o(\pm\omega_2)} V_{in} \quad (3.49)$$

Although the solutions can be written directly by inspection, the above procedure allows for a systematic approach that will be more beneficial for larger circuits.

Second-order Responses

The second-order responses at $2\omega_1$, $2\omega_2$, and $|\omega_1 \pm \omega_2|$ are solved using the same linearized network as used for the first-order responses, except different inputs are used, in a similar fashion as the Volterra series approach. The external signal sources are removed and *second-order nonlinear current sources* are applied for each basic nonlinearity in the circuit and placed in parallel with the linearized equivalent of the nonlinearity. The expressions for the nonlinear current sources at $2\omega_1$, $2\omega_2$, and $|\omega_1 \pm \omega_2|$ for the direct approach differs from those for the Volterra series approach and are listed in Table 3.2. The details for the derivations of the expressions can be found in [7]. The expressions for the current sources for the direct approach that have to be applied depend on the type of the basic nonlinearity, on its second-order nonlinearity coefficient, and on the first-order responses. Hence, the first-order responses must be calculated prior to the second-order responses.

The circuit that has to be solved for the computation of the second-order kernels in the example circuit is shown in Figure 3.7. The original nonlinear circuit of Figure 3.5(b) has one nonlinear transconductance which gives rise to a nonlinear current source that is placed in parallel with the linearized current source.

Figure 3.7: Linearized circuit for calculating second-order kernels.

Table 3.2: Nonlinear second-order current sources to directly compute second-order intermodulation products at $|\omega_1 \pm \omega_2|$ and at $2\omega_1$ (or $2\omega_2$) where $V_{i,m,n}$ is the controlling voltage for the transconductance and $V_{j,m,n}$ is the controlling voltage for the output conductance. Taken from [7].

type of nonlinearity	nonlinear current source for response at $ \omega_1 \pm \omega_2 $	nonlinear current source for response at $2\omega_1$
transconductance	$c_{(2,0,0)} V_{i,1,0} V_{i,0,\pm 1}$	$\frac{1}{2} c_{2,0,0} V_{i,1,0}^2$
output conductance	$c_{(0,2,0)} V_{j,1,0} V_{j,0,\pm 1}$	$\frac{1}{2} c_{0,2,0} V_{j,1,0}^2$
two-dimensional conductance (cross-terms only)	$\frac{1}{2} [c_{(1,1,0)} V_{i,1,0} V_{j,0,\pm 1} + c_{(1,1,0)} V_{i,0,\pm 1} V_{j,1,0}]$	$\frac{1}{2} c_{(1,1,0)} V_{i,1,0} V_{j,1,0}$

The computation of the responses for the intermodulation products at $|\omega_1 \pm \omega_2|$ are determined from

$$\mathbf{Y}(\omega_1 \pm \omega_2) \cdot \mathbf{V}_{1,1} = \mathbf{IN}_{1,1} \quad (3.50)$$

The vector $\mathbf{Y}(\omega_1 \pm \omega_2)$ is the admittance matrix at $|\omega_1 \pm \omega_2|$, $\mathbf{V}_{1,1}$ is the vector of phasors of the components at $|\omega_1 \pm \omega_2|$, and $\mathbf{IN}_{1,1}$ is a vector of second-order nonlinear current sources for the computation at $|\omega_1 \pm \omega_2|$ from Table 3.2.

The computation of the responses at $2\omega_1$ are determined from

$$\mathbf{Y}(2\omega_1) \cdot \mathbf{V}_{2,0} = \mathbf{IN}_{2,0} \quad (3.51)$$

where $\mathbf{V}_{2,0}$ is the vector of phasors of the components at $2\omega_1$, and $\mathbf{IN}_{2,0}$ is a vector of second-order nonlinear current sources for the computation at $2\omega_1$ from the right hand column of Table 3.2. Similarly, the computation for the response at $2\omega_2$ are determined from

$$\mathbf{Y}(2\omega_2) \cdot \mathbf{V}_{0,2} = \mathbf{IN}_{0,2} \quad (3.52)$$

For the response at $|\omega_1 \pm \omega_2|$, Kirchhoff's current law is applied to give

$$\begin{bmatrix} 1 & 0 \\ 0 & Y_o(\omega_1 \pm \omega_2) \end{bmatrix} \cdot \begin{bmatrix} V_{1,1,\pm 1} \\ V_{2,1,\pm 1} \end{bmatrix} = \begin{bmatrix} 0 \\ -i_{NL2ids} \end{bmatrix} \quad (3.53)$$

For a one-dimensional conductance, v_{gs} , or V_1 , controls the nonlinearity and so the value of i_{NL2ids} , from Table 3.2, is

$$i_{NL2ids} = c_{(2,0,0)} V_{1,1,0} V_{1,0,\pm 1} \quad (3.54)$$

which, from equations (3.46) and (3.47), yields

$$i_{NL2ids} = \frac{c_{(2,0,0)} G_s^2}{Y_i(\omega_1) Y_i(\pm \omega_2)} V_{in}^2 \quad (3.55)$$

The solutions to (3.53) are

$$V_{1,1,\pm 1} = 0 \quad (3.56)$$

and

$$V_{2,1,\pm 1} = \frac{c_{(2,0,0)} G_s^2}{Y_i(\omega_1) Y_i(\pm \omega_2) Y_o(\omega_1 \pm \omega_2)} V_{in}^2 \quad (3.57)$$

For the response at $2\omega_1$, Kirchhoff's current law is applied to give

$$\begin{bmatrix} 1 & 0 \\ 0 & Y_o(2\omega_1) \end{bmatrix} \cdot \begin{bmatrix} V_{1,2,0} \\ V_{2,2,0} \end{bmatrix} = \begin{bmatrix} 0 \\ -i_{NL2ids} \end{bmatrix} \quad (3.58)$$

From the right hand column in Table 3.2, i_{NL2ids} is

$$\begin{aligned} i_{NL2ids} &= \frac{c_{(2,0,0)}}{2} V_{1,1,0}^2 \\ &= \frac{c_{(2,0,0)}}{2} \frac{G_s^2}{Y_i^2(\omega_1)} V_{in}^2 \end{aligned} \quad (3.59)$$

The solutions to (3.58) are

$$V_{1,2,0} = 0 \quad (3.60)$$

$$V_{2,2,0} = \frac{c_{(2,0,0)}}{2} \frac{G_s^2}{Y_i^2(\omega_1)Y_o(2\omega_1)} V_{in}^2 \quad (3.61)$$

Similarly, the solutions for the responses at $2\omega_2$ are

$$V_{1,0,2} = 0 \quad (3.62)$$

$$V_{2,0,2} = \frac{c_{(2,0,0)}}{2} \frac{G_s^2}{Y_i^2(\omega_2)Y_o(2\omega_2)} V_{in}^2 \quad (3.63)$$

The second-order harmonic distortion can now be determined as

$$\begin{aligned} HD_2 &= \left| \frac{V_{2,2,0}}{V_{2,1,0}} \right| \\ &= \frac{1}{2} \left| \frac{c_{(2,0,0)}}{g_m} \frac{G_s}{Y_i(\omega_1)} \frac{Y_o(\omega_1)}{Y_o(2\omega_1)} \right| V_{in} \end{aligned} \quad (3.64)$$

from which it can be seen that HD_2 depends on magnitude of V_{in} , the ratio of the second-order nonlinear coefficient, $c_{(2,0,0)}$ and g_m , the ratio of the source admittance and input admittance and the ratio of the output admittance at ω_1 and $2\omega_1$. For low frequencies, where C_{gs} and C_L can be disregarded, (3.64) simplifies to

$$HD_2 = \frac{1}{2} \left| \frac{c_{(2,0,0)}}{g_m} \right| V_{in} \quad (3.65)$$

where HD_2 depends only on the ratio of $c_{(2,0,0)}$ and g_m and the magnitude of V_{in} .

Third-order Responses

The third-order responses at $3\omega_1$, $3\omega_2$, and $2\omega_1 \pm \omega_2$ are solved using exactly the same approach as used for the second-order responses, except $i_{NL2_{ids}}$ is replaced by $i_{NL3_{ids}}$. The expressions for the nonlinear current sources at $2\omega_1 \pm \omega_2$ and $3\omega_1$ (and $3\omega_2$) are listed in Table 3.3 and 3.4. The expressions for the current sources depend on the type of the basic nonlinearity, on its second-order and third-order nonlinearity

coefficients, and on the first-order and second-order responses. Hence, the first-order and second-order responses must be calculated prior to the third-order responses.

Table 3.3: Nonlinear third-order current sources to directly compute third-order intermodulation products at $2\omega_1 \pm \omega_2$ where $V_{i,m,n}$ is the controlling voltage for the transconductance, $V_{j,m,n}$ is the controlling voltage for the output conductance, and $V_{k,m,n}$ is the controlling voltage for the bulk conductance. Taken from [7].

type of nonlinearity	nonlinear current source for response at $2\omega_1 \pm \omega_2$
transconductance	$c_{(2,0,0)}V_{i,1,0}V_{i,1,\pm 1} + c_{(2,0,0)}V_{i,0,\pm 1}V_{i,2,0} + \frac{3}{4}c_{(3,0,0)}V_{i,1,0}^2V_{i,0,\pm 1}$
output conductance	$c_{(0,2,0)}V_{j,1,0}V_{j,1,\pm 1} + c_{(0,2,0)}V_{j,0,\pm 1}V_{j,2,0} + \frac{3}{4}c_{(0,3,0)}V_{j,1,0}^2V_{j,0,\pm 1}$
bulk conductance	$c_{(0,0,2)}V_{k,1,0}V_{k,1,\pm 1} + c_{(0,0,2)}V_{k,0,\pm 1}V_{k,2,0} + \frac{3}{4}c_{(0,0,3)}V_{k,1,0}^2V_{k,0,\pm 1}$
two-dimensional conductance (cross-terms only)	$\frac{1}{2}c_{(1,1,0)}[V_{i,0,\pm 1}V_{j,2,0} + V_{i,1,0}V_{j,1,\pm 1} + V_{i,1,\pm 1}V_{j,1,0} + V_{i,2,0}V_{j,0,\pm 1}]$ $+\frac{1}{4}c_{(2,1,0)}[2V_{i,0,\pm 1}V_{i,1,0}V_{j,1,0} + V_{i,1,0}^2V_{j,0,\pm 1}]$ $+\frac{1}{4}c_{(1,2,0)}[2V_{i,1,0}V_{j,0,\pm 1}V_{j,1,0} + V_{i,0,\pm 1}V_{j,1,0}^2]$
three-dimensional conductance (cross-terms only)	$\frac{1}{4}c_{(1,1,1)}[V_{i,0,\pm 1}V_{j,1,0}V_{k,1,0} + V_{i,1,0}V_{j,0,\pm 1}V_{k,1,0} +$ $V_{i,1,0}V_{j,1,0}V_{k,0,\pm 1}]$

For the response at $2\omega_1 \pm \omega_2$, Kirchhoff's current law is applied to give

$$\begin{bmatrix} 1 & 0 \\ 0 & Y_o(2\omega_1 \pm \omega_2) \end{bmatrix} \cdot \begin{bmatrix} V_{1,2,\pm 1} \\ V_{2,2,\pm 1} \end{bmatrix} = \begin{bmatrix} 0 \\ -i_{NL3_{ids}} \end{bmatrix} \quad (3.66)$$

From Table 3.3, $i_{NL3_{ids}}$ is

$$i_{NL3_{ids}} = c_{(2,0,0)}V_{1,1,0}V_{1,1,\pm 1} + c_{(2,0,0)}V_{1,0,\pm 1}V_{1,2,0} + \frac{3}{4}c_{(3,0,0)}V_{1,1,0}^2V_{1,0,\pm 1} \quad (3.67)$$

Since $V_{1,1,\pm 1} = V_{1,2,0} = 0$, $i_{NL3_{ids}}$ simplifies to

$$\begin{aligned} i_{NL3_{ids}} &= \frac{3}{4}c_{(3,0,0)}V_{1,1,0}^2V_{1,0,\pm 1} \\ &= \frac{3}{4}c_{(3,0,0)}\frac{G_s^3}{Y_i^2(\omega_1)Y_i(\pm\omega_2)}V_{in}^3 \end{aligned} \quad (3.68)$$

Table 3.4: Nonlinear third-order current sources to directly compute third-order harmonics at $3\omega_1$ where $V_{i,m,n}$ is the controlling voltage for the transconductance, $V_{j,m,n}$ is the controlling voltage for the output conductance, and $V_{k,m,n}$ is the controlling voltage for the bulk conductance. Taken from [7].

type of nonlinearity	nonlinear current source for response at $3\omega_1$
transconductance	$c_{(2,0,0)}V_{i,1,0}V_{i,2,0} + \frac{1}{4}c_{(3,0,0)}V_{i,1,0}^3$
output conductance	$c_{(0,2,0)}V_{j,1,0}V_{j,2,0} + \frac{1}{4}c_{(0,3,0)}V_{j,1,0}^3$
bulk conductance	$c_{(0,0,2)}V_{k,1,0}V_{k,2,0} + \frac{1}{4}c_{(0,0,3)}V_{k,1,0}^3$
two-dimensional conductance (cross-terms only)	$\frac{1}{2}c_{(1,1,0)}[V_{i,1,0}V_{j,2,0} + V_{i,2,0}V_{j,1,0}] + \frac{1}{4}c_{(2,1,0)}V_{i,1,0}^2V_{j,1,0} + \frac{1}{4}c_{(1,2,0)}V_{i,1,0}V_{j,1,0}^2$
three-dimensional conductance (only cross-terms)	$\frac{1}{4}c_{(1,1,1)}V_{i,1,0}V_{j,1,0}V_{k,1,0}$

from which the output intermodulation product is

$$V_{2,2,\pm 1} = -\frac{3}{4}c_{(3,0,0)}\frac{G_s^3}{Y_i^2(\omega_1)Y_i(\pm\omega_2)Y_o(2\omega_1 \pm \omega_2)}V_{in}^3 \quad (3.69)$$

For the response at $3\omega_1$, Kirchhoff's current law is applied to give

$$\begin{bmatrix} 1 & 0 \\ 0 & Y_o(3\omega_1) \end{bmatrix} \cdot \begin{bmatrix} V_{1,3,0} \\ V_{2,3,0} \end{bmatrix} = \begin{bmatrix} 0 \\ -i_{NL3ids} \end{bmatrix} \quad (3.70)$$

From Table 3.4, i_{NL3ids} is

$$\begin{aligned} i_{NL3ids} &= c_{(2,0,0)}V_{1,1,0}V_{1,2,0} + \frac{1}{4}c_{(3,0,0)}V_{1,1,0}^3 \\ &= \frac{c_{(3,0,0)}}{4}\frac{G_s^3}{Y_i^3(\omega_1)}V_{in}^3 \end{aligned} \quad (3.71)$$

from which the third-order harmonic tones are

$$V_{2,3,0} = -\frac{c_{(3,0,0)}}{4} \frac{G_s^3}{Y_i^3(\omega_1)Y_o(3\omega_1)} V_{in}^3 \quad (3.72)$$

$$V_{2,0,3} = -\frac{c_{(3,0,0)}}{4} \frac{G_s^3}{Y_i^3(\omega_2)Y_o(3\omega_2)} V_{in}^3 \quad (3.73)$$

The normalized third-order harmonic distortion is

$$\begin{aligned} HD_3 &= \left| \frac{V_{2,3,0}}{V_{2,1,0}} \right| \\ &= \frac{1}{4} \left| \frac{c_{(3,0,0)}}{g_m} \frac{G_s^2}{Y_i^2(\omega_1)} \frac{Y_o(\omega_1)}{Y_o(3\omega_1)} \right| V_{in}^2 \end{aligned} \quad (3.74)$$

Here HD_3 depends on magnitude of V_{in}^2 , the ratio of the third-order nonlinear coefficient, $c_{(3,0,0)}$ and g_m , the ratio of the square of source admittance and input admittance and the ratio of the output admittance at ω_1 and $3\omega_1$. For low frequencies, where C_{gs} and C_L can be disregarded, (3.74) simplifies to

$$HD_3 = \frac{1}{4} \left| \frac{c_{(3,0,0)}}{g_m} \right| V_{in}^2 \quad (3.75)$$

where HD_3 depends only on the ratio of $c_{(3,0,0)}$ and g_m and the magnitude of V_{in}^2 .

The normalized third-order intermodulation product is

$$\begin{aligned} IM_3 &= \left| \frac{V_{2,2,-1}}{V_{2,1,0}} \right| \\ &= \frac{3}{4} \left| \frac{c_{(3,0,0)}}{g_m} \frac{G_s^2}{Y_i(\omega_1)Y_i(\omega_2)} \frac{Y_o(\omega_1)}{Y_o(2\omega_1 - \omega_2)} \right| V_{in}^2 \end{aligned} \quad (3.76)$$

from which yields the third-order intercept point

$$V_{IIP3}^2 = \frac{4}{3} \left| \frac{g_m}{c_{(3,0,0)}} \frac{Y_i(\omega_1)Y_i(\omega_2)}{G_s^2} \frac{Y_o(2\omega_1 - \omega_2)}{Y_o(\omega_1)} \right| \quad (3.77)$$

However, since ω_1 , ω_2 , and $2\omega_2 - \omega_1$ are close together, then $Y_i(\omega_1) \approx Y_i(\omega_2)$ and $Y_o(\omega_1) \approx Y_o(2\omega_2 - \omega_1)$ and *IIP3* simplifies to

$$V_{IIP3}^2 = \frac{4}{3} \left| \frac{g_m}{c_{(3,0,0)}} \frac{Y_i^2(\omega_1)}{G_s^2} \right| \quad (3.78)$$

and for low frequencies, *IIP3* simplifies even further to

$$V_{IIP3}^2 = \frac{4}{3} \left| \frac{g_m}{c_{(3,0,0)}} \right| \quad (3.79)$$

where *IIP3* depends only on the ratio of g_m and $c_{(3,0,0)}$.

3.7 Summary

In this chapter, time domain and frequency domain representations of the Volterra series to third-order were presented. Harmonic and intermodulation products were expressed in terms of Volterra Kernels. The analysis of cascade connection of two nonlinear systems, in terms of Volterra series, was also presented as well as the analysis of a nonlinear system with linear feedback. For a cascaded system, it was shown that the last stage tends to contribute more to the nonlinear distortion and generally needs to be as linear as possible. With a feedback systems, the loop gain generally helps suppress distortion products. However, it was shown that in an ideal square-law device with linear feedback, second-order distortion products can be fed back to the input, mix with the original input signal and create third-order distortion products where none existed without the feedback. This contradiction to the well known notion that feedback reduces nonlinear distortion is only applicable in the case when the third-order nonlinear distortion of the device is very low. The final section presented the "Direct Calculation Method" [7] for determining the harmonic and intermodulation products of a nonlinear circuit and an example was provided. More detail on various circuit configurations and the interpretation of the results will be covered in Chapter 5.

Chapter 4

MOSFET I-V Models

4.1 Introduction

Motivated by higher levels of integration and faster speeds, CMOS transistors have been consistently scaled to smaller and smaller feature sizes and continue to be reduced unabatedly towards 0.1 and sub-0.1 μm gate lengths. At the same time, significant advances with the MOSFET model have been made since the SPICE level 1 model was first introduced. Over 100 models have been reported since 1970 [32] including the more well known MOS level 1, 2, and 3 [33], Phillips MOS level 9 [34], MISNAN [35], PCIM [36], EKV [37], BSIM [38], BSIM2, and BSIM3 [32] and, most recently, BSIM4 [39] models. With such small feature sizes, significantly more parameters are needed to accurately model the modern CMOS device. By way of example, the BSIM3v3 model has over 150 parameters. Such a large number of parameters poses a significant difficulty for nonlinear analysis. It is nearly impossible to derive the nonlinear coefficients of the MOSFET I-V characteristics without the expressions becoming too intractable. The problem is further complicated by the fact that most MOSFET models are piece-wise expressions and many suffer from discontinuities in first-order and higher-order derivatives. Hence, a “relatively” simple *single-piece* expression with only a few parameters is required for an analysis of the nonlinear behaviour of a MOSFET.

The subject of this chapter focuses on an appropriate single-piece MOSFET model and the derivation of the nonlinear coefficients arising from that model. The emphasis will be on the *moderate* inversion region. The nonlinear analysis for both the *weak* and *strong* inversion regions have been extensively addressed in [7]. Section 4.2 gives a qualitative description of the MOSFET transistor operation and the

differences between the operating regions of a MOSFET. Section 4.3 briefly presents, as background information, some of the issues related to MOSFET modelling. It is by no means complete as an entire text book can be written on the subject matter, but is meant to present a “taste” of the issues. In Section 4.4, a single-piece expression based on [17] is introduced and includes the effects of velocity saturation, mobility degradation and channel-length modulation. Expressions for the nonlinear capacitances are presented in Section 4.5 and, finally, expressions for the nonlinear coefficients for the drain current are derived in Section 4.6.

4.2 Overview of a MOSFET Transistor

4.2.1 MOSFET Transistor Operation

The MOSFET transistor is a four terminal device with gate, drain, source and bulk terminals. Figure 4.1 shows an n-channel transistor with external voltages applied. The voltages can be either referred to the source, which is done in most common models such as the SPICE and BSIM models [32], [39], or to the bulk, as is done in the EKV model [37]. The bulk referenced models offer certain advantages over the source referenced models but are less commonly used [40]. Figure 4.1 depicts source referenced voltages, which will be used throughout this work to avoid confusion.

A voltage applied to the gate terminal, V_{GS} is used to form an inversion layer, or *channel*, between the drain and source terminals. The channel conducts the current as V_{DS} is increased (for a fixed source-bulk voltage, V_{SB}). The inversion layer (for an n-channel MOSFET) is an extremely thin layer of electrons at the surface of the p-type bulk (or body) below the gate oxide that forms the conduction channel between the drain and source. As a positive V_{GS} is applied, positive charge is placed on the gate and holes are repelled from the surface, leaving it depleted of holes. The resulting depletion layer contains a number of negatively charged acceptor ions. As V_{GS} is increased, the surface becomes attractive to electrons which are readily available through one or both of the n^+ regions. An inversion layer is formed where the concentration of electrons below the gate oxide increases dramatically after a threshold voltage, V_{th} , has been reached. The channel goes from *weak* inversion, where the V_{GS} is less than V_{th} and the electron concentration is low, through *moderate*

Figure 4.1: An n-channel MOSFET transistor in the inversion region with negative bias between the body and the source.

inversion, where V_{GS} is near V_{th} , to *strong* inversion, when V_{GS} is greater than V_{th} by several times the thermal voltage, $\phi_t = kT/q$, and the electron concentration is high.

A positive potential difference, V_{DS} , across the drain and source causes electron movement from the source, through the channel and out the drain. This corresponds to a positive current, I_{DS} , from the drain, through the channel, to the source. In weak inversion, when the channel is not well formed, a very small amount of current still flows, primarily due to diffusion. In strong inversion, the channel is well formed and drift is the dominant mechanism for current flow resulting in significantly more current than in weak inversion. As V_{DS} is increased, the inversion layer charge density at the drain end decreases until the point where it is approximately zero. At this point the channel is said to be *pinched-off*, although current continues to flow. The drain-source voltage at which pinch-off occurs is $V_{DS(sat)}$. If V_{DS} is increased further, the device is in the *saturation* region and the current remains relatively constant with increasing V_{DS} for a fixed V_{GS} . For lower V_{DS} , the channel is not pinched-off and the transistor is in *nonsaturation*. In nonsaturation, the current tends to increase linearly with increasing V_{DS} for a given V_{GS} . Since the current increases nearly linearly with V_{DS} in nonsaturation, this region is also often referred as the *linear* region, or sometimes the *triode* region.

The transistor operation is also affected by the source-bulk voltage. If a negative voltage is applied to the bulk, the population of the electrons will decrease, resulting in a decrease of I_{DS} . This phenomenon is referred to as *body effect*. The bulk is also sometimes referred to as a *back gate* since the bulk is a conductive structure “insulated” from the channel by the depletion region just as the gate is separated from the channel by the oxide [40].

4.2.2 MOSFET Transistor Characteristics

Figure 4.2 shows a representative curve of I_{DS} versus V_{GS} for a fixed V_{DS} with the various inversion regions marked. In the weak inversion region, diffusion current is dominant and the drain current tends to exhibit an exponential behaviour with respect to V_{GS} , hence, a relatively straight line when plotted on a log scale. Drift current dominates in strong inversion and tends to exhibit a square law behaviour. The moderate inversion region is a narrow region where the mechanism for the current flow changes from drift to diffusion current and both are significant relative to each other. As a result, simplifications of the I-V equations are not possible, making modeling of this region very difficult [32]. Furthermore, the boundaries for the moderate inversion region are difficult to define exactly but are generally within several ϕ_t 's of the threshold voltage [40].

Figure 4.2: $\text{Log}(I_{DS})$ - V_{GS} characteristics of an n-channel MOSFET for a fixed V_{DS} .

Figure 4.3: I_{DS} - V_{DS} characteristics for various V_{GS} . (a) Linear y axis; (b) Logarithmic y axis [40].

The well known, typical output I-V characteristic curve versus V_{DS} , for fixed V_{GS} , is shown in Figure 4.3(a). For low V_{DS} , the current increases fairly dramatically in a linear fashion with V_{DS} and is referred to as the linear region or triode region.

For higher values of V_{DS} , the drain current reaches a maximum point or saturates with increasing V_{DS} . This region is referred to as the saturation region. In addition, since I_{DS} is quadratically dependent on V_{GS} (for a long channel device), the spacing between each step of V_{GS} in Figure 4.3(a) should also increase quadratically. The weak and moderate inversion regions are also shown in Figure 4.3(a) but are more difficult to distinguish than in Figure 4.2. However, if I_{DS} is plotted on the logarithmic axis instead, as in Figure 4.3(b), the weak and moderate regions become easier to discern [40]. In the weak inversion region, the current is exponentially dependent on V_{GS} and so the spacing between each curve in the weak inversion region is almost equal for equal V_{GS} . Furthermore, the current can change several orders of magnitude for a small change in V_{GS} . In the moderate inversion region, the spacing between the curves becomes smaller for increasing V_{GS} , but the current can still vary a couple orders of magnitude. In the strong inversion region, the spacing becomes very small and difficult to discern when plotted on a logarithmic axis.

4.3 MOSFET Modelling Issues

The number one goal of any device model is “A model shall be accurate *and* simple,” [39]. Achieving a balance between simplicity and accuracy is a growing problem for modelling engineers, especially concerning the MOSFET model. MOSFET modelling has largely been driven by the needs of digital circuit designers who required good accuracy for first-order I-V characteristics but small signal parameters were less important [41]. With the advent of RF CMOS, more and more emphasis is being placed on accuracy, especially, with small signal parameters and second and third-order derivatives.

Most CMOS models are, in fact, “piece-wise” models based on separate expressions for the linear and saturation region as well as the weak and strong inversion region. They often suffer from inaccuracies at or near the boundaries of each region. Even more importantly, the first-order and higher-order derivatives of I-V characteristics are often very inaccurate, most especially in the transition region between the two operating regions. Several models exhibit discontinuities at the transition region [41]. It is essential for accurate nonlinear simulations that the models exhibit accurate and continuous derivatives up to, and including, at least third-order. Many

CMOS models do not. More recent models such as the EKV, BSIM3, and BSIM4 provide continuous expressions over all regions of operation and inversion through the use of continuous smoothing functions [39]. However, these functions add significant complexity to the models and, especially the derivatives, may not be necessarily accurate in the transition regions. The accuracy will highly depend on the parameter extraction process where the parameters may have been determined for accurate I-V characteristics but not for higher-order derivatives.

Many physical effects play a much more important role than with the larger, older transistors of the past. Examples are reverse short-channel and narrow-width effects, channel length modulation, drain induced barrier lowering, velocity saturation, mobility degradation due to the vertical electric field, gate induced drain leakage current (GIDL), gate tunnelling currents, impact ionization, velocity overshoot, self-heating, and substrate current body effects, to name but a few [32]. As the CMOS feature size continues to scale down to even smaller sizes, more two and three dimensional effects will have to be incorporated in the models [7]. In addition, RF behaviours, such as non-quasi-static effects, distributed gate and substrate resistance, frequency dependence of transconductances, and more accurate noise models also need to be incorporated as more MOSFET circuits are used for RF applications. As a result, the MOSFET model is likely to continuously change and grow in complexity as device sizes scale down and the existing models become increasingly inaccurate.

It must be pointed out that the nonlinear coefficients for a MOSFET can only be predicted with limited accuracy [7]. There are several reasons for this.

1. The computation of nonlinear coefficients requires sufficiently accurate model equations. There are currently several different MOSFET models, each with varying degrees of accuracy depending on the device size. These models are also continuously changing due to scaling and existing models may become inaccurate for small devices.
2. Many second-order effects occur simultaneously but it is often assumed that the different effects are small enough that they do not interact. These assumptions often result in explicit expression for the model that are computationally efficient. However, the error on the derivatives tends to increase with the order of the derivative [7].

3. Most models use separate equations for the different regions of operation and inversion resulting in discontinuities in the I-V equations. The discontinuities cause significant errors in the derivatives, and hence, errors in the distortion analysis. Several models use smoothing functions for transiting from one region to another, but these smoothing functions may not have continuous derivatives. In addition, those with continuous derivatives may not accurately model the transition region [41].

4.4 A Simple Moderate Inversion Nonlinear Model

The nonlinear equivalent circuit of a MOSFET is shown in Figure 4.4. The model is generally valid for frequencies less than one-tenth of the cut-off frequency [40] which, for a modern deep sub-micron device, can be in excess of 50 - 80 GHz. The model consists of a nonlinear drain current that is a function of V_{gs} , V_{ds} , and V_{sb} and a number of nonlinear capacitances, C_{gs} , C_{gd} , C_{gb} , C_{sb} , and C_{db} . The capacitances will be discussed in Section 4.5.

Figure 4.4: Nonlinear model of a MOSFET.

4.4.1 Long Channel Model

In the moderate inversion region, both drift and diffusion currents are important and must be accounted for in modeling the I_{DS} - V_{DS} characteristics. Unfortunately, convenient simplifications are not known [40], and so analytical solutions tend to be quite complicated. Most compact models, such as BSIM3 [32] or PCIM [36], tend to

model the weak and strong inversion quite accurately and use smoothing functions to ensure the drain current is continuous when transitioning from weak inversion to strong inversion. If the derivatives of the drain current with respect to any terminal voltage is not continuous, severe errors can occur at the boundaries and convergence difficulties can occur during computer simulations [40].

A semi-empirical, “single-piece” expression that provides good accuracy in moderate inversion and acceptable accuracy in weak and strong inversion is used [17] and is given by

$$I_{ds} = I_{d0} \left\{ \left[\ln \left(1 + e^{V_{od}/(2n\phi_t)} \right) \right]^2 - \left[\ln \left(1 + e^{(V_{od}-nV_{ds})/(2n\phi_t)} \right) \right]^2 \right\} \quad (4.1)$$

where

$$I_{d0} = \mu_0 C'_{ox} 2n\phi_t^2 \frac{W}{L} \quad (4.2)$$

$V_{od} = V_{gs} - V_{th}$ is the overdrive voltage, C'_{ox} is the gate oxide capacitance per unit area, μ_0 is the low electric field electron surface mobility, $\phi_t = kT/q$ is the thermal voltage, n is the subthreshold slope factor, V_{th} is the threshold voltage, V_{gs} is the gate-source voltage, and V_{ds} is the drain-source voltage.

The threshold voltage is modeled using a long-channel model where

$$V_{th} = V_{T0} + \gamma(\sqrt{\phi + V_{sb}} - \sqrt{\phi}) \quad (4.3)$$

with

$$V_{T0} = V_{FB} + \phi + \gamma\sqrt{\phi} \quad (4.4)$$

and

$$\gamma = \frac{\sqrt{2q\epsilon_s N_A}}{C'_{ox}} \quad (4.5)$$

V_{T0} is the threshold voltage at $V_{SB} = 0$, γ is the body effect coefficient, V_{sb} is the source-bulk voltage, V_{FB} is the flat band voltage, ϕ is the surface inversion potential, N_A is the acceptor concentration and ϵ_s is the permittivity of silicon, which is about 104 pF/m.

The parameter n is related to the subthreshold slope and describes the rate of the exponential increase of I_{ds} with V_{gs} in the subthreshold region. The value of n typically ranges from 1 - 2, with the higher values when short-channel effects are

present. It is given by

$$n = 1 + \frac{\gamma}{2\sqrt{\phi + V_{sb}}} \quad (4.6)$$

The parameter ϕ is difficult to define exactly [40]. It can vary from two times the Fermi level, ϕ_F , in the weak inversion region to $2\phi_F + \Delta\phi$ in strong inversion [17]. Typically $\Delta\phi \approx 6\phi_t$ in strong inversion for uniform substrates. However, ϕ is often incorrectly pinned to a constant value of $2\phi_F$ even for the strong inversion region [40]. Several expressions can be found to describe ϕ but most are only valid for the strong inversion region. An empirical expression suggested in [17] is valid for all regions and is given by

$$\phi = 2\phi_F + \Delta\phi \left[1 - \left(1 + e^{\frac{V_{od}}{\kappa\phi_t}} \right)^{-1} \right] \quad (4.7)$$

where κ controls the sharpness of the transition from the weak to the strong inversion value of ϕ . Setting

$$\kappa = 2 \left(1 + \frac{\gamma}{2\sqrt{2\phi_F + V_{sb}}} \right) \quad (4.8)$$

has been found to provide satisfactory results in lieu of using κ as a fitting parameter [17].

Several short-channel effects, such as drain induced barrier lowering (DIBL), narrow width effects, and non-uniform doping effects can further complicate the expression for the threshold voltage. In addition, evaluation of the derivatives of (4.3) are significantly complicated by the fact that (4.7) is weakly dependent on V_{gs} [17]. However, reasonable accuracy between the equations and measurements can be achieved if ϕ is used as a fitting parameter instead [42].

In weak inversion, both exponential terms in (4.1) are small, and, using the approximation $\log(1 + x) \approx x$ for $x \ll 1$, equation (4.1) reduces to

$$I_{ds} \approx I_{d0} (1 - e^{-V_{ds}/\phi_t}) e^{V_{od}/(n\phi_t)} \quad (4.9)$$

Equation (4.9) has the form of the well known weak inversion model as given in [40], except for a difference in the constant of proportionality I_{d0} . The disparity between (4.2) can cause an error in I_{ds} , however, V_{th} can be adjusted slightly (since it appears in the exponential) to bring I_{ds} to the correct value [40].

In strong inversion, the first exponential in (4.1) is much greater than 1. The second exponential is also much greater than 1 when operating in the linear region, since V_{ds} is small. Using the approximation $[\log(1 + e^y)]^2 \approx [\log(e^y)]^2 = y^2$, equation (4.1) reduces to

$$I_{ds} = \mu_0 C'_{ox} \frac{W}{L} \left[V_{od} V_{ds} - \frac{1}{2} n V_{ds}^2 \right] \quad (4.10)$$

for the linear region. As V_{ds} increases, the transistor enters the saturation region and the second exponential in (4.1) becomes negligible. Equation (4.1) then reduces to the simpler form

$$I_{ds} = I_{d0} \left[\ln \left(1 + e^{V_{od}/(2n\phi_t)} \right) \right]^2 \quad (4.11)$$

which is independent of the region of inversion. In strong inversion (4.11) further simplifies to

$$I_{ds} = \mu_0 C'_{ox} \frac{W}{L} \frac{V_{od}^2}{2n} \quad (4.12)$$

Both equations (4.10) and (4.12) are the well know quadratic models and form the basis of the SPICE level 3 model.

Unfortunately, the expression given by 4.1 is not sufficient for a device with a short gate length and several other effects must be included to obtain reasonable accuracy

4.4.2 Short Channel Effects

Many short channel effects must be accounted for in a model for a modern deep sub-micron MOSFET. Examples are mobility degradation, velocity saturation, channel length modulation, drain induced barrier lowering, substrate current body effect, gate induced drain leakage currents and gate tunnelling currents, to name a few. Of these, mobility degradation, velocity saturation and channel length modulation are among the most important short channel effects [32], [39]. However, it should be pointed out that the other effects can become more dominant as the MOS transistor continues to scale down to smaller sizes.

Mobility Degradation

Carrier surface mobility is a key parameter in MOSFET modeling [32]. It is a measure of the ease of the carrier motion in the channel, which takes place mainly along the surface interface between silicon and the gate oxide. Carrier mobility depends on both the transverse and longitudinal electric fields and several scattering mechanisms such as

1. electron-phonon scattering caused by the crystal lattice vibrations [40];
2. coulomb scattering caused by ionized impurity ions [40]; and,
3. surface roughness scattering [43].

In older CMOS technologies, the longitudinal fields dominated and the transverse field effects could be ignored, however, in today's short channel devices, the gate oxide is much thinner, such that the transverse fields can no longer be ignored. A typical behaviour of the surface mobility versus the transverse electric field is depicted in Figure 4.5 along with the limiting scattering effects. This figure, however, is only representative, as other effects, such as temperature and doping concentration will cause a shift in the curves. In general, mobility depends on a number of factors such as doping concentration, temperature, gate oxide thickness, threshold voltage, gate voltage, etc.

Figure 4.5: Surface mobility vs. transverse field (from [32]).

A widely used empirical model for the effective mobility is

$$\mu_{eff} = \frac{\mu_0}{1 + \theta V_{od} + \theta_B V_{sb}} \quad (4.13)$$

where μ_0 is the low-field surface mobility, θ and θ_B are fitting parameters for mobility degradation coefficient [44] [40]. The value for θ_B is typically small ($0.005V^{-1}$) and is often neglected [43]. However, this can cause problems, since an increasing V_{sb} would imply an increasing μ_{eff} due to body effect, yet μ_{eff} should decrease with increasing V_{sb} [40].

Velocity Saturation

Velocity saturation is another important parameter that significantly affects the characteristics of a short channel MOSFET. As device dimensions continue to decrease, longitudinal and transverse electric fields become more prominent at lower voltages and can significantly affect the transistor's performance. In a long channel device, the current saturates when the channel is pinched-off at the drain end. However, in a short channel transistor, the current saturates when the carrier reaches velocity saturation before pinch off. At low longitudinal electric fields, the carrier drift velocity, v , is linearly dependent on the electric field. However, as the electric field intensity, $\mathcal{E}_x$, increases, the carriers reach a maximum velocity, v_{sat} , and the device is in velocity saturation, as illustrated in Figure 4.6, which shows typical values [40].

This behaviour can be easily modeled in a piece-wise manner as

$$v = \mu \mathcal{E}_x, \quad \mathcal{E}_x < \mathcal{E}_{sat} \quad (4.14)$$

$$v = v_{sat}, \quad \mathcal{E}_x > \mathcal{E}_{sat} \quad (4.15)$$

A continuous velocity model is given by

$$v = \frac{\mu_{eff} \mathcal{E}_x}{1 + (\mathcal{E}_x / \mathcal{E}_{sat})} \quad (4.16)$$

where $\mathcal{E}_{sat} = 2v_{sat}/\mu_{eff}$ is the critical electric field at which the carrier velocity become saturated.

Figure 4.6: Typical measured electron drift velocity versus longitudinal electric field.

It has been shown that in strong inversion, the drain current in the linear region before the carrier velocity saturates is identical to the linear drain current without velocity saturation except for a multiplicative factor $1/[1 + V_{ds}/(L\mathcal{E}_{sat})]$ [40], [32] such that

$$I_{ds(linear)} = \mu_{eff} C'_{ox} \frac{W}{L} \frac{(V_{od} - 0.5nV_{ds})V_{ds}}{1 + V_{ds}/(L\mathcal{E}_{sat})}, \quad V_{ds} < V_{DS(sat)} \quad (4.17)$$

In strong inversion, $V_{DS(sat)}$ is given by [32]

$$V_{DS(sat)} = \frac{\mathcal{E}_{sat} L V_{od}}{A_{bulk} \mathcal{E}_{sat} L + V_{od}} \quad (4.18)$$

where $A_{bulk} \approx n$ is the coefficient accounting for the bulk charge effect. The bulk charge effect is closely related to the body effect and refers to the changing threshold voltage along the channel when $V_{ds} > 0$ [32]. The saturation drain current is obtained by replacing V_{ds} in (4.17) by $V_{DS(sat)}$ to yield

$$I_{ds(sat)} = \frac{\mu_{eff} C'_{ox} W}{2n L} \frac{V_{od}^2}{1 + V_{od}/(A_{bulk} L \mathcal{E}_{sat})} \quad V_{ds} > V_{DS(sat)} \quad (4.19)$$

Velocity saturation effects can have a significant impact on the I-V characteristics. Most importantly, the saturation current becomes more linearly dependent on

the over-drive voltage, especially for small gate lengths. For small lengths, (4.19) can be approximated by

$$I_{ds(sat)} \approx \frac{\mu_{eff} C'_{ox}}{2n} \frac{W}{L} \frac{V_{od}^2}{V_{od}/nL\mathcal{E}_{sat}} \quad (4.20)$$

$$I_{ds(sat)} \approx v_{sat} C'_{ox} W V_{od} \quad (4.21)$$

The linear dependence of I_{ds} on V_{od} is apparent in (4.21) where the square law relationship has disappeared. In addition, (4.21) reveals that the channel current is independent of the gate length, a second major effect of channel velocity saturation [40].

Velocity Saturation and Mobility Degradation Combined

Often the correction factors for mobility reduction and velocity saturation are merged together since they are both of the form $1/(1+x)$ [7]. The products can be represented as

$$\frac{1}{(1+x)} \cdot \frac{1}{(1+y)} \quad (4.22)$$

If both x and y are much smaller than one, then

$$\frac{1}{(1+x)} \cdot \frac{1}{(1+y)} = \frac{1}{1+x+y+xy} \approx \frac{1}{1+x+y} \quad (4.23)$$

The corresponding drain current is then given by

$$I_{ds(sat)} = \frac{\mu_0 C'_{ox}}{2n} \frac{W}{L} \frac{V_{od}^2}{1 + \left(\theta + \frac{\mu_0}{nL}\right) V_{od}}, \quad V_{ds} > V_{DS(sat)} \quad (4.24)$$

It should be noted that the form used in [11] and [45], repeated here

$$I = K \frac{(V_{GS} - V_T)^2}{1 + \theta(V_{GS} - V_T)} \quad (4.25)$$

represents the merging of the mobility reduction and velocity saturation effects and that the parameter θ in (4.25) is equivalent to $(\theta + \mu_0/nL)$ in (4.24) and (4.13).

Channel Length Modulation

The output conductance of the MOSFET in the saturation region is predominantly due to channel length modulation [46]. However, the output conductance is traditionally not well modelled and suffers from discontinuities in most models [41]. This primarily occurs because channel length modulation occurs in the saturation region but not in the linear region. Traditionally, a correction factor is applied to the saturation current equation only, resulting in a discontinuity in the output conductance when transitioning from the linear to the saturation region. BSIM3, BSIM4 and EKV models use a smoothing function to avoid the discontinuity [32], [37], [39].

Channel length modulation occurs when $V_{ds} > V_{DS(sat)}$. At $V_{ds} = V_{DS(sat)}$, the channel is said to be at pinch off, as depicted in Figure 4.7(a). If V_{ds} is increased further, the effective length of the channel decreases by the length of the depletion region, l_p , (shown in Figure 4.7(b)).

Figure 4.7: Channel at (a) pinch off; (b) above pinch-off.

Since I_{ds} is proportional to $1/L$, the drain current, including channel length modulation, has the form

$$I_{ds} = \frac{I_{ds(sat)}}{1 - l_p/L} \quad (4.26)$$

where $I_{ds(sat)}$ is the saturation drain current at $V_{ds} = V_{DS(sat)}$. The pinch off length, l_p , can be predicted by

$$l_p = \sqrt{\frac{2\epsilon_{si}}{qN_A}} \left[\sqrt{\phi_D + (V_{ds} - V_{DS(sat)})} - \sqrt{\phi_D} \right] \quad (4.27)$$

where

$$\phi_D = \frac{\epsilon_{si} E_p^2}{2qN_A} \quad (4.28)$$

E_p is the lateral field intensity at the beginning of the depletion region, ϵ_{si} is the permittivity of silicon, q is the electron charge magnitude. In practice, values of ϕ_D are chosen empirically for matching the model with experimental data [40].

A simpler model is often used for computation speed and is analogous to the Early voltage, V_A , in bipolar transistors where

$$I_{ds} = I_{ds(sat)} \left(1 + \frac{V_{ds} - V_{DS(sat)}}{V_A} \right) \quad (4.29)$$

The plot of this equation, shown in Figure 4.8, intercepts the horizontal axis at $-V_A$. The saturation curves for various V_{gs} are assumed to be straight lines that all intercept at $-V_A$. In reality, V_A is dependent on V_{gs} and the assumption that all curves intercept at $-V_A$ is not valid. In addition, equation (4.29) predicts a nonzero slope at $V_{DS(sat)}$ whereas the non-saturation equation predicts a zero slope at the same point. As a result, equation (4.29) is only good for approximation purposes [40]. Furthermore DIBL, and substrate current induced body effect (SCBE) also influence V_A [32], [39].

Figure 4.8: First-order modeling of the saturation region analogous to Early voltage.

BSIM3 models the total Early voltage (without consideration of SCBE) as

$$V_A = V_{Asat} + \left(\frac{1}{V_{ACLM}} + \frac{1}{V_{ADIBL}} \right)^{-1} \quad (4.30)$$

where V_{Asat} is the Early voltage at $V_{ds} = V_{DS(sat)}$, thereby providing a continuous drain current and output conductance at the transition point from linear to saturation region [32]. V_{ACLM} is the Early voltage due to channel length modulation and V_{ADIBL} is the Early voltage due to drain induced barrier lowering. All three parameters, V_{Asat} , V_{ACLM} and V_{ADIBL} , are dependent on V_{gs} and V_{ds} (V_{ACLM} only) and have complex

expressions that are not repeated here. Details of the expressions can be found in [32]. In addition to their dependence on V_{gs} , the parameter P_{VAG} was introduced to (4.30) to improve the accuracy of the gate bias dependence of V_A [32]. The modified model is given by

$$V_A = V_{Asat} + \left(1 + \frac{P_{VAG}X}{\mathcal{E}_{sat}L}\right) \left(\frac{1}{V_{ACLM}} + \frac{1}{V_{ADIBL}}\right)^{-1} \quad (4.31)$$

4.4.3 Source-Drain Resistance

A parasitic source and drain resistance, in series with the MOSFET channel, can have a significant impact on the performance of the MOSFET. As MOSFET devices scale down, the parasitic resistances become increasingly important [7]. As shown in Figure 4.9, the resistances consist of three parts:

1. the contact resistance between the $n+$ region and metal ($R_{contact}$);
2. the resistance of the main body of the $n+$ region (R_{shunt}); and,
3. the resistance associated with crowding of the current from the $n+$ region to the inversion layer (R_{spread} , often referred to as *spreading resistance*).

Figure 4.9: (a) Components of the series resistance in a MOSFET, and (b) equivalent lumped model.

Lightly doped drain (LDD) structures, such as depicted in Figure 4.10, involve an additional series resistance that further impacts the MOSFET's performance. However, the resistance of the LDD structures depends on the terminal voltages and is nonlinear. For simplicity's sake, the resistance will be assumed to be linear.

Figure 4.10: A lightly doped drain structure.

The source and drain resistors are in series with the channel, shown in Figure 4.11 and reduce the effective V_{ds} of the *intrinsic* model by the voltage drop across the series device:

$$V'_{ds} = V_{ds} - 2RI_{ds} \quad (4.32)$$

Figure 4.11: A MOS transistor including parasitic source and drain resistances.

In many cases RI_{ds} is much smaller than $V_{gs} - V_{th}$, and, if V_{ds} is also much smaller than $V_{gs} - V_{th}$, then the approximate expression for the linear region in strong inversion can be approximated as

$$I_{ds} \approx \mu_{eff} C'_{ox} \frac{W}{L} V_{od} V'_{ds} \quad (4.33)$$

Combining (4.32) and (4.33) yields

$$I_{ds} = \frac{W}{L} \frac{\mu_{eff} C'_{ox}}{1 + \beta_R V_{od}} V_{od} V_{ds} \quad (4.34)$$

with

$$\beta_R = \frac{2\mu_{eff}C'_{ox}RW}{L} \quad (4.35)$$

The reduction to the drain current has the same form as the simplified model of mobility reduction $\mu_{eff} = \mu_0/(1 + \theta(V_{gs} - V_{th}))$. If both effects are small, then they can be combined, as previously discussed by (4.23), to yield

$$I_{ds(linear)} = \mu_0 C'_{ox} \frac{W}{L} \frac{V_{od}}{1 + (\theta + \beta_R) V_{od}} V_{ds}, \quad V_{ds} < V_{DS(sat)} \quad (4.36)$$

$$I_{ds(sat)} = \frac{\mu_0 C'_{ox}}{2n} \frac{W}{L} \frac{V_{od}^2}{1 + (\theta + \beta_R) V_{od}}, \quad V_{ds} > V_{DS(sat)} \quad (4.37)$$

Note though, that the above expressions have sometimes led to the erroneous usage of the term “mobility degradation due to series resistance”. As pointed out in [40] and [7] such a term does not properly describe what happens because the two effects are totally separate.

A similar analysis can be carried out without making the assumptions of the relative magnitude of the terminal voltages. Even in such cases, though, it is often sufficient to model the series resistance by adding β_R to θ [40].

4.4.4 Final Model for Saturation Region in Moderate Inversion

The saturation region is of particular importance for RF and analog design since most circuits operate in this region. As such, the final model will concentrate only on the saturation region over all regions of inversion. This also helps avoid the difficulties of encountering discontinuities while transitioning from the linear region to the saturation or using a smoothing function that would significantly complicate the analysis. However, the effects of velocity saturation, mobility degradation, series drain-source resistance, and channel length modulation need to be included for reasonable accuracy. Hence, the following expression is used

$$I_{ds} = \frac{\mu_0 C'_{ox}}{2n} \frac{W_{eff}}{L_{eff}} \frac{X^2}{1 + \alpha X} \left(1 + \frac{V_{ds} - V_{DS(sat)}}{V_A} \right) \quad (4.38)$$

with

$$X = 2n\phi_t \ln(1 + e^{(V_{gs}-V_{th})/(2n\phi_t)}) \quad (4.39)$$

$$V_{DS(sat)} = \frac{\mathcal{E}_{sat} L_{eff} X}{A_{bulk} \mathcal{E}_{sat} L_{eff} + X} \quad (4.40)$$

$$V_A = V_{Asat} + \left(1 + \frac{P_{VAG} X}{\mathcal{E}_{sat} L_{eff}}\right) V_{ACLM} \quad (4.41)$$

where L_{eff} and W_{eff} are the effective gate length and width respectively. The parameters n , V_{th} , α , V_{Asat} , V_{ACLM} , and P_{VAG} will be used as parameters to fit measured I-V data to the model. The parameter, α , will be assumed to model the combined mobility degradation, series drain-source resistance, and velocity saturation effects with

$$\alpha = \theta + \beta_R + \frac{\mu_0}{2nv_{sat}L_{eff}} \quad (4.42)$$

For modern deep sub-micron devices, α typically has values of 2 to 3 V^{-1} or higher, depending on channel-length and velocity saturation effects. V_{Asat} will be assumed to be a constant as will V_{ACLM} . It is further assumed that V_{ACLM} will also account for DIBL effects.

The model given by (4.38) and (4.39) is based on the single-piece expression given in [17] and is valid and continuous over all regions of inversion in the saturation region. The model is also of the same form given in [11] but includes channel length modulation.

Comparison of Measurements and Analytical solution

The model given by (4.38) and (4.39) provides adequate accuracy in the moderate inversion region, as can be seen in Figure 4.12, which compares the measured and calculated I_{DS} versus V_{GS} with $V_{DS} = 1V$ and $V_{sb} = 0V$. The parameters V_{th} , n , α , V_{Asat} , V_{ACLM} and P_{VAG} were used as fitting parameters to obtain an accurate fit between (4.38) and the measured data. A least squares fitting function was used to obtain the fit. The parameters L_{eff} , W_{eff} , C'_{ox} , μ_0 , $\mathcal{E}_{sat}$, and A_{bulk} were calculated from foundry supplied BSIM3v3 model files. The parameters used are listed in Table 4.1. Good agreement is obtained with the measured, calculated, and simulated first, second, and third derivatives with respect to V_{GS} , shown in Figure 4.13. The

simulation data was obtained from foundry supplied BSIM3v3 models which provide a continuous third-order derivative curve. The measured and simulated derivatives were actually calculated by numerically differentiating the drain current curve. Due to the measurement precision, the third derivative of the measured data was not available and so only the simulated and calculated data are shown. The plots show good agreement for the derivatives with only a slight discrepancy between the measured, simulated and calculated derivatives that becomes more noticeable by the third derivative. None-the-less, the accuracy is considered sufficient as previously discussed in Section 4.3. More accurate models for the threshold voltage, channel-length modulation, mobility degradation and channel velocity saturation effects can be used to obtain even better accuracy, however, the resulting equations become unwieldy for obtaining a basic understanding of the effects of nonlinearity in a CMOS transistor.

Figure 4.12: Comparison of measured and calculated I_{DS} for a typical $0.18\mu\text{m}$ transistor using (4.38) with $V_{DS} = 1\text{V}$, $V_{SB} = 0\text{V}$ and with the parameters listed in Table 4.1.

Good agreement is also obtained in the saturation region with respect to V_{DS} . Figures 4.14(a) and 4.14(b) compare the drain current versus V_{DS} in the saturation

Table 4.1: Parameters used for fitting (4.38) to measured data and calculated parameters from foundry supplied BSIM3v3 model files for a 0.18 μm process.

Parameter	Value	Parameter	Value
V_{th}	0.485 V	V_{Asat}	0.39 V
α	2.199 V^{-1}	V_{ACLM}	0.34 V
n	1.32	P_{VAG}	21
L_{eff}	100nm	W_{eff}	$127.47 \mu\text{m}$
μ_0	$0.0385 \text{ m}^2/(\text{V}\cdot\text{s})$	C'_{ox}	$0.0095 \text{ fF}/\mu\text{m}^2$
$\mathcal{E}_{sat}$	$6.399 \cdot 10^6 \text{ V/m}$	A_{bulk}	1.143
V_{SB}	0 V

Figure 4.13: Comparison of measured, simulated, and calculated $\log I_{DS}$ and derivatives versus V_{GS} for a typical 0.18 μm transistor using (4.38) with the parameters listed in Table 4.1.

region only for values of V_{GS} that span weak, moderate and strong inversion. Comparison for the linear region is not shown since (4.38) is valid only for the saturation region.

Figure 4.14: Comparison of measured and analytical solution of I_{DS} versus V_{DS} using (4.38) with $V_{GS} = 0, 4, 0.5, 0.6$ and 0.7 V, $V_{SB}=0V$ and with the parameters listed in Table 4.1 on (a) linear scale and (b) on logarithmic scale.

The Functions *long*, *mobhot*, and *clm*

Equation (4.38) is a product of three factors and is rewritten in a similar form as in [7] where

$$I_{ds} = long(V_{gs}, V_{sb}) \cdot mobhot(V_{gs}, V_{sb}) \cdot clm(V_{gs}, V_{ds}, V_{sb}) \quad (4.43)$$

The functions *long*, for long channel model, *mobhot*, for mobility degradation and hot carrier effects, and *clm*, for channel length modulation, will be used later in Section 4.6 to simplify the evaluation of the nonlinear coefficients. They also allow for the evaluation of the relative strength of each function on the nonlinear coefficients.

The first factor

$$long(V_{gs}, V_{sb}) = \frac{\mu_0 C'_{ox} W}{2n} X^2 \quad (4.44)$$

models the drain current of a long channel device without taking into account mobility reduction, velocity saturation, and channel length modulation effects. As can be seen in Figure 4.15, *long* is a strong function of V_{GS} . It is also a function of V_{SB} , since n and V_{th} are both dependent on V_{SB} . Although *long* is independent of V_{DS} in the saturation region, as currently modeled, it is, in reality, weakly dependent on V_{DS} . The threshold voltage can be affected by drain-induced barrier lowering which is dependent on V_{DS} . The effect on the nonlinear coefficients is small, though, and will not be considered here.

Figure 4.15: Comparison of *long* versus V_{GS} for $V_{DS} = -0.3, 0$ and $+0.3V$ using the parameters listed in Table 4.1.

The second term of (4.43)

$$mobhot(V_{gs}, V_{sb}) = \frac{1}{1 + \alpha X} \quad (4.45)$$

models the combined effect of mobility reduction, velocity saturation (or hot carrier effects), and drain source resistance. From Figure 4.16, it can be seen that $mobhot$ is a strong function of V_{GS} , depending on the value of α . For $\alpha = 0$ there are no short channel effects and $mobhot$ is a constant versus V_{GS} . For larger values of α , $mobhot$ is constant at unity in the weak inversion region but then decreases for increasing V_{GS} in the strong inversion region. The rate of the decrease increases with increasing α . Since $mobhot$ decreases with increasing V_{GS} , the drain current will be reduced from the case with no short channel effects. This linearizes I_{DS} as discussed in Section 4.4.2. As in the case with *long*, $mobhot$ is also a function of V_{SB} since n and V_{th} are dependent on V_{SB} . Figure 4.17 shows $mobhot$ versus V_{GS} for different V_{SB} with $\alpha = 2\text{ V}^{-1}$. As V_{SB} increases, the threshold voltage also increases and so the point at which $mobhot$ starts to decrease shifts to higher bias voltage. However, the rate of change of $mobhot$ is approximately the same. $mobhot$ is also weakly dependent on V_{DS} in the saturation region due to V_{th} 's slight dependence on V_{DS} but is not considered here.

Figure 4.16: Comparison of $mobhot$ versus V_{GS} for $\alpha = 0$ (no short channel effects), 0.5, 1, and 2 V^{-1} using the parameters listed in Table 4.1.

Figure 4.17: Comparison of $mobhot$ versus V_{GS} for $V_{SB} = -0.3, 0$, and 0.3 V with $\alpha = 2.0V^{-1}$ and using the parameters listed in Table 4.1.

The third term of (4.43)

$$clm(V_{gs}, V_{ds}, V_{sb}) = 1 + \frac{V_{ds} - V_{DS(sat)}}{V_A} \quad (4.46)$$

models the effects of channel length modulation. From (4.46), clm is a linear function of V_{DS} as currently modeled. The early voltage, V_A , given by (4.41), and the saturation voltage, $V_{DS(sat)}$, given by (4.40), are both dependent on V_{GS} and V_{SB} , however, the dependency of V_A on V_{GS} also depends on P_{VAG} . Figure 4.18 shows V_A versus V_{GS} for different values of V_{Asat} , V_{ACLM} , and P_{VAG} . For $P_{VAG} = 0$, $V_A = V_{Asat} + V_{ACLM}$ and is constant versus V_{GS} but for $P_{VAG} > 0$, V_A and the rate of change of V_A with V_{GS} increases with increasing P_{VAG} . V_A also increases with increasing V_{Asat} and V_{ACLM} . However, since clm is inversely proportional to V_A , clm decreases with increasing values of P_{VAG} , V_{Asat} , and V_{ACLM} , as shown in Figure 4.19. More importantly though, the rate of change of clm with V_{GS} increases with increasing P_{VAG} but reduces with increasing V_{Asat} and V_{ACLM} . Generally, clm is only weakly dependent on V_{GS} for $V_A > 2$ to $3V$. This, however, is not likely going to be the case for *minimum length*

modern sub-micron devices where channel length modulation becomes increasingly more dominant.

Figure 4.18: Comparison of V_A versus V_{GS} for different V_{Asat} , V_{ACLM} , and P_{VAG} with $V_{DS} = 600\text{mV}$ and using the parameters listed in Table 4.1.

4.5 Capacitors in a MOSFET

In this section the capacitors in a MOSFET for all regions of inversion in the saturation region are briefly discussed. The modelling of the MOSFET capacitors is a complex subject in itself and has been addressed in detail in several studies. A good overview with extensive references can be found in [40]. The capacitors that are considered here are shown in Figure 4.4. Each capacitor is comprised of two components: an extrinsic capacitor and an intrinsic capacitor.

4.5.1 Extrinsic Capacitors

Extrinsic capacitors arise from two main sources. One source is the oxide capacitances from the gate to the source and drain, $C_{gs,e}$ and $C_{gd,e}$ respectively. The subscript e

Figure 4.19: Comparison of cIm versus V_{GS} for different V_{Asat} and V_{ACLM} with $V_{DS} = 600\text{mV}$ and using the parameters listed in Table 4.1.

is used to denote the extrinsic capacitance. Each oxide capacitor consists of three components, the overlap capacitance, C_{ov} , the fringing capacitance, C_{fr} , and the top plate capacitance, C_{top} . All three capacitances are proportional to W and can be written in the form

$$C_{gs,e} = C_{gd,e} = W_{eff} C_o'' \quad (4.47)$$

where C_o'' can be approximated by

$$C_o'' \approx L_{ov} C_{ox}' = \frac{2}{\pi} \epsilon_{ox} \ln \left(1 + \frac{t_{poly}}{t_{ox}} \right) + C_{top}'' \quad (4.48)$$

L_{ov} is the overlap distance, t_{poly} is the thickness of the poly gate and C_{top}'' is about 10 percent of the total of C_o'' [40]. These capacitances are considered linear.

The second class of extrinsic capacitors consist of the junction capacitance between the n^+ region of the source and drain and the bulk and form part of C_{sb} and C_{db} . The junction capacitance per unit area, C_j , is nonlinear and can be approximated

by

$$C_j(V_R) = \frac{C_{j0}}{\left(1 + \frac{V_R}{V_j}\right)^{m_j}} \quad (4.49)$$

where m_j is the junction grading coefficient, V_R is the reverse bias voltage at the source (or drain), and C_{j0} is the junction capacitance at zero bias and is given by

$$C_{j0} = \frac{q\epsilon_{si}N_n}{\sqrt{2V_j}} \quad (4.50)$$

Here V_j is the junction potential and depends on the doping concentration of the n^+ region, N_n , and the p -type substrate, N_p , where

$$V_j = \frac{kT}{q} \ln \frac{N_n N_p}{n_i^2} \quad (4.51)$$

Hence the extrinsic source and drain to bulk capacitors are given by

$$C_{sb,e} = A_s C_j(V_{sb}) \quad (4.52)$$

$$C_{db,e} = A_s C_j(V_{db}) \quad (4.53)$$

in which A_s is the area of the n^+ region.

4.5.2 Intrinsic Capacitances

The intrinsic capacitors of a MOSFET are always a fraction of the total intrinsic oxide capacitance $C_{ox} = W_{eff}L_{eff}C'_{ox}$ [7] and are bias dependent. As a result they are nonlinear. Most often, though, they are assumed to be linear, especially when operating in the saturation region in strong inversion. This is certainly not the case in moderate inversion which sometimes has been neglected in developing expressions for the various capacitances in a MOSFET [40]. In such cases, expressions for the weak and strong inversion region are considered valid in adjacent regions but the expressions can result in large error in predicting capacitances in the moderate inversion region [40]. General expressions, valid for all regions of inversion, have been developed for a simplified charge-sheet model and are used here [37], [40], [47]. Quasi-static operation

of the MOSFET is assumed. The expressions are given by

$$C_{gs,i} = WLC'_{ox} \left(\frac{3}{2} + \frac{4n^2\phi_t^2 f(X)}{X^2} \right)^{-1} \quad (4.54)$$

$$C_{gb,i} = WLC'_{ox} \frac{n-1}{n} \left[1 - \left(\frac{3}{2} + \frac{4n^2\phi_t^2 f(X)}{X^2} \right)^{-1} \right] \quad (4.55)$$

$$C_{sb,i} = (n-1)C_{gs} \quad (4.56)$$

$$C_{gd,i} = 0 \quad (4.57)$$

$$C_{db,i} = 0 \quad (4.58)$$

where

$$f(X) = \frac{1}{2} \left(\sqrt{1 + \frac{X^2}{n^2\phi_t^2}} + 1 \right) \quad (4.59)$$

The subscript i is used to denote the intrinsic capacitance.

Figure 4.20 plots $C_{gs,i}$, $C_{gb,i}$, and $C_{sb,i}$ versus V_{GS} when biased in the saturation region. The plots show that $C_{gs,i}$ is the dominant capacitor and is nonlinear in the moderate inversion region. For higher gate voltages, or strong inversion, $C_{gs,i}$ approaches a constant value of approximately $\frac{2}{3}C_{ox}$. Similarly, $C_{sb,i}$ approaches a constant value in strong inversion of $\frac{2}{3}(n-1)C_{ox}$. On the other hand, $C_{gb,i}$ starts at a constant value in the weak inversion and approaches $\frac{n-1}{3n}C_{ox}$ in the strong inversion region. The capacitors $C_{gd,i}$ and $C_{db,i}$ are modelled as zero since the drain is pinched off and any changes in bias levels at the drain do not affect its charge [40].

4.6 Evaluation of Nonlinear Coefficients

This section derives and evaluates the nonlinear coefficients for the drain current expression, given by (4.38), using the three-dimensional power series expansion as defined by (2.57). The coefficients will be used later in Chapter 5 in conjunction with the Volterra series analysis of various circuit configurations. In Chapter 5, it will be shown that the second and third-order harmonic products, as well as the third-order intermodulation product, strongly depend on several nonlinear coefficients. The expressions for the nonlinear coefficients model different effects, namely:

Figure 4.20: Intrinsic capacitances $C_{gs,i}$, $C_{gb,i}$, and $C_{sb,i}$ versus V_{GS} with $V_{DS} = 600\text{mV}$, $V_{SB} = 0\text{V}$ and using the parameter listed in Table 4.1.

1. the nonlinear relationship between the drain current and the terminal voltages;
2. the transition from weak inversion through moderate inversion to strong inversion;
3. mobility reduction, velocity saturation, and drain-source resistance; and,
4. channel length modulation.

Some nonlinear coefficients are primarily determined by one of the above effects, whereas, other coefficients are influenced by more than one effect. This section will determine the dominant terms of the expressions for the nonlinear coefficients but will primarily focus on the nonlinear coefficients with respect to V_{GS} ($c_{(1,0,0)}$, $c_{(2,0,0)}$, and $c_{(3,0,0)}$) and to a lesser extent, the coefficients with respect to V_{DS} ($c_{(0,1,0)}$, $c_{(0,2,0)}$, and $c_{(0,3,0)}$) and V_{SB} ($c_{(0,0,1)}$, $c_{(0,0,2)}$, and $c_{(0,0,3)}$). As will be shown later, the coefficients with respect to V_{GS} dominate, especially for third-order nonlinearities, $HD3$ and $IIP3$. Unfortunately, the expressions for the nonlinear coefficient can be very lengthy, making evaluation difficult and so the full expressions are expressed in terms

of the partial derivatives of *long*, *mobhot*, and *clm*. This also reveals the relative influence of the various effects of channel length modulation and mobility degradation on the nonlinear behaviour of the MOSFET. The full expressions for the various partial derivatives can be found in Appendix C.

4.6.1 Expressions for $c_{(1,0,0)}$, $c_{(2,0,0)}$, and $c_{(3,0,0)}$

Evaluation of $c_{(1,0,0)}$ (g_m)

The first-order derivative of the drain current with respect to V_{GS} is $c_{(1,0,0)}$ and is also commonly referred to as the transconductance or g_m . It is given by

$$c_{(1,0,0)} = \frac{\partial I_{DS}}{\partial V_{GS}} \quad (4.60)$$

Differentiating (4.43) yields

$$\begin{aligned} c_{(1,0,0)} = & \frac{\partial long}{\partial V_{GS}} \cdot mobhot \cdot clm + \frac{\partial mobhot}{\partial V_{GS}} \cdot long \cdot clm \\ & + \frac{\partial clm}{\partial V_{GS}} \cdot long \cdot mobhot \end{aligned} \quad (4.61)$$

An evaluation of the three terms in (4.61) versus V_{GS} for parameters listed in Table 4.1 is plotted in Figure 4.21 and shows that the first two terms in (4.61) are the dominant terms and $c_{(1,0,0)}$ can be approximated by

$$c_{(1,0,0)} \approx \frac{\partial long}{\partial V_{GS}} \cdot mobhot \cdot clm + \frac{\partial mobhot}{\partial V_{GS}} \cdot long \cdot clm \quad (4.62)$$

It is seen that g_m is primarily determined by the dependence of the function *long* on V_{GS} . The dependence of *mobhot* on V_{GS} is also of some importance, especially at higher gate bias levels, which tend to limit g_m . The third term, however, does not play a significant role since the rate of change of *clm* with V_{GS} is significantly smaller than $\partial long / \partial V_{GS}$ and $\partial mobhot / \partial V_{GS}$. The relative error for including only the first two terms is less than 10%.

The nonlinear coefficient $c_{(1,0,0)}$ is obviously strongly dependent on the bias voltages V_{GS} as well as V_{SB} . Figure 4.22(a) plots (4.61) versus V_{GS} for different V_{SB} . Since the threshold voltage is dependent on V_{SB} , the source-bulk bias voltage affects

Figure 4.21: An evaluation of the three terms in (4.61) versus V_{GS} with $V_{DS} = 600\text{mV}$ and $V_{SB} = 0$ for parameters listed in Table 4.1 .

the gate voltage at which I_{DS} and g_m start to increase. The drain-source voltage, however, has very little effect on $c_{(1,0,0)}$, as seen in Figure 4.22(b). This is only true, though, in the saturation region where I_{DS} is relatively constant with respect to V_{DS} for a given V_{GS} . In the linear region, I_{DS} varies significantly with V_{DS} but is not considered here.

The ratio L_{eff}/W_{eff} acts as a scaling factor for I_{DS} . This is also true for g_m since L_{eff} and W_{eff} are constants. Figure 4.23 plots $c_{(1,0,0)}$ for different L_{eff}/W_{eff} ratios and it can be seen that g_m scales with L_{eff}/W_{eff} as expected. It should be noted though that W_{eff} does have some bias dependencies that become more dominant for narrow channel widths. However, typically, very wide channel widths are used for RF applications and the bias dependencies of W_{eff} have negligible effect.

The parameters V_{th} , α , and, to a lesser extent, n play a significant role in (4.61). Both $long$ and $\partial long/\partial V_{GS}$ are strong functions of V_{th} which sets the gate voltage at which the drain current and its derivative starts to increase sharply. This is evident in Figure 4.24 which shows $c_{(1,0,0)}$ versus V_{GS} for different values of V_{th} . $long$ and $\partial long/\partial V_{GS}$ are also dependent on n . Figure 4.25 plots $c_{(1,0,0)}$ versus V_{GS} for

(a)

(b)

Figure 4.22: $c_{(1,0,0)}$ versus V_{GS} using the parameters listed in Table 4.1 for (a) $V_{SB} = -300, 0$, and $+300$ mV with $V_{DS} = 600$ mV and (b) $V_{DS} = 400, 600$, and 800 mV with $V_{SB} = 0$.

different values of n and shows $c_{(1,0,0)}$ decreasing with increasing n . The parameter, α , is also significant. Since mob_{hot} decreases with increasing V_{GS} , $\partial mob_{hot} / \partial V_{GS}$ is negative and, hence, the second term is negative as well. This limits $c_{(1,0,0)}$ as evident in Figure 4.26. An alternative view is that mob_{hot} tends to linearize the drain current for large α and, as a result, the rate of change of I_{DS} , or g_m , will reach a maximum value and flatten out, as shown in Figure 4.26.

Figure 4.23: $c_{(1,0,0)}$ versus V_{GS} for different L_{eff}/W_{eff} with $V_{DS} = 600\text{mV}$, $V_{SB} = 0$ and using the parameters listed in Table 4.1.

Figure 4.24: $c_{(1,0,0)}$ versus V_{GS} for different values of V_{th} with $V_{DS} = 600\text{mV}$ and $V_{SB} = 0$ using the parameters listed in Table 4.1.

Figure 4.25: $c_{(1,0,0)}$ versus V_{GS} for different values of n with $V_{DS} = 600\text{mV}$ and $V_{SB} = 0$ using the parameters listed in Table 4.1.

Figure 4.26: $c_{(1,0,0)}$ versus V_{GS} for different values of α with $V_{DS} = 600\text{mV}$ and $V_{SB} = 0$ using the parameters listed in Table 4.1.

Evaluation of $c_{(2,0,0)}$

The second derivative of the drain current with respect to V_{GS} is $c_{(2,0,0)}$ and is the dominant factor in second-order distortion effects. $c_{(2,0,0)}$ is given by

$$c_{(2,0,0)} = \frac{1}{2} \frac{\partial^2 I_{DS}}{\partial V_{GS}^2} \quad (4.63)$$

Differentiating (4.43) yields

$$\begin{aligned} c_{(2,0,0)} = \frac{1}{2} \left[\frac{\partial^2 long}{\partial V_{GS}^2} \cdot mobhot \cdot clm + \frac{\partial^2 mobhot}{\partial V_{GS}^2} \cdot long \cdot clm \right. \\ \left. + \frac{\partial^2 clm}{\partial V_{GS}^2} \cdot long \cdot mobhot + 2 \frac{\partial long}{\partial V_{GS}} \cdot \frac{\partial mobhot}{\partial V_{GS}} \cdot clm \right. \\ \left. + 2 \cdot \frac{\partial long}{\partial V_{GS}} \cdot \frac{\partial clm}{\partial V_{GS}} \cdot mobhot + 2 \cdot \frac{\partial mobhot}{\partial V_{GS}} \cdot \frac{\partial clm}{\partial V_{GS}} \cdot long \right] \end{aligned} \quad (4.64)$$

An evaluation of the terms in (4.64) versus V_{GS} for parameters listed in Table 4.1 is plotted in Figure 4.27 and shows that the first, second, fourth and fifth terms in (4.64) are the dominant terms. $c_{(2,0,0)}$ can be approximated by

$$\begin{aligned} c_{(2,0,0)} \approx \frac{1}{2} \left[\frac{\partial^2 long}{\partial V_{GS}^2} \cdot mobhot \cdot clm + \frac{\partial^2 mobhot}{\partial V_{GS}^2} \cdot long \cdot clm \right. \\ \left. + 2 \frac{\partial long}{\partial V_{GS}} \cdot \frac{\partial mobhot}{\partial V_{GS}} \cdot clm + 2 \cdot \frac{\partial long}{\partial V_{GS}} \cdot \frac{\partial clm}{\partial V_{GS}} \cdot mobhot \right] \end{aligned} \quad (4.65)$$

$c_{(2,0,0)}$ is primarily determined by the dependence of the function *long* on V_{GS} . However, *mobhot* and, to a lesser extent, *clm*, also have some significance. Since *mobhot* and *clm* decrease with increasing V_{GS} , the fourth and fifth terms are negative and limit the peak value of $c_{(2,0,0)}$. If $V_A \approx 2V$ though, the variation of *clm* with V_{GS} becomes very small and all terms with derivatives of *clm* become negligible.

The plot of $c_{(2,0,0)}$ shows that the rate of change of g_m with respect to V_{GS} reaches a maximum value slightly above the threshold voltage which corresponds to the moderate inversion region. At this point the second-order distortion components can be expected to be at their maximum. For higher values of V_{GS} , the rate of change of g_m reduces and starts to flatten, primarily due to channel velocity saturation effects which linearize the drain current. $c_{(2,0,0)}$ can reach zero for high values of V_{GS} and

Figure 4.27: An evaluation of the six terms in (4.64) versus V_{GS} with $V_{DS} = 600\text{mV}$ and $V_{SB} = 0$ for parameters listed in Table 4.1 .

even go negative for even higher values of V_{GS} . However, the drain current will be at very high levels and is well beyond the range of interest for low power RF design.

The drain-source voltage has very little effect on $c_{(2,0,0)}$, as was the case with $c_{(1,0,0)}$, and is not plotted. Similarly with $c_{(1,0,0)}$, the ratio of W_{eff}/L_{eff} acts as a scaling function on $c_{(2,0,0)}$. The parameters V_{th} , α , and, n play a significant role in (4.64), as was the case with $c_{(1,0,0)}$. Figure 4.28 shows $c_{(2,0,0)}$ versus V_{GS} for different values of V_{th} . The peak value of $c_{(2,0,0)}$ is relatively constant for different V_{th} , however, the location of the peak with respect to V_{GS} increases with increasing V_{th} . A similar type of behaviour can be observed for different values of V_{SB} since $V_{th} \propto V_{SB}$, but is not plotted here. Figure 4.29 plots $c_{(2,0,0)}$ versus V_{GS} for different values of n . The peak value of $c_{(2,0,0)}$ increases with decreasing n since the rate of change of g_m increases with decreasing n . The location of the peak does not shift with respect to V_{GS} since V_{th} is not dependent on n . The behaviour of $c_{(2,0,0)}$ with α is somewhat similar and is plotted in Figure 4.30. The peak value of $c_{(2,0,0)}$ increases with decreasing α since the rate of change of g_m increases with decreasing α . However, g_m tends to flatten for larger values of α and so $c_{(2,0,0)}$ decreases at a faster rate.

Figure 4.28: $c_{(2,0,0)}$ versus V_{GS} for different values of V_{th} with $V_{DS} = 600$ mV and $V_{SB} = 0$ using the parameters listed in Table 4.1.

Figure 4.29: $c_{(2,0,0)}$ versus V_{GS} for different values of n with $V_{DS} = 600$ mV and $V_{SB} = 0$ using the parameters listed in Table 4.1.

Figure 4.30: $c_{(2,0,0)}$ versus V_{GS} for different values of α with $V_{DS} = 600\text{mV}$ and $V_{SB} = 0$ using the parameters listed in Table 4.1.

Evaluation of $c_{(3,0,0)}$

The third derivative of the drain current with respect to V_{GS} is $c_{(3,0,0)}$ and is the most dominant source for third-order distortion products. $c_{(3,0,0)}$ is given by

$$c_{(3,0,0)} = \frac{1}{6} \frac{\partial^3 I_{DS}}{\partial V_{GS}^3} \quad (4.66)$$

Differentiating (4.43) yields

$$\begin{aligned} c_{(3,0,0)} = \frac{1}{6} & \left[\frac{\partial^3 \text{long}}{\partial V_{GS}^3} \cdot \text{mobhot} \cdot \text{clm} + \frac{\partial^3 \text{mobhot}}{\partial V_{GS}^3} \cdot \text{long} \cdot \text{clm} \right. \\ & + \frac{\partial^3 \text{clm}}{\partial V_{GS}^3} \cdot \text{long} \cdot \text{mobhot} + 3 \cdot \frac{\partial^2 \text{long}}{\partial V_{GS}^2} \cdot \frac{\partial \text{mobhot}}{\partial V_{GS}} \cdot \text{clm} \\ & + 3 \cdot \frac{\partial^2 \text{long}}{\partial V_{GS}^2} \cdot \frac{\partial \text{clm}}{\partial V_{GS}} \cdot \text{mobhot} + 3 \cdot \frac{\partial \text{long}}{\partial V_{GS}} \cdot \frac{\partial^2 \text{mobhot}}{\partial V_{GS}^2} \cdot \text{clm} \\ & + 3 \cdot \frac{\partial \text{long}}{\partial V_{GS}} \cdot \frac{\partial^2 \text{clm}}{\partial V_{GS}^2} \cdot \text{mobhot} + 3 \cdot \text{long} \cdot \frac{\partial^2 \text{mobhot}}{\partial V_{GS}^2} \cdot \frac{\partial \text{clm}}{\partial V_{GS}} \\ & \left. + 3 \cdot \text{long} \cdot \frac{\partial \text{mobhot}}{\partial V_{GS}} \cdot \frac{\partial^2 \text{clm}}{\partial V_{GS}^2} + 6 \cdot \frac{\partial \text{long}}{\partial V_{GS}} \cdot \frac{\partial \text{mobhot}}{\partial V_{GS}} \cdot \frac{\partial \text{clm}}{\partial V_{GS}} \right] \end{aligned} \quad (4.67)$$

An evaluation of the terms in (4.67) versus V_{GS} for parameters listed in Table 4.1 is plotted in Figure 4.31 and shows that the first, fourth, fifth, seventh and last terms in (4.67) are the dominant terms. Hence, $c_{(3,0,0)}$ can be approximated by

$$c_{(3,0,0)} \approx \frac{1}{6} \left[\frac{\partial^3 long}{\partial V_{GS}^3} \cdot mobhot \cdot clm + 3 \cdot \frac{\partial^2 long}{\partial V_{GS}^2} \cdot \frac{\partial mobhot}{\partial V_{GS}} \cdot clm \right. \\ + 3 \cdot \frac{\partial^2 long}{\partial V_{GS}^2} \cdot \frac{\partial clm}{\partial V_{GS}} \cdot mobhot + 3 \cdot \frac{\partial long}{\partial V_{GS}} \cdot \frac{\partial^2 clm}{\partial V_{GS}^2} \cdot mobhot \quad (4.68) \\ \left. + 6 \cdot \frac{\partial long}{\partial V_{GS}} \cdot \frac{\partial mobhot}{\partial V_{GS}} \cdot \frac{\partial clm}{\partial V_{GS}} \right]$$

and is compared to the full expression in Figure 4.32. The relative error is less than 10% for $V_{GS} < 0.6V$ and reasonably predicts the zero crossing. The relative error does become large for higher V_{GS} but the accuracy improves if the second, third and sixth terms are also included. Otherwise, all terms must be included.

Figure 4.31: An evaluation of the terms in (4.67) versus V_{GS} with $V_{DS} = 600mV$ and $V_{SB} = 0$ for parameters listed in Table 4.1 .

$c_{(3,0,0)}$ is primarily determined by the dependence of the function *long* on V_{GS} , as all terms with $\partial long / \partial V_{GS}$, $\partial^2 long / \partial V_{GS}^2$, and $\partial^3 long / \partial V_{GS}^3$ are significant. If the

Figure 4.32: A comparison of (4.68) to the full expression (4.67) versus V_{GS} with $V_{DS} = 600\text{mV}$ and $V_{SB} = 0$ for parameters listed in Table 4.1 .

sixth term is included though, the approximation over estimates $c_{(3,0,0)}$ for $V_{GS} > 0.6$ V and shows $c_{(3,0,0)}$ crossing at zero a second time and the second and third terms must be included to compensate. The eighth and ninth terms are negligible due to their dependence on higher-order derivatives of mob_{hot} and clm which are much smaller than derivatives of $long$. However, the dependence of mob_{hot} and clm on V_{GS} is of some importance, especially at higher gate bias levels. The fourth and fifth terms are negative since mob_{hot} and clm decrease with increasing V_{GS} and limit the peak value of $c_{(3,0,0)}$. For longer gate lengths, however, channel length modulation becomes less of a factor and all terms with $\partial^n clm / \partial V_{GS}^n$ become negligible (generally for $V_A > 2 - 3\text{V}$). Equation (4.67) then greatly simplifies to the first, fourth and sixth terms.

Figure 4.32 shows $c_{(3,0,0)}$ initially positive for low values of V_{GS} , corresponding to weak inversion, reaching a peak positive value, then decreasing, crossing zero and reaching a peak negative value in the strong inversion region ($V_{GS} > V_{th}$). $c_{(3,0,0)}$ then slowly increases and approaches zero for higher values of V_{GS} . In the weak inversion region, the mechanism for drain current flow is predominantly due to diffusion, which

follows an exponential type behaviour with respect to the gate voltage. From Figure 4.32 for $V_{GS} < 0.55V$, $c_{(3,0,0)}$ is positive in the weak inversion region, as expected for an exponential I-V characteristic curve, and the device will experience gain expansion. However, in the strong inversion region, the mechanism for drain current flow is mainly due to drift and the drain current characteristics follow something less than a square-law behaviour (i.e. the exponent is less than two). Although the CMOS transistor is often characterized as a square law device, the modern, short channel device deviates from the ideal square law operation and the nonlinear coefficient $c_{(3,0,0)}$ is non-zero and negative in strong inversion. As a result, the device will experience gain compression. Hence, as the operation of the transistor changes from weak inversion to strong inversion, $c_{(3,0,0)}$ is initially positive, passes through zero and becomes negative in the strong inversion region. However, as $c_{(3,0,0)}$ passes through zero in the moderate inversion region, the device acts as an ideal square-law device with no third-order distortion. Since there are no third-order distortion products, $IIP3$ can be expected to be infinite, giving rise to the linearity sweet spot. In reality, higher order distortion components such as fifth- and seventh-order distortion become dominant, since the third-order distortion component is minimized, and limit the $IIP3$ of the device.

The existence of the enhanced linearity region is inherent with the MOSFET and is independent of short channel effects. This is evident in Figure 4.33 which plots $c_{(3,0,0)}$ for *long* (equation (4.44)) without any short channel effects included. Although the bias point at which $c_{(3,0,0)}$ passes through zero, as well as the peak positive and negative magnitude, is different than in Figure 4.32, the same general behaviour can be observed: $c_{(3,0,0)}$ is positive for low V_{GS} due to exponential drain current behaviour, passes through zero, and remains negative for higher values of V_{GS} due to less than square-law behaviour.

While the existence of the linearity sweet spot (ie, where $c_{(3,0,0)}$ crosses zero) is independent of short channel effects, the location of the sweet spot and the peak positive and negative magnitudes are affected by velocity saturation and channel mobility effects. Figure 4.34 plots $c_{(3,0,0)}$ versus V_{GS} for different values of α . $\alpha = 0$ corresponds to no velocity saturation and no mobility degradation effects and the results match fairly closely to $\partial^3 long / \partial V_{GS}^3$. Higher values of α result in more linear behaviour of the drain current and the MOSFET behaves less like a square law device.

Figure 4.33: $\partial^3 long / \partial V_{GS}$ versus V_{GS} using the parameters listed in Table 4.1.

Hence the peak negative value of $c_{(3,0,0)}$ increases for increasing α and the zero crossing shifts to lower V_{GS} .

The parameters V_{th} , and, n also play a significant role in $c_{(3,0,0)}$ since $long$ is highly dependent on these parameters. Figure 4.35 shows $c_{(3,0,0)}$ versus V_{GS} for different values of V_{th} . The peak value of $c_{(3,0,0)}$ is relatively constant for different V_{th} , however, the location of the zero crossing increases with increasing V_{th} . A similar type of behaviour can be observed for different values of V_{SB} since $V_{th} \propto V_{SB}$, but is not plotted here. Figure 4.36 plots $c_{(3,0,0)}$ versus V_{GS} for different values of n . The peak positive and negative value of $c_{(3,0,0)}$ increases with decreasing n since the rate of change of g_m increases with decreasing n . The location of the peaks and the zero crossing does not shift with respect to V_{GS} since V_{th} is not dependent on n .

The bias voltage V_{DS} does not significantly affect $c_{(3,0,0)}$ since $long$ and $mobhot$ are independent of V_{DS} . The ratio W_{eff}/L_{eff} is significant, but merely scales the value of $c_{(3,0,0)}$ as expected and is not plotted here. While the parameter P_{VAG} had negligible effect for $c_{(1,0,0)}$ and $c_{(2,0,0)}$, it does have more effect with $c_{(3,0,0)}$ for $V_A < 2 - 3V$ and is plotted in Figure 4.37. This occurs because $\partial^3 clm / \partial V_{GS}^3$ has much larger peak values than $\partial^2 clm / \partial V_{GS}^2$ and $\partial clm / \partial V_{GS}$ in the moderate inversion region. If $V_A > 2$ though, $\partial^3 clm / \partial V_{GS}^3$ is significantly reduced and P_{VAG} has negligible effect on $c_{(3,0,0)}$.

Figure 4.34: $c_{(3,0,0)}$ versus V_{GS} for different values of α with $V_{DS} = 600\text{mV}$ and $V_{SB} = 0$ using the parameters listed in Table 4.1.

Figure 4.35: $c_{(3,0,0)}$ versus V_{GS} for different values of V_{th} with $V_{DS} = 600\text{mV}$ and $V_{SB} = 0$ using the parameters listed in Table 4.1.

Figure 4.36: $c_{(3,0,0)}$ versus V_{GS} for different values of n with $V_{DS} = 600\text{mV}$ and $V_{SB} = 0$ using the parameters listed in Table 4.1.

Figure 4.37: $c_{(3,0,0)}$ versus V_{GS} for different values of P_{VAG} with $V_{DS} = 600\text{mV}$, $V_{SB} = 0$ and using the parameters listed in Table 4.1.

4.6.2 Expressions for $c_{(0,1,0)}$, $c_{(0,2,0)}$, and $c_{(0,3,0)}$

The first-order derivative of the drain current with respect to V_{DS} is $c_{(0,1,0)}$ and is also commonly referred to as the output conductance or g_o . It is given by

$$c_{(0,1,0)} = \frac{\partial I_{DS}}{\partial V_{DS}} \quad (4.69)$$

Differentiating (4.43) yields

$$\begin{aligned} c_{(0,1,0)} = & \frac{\partial long}{\partial V_{DS}} \cdot mobhot \cdot clm + \frac{\partial mobhot}{\partial V_{DS}} \cdot long \cdot clm \\ & + \frac{\partial clm}{\partial V_{DS}} \cdot long \cdot mobhot \end{aligned} \quad (4.70)$$

Equation (4.70) simplifies to

$$c_{(0,1,0)} = \frac{\partial clm}{\partial V_{DS}} \cdot long \cdot mobhot \quad (4.71)$$

$$= \frac{1}{V_A(V_{GS}, V_{SB})} \cdot long \cdot mobhot \quad (4.72)$$

since *long* and *mobhot* are independent of V_{DS} . As previously mentioned, this is not entirely accurate primarily due to the threshold voltage dependence on V_{DS} . Furthermore, the expressions given by (4.38) to (4.40) are valid only for the saturation region and become increasingly inaccurate from the transition from saturation region to the linear region. Figure 4.38 compares the measured output conductance to the calculated output conductance using the parameters listed in Table 4.1 and shows good agreement for $V_{DS} > V_{DS(sat)}$. Obviously, the error becomes significant for $V_{DS} < V_{DS(sat)}$.

From (4.72), g_o is inversely proportional to V_A . As V_A increases, the slope of I_{DS} versus V_{DS} flattens and g_o decreases, as expected. g_o is also proportional to *long* and *modhot* and so any increase in either *long* or *mobhot* will result in an increase in g_o (or a decrease in the output resistance).

Figure 4.38: A comparison of the measured and calculated output conductance versus V_{DS} with $V_{SB} = 0$ for parameters listed in Table 4.1 .

Since clm is modelled as linear with respect to V_{DS} , the higher-order derivatives with respect to V_{DS} will be zero. Hence

$$c_{(0,2,0)} = 0 \quad (4.73)$$

$$c_{(0,3,0)} = 0 \quad (4.74)$$

However, from Figure 4.39, which plots the measured $c_{(0,2,0)}$ versus V_{DS} , this is not entirely accurate. For $V_{GS} = 400\text{mV}$ and 500mV , $c_{(2,0,0)}$ is nearly zero. For larger V_{GS} though, $c_{(0,2,0)}$ is not zero but is very small and approaches zero for increasing V_{DS} . The fluctuations in the plot are due to fluctuations in the measured data which get highly exaggerated, especially after the second and third derivative. A Savitzky Golay smoothing filtering algorithm [48] was used to smoothen data as much as possible before calculating the second derivative. Although the fluctuations in the measured data are very small, it was not possible to derive the third derivative from

the measured data. However, it is evident from $c_{(0,2,0)}$ in Figure 4.39 that the third derivative will be very small but not exactly zero, especially for $V_{GS} > 500\text{mV}$.

Figure 4.39: Measured $c_{(0,2,0)}$ versus V_{DS} with $V_{SB} = 0$ using a Savitzky Golay smoothing filter [48] of fourth order with 20 coefficients.

4.6.3 Expressions for $c_{(0,0,1)}$, $c_{(0,0,2)}$, and $c_{(0,0,3)}$

The magnitudes of nonlinear coefficients $c_{(0,0,1)}$, $c_{(0,0,2)}$, and $c_{(0,0,3)}$ (and $c_{(0,1,0)}$) are much less than those of $c_{(1,0,0)}$, $c_{(2,0,0)}$, and $c_{(3,0,0)}$ and generally do not dominate the distortion components. As a result, the evaluation of the expressions for these nonlinear coefficients will not be discussed in great detail. However, depending on circuit parameters, the dependencies of the distortion components on $c_{(0,0,1)}$, $c_{(0,0,2)}$, $c_{(0,0,3)}$, and $c_{(0,1,0)}$ can become more significant and will be discussed in more detail in Chapter 5.

The bias voltage V_{DS} does not significantly affect $c_{(0,0,1)}$, $c_{(0,0,2)}$, and $c_{(0,0,3)}$ since *long* and *mobhot* are independent of V_{DS} and is not plotted here. The other parameters, such as W_{eff}/L_{eff} , V_{T0} , α , and P_{VAG} , affect the nonlinear coefficients of

V_{SB} in a very similar fashion as $c_{(1,0,0)}$, $c_{(2,0,0)}$, and $c_{(3,0,0)}$ and are also not plotted here for brevity.

Evaluation of $c_{(0,0,1)}$ (g_{mb})

The first-order derivative of the drain current with respect to V_{SB} is $c_{(0,0,1)}$ and is commonly referred to as the output bulk conductance or g_{mb} . It is given by

$$c_{(0,0,1)} = \frac{\partial I_{DS}}{\partial V_{SB}} \quad (4.75)$$

Differentiating (4.43) yields

$$\begin{aligned} c_{(0,0,1)} = & \frac{\partial long}{\partial V_{SB}} \cdot mobhot \cdot clm + \frac{\partial mobhot}{\partial V_{SB}} \cdot long \cdot clm \\ & + \frac{\partial clm}{\partial V_{SB}} \cdot long \cdot mobhot \end{aligned} \quad (4.76)$$

An evaluation of the three terms in (4.76) versus V_{GS} and V_{SB} for parameters listed in Table 4.1 shows that the first two terms in (4.76) are the dominant terms in a similar fashion as g_m . g_{mb} can be approximated by

$$c_{(0,0,1)} \approx \frac{\partial long}{\partial V_{SB}} \cdot mobhot \cdot clm + \frac{\partial mobhot}{\partial V_{SB}} \cdot long \cdot clm \quad (4.77)$$

g_{mb} , plotted in Figure 4.40, is primarily determined by the dependence of the function $long$ on V_{SB} but since $long$ decreases with increasing V_{SB} (Figure 4.15), $\partial long / \partial V_{SB}$ is negative. On the other hand, $mobhot$ increases with increasing V_{SB} (Figure 4.17) and the second term limits the magnitude of $c_{(0,0,1)}$.

Evaluation of $c_{(0,0,2)}$

The second derivative of the drain current with respect to V_{SB} is $c_{(0,0,2)}$ and given by

$$c_{(0,0,2)} = \frac{1}{2} \frac{\partial^2 I_{DS}}{\partial V_{SB}^2} \quad (4.78)$$

Figure 4.40: g_{mb} versus V_{GS} for $V_{SB} = 0, 0.3$, and 0.6 V using the parameters listed in Table 4.1.

Differentiating (4.43) yields

$$\begin{aligned}
 c_{(0,0,2)} = \frac{1}{2} & \left[\frac{\partial^2 long}{\partial V_{SB}^2} \cdot mobhot \cdot clm + \frac{\partial^2 mobhot}{\partial V_{SB}^2} \cdot long \cdot clm \right. \\
 & + \frac{\partial^2 clm}{\partial V_{SB}^2} \cdot long \cdot mobhot + 2 \frac{\partial long}{\partial V_{SB}} \cdot \frac{\partial mobhot}{\partial V_{SB}} \cdot clm \\
 & \left. + 2 \cdot \frac{\partial long}{\partial V_{SB}} \cdot \frac{\partial clm}{\partial V_{SB}} \cdot mobhot + 2 \cdot \frac{\partial mobhot}{\partial V_{SB}} \cdot \frac{\partial clm}{\partial V_{SB}} \cdot long \right] \quad (4.79)
 \end{aligned}$$

An evaluation of the terms in (4.79) versus V_{GS} and V_{SB} for parameters listed in Table 4.1 shows that the first, fourth and fifth terms in (4.79) are the dominant terms. $c_{(0,0,2)}$ can be approximated by

$$\begin{aligned}
 c_{(0,0,2)} \approx \frac{1}{2} & \left[\frac{\partial^2 long}{\partial V_{SB}^2} \cdot mobhot \cdot clm + 2 \frac{\partial long}{\partial V_{SB}} \cdot \frac{\partial mobhot}{\partial V_{SB}} \cdot clm \right. \\
 & \left. + 2 \cdot \frac{\partial long}{\partial V_{SB}} \cdot \frac{\partial clm}{\partial V_{SB}} \cdot mobhot \right] \quad (4.80)
 \end{aligned}$$

$c_{(0,0,2)}$, plotted in Figure 4.41, is primarily determined by the dependence of the function *long* on V_{SB} . The absolute value of $c_{(0,0,2)}$ is reduced by mobility reduction and velocity saturation as well as channel length modulation (second and third terms), which can be considered a linearizing effect.

Figure 4.41: $c_{(0,0,2)}$ versus V_{GS} for $V_{SB} = 0, 0.3$, and 0.6 V using the parameters listed in Table 4.1.

Evaluation of $c_{(0,0,3)}$

The third derivative of the drain current with respect to V_{SB} is $c_{(0,0,3)}$ and is given by

$$c_{(0,0,3)} = \frac{1}{6} \frac{\partial^3 I_{DS}}{\partial V_{SB}^3} \quad (4.81)$$

Differentiating (4.43) yields

$$\begin{aligned}
c_{(0,0,3)} = \frac{1}{6} & \left[\frac{\partial^3 long}{\partial V_{SB}^3} \cdot mobhot \cdot clm + \frac{\partial^3 mobhot}{\partial V_{SB}^3} \cdot long \cdot clm \right. \\
& + \frac{\partial^3 clm}{\partial V_{SB}^3} \cdot long \cdot mobhot + 3 \cdot \frac{\partial^2 long}{\partial V_{SB}^2} \cdot \frac{\partial mobhot}{\partial V_{SB}} \cdot clm \\
& + 3 \cdot \frac{\partial^2 long}{\partial V_{SB}^2} \cdot \frac{\partial clm}{\partial V_{SB}} \cdot mobhot + 3 \cdot \frac{\partial long}{\partial V_{SB}} \cdot \frac{\partial^2 mobhot}{\partial V_{SB}^2} \cdot clm \\
& + 3 \cdot \frac{\partial long}{\partial V_{SB}} \cdot \frac{\partial^2 clm}{\partial V_{SB}^2} \cdot mobhot + 3 \cdot long \cdot \frac{\partial^2 mobhot}{\partial V_{SB}^2} \cdot \frac{\partial clm}{\partial V_{SB}} \\
& \left. + 3 \cdot long \cdot \frac{\partial mobhot}{\partial V_{SB}} \cdot \frac{\partial^2 clm}{\partial V_{SB}^2} + 6 \cdot \frac{\partial long}{\partial V_{SB}} \cdot \frac{\partial mobhot}{\partial V_{SB}} \cdot \frac{\partial clm}{\partial V_{SB}} \right] \quad (4.82)
\end{aligned}$$

An evaluation of the terms in (4.82) versus V_{SB} for parameters listed in Table 4.1 shows that the first, fourth, and fifth terms in (4.82) are the dominant terms. Hence, $c_{(0,0,3)}$ can be approximated by

$$\begin{aligned}
c_{(0,0,3)} \approx \frac{1}{6} & \left[\frac{\partial^3 long}{\partial V_{SB}^3} \cdot mobhot \cdot clm + 3 \cdot \frac{\partial^2 long}{\partial V_{SB}^2} \cdot \frac{\partial mobhot}{\partial V_{SB}} \cdot clm \right. \\
& \left. + 3 \cdot \frac{\partial^2 long}{\partial V_{SB}^2} \cdot \frac{\partial clm}{\partial V_{SB}} \cdot mobhot \right] \quad (4.83)
\end{aligned}$$

As with $c_{(0,0,1)}$ and $c_{(0,0,2)}$, $c_{(0,0,3)}$ is primarily determined by the dependence of the function *long* on V_{SB} and is plotted in Figure 4.42. However, unlike $c_{(3,0,0)}$ and V_{GS} , the dependencies of *mobhot* and *clm* on V_{SB} are much less significant than on V_{GS} in $c_{(3,0,0)}$ and, as a result, all terms in (4.82) with second-order and third-order derivatives of *mobhot* and *clm* are negligible.

4.6.4 Expressions for the cross-modulation terms

The cross-modulation terms $c_{(1,1,0)}$, $c_{(1,2,0)}$, $c_{(2,1,0)}$, $c_{(0,1,1)}$, $c_{(0,1,2)}$, $c_{(0,2,1)}$, $c_{(1,0,1)}$, $c_{(1,0,2)}$, $c_{(2,0,1)}$, and $c_{(1,1,1)}$ describe the dependence of a nonlinear coefficient on a second controlling voltage. For example, $c_{(1,1,0)}$ describes the dependence of g_m with V_{DS} or the dependence of g_o with V_{GS} . Similarly, $c_{(2,1,0)}$ describes the second-order dependence of g_o with V_{GS} while $c_{(1,2,0)}$ describes the second-order dependence of g_m to V_{DS} . The magnitude of the cross-modulation terms is also much less than those of $c_{(1,0,0)}$, $c_{(2,0,0)}$,

Figure 4.42: $c_{(0,0,3)}$ versus V_{GS} for $V_{SB} = 0, 0.3$, and 0.6 V using the parameters listed in Table 4.1.

and $c_{(3,0,0)}$ and generally does not dominate the distortion components. As a result, the evaluation of the expressions for these nonlinear coefficients will not be discussed in great detail. In addition, since clm is modelled as linear with respect to V_{DS} and $long$ and $mobhot$ are modelled as independent of V_{DS} , all cross modulation terms with a second-order dependence on V_{DS} , or $c_{(n,2,n)}$ are zero. Furthermore, for most circuit applications, $c_{(1,1,0)}$ and $c_{(2,1,0)}$ are the most dominant of the cross-modulation terms. Depending on circuit parameters, the dependencies of the distortion components on the cross modulation terms can become more significant and will be discussed in more detail in Chapter 5.

Evaluation of $c_{(1,1,0)}$ and $c_{(2,1,0)}$

$c_{(1,1,0)}$ is given by

$$c_{(1,1,0)} = \frac{\partial^2 I_{DS}}{\partial V_{GS} \partial V_{DS}} \quad (4.84)$$

Differentiating (4.43) yields

$$c_{(1,1,0)} = \frac{\partial long}{\partial V_{GS}} \cdot \frac{\partial clm}{\partial V_{DS}} \cdot mobhot + long \cdot \frac{\partial mobhot}{\partial V_{GS}} \cdot \frac{\partial clm}{\partial V_{DS}} + long \cdot mobhot \cdot \frac{\partial^2 clm}{\partial V_{GS} \partial V_{DS}} \quad (4.85)$$

$c_{(2,1,0)}$ is given by

$$c_{(2,1,0)} = \frac{\partial^3 I_{DS}}{\partial V_{GS}^2 \partial V_{DS}} \quad (4.86)$$

and differentiating (4.43) yields

$$c_{(2,1,0)} = \frac{1}{2} \left[\frac{\partial^2 long}{\partial V_{GS}^2} \cdot mobhot \cdot \frac{\partial clm}{\partial V_{DS}} + long \cdot \frac{\partial^2 mobhot}{\partial V_{GS}^2} \cdot \frac{\partial clm}{\partial V_{DS}} + 2 \cdot long \cdot \frac{\partial mobhot}{\partial V_{GS}} \cdot \frac{\partial^2 clm}{\partial V_{GS} \partial V_{DS}} + 2 \cdot \frac{\partial long}{\partial V_{GS}} \cdot mobhot \cdot \frac{\partial^2 clm}{\partial V_{GS} \partial V_{DS}} + 2 \cdot \frac{\partial long}{\partial V_{GS}} \cdot \frac{\partial mobhot}{\partial V_{GS}} \cdot \frac{\partial clm}{\partial V_{DS}} + long \cdot mobhot \cdot \frac{\partial^3 clm}{\partial V_{GS}^2 \partial V_{DS}} \right] \quad (4.87)$$

No simplifications are possible with either (4.85) or (4.87) as all terms are significant relative to each other with the exception of the second term in (4.87). Furthermore, no simplifications are possible for a long channel device since all terms are dependent on $\partial clm / \partial V_{DS}$. As a result, both $c_{(1,1,0)}$ and (4.87) are inversely proportional to $1/V_A$ and their magnitudes are highly dependent on channel length modulation effects. Figure 4.43(a) shows $c_{(1,1,0)}$ versus V_{GS} for different V_{SB} and Figure 4.43(b) shows $c_{(2,1,0)}$.

4.7 Summary

In this chapter a “single-piece” MOSFET model is presented that is continuous over all regions of inversion and provides good accuracy in the moderate inversion region. Mobility reduction, velocity saturation and channel length modulation have significant effects on modern deep sub-micron MOSFET and must be included in the models as well. Although fairly complex models are available for these effects, simpler models were used that provided reasonable accuracy for the drain current and, perhaps more importantly, for the higher order derivatives.

(a)

(b)

Figure 4.43: (a) $c_{(1,1,0)}$ and (b) $c_{(2,1,0)}$ versus V_{GS} with $V_{DS} = 600\text{mV}$, $V_{SB}=0\text{V}$, 300mV , and 600mV and with the parameters listed in Table 4.1.

Since the drain current is dependent on three terminal voltages, a total of nineteen nonlinear coefficients up to third-order are required to describe the drain current's nonlinearities. These coefficients are needed for the derivation of the harmonic and intermodulation distortion products of various circuit configurations that will be addressed in the next chapter. The nonlinear coefficients were derived for the saturation region only as functions of $long$, $mobhot$, and clm , which model long channel effects, mobility degradation and hot carrier effects and channel length modulation effects respectively. The dependency of $long$ with respect to V_{GS} was the dominant factor in the nonlinear coefficients. $mobhot$ was the next dominant factor and tends to

linearize the drain current. For a minimum length, modern deep sub-micron device, clm also plays an important role since the output conductance is more dependent on V_{GS} than with longer channel devices. The nonlinear coefficients were also shown to be highly dependent on the parameters V_{th} , n , and α and the bias voltages V_{GS} and V_{SB} .

The importance of the coefficient $c_{(3,0,0)}$ was also discussed in detail and how its behaviour affects third-order distortion products. As the operation of the transistor changes from weak inversion to strong inversion, $c_{(3,0,0)}$ is initially positive, passes through zero and becomes negative in strong inversion. As $c_{(3,0,0)}$ passes through zero, the MOSFET acts as an ideal square-law device with no third-order distortion. Since there are no third-order distortion products, the third-order intercept point can be expected to be very high (ideally approaches infinity), giving rise to a linearity sweet spot. It was also shown that the location of the sweet spot with respect to V_{GS} is primarily dependent on the parameter V_{th} and α .

Finally, it should be evident that the modelling of a MOSFET is a very difficult task and additional effects need to be taken into account as the MOSFET scales to even smaller device sizes. This will obviously influence the computations of the nonlinear coefficients.

Chapter 5

Distortion Analysis and Measurement of Weakly Nonlinear Amplifiers in Moderate Inversion

5.1 Introduction

In this chapter, the nonlinear distortion of some basic RF building blocks for MOS-FETs are studied using the Direct Calculation Method for Volterra series analysis. The analysis will make use of the nonlinear coefficients derived in Chapter 4. This study first focuses on a resistively loaded common-source amplifier. This circuit is studied in depth from which conclusions are drawn and applied to other more complicated circuits. Measurements are also presented both at RF frequencies (up to 2.5 GHz) and at low frequencies (10 MHz). It is shown that the results at RF frequencies do not differ much from the low frequency measurements thereby allowing low frequency measurements to be used throughout the remainder of the chapter. This simplifies the measurements considerably. Other circuit configurations are also considered such as inductive loads, inductive source degeneration, common-gate amplifiers and cascode amplifiers. Although the circuits considered are relatively simple circuits, it will be seen that the analysis can be quite complicated and lengthy.

The analysis will emphasize the moderate inversion region since little to no analysis for this region has been previously presented [53]. While the cut-off frequency, f_T , of a transistor biased in weak inversion is too low for RF applications, there is sufficient current gain for most RF applications while biased in the moderate inversion region, especially with modern sub-micron devices. As CMOS technology continues

to scale to smaller device sizes, and with the quest for lower power consumption, more RF circuits will be designed for operation in the moderate inversion region. As such, a differential, common-source LNA design, suitable for a Bluetooth application is presented that demonstrates operation in the moderate inversion region is very feasible. The measured results compare favourably to several other CMOS LNA designs.

5.2 Common Source Amplifier

Figure 5.1(a) shows a common source amplifier. The source and bulk are grounded and the input to the amplifier is a voltage source with a source resistance R_S . The output is loaded with a resistance R_L and a capacitance C_L . A simplified AC-equivalent circuit is shown in Figure 5.1(b). This is the same circuit configuration that was shown in Section 3.6. The results from Section 3.6 will be used as a starting point from which additional effects will be added and analyzed. Both C_{gs} and C_{db} are initially assumed to be constant after which the effect of the nonlinear capacitances will be considered. Since C_L is in parallel with C_{db} , the two capacitors can be lumped together. A cascode transistor is often used in RF applications to reduce the Miller effect of the gate-drain capacitance, C_{gd} , and so C_{gd} is neglected for now for simplicity. The drain current is also initially assumed to be a function of only V_{gs} after which the nonlinearities due to the output conductance will be addressed.

5.2.1 Influence of g_m Only

From equation (3.79), the third-order intercept point for low frequencies, repeated here for convenience, was found to be

$$V_{IIP3}^2 = \frac{4}{3} \left| \frac{g_m}{c_{(3,0,0)}} \right| \quad (5.1)$$

Since only the influence of g_m is being considered for now, g_m and $c_{(3,0,0)}$ can be simplified considerably with $clm = 1$ to yield a closed form expression for V_{IIP3}^2 given by

$$V_{IIP3}^2 = \left| \frac{16n^2\phi_t^2(z+1)^2X(2+\alpha X)}{6zn\phi_t(2n\phi_t\alpha z-1) + (z-1)X + \alpha(z-1)X^2} \right| \quad (5.2)$$

Figure 5.1: (a) A common-source amplifier. (b) AC-equivalent circuit.

where

$$z = e^{(V_{GS} - V_{th})/(2n\phi_t)} \quad (5.3)$$

Equation (5.2) is plotted in Figure 5.2. Since typical input signal levels for RF applications are in the μV range, $IIP3$ is plotted with respect to $\text{dB}\mu\text{V}$. The plot shows a significant peaking, or a "sweet spot", in the moderate inversion region where $c_{(3,0,0)}$ passes through zero. Ideally, $IIP3$ will be infinite when $c_{(3,0,0)}$ passes through zero, however, other effects limit the peak value. From (5.2), $IIP3$ is dependent only on V_{GS} , V_{th} , $n\phi_t$, and α and is independent of device size and V_{DS} , since $clm = 1$. However, $IIP3$ can be expected to be somewhat dependent on device size due to variations in the threshold voltage and short channel effects with width. Since the zero crossing point of $c_{(3,0,0)}$ largely depends on only V_{th} and α , as discussed in Section 4.6.1, the same is true for the peak $IIP3$.

Figure 5.2: V_{IP3}^2 , given by (5.2), versus V_{GS} when considering only the influence of g_m . The parameters used for (5.2) are listed in Table 4.1 and $V_{SB} = 0V$.

5.2.2 Influence of Output Conductance

The simple AC-equivalent circuit in Figure 5.1(b) is now extended to include the influence of the output conductance, which is shown in Figure 5.3(a). Since the influence of the output conductance is included, the full expression for I_{ds} , given by (4.43), must be now considered.

Fundamental Response

To obtain the fundamental response, the linearized equivalent circuit, shown in Figure 5.3(a), is analyzed. The fundamental response at ω_1 at the gate is given by

$$V_{g,1,0} = -\frac{G_S}{Y_i(\omega_1)} v_s \quad (5.4)$$

Figure 5.3: Equivalent circuits of the common source amplifier of Figure 5.1 including the influence of g_o for determining (a) the fundamental response, (b) the second-order response, and (c) the third-order response.

which will be needed later for determining the second and third-order responses. The response at the output node is given by

$$V_{d,1,0} = -\frac{g_m G_S}{Y_o(\omega_1) Y_i(\omega_1)} v_s \quad (5.5)$$

where

$$Y_i(\omega) = G_s + j\omega C_{gs} \quad (5.6)$$

is the input admittance,

$$Y_o(\omega) = G_L + g_o + j\omega(C_L + C_{db}) \quad (5.7)$$

$$= G'_L + j\omega C'_L \quad (5.8)$$

is the output admittance, $G_s = 1/R_s$, and $G_L = 1/R_L$. The fundamental responses at $\pm\omega_2$, which will be needed for the second and third-order responses, are simply

$$V_{g,0,\pm 1} = -\frac{G_s}{Y_i(\pm\omega_2)} v_s \quad (5.9)$$

$$V_{d,0,\pm 1} = -\frac{g_m G_s}{Y_o(\pm\omega_2) Y_i(\pm\omega_2)} v_s \quad (5.10)$$

Second-order Responses

The second-order responses at $2\omega_1$, $2\omega_2$, and $|\omega_1 \pm \omega_2|$, are solved using the same linearized network as used for the first-order responses, except different inputs are used. The external signal sources are removed and the *second-order nonlinear current source*, i_{NL2ids} is placed in parallel with the linearized equivalent of the nonlinearity. In this case, the current source i_{NL2ids} is comprised of three components:

- i_{NL2gm} - the second-order nonlinear current source due to g_m only;
- i_{NL2go} - the second-order nonlinear current source due to g_o only; and,
- $i_{NL2gm\&go}$ - the second-order nonlinear current source due to g_m and g_o ,

such that $i_{NL2ids} = i_{NL2gm} + i_{NL2go} + i_{NL2gm\&go}$.

The expressions for each current source for the responses at $2\omega_1$ are obtained from Table 3.2 and are dependent on the responses at ω_1 where

$$i_{NL2gm}(2\omega_1) = \frac{c(2,0,0)}{2} V_{g,1,0}^2 \quad (5.11)$$

$$= \frac{c(2,0,0)}{2} \left(\frac{G_s}{Y_i(\omega_1)} \right)^2 v_s^2 \quad (5.12)$$

$$i_{NL2go}(2\omega_1) = \frac{c(0,2,0)}{2} V_{d,1,0}^2 \quad (5.13)$$

$$= \frac{c(0,2,0)}{2} \left(\frac{g_m G_s}{Y_o(\omega_1) Y_i(\omega_1)} \right)^2 v_s^2 \quad (5.14)$$

$$i_{NL2gm\&go}(2\omega_1) = \frac{c(1,1,0)}{2} V_{g,1,0} V_{d,1,0} \quad (5.15)$$

$$= \frac{c(1,1,0)}{2} \frac{g_m G_s^2}{Y_o(\omega_1) Y_i^2(\omega_1)} v_s^2 \quad (5.16)$$

The expressions for the current sources at $2\omega_2$ are the same as above except ω_1 is replaced by ω_2 .

The expressions for each current source for $\omega_1 \pm \omega_2$ are slightly different though and are also obtained from Table 3.2.

$$i_{NL2_{gm}}(\omega_1 \pm \omega_2) = c_{(2,0,0)} V_{g,1,0} V_{g,0,\pm 1} \quad (5.17)$$

$$= c_{(2,0,0)} \frac{G_S^2}{Y_i(\omega_1) Y_i(\pm \omega_2)} v_s^2 \quad (5.18)$$

$$i_{NL2_{go}}(\omega_1 \pm \omega_2) = c_{(0,2,0)} V_{d,1,0} V_{d,0,\pm 1} \quad (5.19)$$

$$= c_{(0,2,0)} \frac{g_m^2 G_S^2}{Y_o(\omega_1) Y_o(\pm \omega_2) Y_i(\omega_1) Y_i(\pm \omega_2)} v_s^2 \quad (5.20)$$

$$i_{NL2_{gm \& go}}(\omega_1 \pm \omega_2) = \frac{c_{(1,1,0)}}{2} (V_{g,1,0} V_{d,0,\pm 1} + V_{g,0,\pm 1} V_{d,1,0}) \quad (5.21)$$

$$= \frac{c_{(1,1,0)}}{2} \frac{g_m G_S^2 (Y_o(\omega_1) + Y_o(\pm \omega_2))}{Y_o(\omega_1) Y_o(\pm \omega_2) Y_i(\omega_1) Y_i(\pm \omega_2)} v_s^2 \quad (5.22)$$

The second-order responses at the gate node are trivial with $V_{g,2,0} = V_{g,0,\pm 2} = V_{g,1,\pm 1} = 0$. However, the response at the output drain node at $2\omega_1$ is

$$V_{d,2,0} = \frac{-i_{NL2_{ids}}(2\omega_1)}{Y_o(2\omega_1)} \quad (5.23)$$

$$= -\frac{v_s^2}{2} \frac{G_S^2}{Y_i^2(\omega_1) Y_o(2\omega_1)} \left(c_{(2,0,0)} - c_{(1,1,0)} \frac{g_m}{Y_o(\omega_1)} + c_{(0,2,0)} \frac{g_m^2}{Y_o^2(\omega_1)} \right) \quad (5.24)$$

The factor $g_m/Y_o(\omega_1)$ is the voltage gain, A_v , at ω_1 from the gate input to the output and so

$$V_{d,2,0} = -\frac{v_s^2}{2} \frac{G_S^2}{Y_i^2(\omega_1) Y_o(2\omega_1)} (c_{(2,0,0)} - c_{(1,1,0)} A_v + c_{(0,2,0)} A_v^2) \quad (5.25)$$

The response at $\omega_1 \pm \omega_2$ is very similar to (5.25) with

$$V_{d,1,\pm 1} = -v_s^2 \frac{G_S^2}{Y_i^2(\omega_1) Y_o(\omega_1 \pm \omega_2)} (c_{(2,0,0)} - c_{(1,1,0)} A_v + c_{(0,2,0)} A_v^2) \quad (5.26)$$

assuming $Y_o(\omega_2) \approx Y_o(\omega_1)$. Hence, the variation of g_o with V_{DS} , given by $c_{(0,2,0)}$, and the variation of g_o with V_{GS} , (or g_m with V_{DS}), given by $c_{(1,1,0)}$, contributes to the second harmonic distortion components. If the voltage gain is low, their contribution is generally negligible since $c_{(2,0,0)} \gg c_{(1,1,0)} \gg c_{(0,2,0)}$. If the voltage gain is large

though, the contribution by $c_{(1,1,0)}$ can become significant since it is multiplied by A_v . Even though clm has been modelled as linear with respect to V_{ds} and $c_{(0,2,0)} = 0$, $c_{(0,2,0)}$ is not exactly zero in a real device as discussed in Section 4.6.2. If the voltage gain is very large then the contribution from $c_{(0,2,0)}$ may also become significant since it is multiplied by A_v^2 .

Third-order Response

The third-order responses at $3\omega_1$, $3\omega_2$, and $2\omega_1 - \omega_2$ are solved using the same approach for the second-order responses, except $i_{NL2_{ids}}$ is replaced by a *third-order nonlinear current source*, $i_{NL3_{ids}}$. The current source $i_{NL3_{ids}}$ is also comprised of three components in this case:

- $i_{NL3_{gm}}$ - the third-order nonlinear current source due to g_m only;
- $i_{NL3_{go}}$ - the third-order nonlinear current source due to g_o only; and
- $i_{NL3_{gm\&go}}$ - the third-order nonlinear current source due to the two-dimensional cross products.

The expressions for each current source for the responses at $3\omega_1$ (or $3\omega_2$) are obtained from Table 3.4 and the expressions for each current source for the responses at $2\omega_1 - \omega_2$ are obtained from Table 3.3. Since $V_{g,2,0} = V_{g,1,-1} = 0$, the third-order nonlinear current sources for calculating the response at $3\omega_1$ become

$$i_{NL3_{gm}}(3\omega_1) = \frac{1}{4}c_{(3,0,0)}V_{g,1,0}^3 \quad (5.27)$$

$$= \frac{c_{(3,0,0)}}{4} \frac{G_s^3}{Y_i^3(w_1)} v_s^3 \quad (5.28)$$

$$i_{NL3_{go}}(3\omega_1) = c_{(2,0,0)}V_{d,1,0}V_{d,2,0} + \frac{1}{4}c_{(0,3,0)}V_{d,1,0}^3 \quad (5.29)$$

$$= \frac{1}{4} \frac{G_s^3}{Y_i^3(w_1)} v_s^3 \left[2c_{(0,2,0)} c_{(2,0,0)} \frac{A_v}{Y_o(2w_1)} - c_{(0,3,0)} A_v^3 \right. \\ \left. + 2c_{(0,2,0)}^2 \frac{A_v^3}{Y_o(2w_1)} - 2c_{(0,2,0)} c_{(1,1,0)} \frac{A_v^2}{Y_o(2w_1)} \right] \quad (5.30)$$

and

$$i_{NL3_{gm\&go}}(3\omega_1) = \frac{1}{2}c_{(1,1,0)}V_{g,1,0}V_{d,2,0} + \frac{1}{4}c_{(2,1,0)}V_{g,1,0}^2V_{d,1,0} + \frac{1}{4}c_{(1,2,0)}V_{g,1,0}V_{d,1,0}^2 \quad (5.31)$$

$$= \frac{1}{4} \frac{G_s^3}{Y_i^3(w_1)} v_s^3 \left[c_{(1,2,0)} A_v^2 c_{(1,1,0)} - c_{(2,1,0)} A_v + c_{(1,1,0)}^2 \frac{A_v}{Y_o(2w_1)} - \frac{c_{(1,1,0)} c_{(2,0,0)}}{Y_o(2w_1)} - c_{(0,2,0)} c_{(1,1,0)} \frac{A_v^2}{Y_o(2w_1)} \right] \quad (5.32)$$

The third-order responses at the gate node are $V_{g,3,0} = V_{g,0,\pm 3} = V_{g,2,\pm 1} = 0$. The response at the output drain node at $3\omega_1$ is

$$V_{d,3,0} = \frac{-i_{NL3_{ids}}(3\omega_1)}{Y_o(3\omega_1)} \quad (5.33)$$

$$= -\frac{v_s^3}{4} \frac{G_s^3}{Y_i^3(\omega_1)Y_o(3\omega_1)} \left[c_{(3,0,0)} - c_{(2,1,0)}A_v + c_{(1,2,0)}A_v^2 - c_{(0,3,0)}A_v^3 + c_{(1,1,0)}^2 \frac{A_v}{Y_o(2w_1)} - \frac{c_{(1,1,0)} c_{(2,0,0)}}{Y_o(2w_1)} + 2c_{(0,2,0)}^2 \frac{A_v^3}{Y_o(2w_1)} + 2c_{(0,2,0)} c_{(2,0,0)} \frac{A_v}{Y_o(2w_1)} - 3c_{(0,2,0)} c_{(1,1,0)} \frac{A_v^2}{Y_o(2w_1)} \right] \quad (5.34)$$

where $A_v = g_m/Y_o(\omega_1)$ was previously defined as the voltage gain of the transistor at ω_1 .

The expressions for the third-order nonlinear current source for $2\omega_1 - \omega_2$ are obtained from Table 3.3 to give

$$i_{NL3_{gm}} = \frac{3}{4}c_{(3,0,0)}V_{g,1,0}^2V_{g,0,-1} \quad (5.35)$$

$$i_{NL3_{go}} = c_{(2,0,0)}(V_{d,1,0}V_{d,1,-1} + V_{d,0,-1}V_{d,2,0}) + \frac{3}{4}c_{(0,3,0)}V_{d,1,0}^2V_{d,0,-1} \quad (5.36)$$

$$i_{NL3_{gm\&go}} = \frac{1}{2}c_{(1,1,0)}[V_{g,0,-1}V_{d,2,0} + V_{g,1,0}V_{d,1,-1}] + \frac{1}{4}c_{(2,1,0)}[2V_{g,0,-1}V_{g,1,0}V_{d,1,0} + V_{g,1,0}^2V_{d,0,-1}] + \frac{1}{4}c_{(1,2,0)}[2V_{g,1,0}V_{d,0,-1}V_{d,1,0} + V_{g,0,-1}V_{d,1,0}^2] \quad (5.37)$$

but are not expanded here.

The response of the intermodulation product at $2\omega_1 - \omega_2$ is similar to the third harmonic, where

$$\begin{aligned}
 V_{d,2,-1} \approx & -\frac{3}{4}v_s^3 \frac{G_S^3}{Y_i^3(\omega_1)Y_o(\omega_1)} \left[c_{(3,0,0)} - c_{(2,1,0)}A_v + c_{(1,2,0)}A_v^2 - c_{(0,3,0)}A_v^3 \right. \\
 & + c_{(1,1,0)}^2 A_v Z_\Sigma - c_{(1,1,0)} c_{(2,0,0)} Z_\Sigma + 2 c_{(0,2,0)}^2 A_v^3 Z_\Sigma \\
 & \left. + 2 c_{(0,2,0)} c_{(2,0,0)} A_v Z_\Sigma - 3 c_{(0,2,0)} c_{(1,1,0)} A_v^2 Z_\Sigma \right] \quad (5.38)
 \end{aligned}$$

assuming $|Y_i(-\omega_2)| \approx |Y_i(2\omega_1 - \omega_2)| \approx |Y_i(\omega_1)|$ and $|Y_o(-\omega_2)| \approx |Y_o(2\omega_1 - \omega_2)| \approx |Y_o(\omega_1)|$, which is often the case if $\omega_1 \approx \omega_2$. The term Z_Σ in (5.38) represents the sum of the output impedance at $2\omega_1$ and $\omega_1 - \omega_2$ with

$$Z_\Sigma = \frac{1}{3} \left(\frac{1}{Y_o(2\omega_1)} + \frac{2}{Y_o(\omega_1 - \omega_2)} \right) \quad (5.39)$$

From equations (5.34) and (5.38), the third harmonic distortion and the intermodulation distortion products depend on the device nonlinear coefficients, and the input and output termination admittances at the fundamental frequency, the second and third harmonic frequencies, as well as the difference frequencies. The expression contains terms arising from the third-order nonlinear coefficients, $c_{(3,0,0)}$, $c_{(0,3,0)}$, $c_{(2,1,0)}$ and $c_{(1,2,0)}$, as well as the second-order nonlinear coefficients, $c_{(2,0,0)}$, $c_{(0,2,0)}$, and $c_{(1,1,0)}$. The coefficient $c_{(3,0,0)}$ is generally the most dominant nonlinear coefficient but since the other coefficients are multiplied by the voltage gain, or in some cases by the square or the cube of A_v , the other terms in (5.34) and (5.38) may become significant if the voltage gain is large. Even though all second and third-order derivatives with respect to V_{DS} are zero since clm has been modelled as linear with V_{DS} , they are not, in reality, exactly zero. As a result, the terms with $c_{(0,3,0)}$, $c_{(1,2,0)}$, and $c_{(0,2,0)}$ could be significant if A_v is very large. With $c_{(0,3,0)}$, $c_{(1,2,0)}$, and $c_{(0,2,0)}$ set to zero though, (5.38) simplifies to

$$\begin{aligned}
 V_{d,2,-1} \approx & -\frac{3}{4}v_s^3 \frac{G_S^3}{Y_i^3(\omega_1)Y_o(\omega_1)} \left[c_{(3,0,0)} - c_{(2,1,0)}A_v \right. \\
 & \left. + c_{(1,1,0)}^2 A_v Z_\Sigma - c_{(1,1,0)} c_{(2,0,0)} Z_\Sigma \right] \quad (5.40)
 \end{aligned}$$

5.2.3 Influence of Load Resistance

Generally $c_{(3,0,0)}$, is the dominant term in (5.40) when the drain is a low impedance point. However, for large A_v (and R_L), the contributions from the other nonlinear coefficients in (5.40) become more significant. This is evident in Figure 5.4 which plots the magnitude of normalized intermodulation product, $|V_{d,2,-1}/V_{d,1,0}|$, versus R_L and the corresponding voltage gain, for a fixed V_{GS} and a fixed V_{DS} . A fixed V_{DS} implies that the power supply is varied to maintain a constant V_{DS} . Admittedly, this is a bit unrealistic for practical applications, but none-the-less, provides some useful insights. The plot shows that $c_{(3,0,0)}$ is the dominant factor for low R_L . As R_L and A_v increases, the other nonlinear terms become more significant. At realistic RF load impedances of a few hundred ohms, the nonlinearities due to the output conductance will dominate rather than the nonlinearities due to the transconductance.

Figure 5.4: Normalized third-order intermodulation product of the output voltage of a common source amplifier versus R_L and the corresponding A_v with $f = 0.01f_T$, $V_{GS} = 600\text{mV}$, $V_{DS} = 600\text{mV}$, $V_{SB} = 0\text{V}$, and using the parameters listed in Table 4.1.

The frequency and phase response of the terms in (5.40) are shown in Figure 5.5(a) and 5.5(b) with a load resistance of 500Ω . The contributions due to $c_{(1,1,0)}^2$ and $c_{(1,1,0)} \cdot c_{(2,0,0)}$ (terms 3 and 4) dominate for frequencies well below the cut-off

frequency, f_T , where

$$f_T = \frac{g_m}{2\pi C_{gs}} \quad (5.41)$$

Term 4 is approximately 180 degrees out of phase with term 3, but since term 4 is subtracted from term 3, the two terms sum together. For frequencies greater than $0.1f_T$, the contributions due to $c_{(1,1,0)}$ and $c_{(2,1,0)}$ (terms 2 and 3) become less significant since their amplitudes roll off at a faster rate due to the reduced voltage gain at higher frequencies. The contribution due to $c_{(1,1,0)} \cdot c_{(2,0,0)}$ (term 4) starts to roll off sooner than $c_{(3,0,0)}$ due to the effect of Z_Σ and its contribution is reduced somewhat but still plays a significant roll.

Figure 5.5: (a) Amplitude and (b) phase response of the normalized $IM3$ versus f/f_T with $V_{GS} = 600\text{mV}$, $V_{DS} = 600\text{mV}$, $V_{SB} = 0\text{V}$, $R_L = 500\Omega$ and with the parameters listed in Table 4.1.

The load resistance also influences the nonlinear behaviour around the sweet spot. Since ideally $c_{(3,0,0)} = 0$ at the sweet spot, the other terms in (5.40) will

obviously dominate. However, in Figure 5.6, which plots the different contributions in (5.40) as a function of V_{GS} with R_L fixed at 100Ω , the sweet spot still exists. It has only been slightly shifted to a lower V_{GS} . The shift occurs because as $c_{(3,0,0)}$ approaches zero, its magnitude approaches the opposite magnitude of the sum of the other terms and they all cancel. Since the other terms in (5.40) depend on R_L , the location of the linearity sweet spot can be expected to be dependent on R_L . After $c_{(3,0,0)}$ passes through zero it may become the dominant term again until the gain, which is proportional to V_{GS} , increases to the point that the other terms become dominant.

Figure 5.6: Normalized third-order intermodulation product of the output voltage of a common source amplifier versus V_{GS} with $R_L = 100\Omega$, $V_{DS} = 600\text{mV}$, $V_{SB} = 0\text{V}$, and $v_s = 1\text{V}$. The parameters listed in Table 4.1 are used for determining the nonlinear coefficients.

Figure 5.7 shows $IIP3$ versus V_{GS} for different values of R_L . In all cases the peak shifts to lower values of V_{GS} as R_L increases. Furthermore, the peak value decreases with increasing R_L . For realistic RF loads, however, the peak corresponds closely to the zero crossing of $c_{(3,0,0)}$. In addition to the shift in the peak location, $IIP3$ decreases with increasing R_L for V_{GS} higher than the location of the sweet spot.

Figure 5.7: $IIP3$ of a common source amplifier versus V_{GS} for different R_L with $V_{DS} = 600\text{mV}$, $V_{SB} = 0\text{V}$, and using the parameters listed in Table 4.1 .

5.2.4 Influence of a Tuned Load

Most often, a tuned load, rather than a resistive load, is used for RF applications to filter the desired signal, increase the voltage head room for biasing and to allow for increased signal swing. The tuned load, such as shown in Figure 5.8, consists of a parallel LC network tuned to the desired frequency of operation. For frequencies above 1 GHz, the inductor and capacitor are generally integrated on-chip, however, the quality factor of integrated inductors are typically below ten. This translates into an equivalent parallel resistor, R_p of a “few” hundred ohms for typical sized integrated inductors and will limit the maximum power transfer to the “real” load.

A tuned load also affects the nonlinear response of the amplifier. Equations (5.38), (5.39), and (5.40) are still valid except $Y_o(\omega)$ is modified to be

$$Y_o(\omega) = G_L + g_o + G_p + j\omega(C_L + C_{db}) - j\frac{1}{\omega L_o} \quad (5.42)$$

$$= G'_L + j\omega C'_L - j\frac{1}{\omega L_o} \quad (5.43)$$

Figure 5.8: A common source amplifier with a tuned load.

where $G_p = 1/R_p$.

Figure 5.9 shows the contribution of the terms in (5.40) versus R_L at f_0 for an amplifier with a tuned load and contrasts sharply to Figure 5.4. Figure 5.9 shows that only term 2 becomes dominant as R_L increases. However, the $c_{(2,1,0)}$ term is in opposite phase with $c_{(3,0,0)}$ (see equation (5.40)) and, therefore cancels with $c_{(3,0,0)}$ as its magnitude increases with R_L . Hence, the total *IM3* reduces as R_L increases until $R_L \approx 800\Omega$. The magnitude is ultimately limited by the contribution from $c_{(1,1,0)}^2$ (term 3) and from $c_{(1,1,0)} \cdot c_{(2,0,0)}$ (term 4), which are filtered significantly at f_0 , since they depend on the response at $2f_0$.

The contribution of all the terms in (5.38) or (5.40), except for $c_{(3,0,0)}$, are reduced due to the filtering of the tuned load. As a result, the voltage gain of the amplifier is reduced at all frequencies other than f_0 . Furthermore, the contribution from the terms containing Z_Σ are reduced even further, even at f_0 , since Z_Σ is dependent on the response at $2\omega_1$ and $\omega_1 - \omega_2$. This is evident in Figure 5.10 which plots the contribution of the various terms in (5.40) versus frequency for an amplifier with a tuned load centred at 1 GHz. The total is dominated by the contribution from $c_{(3,0,0)}$ at all frequencies, except $f_0/2$ and f_0 . At $f_0/2$, the contribution by $c_{(1,1,0)} \cdot c_{(2,0,0)}$ (term 4) becomes somewhat significant due to the response of Z_Σ at 2ω . However, at the desired frequency of f_0 , only the contribution from $c_{(3,0,0)}$ and $c_{(2,1,0)}$ (term 2) are dominant. Figure 5.10 also shows the effect of de-queing the inductor with a lower load resistance of 100Ω , which reduces the contribution from all terms even further, as indicated in Figure 5.9.

Figure 5.9: Normalized third-order intermodulation product of the output voltage of a common source amplifier with a tuned load centred at 1 GHz versus R_L with $V_{GS} = 600\text{mV}$, $V_{DS} = 600\text{mV}$, $V_{SB} = 0\text{V}$, and using the parameters listed in Table 4.1.

Figure 5.10: Amplitude response of the normalized $IM3$ versus frequency with $V_{GS} = 600\text{mV}$, $V_{DS} = 600\text{mV}$, $V_{SB} = 0\text{V}$, $R_L = 100$ and 500Ω and with the parameters listed in Table 4.1.

5.2.5 Influence of C_{gd}

In a single transistor amplifier, the gate-drain capacitor, C_{gd} , gives rise to the Miller effect and can have a significant effect on nonlinear behaviour of the amplifier. C_{gd} actually constitutes a shunt-shunt feedback, which, from Section 3.5, generally lowers distortion in an amplifier. However, as will be seen, this is not necessarily true in the vicinity of the sweet spot.

Figure 5.11: Equivalent circuits of the common source amplifier of Figure 5.1 including the influence of C_{gd} for determining (a) the fundamental response, (b) the second-order response, and (c) the third-order response.

The equivalent AC circuit for determining the first, second and third-order responses is shown in Figure 5.11 and is the same as Figure 5.3 except for C_{gd} between the gate and drain nodes. In the case of a MOSFET, C_{gd} can be considered a constant when operating in the saturation region. The intrinsic portion of C_{gd} is zero but the extrinsic portion of C_{gd} is comprised of the overlap capacitance between the gate and

drain, which is constant with respect to bias voltages. Based on typical parameters from the BSIM3 model ($C_{GD0} = 42\text{fF}/\mu\text{m}$ of gate width) and studies such as [54], typical values of C_{gd} vary between $0.1 C_{ox}$ and $0.3 C_{ox}$.

With the inclusion of C_{gd} , the transistor cutoff frequency, f_T is now

$$f_T = \frac{g_m}{2\pi(C_{gs} + C_{gd})} = \frac{g_m}{2\pi C_i} \quad (5.44)$$

With the parameters listed in Table 4.1, and with $V_{GS} = 600\text{mV}$, $V_{DS} = 600\text{mV}$, and $C_{gd} = 0.25C_{ox}$, the cutoff frequency is $f_T = 42.4\text{GHz}$.

Fundamental Responses

The fundamental input and output response of Figure 5.11(a) are modified from Section 5.2.2 with

$$V_{g,1,0} = -\frac{G_S Y_o(\omega_1)}{\det(\omega_1)} v_s \quad (5.45)$$

$$V_{d,1,0} = -\frac{G_S (g_m - Y_f(\omega_1))}{\det(\omega_1)} v_s \quad (5.46)$$

where

$$Y_i(\omega) = G_S + j\omega(C_{gs} + C_{gd}) \quad (5.47)$$

$$= G_S + j\omega C_i \quad (5.48)$$

$$Y_o(\omega) = G_L + g_o + j\omega(C_L + C_{db} + C_{gd}) \quad (5.49)$$

$$= G'_L + j\omega C_o \quad (5.50)$$

$$Y_f(\omega) = j\omega C_{gd} \quad (5.51)$$

and

$$\det(\omega) = Y_i(\omega)Y_o(\omega) + g_m Y_f(\omega) - Y_f^2(\omega) \quad (5.52)$$

$$= G_S G'_L + \omega^2 (C_i C_o - C_{gd}^2) + j\omega (C_i G'_L + C_o G_S + C_{gd} g_m) \quad (5.53)$$

is the determinant of the admittance matrix. If $C_{gd} = 0$ then equations (5.45), (5.46), (5.47), and (5.49), reduce to (5.4), (5.5), (5.6), and (5.7) respectively.

The voltage gain from the gate input to the drain becomes

$$A_v = \frac{V_{d,1,0}}{V_{g,1,0}} = -\frac{g_m - j\omega_1 C_{gd}}{Y_o(\omega_1)} \quad (5.54)$$

which reduces to $g_m R'_L$ for low frequencies.

Figure 5.12: Shunt-shunt feedback configuration using y-parameter representation [45].

Since C_{gd} constitutes a shunt-shunt feedback, it is useful to look at the closed loop gain expression in terms of the open loop gain, $H(\omega)$, and the approximate inverse loop gain, $R(\omega)$, as done in Section 3.5. The equivalent circuit in Figure 5.11(a) is redrawn in Figure 5.12 as a two-port representation with a feedback network consisting of C_{gd} only. The y-parameters for the feedback network are

$$Y_{11f} = \left. \frac{i_1}{v_1} \right|_{v_2=0} = j\omega C_{gd} \quad (5.55)$$

$$Y_{12f} = \left. \frac{i_1}{v_2} \right|_{v_1=0} = -j\omega C_{gd} \quad (5.56)$$

$$Y_{21f} = \left. \frac{i_2}{v_1} \right|_{v_2=0} = -j\omega C_{gd} \quad (5.57)$$

$$Y_{22f} = \left. \frac{i_2}{v_2} \right|_{v_1=0} = j\omega C_{gd} \quad (5.58)$$

Following the method in [45], the *closed loop gain* of the feedback configuration in Figure 5.12 is given by

$$\frac{V_{d,1,0}}{v_s} = \frac{G_S H_1(\omega)}{1 + H_1(\omega) F(\omega)} \quad (5.59)$$

$H_1(\omega)$ is the first-order *open loop* gain and $F(\omega)$ is the feedback gain and are given by

$$H_1(\omega) = -\frac{g_m - Y_f(\omega)}{Y_i(\omega) Y_o(\omega)} \quad (5.60)$$

$$F(\omega) = Y_{12f} = -Y_f(\omega) = -j\omega C_{gd} \quad (5.61)$$

The *loop gain*, $T(\omega)$, is defined as the total gain around the feedback loop [45] with

$$T(\omega) = H_1(\omega) F(\omega) \quad (5.62)$$

Equating (5.59) to (5.46) yields

$$R(\omega) = \frac{1}{1 + T(\omega)} = \frac{Y_i(\omega) Y_o(\omega)}{\det(\omega)} \quad (5.63)$$

which is approximately the inverse of the loop gain, especially if $T(\omega) \gg 1$. It will be shown later that the third-order intermodulation product will be reduced by $R(\omega)$.

The final block diagram representation of the amplifier is shown in Figure 5.13.

Figure 5.13: Final block diagram representation of the overall feedback system.

Figure 5.14: Frequency response of $T(\omega)$ for different values of C_{gd} and R_L with g_o omitted, $V_{GS} = 600\text{mV}$, $V_{DS} = 600\text{mV}$, and with the parameters listed in Table 4.1.

From (5.60) and (5.61), if $g_m \gg |Y_f(\omega)|$, then the loop gain can be approximated as

$$T(\omega) \approx \frac{j\omega g_m R'_L R_S C_{gd}}{(1 + j\omega R_S C_i)(1 + j\omega R'_L C_o)} \quad (5.64)$$

where the loop gain is directly proportional to g_m , R'_L , R_S and C_{gd} . For large R_L though, the loop gain will be shunted by r_o . Figure 5.14 plots the loop gain for various R_L , with r_o omitted, and various C_{gd} . The loop gain is low for low frequencies since C_{gd} is a high impedance and does not couple the drain node to the gate very strongly. As the frequency increases, however, so does $T(\omega)$ until a maximum value at the first pole frequency

$$\frac{f_{p1}}{f_T} = \frac{1}{g_m R'_L} \frac{C_i}{C_o} \quad (5.65)$$

However, as mentioned earlier, if the output conductance is included, the pole frequency will be limited by r_o as R_L increases. The loop gain then remains constant

until the second pole frequency at

$$\frac{f_{p2}}{f_T} = \frac{1}{g_m R_S} \quad (5.66)$$

after which, the loop gain decreases with increasing frequency until the zero frequency at

$$\frac{f_z}{f_T} = \frac{C_{gs} + C_{gd}}{C_{gd}} \quad (5.67)$$

The loop gain is not high for this particular case and, as will be shown, does not impact dramatically on the distortion products. However, in other cases, such as source degeneration, the loop gain can be significantly greater and play a crucial role in suppressing distortion products, as will be seen in Section 5.3.

Third-order Intermodulation Response

The second and third-order responses are determined in the same manner as in Section 5.2.2, except the equivalent circuit in Figures 5.11(b) and (c) are used instead. The second-order responses are not derived here but it should be noted that the second-order responses at the gate, $V_{g,2,0}$ and $V_{g,1,-1}$, are not zero in this case due to the feedback from C_{gd} . The third harmonic distortion expression is also not derived here; only the third-order intermodulation product is considered. As seen previously, the expressions for third harmonic distortion and the third-order intermodulation product are very similar and the conclusions apply equally to both cases.

Since $V_{g,2,0}$ and $V_{g,1,-1}$ are not zero, the expressions for the third-order nonlinear current source for $2\omega_1 - \omega_2$, obtained from Table 3.3, are

$$i_{NL3_{gm}} = c_{(2,0,0)}(V_{g,1,0}V_{g,1,-1} + V_{g,0,-1}V_{g,2,0}) + \frac{3}{4}c_{(3,0,0)}V_{g,1,0}^2V_{g,0,-1} \quad (5.68)$$

$$i_{NL3_{go}} = c_{(2,0,0)}(V_{d,1,0}V_{d,1,-1} + V_{d,0,-1}V_{d,2,0}) + \frac{3}{4}c_{(0,3,0)}V_{d,1,0}^2V_{d,0,-1} \quad (5.69)$$

$$\begin{aligned} i_{NL3_{gm\&go}} = & \frac{1}{2}c_{(1,1,0)}[V_{g,0,-1}V_{d,2,0} + V_{g,1,0}V_{d,1,-1} + V_{g,1,-1}V_{d,1,0} + V_{g,2,0}V_{d,0,-1}] \\ & + \frac{1}{4}c_{(2,1,0)}[2V_{g,0,-1}V_{g,1,0}V_{d,1,0} + V_{g,1,0}^2V_{d,0,-1}] \\ & + \frac{1}{4}c_{(1,2,0)}[2V_{g,1,0}V_{d,0,-1}V_{d,1,0} + V_{g,0,-1}V_{d,1,0}^2] \end{aligned} \quad (5.70)$$

but are not expanded here.

The final expression for the third-order intermodulation product is given by the approximation

$$V_{d,2,-1} \approx -\frac{3}{4}v_s^3 \frac{G_S^3 Y_i(\omega_1) Y_o^3(\omega_1)}{\det^4(\omega_1)} \left[c_{(3,0,0)} - c_{(2,1,0)} A_v + c_{(1,1,0)}^2 A_v Z_\Sigma - c_{(1,1,0)} c_{(2,0,0)} Z_\Sigma - c_{(2,0,0)}^2 \frac{2 Y_f(2\omega_1)}{3 \det(2\omega_1)} \right] \quad (5.71)$$

where

$$Z_\Sigma = \frac{1}{3} \left(\frac{A_v Y_f(2\omega_1)}{\det(2\omega_1)} + \frac{Y_i(2\omega_1)}{\det(2\omega_1)} + \frac{2Y_i(\omega_1 - \omega_2)}{\det(\omega_1 - \omega_2)} \right) \quad (5.72)$$

assuming $|Y_i(-\omega_2)| \approx |Y_i(2\omega_1 - \omega_2)| \approx |Y_i(\omega_1)|$ and $|Y_o(-\omega_2)| \approx |Y_o(2\omega_1 - \omega_2)| \approx |Y_o(\omega_1)|$ as before. Numerical evaluation shows that the first term in Z_Σ is significantly less than the other two terms, even for large A_v , and can be approximated by

$$Z_\Sigma \approx \frac{1}{3} \left(\frac{Y_i(2\omega_1)}{\det(2\omega_1)} + \frac{2Y_i(\omega_1 - \omega_2)}{\det(\omega_1 - \omega_2)} \right) \quad (5.73)$$

Equation (5.71) is rewritten in terms of the two-port network parameter in which

$$V_{d,2,-1} \approx -\frac{3}{4}v_s^3 \frac{G_S^3}{Y_i^3(\omega_1) Y_o(\omega_1)} R^4(\omega_1) \left[c_{(3,0,0)} - c_{(2,1,0)} A_v + c_{(1,1,0)}^2 A_v Z_\Sigma - c_{(1,1,0)} c_{(2,0,0)} Z_\Sigma - j\omega_1 C_{gd} \frac{4}{3} \frac{c_{(2,0,0)}^2}{Y_i(2\omega_1) Y_o(2\omega_1)} R(2\omega_1) \right] \quad (5.74)$$

and

$$Z_\Sigma = \frac{1}{3} \left(\frac{R(2\omega_1)}{Y_o(2\omega_1)} + \frac{2R(\omega_1 - \omega_2)}{Y_o(\omega_1 - \omega_2)} \right) \quad (5.75)$$

Comparing (5.74) to (5.40) shows several effects that feedback has on the third order distortion components.

1. The overall magnitude of $V_{d,2,-1}$ is reduced by a factor of $R^4(\omega_1)$ from the case with no C_{gd} , since $R(\omega_1)$ is inversely proportional to $1 + T(\omega_1)$. If $T(\omega_1) \gg 1$, $IIP3$ can be expected to improve for high frequencies since C_{gd} becomes a low impedance and strongly couples the output and input nodes.

2. The terms in Z_Σ are reduced by a factor of $R(2\omega_1)$ for the first term and by $R(\omega_1 - \omega_2)$ for the last term.
3. An additional term is introduced that depends solely on second-order nonlinearities, namely, $c_{(2,0,0)}$, the feedback gain at the second harmonic, and the input and output admittances at the second harmonic. The second-order nonlinearities manifest as third-order nonlinearities through the feedback system. Second-order tones that are generated by second-order nonlinearities are fed back to mix with the fundamental tones at the input thereby creating additional third-order distortion.

The last term in (5.74) can have a significant impact on the linearity sweet spot depending on C_{gs} , $R(2\omega_1)$ and the frequency of the input tones. As discussed in Section 5.2.3, $c_{(3,0,0)}$ is the dominate source of the third-order distortion for low A_v . The second, third and fourth terms in (5.74) cause a slight shift in the location of the sweet spot due to cancelling of the terms. The last term in (5.74) does not cause a shift in the location of the sweet spot but reduces the peak magnitude of the sweet spot. This can be seen in Figure 5.15 which shows $IIP3$ at the sweet spot for different values of C_{gs} at a fixed frequency. The same effect occurs for a fixed C_{gd} and different values of f . Hence, feedback degrades the third-order distortion at the sweet spot.

An alternative view is to consider the simple case of a memoryless, linear feedback system discussed in Section 3.5. From equation (3.37), repeated here for convenience, the third-order transfer function for such a system is

$$G_3 = \left[H_3 - \frac{2H_2^2}{H_1} \right] \frac{1}{T^4} \quad (5.76)$$

where it can be seen that the third-order distortion created by H_3 is reduced by H_2 . But if $H_3 = 0$, as is the case at the sweet spot, then there would be an apparent degradation to the third-order distortion. This apparent degradation may seem to contradict the well known theory that feedback improves linearity, but the degradation is only applicable at the sweet spot.

Figure 5.15: $IIP3$ of a common source amplifier versus V_{GS} for different C_{gd} with $f_{in} = 1$ GHz, $V_{DS} = 600$ mV, $V_{SB} = 0$ V, $R_L = 100 \Omega$ and using the parameters listed in Table 4.1 .

5.2.6 Influence of a Nonlinear C_{gs} and C_{db}

Up till now, the analysis and results presented have not included the effects of any nonlinear capacitors and so this section will investigate the influence of the nonlinear capacitors C_{gs} and C_{db} . From section 4.5, the intrinsic gate-source capacitance, $C_{gs,i}$, is the dominant intrinsic capacitor whereas $C_{db,e}$ is the dominate extrinsic capacitor. $C_{gd,e}$ is not considered since it is constant and $C_{sb,e}$ is not considered since the source and bulk are assumed to be tied to ground.

The capacitors $C_{gs,i}$ and $C_{db,e}$ influence the nonlinear behaviour of the common-source amplifier in three ways:

1. they are additional nonlinear elements;
2. even if they were linear, their presence affects the values of the nonlinear current sources and the transfer function from the current source to the output, as already seen in (5.40); and,

3. the addition of a nonlinear element and current source at the gate node results in non-zero second and third-order response at the gate, which ultimately results in more nonlinear terms in the final expression for second and third-order harmonic and intermodulation distortion products.

The equivalent AC circuit under consideration is shown in Figure 5.16(a) along with the equivalent circuits for calculating the first, second and third-order responses. The influence of g_m , C_{gs} , and C_{db} only are considered and the influence of velocity saturation and g_o are neglected. In other words, mob_{hot} and clm are assumed to be unity. This keeps the analysis from becoming too intractable, otherwise the final expression for $IM3$ can have over 20 terms. Furthermore, the frequency behaviour is examined beyond the range of validity for quasi-static approximation. Nevertheless, the analysis beyond frequencies higher than $0.1f_o$ provides some useful insights into the high frequency behaviour of the MOSFET.

Fundamental Response

The equivalent circuit for determining the fundamental response, shown in Figure 5.16(b), is the same as Figure 5.3(a) except C_{db} is included but g_o is not, since clm is assumed to be unity. The first-order responses $V_{g,1,0}$, $V_{d,1,0}$, $V_{g,0,-1}$, $V_{d,0,-1}$ are the same as (5.4), (5.5), (5.9), and (5.10) except the output admittance $Y_o(\omega)$ is

$$Y_o(\omega) = G_L + j\omega(C_{db} + C_L) \quad (5.77)$$

Third-order Intermodulation Response

The second and third-order responses are determined in the same manner as previous sections, except the equivalent circuit in Figure 5.16(c) and (d) are used instead. The equivalent circuits differ from Figure 5.3(b) and (c) with the nonlinear current sources, $i_{NL2_{cgs}}$ and $i_{NL3_{cgs}}$ placed in parallel with C_{gs} . As a result of these nonlinear current sources, the response at $V_{g,2,0}$ and $V_{g,1,-1}$ will not be zero and must be accounted for when calculating $V_{d,3,0}$ or $V_{d,2,-1}$. A nonlinear current source is also placed in parallel with C_{db} and is included as a part of $i_{NL2_{ids}}$ and $i_{NL3_{ids}}$ where $i_{NL2_{ids}} = i_{NL2_{gm}} + i_{NL2_{cdb}}$ and $i_{NL3_{ids}} = i_{NL3_{gm}} + i_{NL3_{cdb}}$. The second-order responses and the third harmonic

Figure 5.16: (a) Equivalent AC circuit of the common source amplifier of Figure 5.1 including the influence of a nonlinear C_{gs} and the equivalent circuits for determining (b) the fundamental response, (c) the second-order response, and (d) the third-order response.

distortion expression are not derived here. Only the third-order intermodulation product is considered.

The third-order response at $2\omega_1 - \omega_2$ is solved in a similar manner as in Section 5.2.2 except the equivalent circuit in Figure 5.16(d) is used to determine the expressions. The expressions for the nonlinear current sources $i_{NL3_{gm}}$, $i_{NL3_{cgs}}$ and $i_{NL3_{cgs}}$ are obtained from Table 3.4. The basic expression for $i_{NL3_{gm}}$ is

$$i_{NL3_{gm}} = c_{(2,0,0)} V_{g,1,0} V_{g,1,-1} + c_{(2,0,0)} V_{g,0,-1} V_{g,2,0} + \frac{3}{4} c_{(3,0,0)} V_{g,1,0}^2 V_{g,0,-1} \quad (5.78)$$

and the expressions for $i_{NL3_{cgs}}$ and $i_{NL3_{cgs}}$ are

$$i_{NL3_{cgs}} = j(2\omega_1 - \omega_2) \left[c_{2cgs} V_{d,1,0} V_{d,1,-1} + c_{2cgs} V_{d,0,-1} V_{d,2,0} + \frac{3}{4} c_{3cgs} V_{d,1,0}^2 V_{d,0,-1} \right] \quad (5.79)$$

$$i_{NL3_{cgs}} = j(2\omega_1 - \omega_2) \left[c_{2cgs} V_{g,1,0} V_{g,1,-1} + c_{2cgs} V_{g,0,-1} V_{g,2,0} + \frac{3}{4} c_{3cgs} V_{g,1,0}^2 V_{g,0,-1} \right] \quad (5.80)$$

The final expression for $IM3$ is

$$V_{d,2,-1} = \frac{g_m i_{NL3_{cgs}}(2\omega_1 - \omega_2) - Y_i(2\omega_1 - \omega_2) i_{NL3_{ids}}(2\omega_1 - \omega_2)}{Y_i(2\omega_1 - \omega_2) Y_o(2\omega_1 - \omega_2)} \quad (5.81)$$

If it assumed that $\omega_1 \approx \omega_2$ then (5.81) simplifies to

$$V_{d,2,-1} = -\frac{3}{4} v_s^3 \frac{G_S^3}{Y_i^3(\omega_1) Y_o(\omega_1)} \left[c_{(3,0,0)} - \frac{4}{3} \frac{\omega_1^2 c_{2cgs}^2 g_m}{Y_i(\omega_1) Y_i(2\omega_1)} - j\omega_1 \frac{4}{3} \frac{c_{(2,0,0)} c_{2cgs}}{Y_i(2\omega_1)} - j\omega_1 \frac{g_m c_{3cgs}}{Y_i(\omega_1)} - j\omega_1 \frac{g_m^3 c_{3cdb}}{Y_o^3(\omega_1)} + j\omega_1 \frac{4}{3} \frac{g_m c_{(2,0,0)} c_{2cdb}}{Y_o(0) Y_o(\omega_1)} - \frac{4}{3} \frac{\omega_1^2 g_m^3 c_{2cdb}}{Y_o(2\omega_1) Y_o^3(\omega_1)} + j\omega_1 \frac{2}{3} \frac{g_m c_{(2,0,0)} c_{2cdb}}{Y_o(2\omega_1) Y_o(\omega_1)} + \frac{4}{3} \frac{\omega_1^2 g_m^2 c_{2cgs} c_{2cdb}}{Y_i(2\omega_1) Y_o(2\omega_1) Y_o(\omega_1)} \right] \quad (5.82)$$

Numerical analysis shows that the first, third, fourth, fifth, and sixth terms of (5.82) are dominant. Hence, the third-order intermodulation product is primarily dependent on the third-order nonlinear coefficients $c_{(3,0,0)}$, c_{3cgs} , and c_{3cdb} as well as the second-order nonlinear coefficients $c_{(2,0,0)}$, c_{2cgs} , and c_{2cdb} . The response of the input and output admittances at DC (term 6 of (5.82)), ω_1 and $2\omega_1$ are of significance. Furthermore, $IM3$ is also influenced by the voltage gain $A_v = g_m/Y_o(\omega_1)$, the load resistance R_L , and the frequency of operation, since several terms are directly proportional to ω_1 .

Figure 5.17: The normalized $IM3$ and the dominant terms of (5.82) versus f/f_T with $V_{GS} = 600\text{mV}$, $V_{DS} = 600\text{mV}$, $V_{SB} = 0\text{V}$, $mob_{hot} = 1$, $clm = 1$ and using the parameters listed in Table 4.1.

For low frequencies and low voltage gain, $c_{(3,0,0)}$ is the dominant source for the third-order distortion and is constant versus frequency until the pole frequency

$$f_1 = \frac{1}{2\pi R_s C_{gs}} \approx \frac{f_T}{g_m R_s} \quad (5.83)$$

after which the response for the first term rolls off at 20 dB per decade, and is shown in Figure 5.17. The pole frequency was estimated to be approximately $0.57f_T$ ($g_m \approx 35$

mS at $V_{gs} = 600$ mV and $R_S = 50\Omega$). As the frequency of operation increases, the third and fourth terms of (5.82) become more significant. However, due to the phase difference of the terms, some cancellation occurs. From Figure 5.17 the other terms are initially very small for low frequencies but increase at 10 dB per decade until the pole frequency f_1 , after which all terms roll off at 20 dB per decade. The fifth term, however, rolls off at a faster rate since it is proportional to A_v^3 . Although the fifth and sixth terms are not significant for all frequencies for this particular case, they can be significant if the gain is high.

At the sweet spot, the response of the dominant terms of (5.82), plotted in Figure 5.18, is different from that plotted in Figure 5.17. For very low frequencies, the first term is still the dominant term even though $c_{(3,0,0)}$ is much smaller than in Figure 5.17. However, for $f > 0.01f_T$ the third and fourth terms become dominant and $IM3$ increases until $f \approx 0.2f_T$, after which the magnitude decreases. Again though, due to the phase difference between the two terms, they add as vectors and the total is less than the individual terms. Hence, for low frequencies where $f \ll f_T$ and low voltage gain, $IIP3$ and the sweet spot are not affected by the nonlinear C_{gs} and C_{db} . As the frequency increases though, the peak magnitude of the sweet spot will be significantly reduced. Figure 5.19 plots $IIP3$ versus V_{gs} for different f/f_T ratios. It shows that although the sweet spot is reduced for $f = 0.01f_T$ and $f = 0.1f_T$ it is still very pronounced. However, for $f = f_T$, the sweet spot has disappeared completely and, furthermore, $IIP3$ has increased. This, however, is due to the reduced input signal levels caused by the filtering by $R_S C_{gs}$ at high frequencies.

In Figure 5.17 and 5.18, the fifth and sixth terms are not dominant. However, they do become dominant as the voltage gain increases. This can be seen in Figure 5.20 which plots the dominant terms of (5.82) versus R_L and the corresponding voltage gain with $f = 0.01f_T$. For high voltage gain, both the fifth and sixth terms dominate. A similar type behaviour as shown in Figure 5.19 can also be expected for different R_L . As R_L increases the peak at the sweet spot should decrease and shift to lower bias voltages. This is shown in Figure 5.21. The plot also shows that as R_L increases, the sweet spot becomes less pronounced and $IIP3$ also decreases elsewhere. This occurs because the voltage gain is also proportional to V_{GS} and so the fifth and sixth terms of (5.82) become dominant at higher V_{GS} as well as higher R_L and, hence, $IM3$ increases and $IIP3$ decreases.

Figure 5.18: The normalized $IM3$ and the dominant terms of (5.82) versus f/f_T when biased close to the sweet spot with $V_{GS} = 700\text{mV}$, $V_{SB} = 0\text{V}$, $mob_{hot} = 1$, $clm = 1$ and using the parameters listed in Table 4.1.

Figure 5.19: $IIP3$ of a common source amplifier with nonlinear C_{gs} versus V_{GS} for different f/f_T with $V_{SB} = 0\text{V}$, $mob_{hot} = 1$, $clm = 1$ and using the parameters listed in Table 4.1.

Figure 5.20: The normalized $IM3$ and the dominant terms of (5.82) versus R_L and the corresponding A_v with $f = 0.01f_T$, $R_S = 50\Omega$, $C_L = 75\text{fF}$, $V_{GS} = 700\text{mV}$, $V_{SB} = 0\text{V}$, $mob_{hot} = 1$, $clm = 1$ and using the parameters listed in Table 4.1.

Figure 5.21: $IIP3$ of a common source amplifier with nonlinear C_{gs} versus V_{GS} for different R_L with $f = 0.01f_T$, $R_S = 50\Omega$, $C_L = 75\text{fF}$, $V_{SB} = 0\text{V}$, $mob_{hot} = 1$, $clm = 1$ and using the parameters listed in Table 4.1.

5.3 Source Degenerated Amplifier

Source degeneration is often used in RF circuits for circuit linearization and, especially in RF CMOS circuits, for input impedance matching. In most cases, inductors, rather than resistors, are used for degeneration in RF applications. An amplifier with inductive source degeneration is shown in Figure 5.22. The degeneration inductor has several effects:

1. it reduces the effective RF transconductance of the amplifier, thereby reducing the gain;
2. it allows for more DC voltage headroom than a degeneration resistor since there is very little DC drop across the degeneration inductor;
3. it transforms a portion of the input impedance, which is largely capacitive, into a real portion that can be used for input matching; and,
4. it causes feedback by subtracting the AC voltage across the inductor, which is proportional to the output current, directly from the input voltage [45].

Since degeneration actually constitutes feedback, it generally improves the linearity. However, as in the case with C_{gd} , it will be shown that degeneration degrades the linearity in the vicinity of the sweet spot, again contradicting the well known theory.

Figure 5.22: A common-source amplifier with inductive source degeneration.

One of the primary reasons for using inductive degeneration in RF CMOS circuits is to use the inductor to assist with input matching of the amplifier. The

input impedance of a MOSFET is inherently capacitive making it nearly impossible to match to a resistive source impedance, R_s , for maximum power transfer. The input impedance of a common-source amplifier with inductive source degeneration is

$$Z_{in} = \frac{g_m L_s}{C_{gs}} + j \left(\omega L_s - \frac{1}{\omega C_{gs}} \right) \quad (5.84)$$

where L_s is the source degeneration inductor. The degeneration inductor is transformed at the input to create a resistive portion to the input impedance where practically none existed before. Thus, by proper selection of g_m , C_{gs} , and L_s , the input of the amplifier can be matched to R_s . In addition, an inductor in series with the gate is often used to fully cancel the remainder of the gate-source capacitance from the $j(\omega L_s - \frac{1}{\omega C_{gs}})$ term in (5.84).

5.3.1 First-Order Response

The equivalent AC circuit for determining the first-order responses is shown in Figure 5.23. An ideal inductor is assumed for the source degeneration. It is also assumed that body effect is not present. The nonlinearities due to C_{gs} , C_{db} , and g_o will be neglected for simplicity sake and C_{gd} will also be neglected so that only the effects due to degeneration will be considered.

Figure 5.23: Equivalent AC circuit of a common-source amplifier with inductive source degeneration used for solving the first-order responses.

Source degeneration constitutes a series-series feedback system and so the first-order response at the output is examined in terms of the system parameters $H(\omega)$ and $R(\omega)$ as done in Section 3.5 and Section 5.2.5. The equivalent circuit in Figure

5.23 is redrawn in Figure 5.24 as a two-port representation with a feedback network consisting of L_S only. In this case, z -parameters are the most convenient parameters to consider. Unfortunately, y -parameters are the most convenient parameters to use for the nodal analysis to solve the third-order response later. As a result, z -parameters are first used to solve the feedback network and then converted later to y -parameters for the third-order response.

Figure 5.24: Series-series feedback configuration using z -parameter representation [45].

The z -parameters for the feedback network are

$$Z_{11f} = \left. \frac{v_1}{i_1} \right|_{i_2=0} = j\omega L_S \quad (5.85)$$

$$Z_{12f} = \left. \frac{v_1}{i_2} \right|_{i_1=0} = j\omega L_S \quad (5.86)$$

$$Z_{21f} = \left. \frac{v_2}{i_1} \right|_{i_2=0} = j\omega L_S \quad (5.87)$$

$$Z_{22f} = \left. \frac{v_2}{i_2} \right|_{i_1=0} = j\omega L_S \quad (5.88)$$

while the z -parameters for the transistor are

$$Z_{11a} = \frac{1}{j\omega C_{gs}} \quad (5.89)$$

$$Z_{12a} = 0 \quad (5.90)$$

$$Z_{21a} = \frac{g_m Z_o}{j\omega C_{gs}} \quad (5.91)$$

$$Z_{22a} = 0 \quad (5.92)$$

Following the method in [45], the closed loop gain of the series-series feedback configuration in Figure 5.24 is

$$\frac{v_o}{v_s} = \frac{Z_L(\omega)H_1(\omega)}{1 + H_1(\omega)F(\omega)} \quad (5.93)$$

where

$$H_1(\omega) = -\frac{Z_{21a}}{Z_i(\omega)Z_o(\omega)} \quad (5.94)$$

$$= -\frac{g_m \frac{1}{j\omega C_{gs}}}{Z_i(\omega)} \quad (5.95)$$

is the open loop gain, and

$$F(\omega) = Z_{12f} = j\omega L_S \quad (5.96)$$

is the feedback gain. Z_i and Z_o are given by

$$Z_i = Z_S + Z_{11a} + Z_{11f} \quad (5.97)$$

$$Z_o = Z_L + Z_{22a} + Z_{22f} \quad (5.98)$$

but are not the same as the input and output impedance of the amplifier

Equation (5.93) and (5.94) are rewritten in terms of admittances to yield

$$\frac{v_o}{v_s} = \frac{H_1(\omega)}{Y_o(\omega)(1 + H_1(\omega)F(\omega))} \quad (5.99)$$

where

$$H_1(\omega) = -\frac{g_m G_S}{Y_i(\omega)} \quad (5.100)$$

$$Y_i(\omega) = G_S(1 - \omega^2 L_S C_{gs}) + j\omega C_{gs} \quad (5.101)$$

$$Y_o(\omega) = G_L + j\omega C_L \quad (5.102)$$

If L_S approaches zero, then (5.99) becomes the same as (3.48).

Figure 5.25: Loop gain, $T(\omega)$, given by (5.103), versus f/f_T for different values of L_S and R_S .

The loop gain, $T(\omega)$, is given by

$$T(\omega) = \frac{j\omega g_m G_S L_S}{G_S(1 - \omega^2 L_S C_{gs}) + j\omega C_{gs}} \quad (5.103)$$

and is plotted in Figure 5.25. From (5.103), the loop gain is proportional to g_m , G_S , and L_S . For low frequencies, the gain is low due to the impedance of the degeneration inductor. As the frequency increases so does the gain until it reaches a peak at the

double pole frequency at approximately

$$\frac{f_{2p}}{f_T} \approx \frac{1}{g_m} \sqrt{\frac{C_{gs}}{L_S}} \quad (5.104)$$

In the case of source degeneration, the loop gain is also dependent on source impedance, rather than the load impedance. Figure 5.25 shows the effect of the source impedance on the loop gain as well. Unlike the loop gain with feedback due to C_{gd} , the loop gain due to degeneration can be quite high, and can reduce the third-order intermodulation distortion products significantly.

5.3.2 Third-Order Response

The equivalent AC circuit for determining the third-order responses is shown in Figure 5.26. Since only the nonlinearities due to g_m are being considered, the third-order nonlinear current source, $i_{NL3_{ids}}$ consists of only the nonlinear current $i_{NL3_{gm}}$, whose expression is obtained from Table 3.4 for third-harmonic distortion component and from Table 3.3 for the third-order intermodulation distortion product. Only $IM3$ will be considered here.

Figure 5.26: Equivalent AC circuit of a common-source amplifier with inductive source degeneration used for solving the third-order responses.

The final expression for $IM3$, in terms of the two-port network parameters, is given by the approximation

$$V_{d,2,-1} \approx -\frac{3}{4} v_s^3 \frac{G_S^3}{Y_i^3(\omega_1) Y_o(\omega_1)} R^4(\omega_1) \left[c_{(3,0,0)} + j 2\omega_1 L_S c_{(2,0,0)}^2 \frac{2 G_S R(2\omega_1)}{3 Y_i(2\omega_1)} \right] \quad (5.105)$$

assuming $|Y_i(-\omega_2)| \approx |Y_i(2\omega_1 - \omega_2)| \approx |Y_i(\omega_1)|$ and $|Y_o(-\omega_2)| \approx |Y_o(2\omega_1 - \omega_2)| \approx |Y_o(\omega_1)|$ as before and assuming the impedance of the degeneration inductor is negligible at $\omega_1 - \omega_2$. $R(\omega)$ is approximately the inverse of the loop gain.

From (5.105), it can be seen that the feedback caused by source degeneration inductor has the same effects as the feedback caused by C_{gd} .

1. The overall magnitude of $IM3$ is reduced by a factor of $R^4(\omega_1)$ from the case with no degeneration. For $T(\omega_1) \gg 1$ (see Figure 5.25), then $R(\omega_1) \approx 1/T(\omega_1)$ and $IIP3$ can be expected to improve for $f > 0.001f_T$, depending on the value of the source resistance and degeneration inductor.
2. An additional term is introduced that depends solely on second-order nonlinearities, and second-order responses. As in the case with C_{gd} , the second-order nonlinearities manifest as third-order nonlinearities through the feedback system. Second-order tones generated by second-order nonlinearities are fed back to mix with the fundamental tones at the input thereby creating additional third-order distortion.

Figure 5.27 shows the frequency response of each of the terms in (5.105) as well as the total. For low frequencies, the $c_{(3,0,0)}$ term dominates but the $c_{(2,0,0)}$ term increases with frequency. Both terms add as vectors and the total is greater than each term for higher frequencies. Although both terms add, the $IM3$ product is reduced due to the loop gain $T(\omega_1)$. Since the loop gain for a 5nH degeneration inductor is greater than 1, the $IM3$ product is significantly reduced for 5nH.

Figure 5.28 and 5.29 show the effect inductive degeneration has on the sweet spot. Degeneration can be seen to degrade the linearity performance in the vicinity of the sweet spot in a similar manner as C_{gd} . In the vicinity of the sweet spot $c_{(3,0,0)}$ approaches zero and so the nonlinearity due to $c_{(2,0,0)}$ dominates and $IIP3$ degrades. However, from Figure 5.28, $IIP3$ improves with increasing L_S beyond the sweet spot, as expected. Figure 5.29 also shows that $IIP3$ degrades in the vicinity of the sweet spot with increasing frequency for a fixed L_S . This occurs because the contribution from the $c_{(2,0,0)}$ term increases with increasing frequency, as already seen in Figure 5.27.

Figure 5.27: The frequency response of (5.105) versus f/f_T with $V_{GS} = 700\text{mV}$, $V_{DS} = 600\text{mV}$, $V_{SB} = 0\text{V}$, $R_S = 50\Omega$, $R_L = 100\Omega$, and using the parameters listed in Table 4.1.

Figure 5.28: $IIP3$ of a source degenerated common source amplifier versus V_{GS} for different L_S with $f_{in} = 1\text{GHz}$, $V_{DS} = 600\text{mV}$, $V_{SB} = 0\text{V}$, $R_L = 100\Omega$ and using the parameters listed in Table 4.1 .

Figure 5.29: $IIP3$ of a source degenerated common source amplifier versus V_{GS} for different frequencies with $L_S = 1\text{nH}$, $V_{DS} = 600\text{mV}$, $V_{SB} = 0\text{V}$, $R_L = 100\Omega$ and using the parameters listed in Table 4.1 .

5.4 Measurement and Simulation Results

A number of measurements and simulations were conducted to confirm the analysis and to investigate the behaviour around the sweet spot. The measurement and simulation results presented in this section were made with common source amplifier test circuits using two signals at the input. They were *not* extracted from DC measurements.

The measurement test setup is shown in Figure 5.30. The test setup consists of stand alone transistors of different device sizes using $0.18\mu\text{m}$ and $0.25\mu\text{m}$ technologies on separate wafers. The wafers were probed using ground-signal-ground configured 40 GHz air-coplanar RF probes. The bulk and source nodes of the test transistors were tied to ground on the chip. Two signal generators were used to generate a two tone input signal that was fed to the gate of the test transistor. In order to simplify the measurements, the two tones were set at 10 MHz and 11 MHz to give third-order intermodulation products at 9 and 12 MHz. Measurements were also made at

2.0 GHz and 2.001 GHz to confirm the RF performance. An external combiner was used to combine the tones from the signal generators and an external bias tee was used to bias the gate of the test transistor. Since the input impedance of a CMOS transistor is high, the *delivered* voltage level of the input signal was determined as $v_{in,rms}^2 = v_{s,rms}^2 = 4R_S P_{av}$, where P_{av} is the available power from the signal generator. The insertion loss of the test setup was also accounted for in the calculation. The input power levels were set well below the 1 dB compression point of the test circuit.

Figure 5.30: Measurement setup for measuring $IIP3$ of a single transistor common source amplifier.

At the output node, the drain voltage V_{DS} was held constant through an external bias tee and the AC signal was fed to a spectrum analyzer which also provided a load resistance of 50Ω . An external load resistor, R_{ext} , was connected in series with the spectrum analyzer for measurements where a total load resistance of more than 50Ω was desired. $IIP3$ at the drain node was calculated from the output fundamental and $IM3$ signal levels for various V_{GS} setting with a fixed V_{DS} . The insertion loss of the test setup at the output and the voltage divider effect of R_{ext} were also accounted for in the $IIP3$ calculations.

An similar setup was used for the simulation results using the Spectre PSS simulator. A two tone power source with a source resistance of 50Ω was used to generate the input tones. Ideal bias tee models were used to provided the bias for the transistor and foundry supplied BSIM3v3 models were used for the transistor models.

Input versus output voltage levels were first checked to ensure measurement setup procedures were correct and to ensure the fundamental tone and $IM3$ tones behaved as expected. Figure 5.31 shows the measured input versus output voltage

levels for a $0.18\mu\text{m}$ transistor with V_{GS} fixed at 600mV and V_{DS} fixed at 1V . The fundamental tone increases with a slope of one whereas $IM3$ increases with a slope of three until signal levels approach the one dB compression point. The extrapolated $IIP3$ is $116.65\text{ dB}\mu\text{V}$. The input referred 1 dB compression point is estimated to be $104\text{ dB}\mu\text{V}$. The available power levels of the input signal were then set well below the 1 dB compression point and the extrapolated $IIP3$ was then measured for various V_{GS} .

Figure 5.31: Input versus output voltage levels of the fundamental tone at 10 MHz and $IM3$ at 9 MHz of a single transistor common source amplifier with $W/L = 127.5\mu\text{m}/0.18\mu\text{m}$, $V_{GS} = 600\text{mV}$, $V_{DS} = 1\text{V}$ and $R_L = 50\Omega$.

Figure 5.32 shows the measured and simulated $IIP3$ versus V_{GS} for a device size of $W/L = 127.5\mu\text{m}/0.18\mu\text{m}$ with $R_L = 50\Omega$. The calculated $IIP3$, given by (5.40), is also plotted for comparison. The measured results at both 10 MHz and 2 GHz show a pronounced sweet spot at $V_{GS} \approx 520\text{mV}$. The calculated and simulated results also show a pronounced sweet spot at approximately the same V_{GS} . The difference between the location of the sweet spot for the calculated and simulated results is less than 5% of the measured result. The difference bias location of the sweet spot is primarily attributed to differences in the threshold voltage.

Figure 5.32: Measured, simulated, and calculated $IIP3$ versus V_{GS} for a common source amplifier with $V_{DS} = 1$ V, $R_L = 50\Omega$ and using the parameters listed in Table 4.1.

The peak value of the measured $IIP3$ is reduced some at 2 GHz, primarily due to the effects of a nonlinear C_{gs} in a similar fashion as shown previously in Figure 5.19 with $f/f_T = 0.01$. However, the peak is still quite pronounced and at the same gate bias voltage as when measured at 10 MHz. As such, the measurements can be greatly simplified by measuring $IIP3$ at lower frequencies. The measured peak values for both 10 MHz and 2 GHz are not as pronounced as the simulated or analytical value, but none-the-less, the sweet spot is still very much evident. The difference between the measured and calculated peak values can be attributed to a number of factors such as:

- inaccuracies in the drain current model;
- fifth-order nonlinear effects not included in the analysis;
- the output conductance may be more nonlinear with respect to V_{DS} than modeled;
- measurement inaccuracies and V_{GS} measurement resolution;

- unaccounted nonlinear capacitances; and,
- unaccounted parasitic capacitances.

The peak value of the simulation result is closer to the measured since it does include higher order nonlinear effects, nonlinear capacitance effects and some parasitic capacitors. However, the simulation results do not include all the parasitics associated with the test setup, and although the BSIM3 model is fairly accurate, there are likely some inaccuracies in the drain current model and the output conductance model especially with respect to the higher order derivatives.

It is instructive to consider $IIP3$ versus the drain current as well as the gate bias level. Figure 5.33 shows the same measured, simulated, and calculated $IIP3$ as in Figure 5.32 but with respect to I_{DS} . The difference between the location of the sweet spot with respect to I_{DS} is less than 5%. Figure 5.33 shows the sweet spot at $I_{DS} = 1.54\text{mA}$. *This bias current for a transistor size of $W/L = 127.5\mu/0.18\mu$ falls directly in the region of interest for RF design.*

Figure 5.33: Measured, simulated, and calculated $IIP3$ versus I_{DS} for a common source amplifier with $V_{DS} = 1\text{ V}$, $R_L = 50\Omega$ and using the parameters listed in Table 4.1.

5.4.1 Influence of Device Width

From the previous analysis, such as in Section 5.2.1, the $IIP3$ of a MOSFET has been modelled as independent of device width. It has also been pointed out that this is not quite accurate since the threshold voltage is weakly dependent on the width. The threshold voltage is inversely proportional to the width and so the smaller device should have a slightly higher V_{th} . From Figure 4.35, the zero crossing of $c_{(3,0,0)}$ shifts to higher V_{GS} for higher V_{th} . Figure 5.34 plots the measured $IIP3$ for two different device sizes and shows that the location of the sweet spot has shifted to slightly higher V_{GS} for the smaller width, as expected. The difference in the peak values at the sweet spot is attributed to measurement accuracy and V_{GS} measurement resolution.

Figure 5.34: Comparison of measured $IIP3$ versus V_{GS} for different widths for $0.18\mu\text{m}$ technology with $V_{DS} = 1\text{V}$ and $R_L = 50\Omega$.

5.4.2 Influence of Technology Scaling

As CMOS technology continues to scale down, the location of the sweet spot is expected to shift to higher current per μm of width [2]. This is primarily attributed to

differences in the scaling of the device size and the threshold voltage. According to constant electric field scaling scheme, both the device size and the threshold voltage should scale down in the same proportion, $1/k$, and the gate oxide per unit width should scale by k , where $k > 1$ is an arbitrary scaling factor [49]. As a result, the drain current would also scale by $1/k$ and the drain current per unit width would remain constant. This means that the sweet spot should occur at the same bias current for the same width device as technology scales down. However, V_{th} typically scales at a smaller rate, $1/\beta$, where $1 < \beta < k$ [49]-[52], and the drain current per unit width will scale at k^2/β^2 . Hence, for the same width device, the sweet spot will occur at higher bias current. This is evident in Figure 5.35 which compares measured results of 0.25 with 0.18 μm technology. This trend is also confirmed in [2] which shows measured and simulated results for various technologies.

Figure 5.35: Comparison of measured $IIP3$ versus I_{DS} per μm of width for 0.25 and 0.18 μm technology.

5.4.3 Influence of Load Resistance and Gain

Figure 5.36 shows the effect R_L (and gain) has on the sweet spot. As expected from Section 5.2.3, the location of the sweet spot shifts to lower V_{GS} (and I_{DS}) with increasing R_L and gain. Furthermore the sweet spot becomes less pronounced and the peak value decreases. From Section 5.2.6 and Figure 5.21, the influence of the nonlinear behaviour of C_{db} becomes more dominant as gain increases and tends to reduce the peak value at the sweet spot and may explain the reduced sharpness of the peak in Figure 5.36 for higher values of R_L .

Figure 5.36: Comparison of measured $IIP3$ versus V_{GS} for $0.25\mu\text{m}$ technology with $W/L = 220\mu\text{m}/0.25\mu\text{m}$, and $V_{DS} = 1\text{V}$.

5.4.4 Influence of Drain-Source Bias

From the analysis in Section 5.2.2, $IIP3$ is, to first-order, independent of V_{DS} . This, however, is only valid in the saturation region and is evident in Figure 5.37. For $V_{GS} = 300\text{mV}$, the transistor is biased in the weak inversion and is well in the saturation region for $V_{DS} > 250\text{mV}$. $IIP3$ can be seen to be constant versus V_{DS} as expected.

For $V_{GS} = 570\text{mV}$, the transistor is biased at the sweet spot. For $V_{DS} < 400\text{mV}$, the linearity of the amplifier is significantly affected by the drain-source voltage. However, for $V_{DS} > 400\text{mV}$, the transistor is biased strongly in the saturation region and $IIP3$ does not vary significantly. For higher gate voltages the transition region between the linear and saturation region becomes less obvious. For $V_{GS} = 900\text{mV}$, $IIP3$ becomes relatively constant with for $V_{DS} > 800\text{mV}$. However, the gain is also significantly higher and, as a result, the nonlinear coefficients that are dependent on V_{DS} , such as $c_{(1,1,0)}$ and $c_{(2,1,0)}$, play a more significant role and $IIP3$ is more dependent of V_{DS} than the lower V_{GS} bias levels.

Figure 5.37: Measured $IIP3$ versus V_{DS} and different V_{GS} for $0.25\mu\text{m}$ technology with $W/L = 220\mu\text{m}/0.25\mu\text{m}$, and $V_{DS} = 1\text{V}$.

5.4.5 Influence of Signal Level

Although the analysis presented here is based on the assumption of small signal operation, the linearity sweet spot still exists even for large signals. Figure 5.38 shows the measured optimum bias voltage and the corresponding current to achieve

minimum $IM3$ for a $0.25\mu\text{m}$ device for increasing signal levels. I_{DS} continuously increases as v_{in} increases due to a DC shift caused by second-order distortion effects. The optimum bias voltage, however, remains relatively constant up to about 105 $\text{dB}\mu\text{V}$ after which the optimum bias voltage decreases significantly. The shift in the required bias voltage occurs at large signal levels due to the asymmetry in $c_{(3,0,0)}$. If the circuit is biased directly on the sweet spot, a small signal should experience the same distortion on both the positive and negative swings. However, from Figure 4.34, $c_{(3,0,0)}$ is not symmetrical about the zero crossing and depends heavily on α . Hence, for larger signal levels, the third-order distortion will be different for positive and negative swings and the optimum V_{GS} for minimum $IM3$ can be expected to shift to lower V_{GS} .

Figure 5.38: Measured V_{GS} and corresponding I_{DS} versus v_{in} for $0.25\mu\text{m}$ technology with $W/L = 220\mu\text{m}/0.25\mu\text{m}$, and $V_{DS} = 1\text{V}$.

5.4.6 Additional Results

A comparison of the measured and simulated results in Figure 5.32 and 5.33 shows that the BSIM3v3 model predicts the magnitude and location of the sweet spot reasonably accurately. Hence, some simulation results are used to confirm analysis and

investigate other behaviours that are difficult to measure. The effects considered here are

1. the influence of a tuned load centred at 2.4 GHz;
2. the influence of source degeneration operating at 2.4 GHz; and,
3. the influence of process variation and temperature.

Influence of a Tuned Load (Simulated)

Figure 5.39 compares the simulated $IIP3$ of a common source amplifier with a resistive load to the $IIP3$ of a common source amplifier with a tuned load for different load values. For $R_L = 50\Omega$ there is a slight shift in the location of the linearity sweet spot from a resistive load to a tuned load. However, for $R_L = 500\Omega$, there is a significant shift in the location of the sweet spot and an increase in the peak value as well. From Section 5.2.3 and 5.4.3, the location of the sweet spot shifts to lower values of V_{GS} with increasing R_L since the terms arising from the nonlinearities due to the output conductance become more significant and cancel with $c_{(3,0,0)}$. With a tuned load though, the nonlinear terms due to the output conductance are not as significant compared to a resistive load and shift from the zero crossing of $c_{(3,0,0)}$ alone is not as large.

Figure 5.40 shows the simulated frequency response of the normalized $IM3$ of a common source amplifier with a tuned load centred at 2.4 GHz. For the simulation, the frequency of the two tone inputs were varied but the spacing between the two tones was maintained at 1 MHz. The plot shows a peak at $f_0/2$ and a dip at f_0 . The response at $f_0/2$ occurs as the magnitude of the fourth term of (5.40) reaches a maximum at $f_0/2$ and is sufficiently large enough to add to the $c_{(3,0,0)}$ contribution. The dip at f_0 occurs since the second term of (5.40) is maximum at f_0 but is in opposite phase of $c_{(3,0,0)}$. The simulation results agrees quite closely with the analytical result in Figure 5.10, except the amplitude at $f_0/2$ and f_0 differs somewhat. This is attributed to the accuracy of the analytical model of the output conductance which models g_o as linear with respect to V_{ds} . However, if g_o is not exactly linear with respect to V_{ds} , then, according to (5.38), $c_{(1,2,0)}$, $c_{(0,3,0)}$ will also contribute to $IM3$ at f_0 . Similarly, the other terms in (5.38) containing Z_Σ will contribute at $f_0/2$.

Figure 5.39: Comparison of simulated $IIP3$ of a common source amplifier with a tuned load at 2.4 GHz to a resistive load versus V_{GS} for $R_L = 50\Omega$ and 500Ω .

Figure 5.40: Simulated amplitude response of the normalized $IM3$ versus frequency with $V_{GS} = 600\text{mV}$, $V_{DS} = 600\text{mV}$, $V_{SB} = 0\text{V}$, $R_L = 500\Omega$.

Influence of Source Degeneration (Simulated)

Figure 5.41 shows the simulated effects of inductive source degeneration on a common source amplifier with a tuned load operating at 2.4 GHz. As expected from Section 5.3, degeneration causes an apparent degradation in $IIP3$ in the vicinity of the linearity sweet spot but improves $IIP3$ elsewhere. As the degeneration increases, the existence of the sweet spot can disappear entirely.

Figure 5.41: Simulation results for $IIP3$ of a common source amplifier with inductive source degeneration versus V_{GS} operating at 2.4 GHz

Influence of Process Variation and Temperature (Simulated)

Not surprisingly, process variation and temperature can significantly affect the sweet spot. In Figure 5.42, the simulated location of the sweet spot changes by several hundred millivolts between fast and slow corners, primarily due to variations in V_{th} . However, in Figure 5.43, the sweet spot location does not change significantly in terms of the bias current.

A similar behaviour can also be seen for variations over temperature. In Figure 5.44(a), the location of the sweet spot changes with V_{GS} but in Figure 5.44(b), the

Figure 5.42: Simulated effect of process variation on sweet spot for 0.18 μm single transistor amplifier versus V_{GS} over fast and slow corners .

Figure 5.43: Simulated effect of process variation on sweet spot for 0.18 μm single transistor amplifier versus I_{DS} over fast and slow corners .

location is constant with respect to I_{DS} . This suggests careful attention to the bias will be required to ensure operation at the sweet spot. The designer must ensure constant current bias over process and temperature. Circuit linearization techniques, such as demonstrated in [12], [13], [14], [15], based on the cancellation of the third-order nonlinearities, can create a wider enhanced linear region next to the sweet spot and possibly reduce the susceptibility to process variations.

Figure 5.44: Simulated effect of process variation over fast and slow corners on sweet spot for 0.18 μm single transistor amplifier versus (a) V_{GS} and (b) I_{DS} .

5.5 Common-gate and Cascode Amplifier

The common-gate amplifier, shown in Figure 5.45(a), is an alternative configuration to the common-source amplifier that is sometimes used in RF applications. Its primary advantage is that it provides a broadband low input impedance that is resistive rather than capacitive as in the common-source amplifiers case. However, the common-gate amplifier typically has a higher noise figure than a common-source amplifier and, as a result is not used as often as the common source amplifier. Although the voltage gain of the common-gate amplifier is about the same as the common-source amplifier, the body effect of the common-gate amplifier plays a significant role in its linearity. If there is no body effect, the linearity of the common-gate amplifier is the same as the common-source amplifier. However, if there is body effect, then the bulk nonlinear coefficients must be accounted for and several other additional terms dependent on $c_{(0,0,1)}$ (or g_{mb}), $c_{(0,0,2)}$, and $c_{(0,0,3)}$ will appear in (5.38). As with the common-source amplifier, if the drain is a low impedance point, $c_{(3,0,0)}$ dominates, but as the gain increases the other nonlinear coefficients become more significant and can cause a significant shift in the sweet spot.

Figure 5.45: (a) Common source amplifier with source degeneration, (b) common gate amplifier, and (c) cascode amplifier with biasing details not shown.

A cascode structure, shown in Figure 5.45(b), is often used in RF amplifiers to increase the reverse isolation of a common-source amplifier and to improve stability. At high frequencies, the gate-drain capacitance of M1, C_{gd} , is not negligible and can

cause significant coupling between the input and output ports. The cascode transistor acts as a common-gate stage and can provide for isolation between the input and output ports and greatly suppress the potential for instability. Ideally, the cascode transistor simply passes the current of the input transistor to the load without any loss and without distortion. However, as with the common-gate amplifier, the cascode transistor can be an extra source for distortion mainly due to the body effect. If the drain of the cascode transistor is a low impedance source, the gain of the amplifier is low, the contribution from the bulk nonlinear coefficients are minimal, and the linearity of the amplifier is dominated by the input transistor. However, if the gain is increased, the bulk nonlinear coefficients can dominate and affect the overall amplifier linearity.

5.6 An LNA Design for Bluetooth

This section presents the design and implementation of a 1.2V - 1.8V, 2.5 GHz LNA in a mixed-signal 0.18 μm CMOS technology [55]. A standard differential cascode structure with inductive source degeneration was used with off-chip transmission line transformers (TLT). Inductive degeneration was primarily used for input matching purposes and as such, it was not possible to take advantage of any linearity sweet spots. This design demonstrates that it is in reality quite difficult to take advantage of the linearity sweet spot in a practical LNA. However, as an example of a typical modern CMOS design, it shows that RF designs are being biased more and more in the moderate inversion region and so linearity analysis of the moderate inversion region will become more critical.

5.6.1 Differential LNA Design

Noise figure is typically the most important characteristic of an LNA, however, the requirement for noise figure is not as stringent for Bluetooth applications due to the relaxed receiver sensitivity (-70 dBm with a 0.1% BER which translates to a maximum NF of approximately 5dB). Linearity tends to become the more challenging specification to meet, especially with low voltage supplies. The desired $IIP3$ for a typical LNA for a Bluetooth application is generally greater than -10 dBm. A

differential cascode structure with a tail current source, shown in Figure 5.46, was used to reduce even-order harmonics and to facilitate AC coupling directly to a differential Gilbert mixer (not discussed in this work). However, the cascode topology with tail current limits the minimum supply voltage due to stacking of the transistors. The tail current supply could be removed for lower voltage supplies, but the operation of the LNA becomes very sensitive to the bias levels of the input RF transistors. While a folded cascode structure reduces the number of stacked transistors and allows for a reduced voltage supply, it does not offer power saving since the overall current supply must be increased due to the additional parallel current paths.

Figure 5.46: Basic differential LNA structure.

A 0.18 μm analog/RF CMOS process was used for implementation of the design. The process has 6 metal layers with 2 μm thick top metal layer, MIM capacitors, high resistive poly-resistors, multiple-threshold transistors, and a deep N-well (DNW) formation, depicted in Figure 5.47, to provide separate substrate connections for NMOS transistors. In addition to offering increased transistor isolation for reduced noise operation, body effect can be eliminated with the deep N-well potentially allowing for reduced voltage operation. Two versions of the LNA were implemented,

a standard transistor version and a DNW transistor version, to evaluate any potential improvements by using a DNW structure.

Figure 5.47: A cross section of a deep N-well.

Tuned loads were used instead of resistive loads allowing for a low supply voltage while maintaining the required linearity. Output buffers for the stand alone LNAs were not necessary due to the 4:1 impedance transformation provided by the TLT. On-chip inductive source degeneration was required to meet the linearity specifications and to provide a real valued input resistance for matching purposes. Multiple fingers were used in the RF and cascode transistors to reduce the effect of the distributed gate resistance, minimizing the noise figure of the LNA [56]. Biasing was provided on chip with separate pin connections for adjusting individual bias levels, if required.

5.6.2 LNA Measurements

Figure 5.48 shows the die photograph of the two versions of the stand alone LNAs. The die area for one LNA is 1.44 mm^2 , mostly used by the load and source degeneration inductors. A savings of die area of approximately 60% can be realized by using center-tapped symmetrical inductors.

The dice were mounted on an FR4 board and bonded directly to the board, significantly reducing package parasitics and bond wire inductance, estimated at 0.5 nH. The total current consumption of each LNA was 2.15 mA with a power supply of 1.8V. The measured power gain of the standard version LNA was 8.9 dB and the noise figure was 3.2 dB. Both metrics compared well to the simulated gain and noise figure of 10.0 dB and 2.5 dB respectively. The differences are attributed to unaccounted parasitics and a slight mis-tuning of the tuned load. The input referred third order

Figure 5.48: Microphotograph of standard version LNA and deep Nwell version LNA.

intercept point was measured at -5.6 dBm and 1 dB compression point was -16.1 dBm and is shown in Figure 5.49. Figure 5.50 compares the simulated and measured gains of the standard version and DNW version LNA. From simulations, the DNW version was expected to provide slightly more gain. Unfortunately, measurements show the opposite, probably due to increased mismatching between transistors. With the DNW option, transistors must be placed further apart and interdigitated layout schemes are not possible. The difference in gain can also be seen in Figure 5.51 which compares the gain and noise figure of the standard version LNA to the DNW LNA versus the power supply voltage V_{dd} . The NF of the DNW LNA was expected to be slightly lower than the standard version but because the gain is lower, the noise figure is slightly higher than expected. In both cases, the LNAs are functional down to 0.8 V but performance degrades below 1.2 V.

The $IIP3$ performance versus the core bias current for both versions of LNAs is shown in Figure 5.52 and does not show any linearity sweet spot. This, however, is expected since the LNAs use 2 nH source degeneration inductors, which, from Section 5.3, can eliminate the sweet spot due to the feedback. The required bias current for the sweet spot for the input transistors was approximately 1 mA, or a core bias current of 2 mA. Figure 5.52 also shows that the deep n-well version has a slightly improved $IIP3$. However, simulations showed no improvement in $IIP3$, which indicates that

Figure 5.49: Measured input-referred third order intercept point for the standard version LNA with $V_{dd} = 1.8\text{V}$ and $I_{core} = 2.15\text{mA}$.

Figure 5.50: Comparison of S21 for standard version and deep Nwell version LNAs.

Figure 5.51: Measured gain and noise figure versus V_{dd} for standard version and deep Nwell version LNAs.

the nonlinear influence of the body effect in the standard version LNA is minimal, primarily since the voltage gain of the LNA is relatively low. The improved $IIP3$ in the measurement results may be due to the reduced gain of the deep n-well version LNA rather than reduced body effect.

Table 5.1 summarizes the measured performance which compares favourably to other recently reported designs and yet has at least 30% less power dissipation than any of those reported. The reduced power is primarily attributed to operation in the moderate inversion region. All LNAs reported in Table 5.1 are likely operating in the strong inversion region with possible exception of [57]. Their design is likely operating close to or in the moderate inversion region. The measured results from the LNAs show that operation in the moderate inversion region is very feasible with modern CMOS technology. As such, designers will need to be aware of the existence of the linearity sweet spot in the moderate inversion region, otherwise they may be misled by simulated or measured results.

Figure 5.52: Measured $IIP3$ versus the core bias current for standard version and deep Nwell version LNAs.

Table 5.1: Summary of measurement results for version with standard transistors and comparison to other reported designs.

Gain (dB)	NF (dB)	IIP3 (dBm)	f_0 (GHz)	Vdd (V)	Idd (mA)	Technology	Topology	Ref.
8.9	3.2	-5.5	2.45	1.8	2.15	0.18 μ m CMOS	Diff. D.C.S.	
10	3.5	-4	2.0	1.0	10	0.20 μ m SOI	S.E. D.C.S.	[58]
17.5	2.3	-9	2.5	3.3	4	0.6 μ m CMOS	Diff. D.C.S.	[59]
8	4.7	NA	5.0	3	2	0.25 μ m CMOS	Diff. D.C.S.	[57]
12	2.3	+5.4	1.9	1.5	6.5	0.13 μ m CMOS	S.E.. D.C.S.	[60]
18	2.5	-1.0	1.9	1.8	5	0.18 μ m CMOS	Diff. D.C.S.	[61]
10	2.85	15.6	0.9	2.7	7.8	0.35 μ m CMOS	S.E. D.C.S.	[62]
16	2.5	-15	2.4	1.8	14	0.18 μ m CMOS	Diff. C.S.	[63]

5.7 Summary

In this chapter, the nonlinear behaviour of several basic CMOS amplifiers has been presented with the emphasis on the common source amplifier, the most commonly used circuit topology in RF applications. Moreover, the analysis and measurements focused on the nonlinear behaviour of the amplifier operating in the moderate inversion region. As CMOS technology continues to scale down to smaller feature sizes, and with the push for lower and lower power consumption, future designs will be operating more and more in the moderate inversion region.

The analysis and measurement revealed that for low voltage gains, the third-order distortion of the amplifiers is dominated by the nonlinear behaviour of g_m with respect to V_{gs} . However, for most practical RF applications, the gain of the circuit will likely be sufficiently large enough that the third-order distortion will be dominated by the nonlinear behaviour of the output conductance instead. However, if a tuned load is used instead of a resistive load, the effects of the nonlinear behaviour of the output conductance are reduced somewhat at the centre frequency but still can be a major contributor. At all other frequencies though, the effects of the output conductance are virtually eliminated.

The analysis and measurements also reveal that the nonlinear behaviour of the C_{gs} will become significant at very high frequencies. However, for $f < 0.1f_T$, only the behaviour around the linearity sweet spot is significantly affected.

Inductive source degeneration is often used in RF applications to provide a real portion to the input impedance of a common source amplifier. Source degeneration constitutes series-series feedback, which has a significant effect on the third-order nonlinear behaviour of the common source amplifier. The feedback actually introduces an additional third-order distortion component that is dependent wholly on second-order distortion effects. Second-order distortion tones are fed back to mix with fundamental components that produce additional third-order distortion tones. However, feedback also reduces the overall magnitude of the third-order distortion tones by the loop gain of the feedback system. The net effect is that linearity generally improves with feedback, except in the vicinity of the sweet spot, where the linearity actually degrades. However, this contradiction to the well known theory only applies in the vicinity of

the sweet spot. This type of behaviour is also applicable when the influence of C_{gd} is considered.

Several practical design considerations were also addressed in this chapter. It was shown that the location of linearity sweet spot will vary considerably with respect to the gate bias voltage due to process variation and changes in temperature. The location of the sweet spot, however, does not vary much with respect to the bias current, suggesting that a bias feedback network could be used to take advantage of the sweet spot. A common gate topology and a cascode topology were also addressed. The nonlinear behaviour of both topologies should be similar but if body effect is present, it can influence their nonlinear behaviour.

Finally, an LNA design suitable for Bluetooth application was presented that demonstrated the feasibility of operating in the moderate inversion region. The design compared favourably to several other designs, but with at least a 30% reduction in power consumption. However, the design was not able to take advantage of the linearity sweet spot due to the use of source degeneration inductors. None-the-less, designers need to be aware of the existence of the linearity sweet spot since it highly depends on many factors and could lead to erroneous results if they are not properly accounted for.

Chapter 6

Conclusions

6.1 Summary

The focus of this thesis has been the analysis of the nonlinear behaviour of a MOSFET transistor over all regions of operation, including the moderate inversion region. Third-order intermodulation distortion in a MOSFET is analyzed by means of Volterra Series representation. The analysis, confirmed by measurements, reveals a significant peaking or “sweet spot” of the third-order intercept point in the moderate inversion region. This peaking occurs as the mechanism for the drain current flow changes from diffusion, a largely exponential behaviour, to drift, a slightly less than square-law type behaviour (i.e. the exponent is slightly less than two). Over a very narrow bias range in the moderate inversion region, the current flow follows a nearly ideal square-law behaviour with minimized third-order distortion. However, this also means that second-order distortion products are maximized. While the cut-off frequency of a transistor biased in weak inversion is too low for RF applications, there is sufficient current gain for most RF applications while biased in the moderate inversion region, especially with modern sub-micron devices. As such, the thesis also addresses issues relevant to RF circuits.

A “single-piece” MOSFET model was presented in Chapter 4 that is continuous over all regions of inversion. Mobility reduction, velocity saturation and channel length modulation effects were included in the model as well. Since the drain current is dependent on three terminal voltages, a total of nineteen power series coefficients, or nonlinear coefficients, were required to describe the nonlinearities of the drain current. The coefficients were derived for the saturation region only, as functions of *long*, *mobhot*, and *clm*. The dependency of *long* with respect to V_{gs} was the dominant

factor in the nonlinear coefficients. *mobhot* was the next dominant factor and tends to linearize the drain current. For a minimum length, modern deep sub-micron device, *clm* also plays an important role since the output conductance is more dependent on V_{gs} than with longer channel devices.

The analysis and measurements of the nonlinear behaviour of several basic CMOS amplifiers was presented with the emphasis on the common source amplifier, the most commonly used circuit topology in RF applications. The analysis and measurements revealed that for low voltage gains, the third-order distortion of the amplifiers is dominated by the nonlinear behaviour of g_m with respect to V_{gs} . However, for most practical RF applications, the gain of the circuit will likely be sufficiently large enough that the third-order distortion will be dominated by the nonlinear behaviour of the output conductance instead. However, if a tuned load is used instead of a resistive load, the effects of the nonlinear behaviour of the output conductance are reduced somewhat at the centre frequency but still can be a major contributor. At all other frequencies, the effects of the output conductance are virtually eliminated. The analysis and measurements also reveal that the nonlinear behaviour of the C_{gs} will become significant at very high frequencies. However, for $f < 0.1f_T$, only the behaviour around the linearity sweet spot is significantly affected.

Inductive source degeneration is often used in RF CMOS amplifier designs in order to present a real impedance for optimum power matching and power transfer. Source degeneration also constitutes series-series feedback and introduces an additional third-order distortion component that is dependent wholly on second-order distortion effects. However, feedback reduces the overall magnitude of the third-order distortion tones by the loop gain of the feedback system and the net effect is that linearity generally improves with feedback, except in the vicinity of the sweet spot, where the linearity actually degrades. This type of behaviour is also applicable when the influence of C_{gd} is considered.

Several practical design considerations were also addressed in this thesis. It was shown that the location of linearity sweet spot will vary considerably with respect to the gate bias voltage due to process variation and changes in temperature. The location of the sweet spot, however, does not vary much with respect to the bias current, suggesting that a bias feedback network could be used to take advantage of the sweet spot. A common gate topology and a cascode topology were also addressed.

The nonlinear behaviour of both topologies should be similar but if body effect is present, it can influence their nonlinear behaviour.

Lastly, an LNA design suitable for Bluetooth application was presented in Chapter 5 that demonstrated the feasibility of operating in the moderate inversion region. The design compared favourably to several other designs, but with at least a 30% reduction in power consumption. However, the design was not able to take advantage of the linearity sweet spot due to the use of source degeneration inductors. Nonetheless, designers need to be aware of the existence of the linearity sweet spot since it highly depends on many factors and could lead to erroneous results if they are not properly accounted for.

6.2 Contributions

1. Analysis and measurements of the linearity of the moderate inversion region of a MOSFET that shows a region of significantly reduced third-order distortion, referred to as the “linearity sweet spot”. As far as can be determined, the analysis of the moderate inversion region has not been presented elsewhere.
2. New analysis on the effects of the MOSFET’s nonlinear capacitance on the linearity of various circuits. Several models of the MOSFET’s nonlinear capacitors has been presented elsewhere, but their effects on the nonlinear behaviour of the circuit has not present elsewhere as far as can be determined.
3. Analysis of the implications of the linearity sweet spot on typical RF circuits. The analysis includes the effects of source degeneration and other forms of feedback on the linearity of the MOSFET and shows that feedback can actually degrade the linearity in the vicinity of the sweetspot.
4. Design and implementation of a low power RF CMOS LNA operating in the moderate inversion region that is suitable for Bluetooth applications. By operating in the moderate inversion region, the design achieved at least 30% power savings compared to other recent designs and still achieved comparable performance.

5. Adapted a MOSFET model for linearity analysis that is valid in the saturation region over all regions of inversion. The model was a conglomeration of several existing model but was simplified for ease of analysis, yet provided good accuracy

6.2.1 Publications

Refereed

- Bill Toole, Calvin Plett, Mark Cloutier, "RF Circuit Implications of Moderate Inversion Enhanced Linear Region in MOSFETs," IEEE Transactions on Circuits and Systems I, vol 51, No. 2, pp. 319-328, Feb. 2004.
- Bill Toole, Calvin Plett, Mark Cloutier, "RF Circuit Implications of a Low-current Linearity "Sweet Spot" in MOSFETs" proc. ESSCIRC Florence, Italy, pp. 619-622, Sept 2002.
- W. Toole, C. Plett, "A Low Voltage, Low Power RF CMOS LNA for Bluetooth Applications using Transmission Line Transformers," pp. 444-447, Proc. ESSCIRC 2001, Villach, Austria, Sept 2001.

Other

- W. Toole, C. Plett, "Low Voltage, Low Power RF CMOS for Bluetooth Applications," Proceedings Micronet Annual Workshop, pp. 83,84, Aylmer Quebec, April 20, 2001
- Arif Siddiqi, Mashkoor Baig, Neric Fong, Bill Toole, Ralph Mason, Tad Kwasniewski, Calvin Plett, "0.25 μ m CMOS First Run Design and Results", Micronet 0.25 μ m CMOS First Look Meeting, Ottawa, Ont. Sept. 1999.
- Bill Toole, Calvin Plett, Tad Kwasniewski, "Considerations for Sub-1V RF Design in CMOS," Cito Researcher Retreat, Hamilton, Ont., June, 1999.
- Bill Toole, Calvin Plett, Tad Kwasniewski, "Considerations for Sub-1V RF Design in CMOS," Micronet Annual Workshop, Ottawa, Ont., May, 1999.

- W. Toole, C. Plett, Tad Kwasniewski, "CMOS Technology Considerations for Next Generation GHz Range Transceiver," Internal Publication , 1999.

6.3 Recommendations for Future Research

The following issues that arose in this thesis should be studied in greater detail in future work.

1. Improved MOSFET models. This is an area that will likely always be an issue with MOSFETs. Many advances have been made with models such as the BSIM3, BSIM4 and EKV model but as CMOS continues to scale to smaller feature sizes, the models of today will quickly become inadequate for the smaller sizes. Most models are developed with the primary (and necessary) goal of attaining accurate I-V characteristics. The models also need to give accurate values for all small signal quantities such as g_m , g_o , and g_{mb} across all regions of inversion, and, most importantly they should be continuous with respect to any voltage terminal. Furthermore, for accurate RF modelling, the CMOS models need to give accurate cross-products of the small signal terms as well as higher order derivatives.
2. Better analytical model of output conductance. A simple yet accurate model of the output conductance over all regions of operation and inversion needs to be developed in order to better understand the effect of the second and third-order nonlinearities with respect to V_{gs} and V_{ds} . Even though these terms are very small, they can become significant if the voltage gain is large and dominate the second and third-order distortion products.
3. Investigate effects of nonlinear capacitance more fully. The analysis presented in this thesis used relatively simple models for the intrinsic capacitances of a MOSFET and were based on the assumption of quasi-static operation. More complete intrinsic models based on non-quasi static operation could be investigated as well as more complete extrinsic models. However, it is very likely that the analysis will become overly complicated.

4. Development of a symbolic Volterra Series circuit analysis program. Volterra Series analysis is a very powerful method for analysis of weakly nonlinear circuits, however, hand calculations quickly become very tedious even for very small circuits. Symbolic network analysis programs can calculate closed-form expressions of the AC behaviour of linearized circuits in terms of their small-signal parameters. Being able to extend this capability to the second-order and third-order equivalent nonlinear circuits would be a very useful feature. It will be particularly difficult, though, to keep these expressions in a simple enough format for interpretation.
5. Extend the analysis to larger circuits. The circuits presented in this thesis were in fact very simple circuits. The analysis of the nonlinear behaviour in the moderate inversion region could be extended to other more complex RF circuits such as folded cascode amplifiers, single and double balanced mixers. As pointed out above, though the expressions for the more complex circuits would likely be too tedious for hand calculation, necessitating the need for a symbolic Volterra Series based circuit analysis program.
6. Devise circuit to take advantage of sweet spot. It has been shown that it is very difficult for a circuit to deliberately take advantage of the third-order linearity sweet spot. That is not to preclude the possibility though. It has also been shown that the location of the sweet spot is quite predictable with respect to the drain current and so a bias feedback network may be used to maintain the required bias for the sweet spot. Depending on the application, calibration circuits may also be used during a power-up sequence perhaps to test and fine tune the bias to achieve the linearity sweet spot. Since the third-order distortion products are minimized when the second-order distortion products are at maximum, one possibility is to calibrate for maximum second-order distortion which is generally easier to measure. However, this highly depends on the application. In a direct conversion receiver, maximized second-order distortion products are also highly undesirable.

Appendix A

Taylor Series Expansions

A.1 Taylor Series Expansion of a Single Tone

If a single-tone sinusoidal input, $x(t) = A \cos(\omega t + \theta)$, is applied to a nonlinear system whose output $y(t)$, can be described by

$$y(t) = c_0 + c_1 x(t) + c_2 x(t)^2 + c_3 x(t)^3 + \dots \quad (\text{A.1})$$

where $x(t)$ and $y(t)$ can either be a voltage or a current signal, then the output, to seventh-order, is given by

$$\begin{aligned} y(t) = & c_0 + c_1 A \cos(\omega t + \theta) + c_2 A^2 \cos^2(\omega t + \theta) + \\ & c_3 A^3 \cos^3(\omega t + \theta) + c_4 A^4 \cos^4(\omega t + \theta) + \\ & c_5 A^5 \cos^5(\omega t + \theta) + c_6 A^6 \cos^6(\omega t + \theta) + \\ & c_7 A^7 \cos^7(\omega t + \theta) + \dots \end{aligned} \quad (\text{A.2})$$

Using the trigonometric identities

$$\cos^2 x = \frac{1}{2} + \frac{1}{2} \cos 2x \quad (\text{A.3})$$

$$\cos x \cos y = \frac{1}{2} \cos(x + y) + \frac{1}{2} \cos(x - y) \quad (\text{A.4})$$

(A.2) can be rewritten as

$$\begin{aligned}
 y(t) = & K_0 + K_1 \cos(\omega t + \theta) + K_2 \cos(2\omega t + 2\theta) + \\
 & K_3 \cos(3\omega t + 3\theta) + K_4 \cos(4\omega t + 4\theta) + \\
 & K_5 \cos(5\omega t + 5\theta) + K_6 \cos(6\omega t + 6\theta) + \\
 & K_7 \cos(7\omega t + 7\theta) + \dots
 \end{aligned} \tag{A.5}$$

where K_n is the output amplitude response at $n\omega$ and are given by

$$K_0 = c_0 + \frac{1}{2}A^2c_2 + 38A^4c_4 + \frac{5}{16}A^6c_6 + \dots \tag{A.6}$$

$$= c_0 + \sum_{m=1}^{\infty} \frac{c_{2m}}{2^{2m-1}} \cdot \binom{2m-1}{m-1} \cdot A^{2m} \tag{A.7}$$

$$K_1 = Ac_1 + \frac{3}{4}A^3c_3 + \frac{5}{8}A^5c_5 + \frac{35}{64}A^7c_7 + \dots \tag{A.8}$$

$$= \sum_{m=1}^{\infty} \frac{c_{2m-1}}{2^{2m-2}} \cdot \binom{2m-1}{m-1} \cdot A^{2m-1} \tag{A.9}$$

$$K_2 = \frac{1}{2}A^2c_2 + \frac{1}{2}A^4c_4 + \frac{15}{32}A^6c_6 + \dots \tag{A.10}$$

$$= \sum_{m=1}^{\infty} \frac{c_{2m}}{2^{2m-1}} \cdot \binom{2m}{m-1} \cdot A^{2m} \tag{A.11}$$

$$K_3 = \frac{1}{4}A^3c_3 + \frac{5}{16}A^5c_5 + \frac{21}{64}A^7c_7 \dots \tag{A.12}$$

$$= \sum_{m=1}^{\infty} \frac{c_{2m+1}}{2^{2m}} \cdot \binom{2m+1}{m-1} \cdot A^{2m+1}. \tag{A.13}$$

where the symbol $\binom{m}{n}$ stands for $\frac{m!}{(m-n)!n!}$.

A.2 Taylor Series Expansion of Two Tones

A.2.1 Unequal Amplitudes

If a two-tone sinusoidal input, $x(t) = A \cos(\omega_1 t) + B \cos(\omega_2 t)$, is applied to (A.1) then, using (A.3) and (A.4), the output is given by

$$\begin{aligned} y(t) = & K_{0,0} + K_{1,0} \cos(\omega_1 t) + K_{0,1} \cos(\omega_2 t) + K_{2,0} \cos(2\omega_1 t) + \quad (A.14) \\ & K_{0,2} \cos(2\omega_2 t) + K_{3,0} \cos(3\omega_1 t) + K_{0,3} \cos(3\omega_2 t) + \dots + \\ & K_{1,\pm 1} \cos((\omega_1 \pm \omega_2)t) + K_{\pm 1,1} \cos((\omega_2 \pm \omega_1)t) + \\ & K_{2,\pm 1} \cos((2\omega_1 \pm \omega_2)t) + K_{\pm 1,2} \cos((2\omega_2 \pm \omega_1)t) + \dots \end{aligned}$$

where $K_{n,m}$ is the amplitude response of a tone at $n\omega_1 + m\omega_2$. The responses at the different tones are, to seventh order

$$K_{0,0} = c_0 + \frac{1}{2}(A^2 + B^2) + \frac{3}{8}A^4c_4 + \frac{3}{2}A^2B^2c_4 + \frac{3}{8}B^4c_4 + \dots \quad (A.15)$$

$$\begin{aligned} & \frac{5}{16}A^6c_6 + \frac{45}{16}A^4B^2c_6 + \frac{45}{16}A^2B^4c_6 + \frac{5}{16}B^6c_6 + \dots \\ K_{1,0} = & Ac_1 + \frac{3}{4}A^3c_3 + \frac{3}{2}AB^2c_3 + \frac{5}{8}A^5c_5 + \frac{15}{4}A^3B^2c_5 + \dots \quad (A.16) \\ & \frac{15}{8}AB^4c_5 + \frac{35}{64}A^7c_7 + \frac{105}{16}A^5B^2c_7 + \frac{315}{32}A^3B^4c_7 + \dots \\ & \frac{35}{16}AB^6c_7 + \dots \end{aligned}$$

$$\begin{aligned} K_{0,1} = & Bc_1 + \frac{3}{4}B^3c_3 + \frac{3}{2}A^2Bc_3 + \frac{5}{8}B^5c_5 + \frac{15}{4}A^2B^3c_5 + \dots \quad (A.17) \\ & \frac{15}{8}A^4Bc_5 + \frac{35}{64}B^7c_7 + \frac{105}{16}A^2B^5c_7 + \frac{315}{32}A^4B^3c_7 + \dots \\ & \frac{35}{16}A^6Bc_7 + \dots \end{aligned}$$

$$\begin{aligned} K_{2,0} = & \frac{1}{2}A^2c_2 + \frac{1}{2}A^4c_4 + \frac{3}{2}A^2B^2c_4 + \frac{15}{32}A^6c_6 + \frac{15}{4}A^4B^2c_6 + \dots \quad (A.18) \\ & \frac{45}{16}A^2B^4c_6 + \dots \end{aligned}$$

$$\begin{aligned} K_{0,2} = & \frac{1}{2}B^2c_2 + \frac{1}{2}B^4c_4 + \frac{3}{2}A^2B^2c_4 + \frac{15}{32}B^6c_6 + \frac{15}{4}A^2B^4c_6 + \dots \quad (A.19) \\ & \frac{45}{16}A^4B^2c_6 + \dots \end{aligned}$$

$$K_{3,0} = \frac{1}{4}A^3c_3 + \frac{5}{16}A^5c_5 + \frac{5}{4}A^3B^2c_5 + \frac{21}{64}A^7c_7 + \frac{105}{32}A^5B^2c_7 + \dots \quad (\text{A.20})$$

$$\frac{105}{32}A^3B^4c_7 + \dots$$

$$K_{0,3} = \frac{1}{4}B^3c_3 + \frac{5}{16}B^5c_5 + \frac{5}{4}A^2B^3c_5 + \frac{21}{64}B^7c_7 + \frac{105}{32}A^2B^5c_7 + \dots \quad (\text{A.21})$$

$$\frac{105}{32}A^4B^3c_7 + \dots$$

$$K_{1,\pm 1} = K_{\pm 1,1} = ABc_2 + \frac{3}{2}A^3Bc_4 + \frac{3}{2}AB^3c_4 + \frac{15}{8}A^5Bc_6 + \dots \quad (\text{A.22})$$

$$\frac{45}{8}A^3B^3c_6 + \frac{15}{8}AB^5c_6 + \dots$$

$$K_{2,\pm 1} = \frac{3}{4}A^2Bc_3 + \frac{5}{4}A^4Bc_5 + \frac{15}{8}A^2B^3c_5 + \frac{105}{64}A^6Bc_7 + \dots \quad (\text{A.23})$$

$$\frac{105}{32}A^2B^5c_7 + \dots$$

$$K_{\pm 1,2} = \frac{3}{4}AB^2c_3 + \frac{5}{4}AB^4c_5 + \frac{15}{8}A^3B^2c_5 + \frac{105}{64}AB^6c_7 + \dots \quad (\text{A.24})$$

$$\frac{105}{32}A^5B^2c_7 + \dots$$

A.2.2 Equal Amplitudes

If the two tones are of equal amplitude A , then the hamonic responses are simplified to

$$K_{0,0} = c_0 + A^2c_2 + \frac{9}{4}A^4c_4 + \frac{25}{4}A^6c_6 + \dots \quad (\text{A.25})$$

$$K_{1,0} = K_{0,1} = Ac_1 + \frac{9}{4}A^3c_3 + \frac{25}{4}A^5c_5 + \frac{1225}{64}A^7c_7 + \dots \quad (\text{A.26})$$

$$K_{2,0} = K_{0,2} = \frac{1}{2}A^2c_2 + 2A^4c_4 + \frac{225}{32}A^6c_6 + \dots \quad (\text{A.27})$$

$$K_{3,0} = K_{0,3} = \frac{1}{4}A^3c_3 + \frac{25}{16}A^5c_5 + \frac{441}{64}A^7c_7 + \dots \quad (\text{A.28})$$

$$K_{1,\pm 1} = K_{\pm 1,1} = A^2c_2 + 3A^4c_4 + \frac{75}{8}A^6c_6 + \dots \quad (\text{A.29})$$

$$K_{2,\pm 1} = K_{\pm 1,2} = \frac{3}{4}A^3c_3 + \frac{25}{8}A^5c_5 + \frac{735}{64}A^7c_7 + \dots \quad (\text{A.30})$$

Appendix B

Basics of Volterra Series

B.1 Sinusoidal Response of a Second-Order Volterra Series System

Assume that a second-order system is excited by a sinusoidal input

$$\begin{aligned} x(t) &= A \cos \omega t \\ &= \frac{A}{2} e^{j\omega t} + \frac{A}{2} e^{-j\omega t} \\ &= x_a(t) + x_b(t) \end{aligned} \quad (\text{B.1})$$

where

$$x_a(t) = \frac{A}{2} e^{j\omega t} \quad (\text{B.2})$$

and

$$x_b(t) = x_a^*(t) = \frac{A}{2} e^{-j\omega t} \quad (\text{B.3})$$

The response is then

$$\begin{aligned} y(t) &= \mathbf{H}_2[x(t)] \\ &= \mathbf{H}_2[x_a(t) + x_b(t)] \\ &= \mathbf{H}_2[x_a(t)] + \mathbf{H}_2[x_b(t)] + \mathbf{H}_2\{x_a(t), x_b(t)\} + \mathbf{H}_2\{x_b(t), x_a(t)\} \end{aligned} \quad (\text{B.4})$$

where $\mathbf{H}_2\{\cdot\}$ is called a bilinear operator where

$$\mathbf{H}_2\{x_1(t), x_2(t)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_2(\tau_1, \tau_2) x_1(t - \tau_1) x_2(t - \tau_2) d\tau_1 d\tau_2 \quad (\text{B.5})$$

Although $\mathbf{H}_2$ can be considered symmetric, it is not necessary to let $\mathbf{H}_2\{x_a(t), x_b(t)\} = \mathbf{H}_2\{x_b(t), x_a(t)\}$ for the derivation here. The first term in (B.4) is rewritten in terms of the second-order Volterra kernel where

$$\begin{aligned}
 \mathbf{H}_2[x_a(t)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_2(\tau_1, \tau_2) x_a(t - \tau_1) x_a(t - \tau_2) d\tau_1 d\tau_2 \\
 &= \frac{A^2}{4} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_2(\tau_1, \tau_2) e^{j\omega(t-\tau_1)} e^{j\omega(t-\tau_2)} d\tau_1 d\tau_2 \\
 &= \frac{A^2}{4} e^{j2\omega t} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_2(\tau_1, \tau_2) e^{-j\omega\tau_1} e^{-j\omega\tau_2} d\tau_1 d\tau_2 \\
 &= \frac{A^2}{4} H_2(j\omega, j\omega) e^{j2\omega t}
 \end{aligned} \tag{B.6}$$

Similarly, the second term in (B.4) is

$$\mathbf{H}_2[x_b(t)] = \frac{A^2}{4} H_2(-j\omega, -j\omega) e^{-j2\omega t} \tag{B.7}$$

The third term in (B.4) is

$$\begin{aligned}
 \mathbf{H}_2\{x_a(t), x_b(t)\} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_2(\tau_1, \tau_2) x_a(t - \tau_1) x_b(t - \tau_2) d\tau_1 d\tau_2 \\
 &= \frac{A^2}{4} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_2(\tau_1, \tau_2) e^{j\omega(t-\tau_1)} e^{-j\omega(t-\tau_2)} d\tau_1 d\tau_2 \\
 &= \frac{A^2}{4} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_2(\tau_1, \tau_2) e^{-j\omega\tau_1} e^{j\omega\tau_2} d\tau_1 d\tau_2 \\
 &= \frac{A^2}{4} H_2(j\omega, -j\omega)
 \end{aligned} \tag{B.8}$$

Similarly,

$$\mathbf{H}_2\{x_b(t), x_a(t)\} = \frac{A^2}{4} H_2(-j\omega, j\omega) \tag{B.9}$$

Adding all four terms yields

$$\begin{aligned}
 y(t) &= \frac{A^2}{4} H_2(j\omega, j\omega) e^{j2\omega t} + \frac{A^2}{4} H_2(-j\omega, -j\omega) e^{-j2\omega t} \\
 &\quad + \frac{A^2}{4} H_2(j\omega, -j\omega) + \frac{A^2}{4} H_2(-j\omega, j\omega)
 \end{aligned} \tag{B.10}$$

$$= \frac{A^2}{2} \text{Re} [H_2(j\omega, j\omega) e^{j2\omega t}] + \frac{A^2}{2} \text{Re} [H_2(j\omega, -j\omega)] \tag{B.11}$$

B.2 Sinusoidal Response of a Third-Order Volterra Series System

The response of a third-order operator excited by a sinusoid, as previously defined by (B.1), is

$$\begin{aligned} y_3(t) &= \mathbf{H}_3[x(t)] \\ &= \mathbf{H}_3[x_a(t) + x_b(t)] \end{aligned} \quad (\text{B.12})$$

$$(\text{B.13})$$

The response of a third-order operator to an input that is the sum of functions is equal to the sum of the product operations on the inputs taken three at a time [28]. In other words, if

$$x(t) = \sum_{n=1}^N c_n x_n(t) \quad (\text{B.14})$$

then the corresponding response is

$$\begin{aligned} y_3(t) &= \mathbf{T}_3[x(t)] = \mathbf{T}_3 \left[\sum_{n=1}^N c_n x_n(t) \right] \\ &= \sum_{n_1=1}^N \sum_{n_2=1}^N \sum_{n_3=1}^N \mathbf{T}_3\{c_{n_1} x_{n_1}(t), c_{n_2} x_{n_2}(t), c_{n_3} x_{n_3}(t)\} \\ &= \sum_{n_1=1}^N \sum_{n_2=1}^N \sum_{n_3=1}^N c_{n_1} c_{n_2} c_{n_3} \mathbf{T}_3\{x_{n_1}(t), x_{n_2}(t), x_{n_3}(t)\} \end{aligned} \quad (\text{B.15})$$

The operator $\mathbf{T}\{x_a, x_b, x_c\}$ is called a trilinear operation. Guided by the expression

$$(a + b)^3 = a^3 + b^3 + 3a^2b + 2ab^2 \quad (\text{B.16})$$

which is the sum of all the ways of multiplying a and b together three at a time, the third-order response becomes

$$\begin{aligned}
 y_3(t) &= \mathbf{H}_3[x_a(t) + x_b(t)] \\
 &= \mathbf{H}_3[x_a(t)] + \mathbf{H}_3[x_b(t)] + 3\mathbf{H}_3\{x_a(t), x_a(t), x_b(t)\} \\
 &\quad + 3\mathbf{H}_3\{x_a(t), x_b(t), x_b(t)\}
 \end{aligned} \tag{B.17}$$

Considering each term, the first term is

$$\mathbf{H}_3[x_a(t)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_3(\tau_1, \tau_2, \tau_3) x_a(t - \tau_1) x_a(t - \tau_2) x_a(t - \tau_3) d\tau_1 d\tau_2 d\tau_3 \tag{B.18}$$

Substituting (B.2) yields

$$\begin{aligned}
 \mathbf{H}_3[x_a(t)] &= \left(\frac{A}{2}\right)^3 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_3(\tau_1, \tau_2, \tau_3) e^{j\omega(t-\tau_1)} e^{j\omega(t-\tau_2)} e^{j\omega(t-\tau_3)} d\tau_1 d\tau_2 d\tau_3 \\
 &= \left(\frac{A}{2}\right)^3 e^{j3\omega t} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_3(\tau_1, \tau_2, \tau_3) e^{-j\omega(\tau_1+\tau_2+\tau_3)} d\tau_1 d\tau_2 d\tau_3 \\
 &= \left(\frac{A}{2}\right)^3 H_3(j\omega, j\omega, j\omega) e^{j3\omega t}
 \end{aligned} \tag{B.19}$$

Similarly

$$\mathbf{H}_3[x_b(t)] = \left(\frac{A}{2}\right)^3 H_3(-j\omega, -j\omega, -j\omega) e^{-j3\omega t} \tag{B.20}$$

The third term in (B.17) is

$$\begin{aligned}
 3\mathbf{H}_3\{x_a(t), x_b(t), x_b(t)\} \\
 = 3 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_3(\tau_1, \tau_2, \tau_3) x_a(t - \tau_1) x_a(t - \tau_2) x_b(t - \tau_3) d\tau_1 d\tau_2 d\tau_3
 \end{aligned} \tag{B.21}$$

Substituting the expression for $x_a(t)$ and $x_b(t)$ yields

$$\begin{aligned}
& 3\mathbf{H}_3\{x_a(t), x_b(t), x_b(t)\} \\
&= 3 \left(\frac{A}{2}\right)^3 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_3(\tau_1, \tau_2, \tau_3) e^{j\omega(t-\tau_1)} e^{j\omega(t-\tau_2)} e^{j\omega(t-\tau_3)} d\tau_1 d\tau_2 d\tau_3 \\
&= 3 \left(\frac{A}{2}\right)^3 e^{j\omega t} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_3(\tau_1, \tau_2, \tau_3) e^{-j\omega(\tau_1+\tau_2-\tau_3)} d\tau_1 d\tau_2 d\tau_3 \\
&= 3 \left(\frac{A}{2}\right)^3 H_3(j\omega, j\omega, -j\omega) e^{j\omega t} \tag{B.22}
\end{aligned}$$

Similarly, the fourth term in (B.17) is

$$3\mathbf{H}_3\{x_a(t), x_b(t), x_b(t)\} = 3 \left(\frac{A}{2}\right)^3 H_3(-j\omega, -j\omega, j\omega) e^{j\omega t} \tag{B.23}$$

Since the first and second terms are conjugates and the third and fourth terms are also conjugates, the sum of all four terms yields

$$y_3(t) = \frac{A^3}{4} \text{Re}[H_3(j\omega, j\omega, j\omega)e^{j3\omega t}] + \frac{3A^3}{4} \text{Re}[H_3(j\omega, j\omega, -j\omega)e^{j\omega t}] \tag{B.24}$$

B.3 Synthesis of Volterra Operators

B.3.1 Basic Second-Order System

The most general block diagram of a second-order Volterra operator, which corresponds to the block $\mathbf{H}_2$ in Figure 3.1, is depicted in Figure B.1. This second-order system comprises of two linear blocks with impulse response $k_a(t)$ and $k_b(t)$ and an output of $z_a(t)$ and $z_b(t)$ respectively. The product of $z_a(t)$ $z_b(t)$ is fed to a linear system with an impulse response of $k_c(t)$ to produce an overall output of $y_2(t)$.

The second-order Volterra kernel of this general system is given by [28]

$$h_2(\tau_1, \tau_2) = \int_{-\infty}^{\infty} k_c(\sigma) k_a(\tau_1 - \sigma) k_b(\tau_2 - \sigma) d\sigma \tag{B.25}$$

Figure B.1: Basic second-order system.

The Fourier transform is [7]

$$H_2(j\omega_1, j\omega_2) = K_a(j\omega_1)K_b(j\omega_2)K_c(j\omega_1 + j\omega_2) \quad (\text{B.26})$$

where $K_a(j\omega)$, $K_b(j\omega)$, and $K_c(j\omega)$ are the Fourier transforms of the impulse responses $k_a(t)$, $k_b(t)$, and $k_c(t)$.

In a memory-less system, the block diagram simplifies to that depicted in Figure B.2, where $k_a(t)$ and $k_b(t)$ are replaced by simple through connections and $k_c(t)$ corresponds to a multiplication with a constant scale factor c_2 . As a result, $K_a(j\omega_1) = K_b(j\omega_2) = 1$, and $K_c(j\omega_1 + j\omega_2) = c_2$. Hence, from (B.26), the Fourier transform of the system simplifies to

$$H_2(j\omega_1, j\omega_2) = c_2 \quad (\text{B.27})$$

Figure B.2: Basic second-order memoryless system.

B.3.2 Basic Third-Order System

The most general third-order system that can be obtained with two multipliers is shown in Figure B.3. As depicted, the input, $x(t)$, is common to a second-order system, F_2 with an output of $z_a(t)$ and to the linear system with the impulse response $k_d(t)$ and an output of $z_b(t)$. The second-order system is identical to that depicted in Figure B.1. The product $z_a(t) z_b(t)$ is the input to the linear system with an impulse response of $k_e(t)$ and an output of $y_3(t)$.

Figure B.3: Basic third-order system.

In [28], the third-order transform is proven to be

$$H_3(j\omega_1, j\omega_2, j\omega_3) = K_a(j\omega_1)K_b(j\omega_2)K_c(j\omega_1 + j\omega_2)K_d(j\omega_3)K_e(j\omega_1 + j\omega_2 + j\omega_3) \quad (\text{B.28})$$

As was in the case with a second-order memoryless system, a third-order system is transformed to a memoryless system by replacing $k_a(t)$, $k_b(t)$, $k_c(t)$, and $k_d(t)$ with through connections and by replacing $k_e(t)$ with a constant scaling factor c_3 . The resulting block figure is shown in Figure B.4. The resulting third-order kernel transforms reduces to

$$H_3(j\omega_1, j\omega_2, j\omega_3) = c_3 \quad (\text{B.29})$$

Figure B.4: Basic third-order memoryless system.

B.4 Cascade Connection of Nonlinear Systems

Consider the cascade connection of two nonlinear systems such as shown in Figure B.5.

Figure B.5: The system $\mathbf{G}$ formed by the cascade connection of $\mathbf{H}$ and $\mathbf{F}$.

The response $y(t)$ of the system $\mathbf{H}$ to the input $x(t)$ is

$$\begin{aligned} y(t) &= \mathbf{H}[x(t)] \\ &= \sum_{n=1}^{\infty} \mathbf{H}_n[x(t)] \end{aligned} \quad (\text{B.30})$$

where $\mathbf{H}_n$ is an n th-order Volterra operator with the kernel $h_n(\tau_1, \dots, \tau_n)$. The response $z(t)$ of $\mathbf{F}$ for the input $y(t)$ is

$$\begin{aligned} z(t) &= \mathbf{F}[y(t)] \\ &= \sum_{n=1}^{\infty} \mathbf{F}_n[y(t)] \end{aligned} \quad (\text{B.31})$$

in which $\mathbf{F}_n$ is an n th-order Volterra operator with the kernel $f_n(\tau_1, \dots, \tau_n)$. Similarly, the response of the overall system, $\mathbf{G} = \mathbf{H}\mathbf{F}$, can be given by the Volterra series

$$\begin{aligned} z(t) &= \mathbf{G}[x(t)] \\ &= \sum_{n=1}^{\infty} \mathbf{G}_n[x(t)] \end{aligned} \quad (\text{B.32})$$

where $\mathbf{G}_n$ is an n th-order Volterra operator with the kernel $g_n(\tau_1, \dots, \tau_n)$. To determine the Volterra operators $\mathbf{G}_n$ in terms of $\mathbf{H}_n$ and $\mathbf{F}_n$, $x(t)$ is replaced by $cx(t)$ in equation (B.32), such that

$$z(t) = \mathbf{G}[cx(t)] = \sum_{n=1}^{\infty} c^n \mathbf{G}_n[x(t)] \quad (\text{B.33})$$

Equation (B.33) is a power series in c with $\mathbf{G}_n[x(t)]$ being the coefficients of c^n . Replacing $x(t)$ by $cx(t)$ and substituting (B.30) in (B.31) results in another power series of c . Since both series are equal, the Volterra operators $\mathbf{G}_n$ are determined by equating like powers of c in both expressions. Substituting (B.30) in (B.31) yields

$$z = \sum_{m=1}^{\infty} \mathbf{F}_m \left[\sum_{n=1}^{\infty} c^n y_n \right] \quad (\text{B.34})$$

where the argument t has been dropped for convenience and

$$y_n(t) = \mathbf{H}_n[x(t)] \quad (\text{B.35})$$

From the basic m -linear property of operator $\mathbf{F}_m$, given by [28]

$$\begin{aligned} \mathbf{F}_m \left[\sum_{n=1}^{\infty} c^n x_n \right] &= \sum_{n_1=1}^{\infty} \cdots \sum_{n_m=1}^{\infty} \mathbf{F}_m \{ c^{n_1} x_{n_1}, \dots, c^{n_m} x_{n_m} \} \\ &= \sum_{n_1=1}^{\infty} \cdots \sum_{n_m=1}^{\infty} c^{n_1} \cdots c^{n_m} \mathbf{F}_m \{ x_{n_1}, \dots, x_{n_m} \} \end{aligned} \quad (\text{B.36})$$

where $\mathbf{F}_m\{\cdot\}$ is the m -linear Volterra operator, equation (B.34) becomes

$$z = \sum_{m=1}^{\infty} \sum_{n_1=1}^{\infty} \cdots \sum_{n_m=1}^{\infty} c^{n_1+\cdots+n_m} \mathbf{F}_m\{y_{n_1}, \dots, y_{n_m}\} \quad (\text{B.37})$$

Equating the first power of c in (B.37) and (B.33) occurs for $m = 1$ and $n_1 = 1$ yields

$$\mathbf{G}_1 = \mathbf{F}_1\{y_1\} = \mathbf{F}_1[y_1] = \mathbf{F}_1\mathbf{H}_1[x] \quad (\text{B.38})$$

The coefficients of c^2 are obtained from (B.37) for $m = 1$ with $n_1 = 2$ and for $m = 2$ with $n_1 = n_2 = 1$ to yield

$$\begin{aligned} \mathbf{G}_2[x] &= \mathbf{F}_1\{y_2\} + \mathbf{F}_2\{y_1, y_1\} \\ &= \mathbf{F}_1[y_2] + \mathbf{F}_2[y_1] \\ &= \mathbf{F}_1\mathbf{H}_2[x] + \mathbf{F}_2\mathbf{H}_1[x] \end{aligned} \quad (\text{B.39})$$

The coefficients of c^3 are obtained from (B.37) for $m = 1$ with $n_1 = 3$, for $m = 2$ with $n_1 = 1$ and $n_2 = 2$ as well as $n_1 = 2$ and $n_2 = 1$, and for $m = 3$ with $n_1 = n_2 = n_3 = 1$. Thus these yield

$$\begin{aligned} \mathbf{G}_3[x] &= \mathbf{F}_1\{y_3\} + \mathbf{F}_2\{y_1, y_2\} + \mathbf{F}_2\{y_2, y_1\} + \mathbf{F}_3\{y_1, y_1, y_1\} \\ &= \mathbf{F}_1[y_3] + \mathbf{F}_2\{y_1, y_2\} + \mathbf{F}_2\{y_2, y_1\} + \mathbf{F}_3[y_1] \end{aligned} \quad (\text{B.40})$$

Since a Volterra operator can be considered symmetric with no loss of generality, then

$$\mathbf{F}_2\{y_1, y_2\} = \mathbf{F}_2\{y_2, y_1\} \quad (\text{B.41})$$

and (B.40) can be expressed as

$$\mathbf{G}_3[x] = \mathbf{F}_1[y_3] + 2\mathbf{F}_2\{y_1, y_2\} + \mathbf{F}_3[y_1] \quad (\text{B.42})$$

In addition, $2\mathbf{F}_2\{y_1, y_2\}$ can be rewritten as

$$2\mathbf{F}_2\{y_1, y_2\} = \mathbf{F}_2[y_1 + y_2] - \mathbf{F}_2[y_1] - \mathbf{F}_2[y_2] \quad (\text{B.43})$$

yielding $\mathbf{G}_3[x]$ as

$$\begin{aligned}\mathbf{G}_3[x] &= \mathbf{F}_1[y_3] + \mathbf{F}_2[y_1 + y_2] - \mathbf{F}_2[y_1] - \mathbf{F}_2[y_2] + \mathbf{F}_3[y_1] \\ &= \mathbf{F}_1\mathbf{H}_3[x] + \mathbf{F}_2[\mathbf{H}_1[x] + \mathbf{H}_2[x]] - \mathbf{F}_2\mathbf{H}_1[x] - \mathbf{F}_2\mathbf{H}_2[x] + \mathbf{F}_3\mathbf{H}_1[x]\end{aligned}\quad (\text{B.44})$$

Note that $\mathbf{F}_2[\mathbf{H}_1 + \mathbf{H}_2] \neq \mathbf{F}_2\mathbf{H}_1 + \mathbf{F}_2\mathbf{H}_2$. Summarizing the above results gives

$$\mathbf{G}_1 = \mathbf{F}_1\mathbf{H}_1 \quad (\text{B.45})$$

$$\mathbf{G}_2 = \mathbf{F}_1\mathbf{H}_2 + \mathbf{F}_2\mathbf{H}_1 \quad (\text{B.46})$$

$$\mathbf{G}_3 = \mathbf{F}_1\mathbf{H}_3 + \mathbf{F}_2[\mathbf{H}_1 + \mathbf{H}_2] - \mathbf{F}_2\mathbf{H}_1 - \mathbf{F}_2\mathbf{H}_2 + \mathbf{F}_3\mathbf{H}_1 \quad (\text{B.47})$$

B.4.1 Fourier Transforms

First-Order Transform

An easier interpretation can be obtained from the Fourier transform with the complex input signal

$$x(t) = Ae^{j\omega_1 t} \quad (\text{B.48})$$

Hence, the output at frequency ω_1 occurs only in the linear portion of $z(t)$ where

$$z_1(t) = AG_1(j\omega_1)e^{j\omega_1 t} \quad (\text{B.49})$$

However,

$$\begin{aligned}z_1(t) &= y_1(t)F_1(j\omega) \\ &= AH_1(j\omega_1)F_1(j\omega_1)e^{j\omega_1 t}\end{aligned}\quad (\text{B.50})$$

From equating (B.49) and (B.50), $G_1(j\omega_1)$ is easily determined as

$$G_1(j\omega_1) = H_1(j\omega_1)F_1(j\omega_1) \quad (\text{B.51})$$

Second-Order Transform

To determine the second-order transfer function, $G_2(j\omega_1, j\omega_2)$, the input signal is chosen to be

$$x(t) = A_1 e^{j\omega_1 t} + A_2 e^{j\omega_2 t} \quad (\text{B.52})$$

$G_2(j\omega_1, j\omega_2)$ relates the output component of $z(t)$ at $\omega_1 + \omega_2$ to the two input components at ω_1 and ω_2 . The output frequency at $\omega_1 + \omega_2$ occurs only in the second-order portion of $z(t)$. In [27], it is shown that the total output response of a nonlinear system can be expressed as

$$z(t) = \sum_{n=1}^N z_n(t) \quad (\text{B.53})$$

where $z_n(t)$ is the n th-order component of the response generated by the n th-order portion of the system and is given by

$$z_n(t) = \sum_{q_1=1}^Q \cdots \sum_{q_n=1}^Q A_{q_1} \cdots A_{q_n} H_n(j\omega_{q_1}, \dots, j\omega_{q_n}) e^{j(\omega_{q_1} + \cdots + \omega_{q_n})t} \quad (\text{B.54})$$

for an input defined as

$$x(t) = \sum_{q=1}^Q A_q e^{j\omega_q t} \quad (\text{B.55})$$

Hence, $z_2(t)$ is

$$z_2(t) = \sum_{q_1=1}^2 \sum_{q_2=1}^2 A_{q_1} A_{q_2} G_2(j\omega_{q_1}, j\omega_{q_2}) e^{j(\omega_{q_1} + \omega_{q_2})t} \quad (\text{B.56})$$

from which the term with the frequency component $\omega_1 + \omega_2$, denoted as $z_2(t; \omega_1 + \omega_2)$, is

$$z_2(t; \omega_1 + \omega_2) = 2A_1 A_2 G_2(j\omega_1, j\omega_2) e^{j(\omega_1 + \omega_2)t} \quad (\text{B.57})$$

Working backwards through the system, the output at $\omega_1 + \omega_2$ is due to:

1. the linear response of F to the component of $y(t)$ at $\omega_1 + \omega_2$; and
2. the quadratic response of F to the components of $y(t)$ at ω_1 and ω_2

If Y_{ω_k} is the complex amplitude of the ω_k component of $y(t)$, then $z_2(t : \omega_1 + \omega_2)$ can be expressed as

$$z_2(t : \omega_1 + \omega_2) = [Y_{\omega_1+\omega_2}F_1(j(\omega_1 + \omega_2)) + 2Y_{\omega_1}Y_{\omega_2}F_2(j\omega_1, j\omega_2)] e^{j(\omega_1+\omega_2)t} \quad (\text{B.58})$$

However, in terms of system H

$$\begin{aligned} Y_{\omega_1} &= A_1 H_1(j\omega_1) \\ Y_{\omega_2} &= A_2 H_1(j\omega_2) \\ Y_{\omega_1+\omega_2} &= 2A_1 A_2 H_2(j\omega_1, j\omega_2) \end{aligned} \quad (\text{B.59})$$

Substituting (B.59) into (B.58) yields

$$\begin{aligned} z_2(t : \omega_1 + \omega_2) &= 2A_1 A_2 [H_2(j\omega_1, j\omega_2)F_1(j\omega_1 + j\omega_2) \\ &\quad + H_1(j\omega_1)H_1(j\omega_2)F_2(j\omega_1, j\omega_2)] e^{j(\omega_1+\omega_2)t} \end{aligned} \quad (\text{B.60})$$

Comparing (B.57) with (B.60) it follows that

$$G_2(j\omega_1, j\omega_2) = H_2(j\omega_1, j\omega_2)F_1(j\omega_1 + j\omega_2) + H_1(j\omega_1)H_1(j\omega_2)F_2(j\omega_1, j\omega_2) \quad (\text{B.61})$$

Third-Order Transform

The third-order transfer function, $G_3(j\omega_1, j\omega_2, j\omega_3)$ is obtained in a similar manner as the second-order transfer function. The input signal is assumed to be

$$x(t) = A_1 e^{j\omega_1 t} + A_2 e^{j\omega_2 t} + A_3 e^{j\omega_3 t} \quad (\text{B.62})$$

$G_3(j\omega_1, j\omega_2, j\omega_3)$ is the nonlinear transfer function relating the $\omega_1 + \omega_2 + \omega_3$ frequency component of $z(t)$ to the component of $x(t)$ at ω_1 , ω_2 , and ω_3 . The output frequency at $\omega_1 + \omega_2 + \omega_3$ occurs only in the third-order portion of $z(t)$ in which

$$z_3(t) = \sum_{q_1=1}^3 \sum_{q_2=1}^3 \sum_{q_3=1}^3 A_{q_1} A_{q_2} A_{q_3} G_3(j\omega_{q_1}, j\omega_{q_2}, j\omega_{q_3}) e^{j(\omega_{q_1}+\omega_{q_2}+\omega_{q_3})t} \quad (\text{B.63})$$

The term with the frequency component $\omega_1 + \omega_2 + \omega_3$, denoted as $z_3(t; \omega_1 + \omega_2 + \omega_3)$, is

$$z_3(t; \omega_1 + \omega_2 + \omega_3) = 6A_1A_2A_3G_3(j\omega_1, j\omega_2, j\omega_3) e^{j(\omega_1 + \omega_2 + \omega_3)t} \quad (\text{B.64})$$

With respect to the system F , the output at $\omega_1 + \omega_2 + \omega_3$ is due to:

1. the linear response of F to the component of $y(t)$ at $\omega_1 + \omega_2 + \omega_3$;
2. all quadratic responses of F involving components of $y(t)$ at $\omega_j + \omega_k$ and ω_m , where $j \neq k \neq m$; and,
3. the cubic response of F to the components of $y(t)$ at ω_1, ω_2 , and ω_3 .

Hence, the output is

$$\begin{aligned} z_3(t; \omega_1 + \omega_2 + \omega_3) = & [Y_{\omega_1 + \omega_2 + \omega_3} F_1(j\omega_1 + j\omega_2 + j\omega_3) \\ & + 2Y_{\omega_1 + \omega_2} Y_{\omega_3} F_2(j\omega_1 + j\omega_2, j\omega_3) \\ & + 2Y_{\omega_1 + \omega_3} Y_{\omega_2} F_2(j\omega_1 + j\omega_3, j\omega_2) \\ & + 2Y_{\omega_2 + \omega_3} Y_{\omega_1} F_2(j\omega_2 + j\omega_3, j\omega_1) \\ & + 6Y_{\omega_1} Y_{\omega_2} Y_{\omega_3} F_3(j\omega_1, j\omega_2, j\omega_3)] e^{j(\omega_1 + \omega_2 + \omega_3)t} \end{aligned} \quad (\text{B.65})$$

where

$$\begin{aligned} Y_{\omega_1} &= A_1 H_1(j\omega_1) \\ Y_{\omega_2} &= A_2 H_1(j\omega_2) \\ Y_{\omega_3} &= A_2 H_1(j\omega_3) \\ Y_{\omega_1 + \omega_2} &= 2A_1 A_2 H_2(j\omega_1, j\omega_2) \\ Y_{\omega_1 + \omega_3} &= 2A_1 A_3 H_2(j\omega_1, j\omega_3) \\ Y_{\omega_2 + \omega_3} &= 2A_2 A_3 H_2(j\omega_2, j\omega_3) \\ Y_{\omega_1 + \omega_2 + \omega_3} &= 6A_1 A_2 A_3 H_3(j\omega_1, j\omega_2, j\omega_3) \end{aligned} \quad (\text{B.66})$$

As a result, (B.65) becomes

$$\begin{aligned}
 z_3(t : \omega_1 + \omega_2 + \omega_3) = & 6A_1A_2A_3\{H_3(j\omega_1, j\omega_2, j\omega_3)F_1(j\omega_1 + j\omega_2 + j\omega_3) \\
 & + \frac{2}{3}[H_1(j\omega_3)H_2(j\omega_1, j\omega_2)F_2(j\omega_1 + j\omega_2, j\omega_3) \\
 & + H_1(j\omega_2)H_2(j\omega_1, j\omega_3)F_2(j\omega_1 + j\omega_3, j\omega_2) \\
 & + H_1(j\omega_1)H_2(j\omega_2, j\omega_3)F_2(j\omega_2 + j\omega_3, j\omega_1)] \\
 & + H_1(j\omega_1)H_1(j\omega_2)H_1(j\omega_3)F_3(j\omega_1, j\omega_2, j\omega_3)\} \\
 & \cdot e^{j(\omega_1 + \omega_2 + \omega_3)t}
 \end{aligned} \tag{B.67}$$

Equating (B.64) and (B.67) gives

$$\begin{aligned}
 G_3(j\omega_1, j\omega_2, j\omega_3) = & H_3(j\omega_1, j\omega_2, j\omega_3)F_1(j\omega_1 + j\omega_2 + j\omega_3) \\
 & + H_1(j\omega_1)H_1(j\omega_2)H_1(j\omega_3)F_3(j\omega_1, j\omega_2, j\omega_3) \\
 & + \frac{2}{3}[H_1(j\omega_3)H_2(j\omega_1, j\omega_2)F_2(j\omega_1 + j\omega_2, j\omega_3) \\
 & + H_1(j\omega_2)H_2(j\omega_1, j\omega_3)F_2(j\omega_1 + j\omega_3, j\omega_2) \\
 & + H_1(j\omega_1)H_2(j\omega_2, j\omega_3)F_2(j\omega_2 + j\omega_3, j\omega_1)]
 \end{aligned} \tag{B.68}$$

B.5 Nonlinear Circuit Analysis with Volterra Series

In this section it is explained how to compute first, second, and third order Volterra kernels of voltages and currents in a weakly nonlinear circuit using the "Volterra series" [7] method. The method computes the Volterra kernels in increasing order by each time solving the same linear network that is the linearized equivalent of the original circuit. This linear circuit is used to compute the first-order kernels. The second-order kernels are solved using the same linearized network but with different inputs that depend on the first-order kernels. The third-order kernels are computed similarly, but with inputs that depend on the first and second-order kernels. Higher order kernels are computed in the same manner.

This method is explained using an example common-source amplifier circuit, depicted in Figure B.6(a). The source and bulk are grounded and the input to the amplifier is a voltage source with an output resistance R_S . The output is loaded with a resistance R_L and a capacitance C_L . A simplified AC-equivalent circuit is shown in Figure B.6(b). The transistor is assumed to be operating in the saturation region and so gate-source capacitance, C_{gs} can be assumed to be linear. The gate-drain and drain-bulk capacitances, C_{gd} and C_{db} , are neglected for simplicity. The drain current is assumed to be a function of v_{gs} only.

Figure B.6: (a) A common-source amplifier. (b) AC-equivalent circuit.

B.5.1 First-order Kernels

The first-order kernels are computed with the linearized circuit first with the matrix equation

$$\mathbf{Y}(j\omega) \cdot \mathbf{H}_1(j\omega) = \mathbf{I}_1 \quad (\text{B.69})$$

where $\mathbf{Y}$ is the admittance matrix of the circuit, $\mathbf{H}_1$ is the vector of first-order Volterra kernel transforms of the node voltages, and $\mathbf{I}_1$ is the vector of excitations.

Figure B.7: Linearized circuit for calculating first-order kernels.

The linearized circuit is shown in Figure B.7. By applying Kirchhoff's current law at nodes **1** and **2**, the equation (B.69) yields

$$\begin{bmatrix} g_S + j\omega C_{gs} & 0 \\ g_m & g_L + j\omega C_L \end{bmatrix} \cdot \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} g_S V_{in} \\ 0 \end{bmatrix} \quad (\text{B.70})$$

The node voltages, V_1 and V_2 , reduce to the transfer functions, denoted as $H_{1_1}(j\omega)$ and $H_{1_2}(j\omega)$, at nodes 1 and 2 when $V_{in} = 1$. The first subscript in the two transfer functions indicate the order of the transfer function and the second subscript corresponds to the node number in the circuit. As a result, the transfer function equations are given in the matrix form as

$$\begin{bmatrix} g_S + j\omega C_{gs} & 0 \\ g_m & g_L + j\omega C_L \end{bmatrix} \cdot \begin{bmatrix} H_{1_1}(j\omega) \\ H_{1_2}(j\omega) \end{bmatrix} = \begin{bmatrix} g_S \\ 0 \end{bmatrix} \quad (\text{B.71})$$

The set of equations are solved using Cramer's rule to yield

$$\begin{aligned} H_{1_1}(j\omega) &= \frac{1}{\det(j\omega)} \cdot \begin{vmatrix} g_S & 0 \\ 0 & g_L + j\omega C_L \end{vmatrix} = \frac{g_S(g_L + j\omega C_L)}{(g_S + j\omega C_{gs})(g_L + j\omega C_L)} \\ &= \frac{g_S}{g_S + j\omega C_{gs}} \end{aligned} \quad (\text{B.72})$$

and

$$\begin{aligned}
 H_{1_2}(j\omega) &= \frac{1}{\det(j\omega)} \cdot \begin{vmatrix} g_s + j\omega C_{gs} & g_m \\ g_s & 0 \end{vmatrix} \\
 &= \frac{-g_m g_s}{(g_s + j\omega C_{gs})(g_L + j\omega C_L)} \quad (\text{B.73})
 \end{aligned}$$

Although the solutions can be written directly by inspection, the above procedure allows for a systematic approach that will be more beneficial for larger circuits.

B.5.2 Second-order Kernels

The second-order responses are computed using the information from the first-order responses. The same linearized network is solved but with different inputs. The external signal sources are removed and *second-order nonlinear current sources* are applied for each basic nonlinearity in the circuit and placed in parallel with the linearized equivalent of the nonlinearity. The circuit that has to be solved for the computation of the second-order kernels in the example circuit is shown in Figure B.8. The original nonlinear circuit of Figure B.6(b) has one nonlinear transconductance as shown in Figure B.8, which gives rise to a nonlinear current source.

Figure B.8: Linearized circuit for calculating second-order kernels.

The value of the current sources that have to be applied depends on the type of the basic nonlinearity, on its second-order nonlinearity coefficient, and on the first-order kernel. Hence, the first-order kernels must be calculated prior to the second-order kernels. The expressions for the nonlinear current sources are given in B.1. H_{1_k}

is the first-order transfer function of the voltage that controls a one-dimensional nonlinearity or the first controlling voltage of a two-dimensional conductance. H_{1_m} is the first-order transfer function of the second voltage that controls the two-dimensional conductance. The details for the derivations of the expressions can be found in [7].

Table B.1: Nonlinear second-order current sources for different basic nonlinearities for the calculation of second order Volterra kernels. $H_{1_k}(\cdot)$ is the first-order transfer function of the controlling voltage of the transconductance and $H_{1_m}(\cdot)$ is the first-order transfer function of the controlling voltage of the output conductance. Taken from [7].

type of nonlinearity	expression for the second-order nonlinear current source
(trans)conductance	$c_{(2,0,0)} H_{1_k}(j\omega_1) H_{1_k}(j\omega_2)$
two-dimensional conductance (only cross-terms)	$\frac{1}{2} [c_{(1,1,0)} H_{1_k}(j\omega_1) H_{1_m}(j\omega_2) + c_{(1,1,0)} H_{1_k}(j\omega_2) H_{1_m}(j\omega_1)]$
three-dimensional conductance (only cross-terms)	0

The computation of the second-order kernel transforms are determined from

$$\mathbf{Y}(j\omega_1 + j\omega_2) \cdot \mathbf{H}_2(j\omega_1, j\omega_2) = \mathbf{I}_2 \quad (\text{B.74})$$

The vector $\mathbf{Y}(j\omega_1 + j\omega_2)$ is the admittance matrix at $\omega_1 + \omega_2$, $\mathbf{H}_2(j\omega_1, j\omega_2)$ is the vector of second-order kernel transforms of the node voltages, and $\mathbf{I}_2$ corresponds to the second-order nonlinear current sources. Applying Kirchhoff's current law gives

$$\begin{bmatrix} g_S + (j\omega_1 + j\omega_2)C_{gs} & 0 \\ g_m & g_L + (j\omega_1 + j\omega_2)C_L \end{bmatrix} \cdot \begin{bmatrix} H_{2_1}(j\omega_1, j\omega_2) \\ H_{2_2}(j\omega_1, j\omega_2) \end{bmatrix} = \begin{bmatrix} 0 \\ -i_{NL2_{ids}} \end{bmatrix} \quad (\text{B.75})$$

It can be observed that the left most matrix is the same as the matrix in (B.69) but evaluated at $(j\omega_1 + j\omega_2)$ and that $H_{2_1}(j\omega_1, j\omega_2)$ and $H_{2_2}(j\omega_1, j\omega_2)$ are the unknowns.

Since the external excitation was shorted out, and there are no other sources at the input, $v_{gs} = 0$ and the trivial solution $H_{2_1} = 0$ is obtained. This leaves

$$H_{2_2}(j\omega_1, j\omega_2) = \frac{-i_{NL2_{ids}}}{g_L + (j\omega_1 + j\omega_2)C_L} \quad (\text{B.76})$$

For a one-dimensional conductance, v_{gs} or V_1 controls the nonlinearity and so the value of $i_{NL2_{ids}}$, from Table B.1, is

$$i_{NL2_{ids}} = c_{(2,0,0)} H_{1_1}(j\omega_1) H_{1_1}(j\omega_2) \quad (\text{B.77})$$

Substituting (B.72), this yields

$$i_{NL2_{ids}} = \frac{c_{(2,0,0)} g_S^2}{(g_S + j\omega_1 C_{gs})(g_S + j\omega_2 C_{gs})} \quad (\text{B.78})$$

resulting in

$$H_{2_2}(j\omega_1, j\omega_2) = \frac{-c_{(2,0,0)} g_S^2}{(g_S + j\omega_1 C_{gs})(g_S + j\omega_2 C_{gs})(g_L + (j\omega_1 + j\omega_2)C_L)} \quad (\text{B.79})$$

B.5.3 Third-order Kernels

As in the case of second-order kernels, the third-order kernels are computed as the response of third-order nonlinear current sources and a similar matrix must be solved where

$$\mathbf{Y}(j\omega_1 + j\omega_2 + j\omega_3) \cdot \mathbf{H}_3(j\omega_1, j\omega_2, j\omega_3) = \mathbf{I}_3 \quad (\text{B.80})$$

The expressions for the third-order nonlinear current sources are given in Table B.2. H_{1_k} and H_{2_k} are the first and second-order transfer functions of the voltage that controls the one-dimensional or the first controlling voltage of a two-dimensional conductance. H_{1_m} and H_{2_m} are the first-order transfer functions of the second voltage that controls the two and three-dimensional conductances. H_{1_n} is the first-order transfer function of the third voltage that controls the three-dimensional conductance. As can be clearly seen, the expressions for the the third-order nonlinear current source are much more involved than the second-order.

Table B.2: Nonlinear third-order current sources for different basic nonlinearities for the calculation of second order Volterra kernels. $H_{1_k}(\cdot)$ and $H_{2_k}(\cdot)$ are the first-order and second-order transfer function of the controlling voltage of the transconductance. $H_{1_m}(\cdot)$ and $H_{2_m}(\cdot)$ are the first-order and second-order transfer function of the controlling voltage of the output conductance. $H_{1_n}(\cdot)$ and $H_{2_n}(\cdot)$ are the first-order and second-order transfer function of the controlling voltage of the bulk conductance. Taken from [7].

type of nonlinearity	expression for the second-order nonlinear current source
(trans)conductance	$c_{(3,0,0)}H_{1_k}(j\omega_1)H_{1_k}(j\omega_2)H_{1_k}(j\omega_3)$ $+\frac{2}{3}c_{(2,0,0)}[H_{1_k}(j\omega_1)H_{2_k}(j\omega_2, j\omega_3)$ $+H_{1_k}(j\omega_2)H_{2_k}(j\omega_1, j\omega_3)$ $+H_{1_k}(j\omega_3)H_{2_k}(j\omega_1, j\omega_2)]$
two-dimensional conductance (only cross-terms)	$\frac{1}{3}c_{(1,1,0)}[H_{1_K}(j\omega_1)H_{2_m}(j\omega_2, j\omega_3)$ $+H_{1_K}(j\omega_2)H_{2_m}(j\omega_1, j\omega_3)$ $+H_{1_K}(j\omega_3)H_{2_m}(j\omega_1, j\omega_2) + H_{2_K}(j\omega_1, j\omega_2)H_{1_m}(j\omega_3)$ $+H_{2_K}(j\omega_1, j\omega_3)H_{1_m}(j\omega_2) + H_{2_K}(j\omega_2, j\omega_3)H_{1_m}(j\omega_1)]$ $\frac{1}{3}c_{(2,1,0)}[H_{1_K}(j\omega_1)H_{1_K}(j\omega_2)H_{1_m}(j\omega_3)$ $+H_{1_K}(j\omega_1)H_{1_K}(j\omega_3)H_{1_m}(j\omega_2)$ $+H_{1_K}(j\omega_2)H_{1_K}(j\omega_3)H_{1_m}(j\omega_1)]$ $\frac{1}{3}c_{(1,2,0)}[H_{1_K}(j\omega_1)H_{1_m}(j\omega_2)H_{1_m}(j\omega_3)$ $+H_{1_K}(j\omega_2)H_{1_m}(j\omega_1)H_{1_m}(j\omega_3)$ $+H_{1_K}(j\omega_3)H_{1_m}(j\omega_1)H_{1_m}(j\omega_2)]$
three-dimensional conductance (only cross-terms)	$\frac{1}{3}c_{(1,1,1)}[H_{1_K}(j\omega_1)H_{1_m}(j\omega_2)H_{1_n}(j\omega_3)$ $+H_{1_K}(j\omega_1)H_{1_m}(j\omega_3)H_{1_n}(j\omega_2)$ $+H_{1_K}(j\omega_2)H_{1_m}(j\omega_1)H_{1_n}(j\omega_3)$ $+H_{1_K}(j\omega_2)H_{1_m}(j\omega_3)H_{1_n}(j\omega_1)$ $+H_{1_K}(j\omega_3)H_{1_m}(j\omega_1)H_{1_n}(j\omega_2)$ $+H_{1_K}(j\omega_3)H_{1_m}(j\omega_2)H_{1_n}(j\omega_3)]$

Applying Kirchhoff's current law gives

$$\begin{bmatrix} g_S + j(\omega_1 + \omega_2 + \omega_3)C_{gs} & 0 \\ g_m & g_L + j(\omega_1 + \omega_2 + \omega_3)C_L \end{bmatrix} \cdot \begin{bmatrix} H_{3_1}(j\omega_1, j\omega_2, j\omega_3) \\ H_{3_2}(j\omega_1, j\omega_2, j\omega_3) \end{bmatrix} = \begin{bmatrix} 0 \\ -i_{NL3_{ids}} \end{bmatrix} \quad (\text{B.81})$$

Since V_1 controls the nonlinearity and $H_{3_1} = 0$, the value of $i_{NL3_{ids}}$, from Table B.2, is

$$\begin{aligned} i_{NL3_{ids}} &= c_{(3,0,0)} H_{1_1}(j\omega_1) H_{1_1}(j\omega_2) H_{1_1}(j\omega_3) \\ &= \frac{c_{(3,0,0)} g_S^3}{(g_S + j\omega_1 C_{gs})(g_S + j\omega_2 C_{gs})(g_S + j\omega_3 C_{gs})} \end{aligned}$$

resulting in

$$H_{3_2}(j\omega_1, j\omega_2, j\omega_3) = \frac{i_{NL3_{ids}}}{(g_L + (j\omega_1 + j\omega_2 + j\omega_3)C_L)} \quad (\text{B.82})$$

Hence, the third-order intermodulation product can be found from (3.18) as

$$IM_3 = \frac{3}{4} \frac{c_{(3,0,0)} g_S^2 (g_L + j\omega_1 C_l)}{g_m (g_L + j(\omega_1 + \omega_2)C_L) (g_S - j\omega_1 C_{gs})^2} \quad (\text{B.83})$$

Appendix C

Evaluation of Derivatives of I_{ds}

C.1 Derivatives for $c_{(1,0,0)}$, $c_{(2,0,0)}$, and $c_{(3,0,0)}$

The following are the expressions for the derivatives with respect to V_{gs} of *long*, given by (4.44), *mobhot*, given by (4.45) and *clm*, given by (4.46), required to calculate $c_{(1,0,0)}$, $c_{(2,0,0)}$, and $c_{(3,0,0)}$. To reduce the possibilities of errors, the derivatives were calculated symbolically using Mathematica 5.0 and exported to Latex format using the "TeXForm" function.

$$I_{d0} = \mu_0 C'_{ox} 2n\phi_t^2 \frac{W_{eff}}{L_{eff}} \quad (C.1)$$

$$z = e^{(V_{gs} - V_{th})/(2n\phi_t)} \quad (C.2)$$

$$\begin{aligned} V_{DS(sat)}^{(1,0)} &= \frac{\partial V_{DS(sat)}}{\partial V_{gs}} \\ &= \frac{A_{bulk} \mathcal{E}_{sat}^2 L_{eff}^2 z}{(1+z)(A_{bulk} \mathcal{E}_{sat} L_{eff} + 2n\phi_t \log(1+z))^2} \end{aligned} \quad (C.3)$$

$$\begin{aligned} V_{DS(sat)}^{(2,0)} &= \frac{\partial^2 V_{DS(sat)}}{\partial V_{gs}^2} \\ &= \frac{A_{bulk} \mathcal{E}_{sat}^2 L_{eff}^2 z (A_{bulk} \mathcal{E}_{sat} L_{eff} - 4n\phi_t z + 2n\phi_t \log(1+z))}{2n\phi_t (1+z)^2 (A_{bulk} \mathcal{E}_{sat} L_{eff} + 2n\phi_t \log(1+z))^3} \end{aligned} \quad (C.4)$$

$$\begin{aligned}
V_{DS(sat)}^{(3,0)} &= \frac{\partial^3 V_{DS(sat)}}{\partial V_{gs}^3} \\
&= \frac{-A_{bulk} \mathcal{E}_{sat}^2 L_{eff}^2 z}{4n^2 \phi_t^2 (1+z)^3 (A_{bulk} \mathcal{E}_{sat} L_{eff} + 2n\phi_t \log(1+z))^4} \\
&\quad \cdot \left(A_{bulk}^2 \mathcal{E}_{sat}^2 L_{eff}^2 (z-1) + 12A_{bulk} \mathcal{E}_{sat} L_{eff} n\phi_t z - 24n^2 \phi_t^2 z^2 \right. \\
&\quad + 4n\phi_t (A_{bulk} \mathcal{E}_{sat} L_{eff} (z-1) + 6n\phi_t z) \log(1+z) \\
&\quad \left. + 4n^2 \phi_t^2 (z-1) \log(1+z)^2 \right) \quad (C.5)
\end{aligned}$$

$$\begin{aligned}
V_A^{(1,0)} &= \frac{\partial V_A}{\partial V_{gs}} \\
&= \frac{z P_{VAG} V_{ACLM}}{\mathcal{E}_{sat} L_{eff} (1+z)} \quad (C.6)
\end{aligned}$$

$$\begin{aligned}
V_A^{(2,0)} &= \frac{\partial^2 V_A}{\partial V_{gs}^2} \\
&= \frac{z P_{VAG} V_{ACLM}}{2\mathcal{E}_{sat} L_{eff} n\phi_t (1+z)^2} \quad (C.7)
\end{aligned}$$

$$\begin{aligned}
V_A^{(3,0)} &= \frac{\partial^3 V_A}{\partial V_{gs}^3} \\
&= \frac{-z P_{VAG} V_{ACLM} (z-1)}{4\mathcal{E}_{sat} L_{eff} n^2 \phi_t^2 (1+z)^3} \quad (C.8)
\end{aligned}$$

$$\frac{\partial long}{\partial V_{gs}} = I_{d0} \frac{z \log(1+z)}{n\phi_t (1+z)} \quad (C.9)$$

$$\frac{\partial^2 long}{\partial V_{gs}^2} = I_{d0} \frac{z(z + \log(1+z))}{2n^2 \phi_t^2 (1+z)^2} \quad (C.10)$$

$$\frac{\partial^3 long}{\partial V_{gs}^3} = I_{d0} \frac{z(3z - (z-1)\log(1+z))}{4n^3 \phi_t^3 (1+z)^3} \quad (C.11)$$

$$\frac{\partial mobhot}{\partial V_{gs}} = -\frac{z\alpha}{(1+z)(1+2n\phi_t\alpha\log(1+z))^2} \quad (C.12)$$

$$\frac{\partial^2 mobhot}{\partial V_{gs}^2} = \frac{z\alpha(-1 + 4nz\alpha\phi_t - 2n\phi_t\alpha\log(1+z))}{2n\phi_t(1+z)^2(1+2n\phi_t\alpha\log(1+z))^3} \quad (C.13)$$

$$\begin{aligned} \frac{\partial^3 mobhot}{\partial V_{gs}^3} = & \frac{z\alpha}{4n^2\phi_t^2(1+z)^3(1+2n\phi_t\alpha\log(1+z))^4} \left(z - 1 + 12n\phi_t\alpha z \right. \\ & - 24n^2\phi_t^2\alpha^2z^2 + 4n\phi_t\alpha(z-1+6n\phi_t\alpha z)\log(1+z) \\ & \left. + 4n^2\phi_t^2(z-1)\alpha^2\log(1+z)^2 \right) \end{aligned} \quad (C.14)$$

$$\frac{\partial clm}{\partial V_{gs}} = \frac{(-V_{ds} + V_{DS(sat)}) V_A^{(1,0)} - V_A V_{DS(sat)}^{(1,0)}}{V_A^2} \quad (C.15)$$

$$\begin{aligned} \frac{\partial^2 clm}{\partial V_{gs}^2} = & \frac{1}{V_A^3} \left(2(V_{ds} - V_{DS(sat)}) V_A^{(1,0)2} + 2V_A V_A^{(1,0)} V_{DS(sat)}^{(1,0)} \right. \\ & \left. - V_A \left((V_{ds} - V_{DS(sat)}) V_A^{(2,0)} + V_A V_{DS(sat)}^{(2,0)} \right) \right) \end{aligned} \quad (C.16)$$

$$\begin{aligned} \frac{\partial^3 clm}{\partial V_{gs}^3} = & \frac{1}{V_A^4} \left(-6(V_{ds} - V_{DS(sat)}) V_A^{(1,0)3} - 6V_A V_A^{(1,0)2} V_{DS(sat)}^{(1,0)} \right. \\ & + 3V_A V_A^{(1,0)} \left(2(V_{ds} - V_{DS(sat)}) V_A^{(2,0)} + V_A V_{DS(sat)}^{(2,0)} \right) \\ & + V_A^2 \left(3V_{DS(sat)}^{(1,0)} V_A^{(2,0)} + (-V_{ds} + V_{DS(sat)}) V_A^{(3,0)} \right. \\ & \left. \left. - V_A V_{DS(sat)}^{(3,0)} \right) \right) \end{aligned} \quad (C.17)$$

C.2 Derivatives for $c_{(0,1,0)}$, $c_{(0,2,0)}$, and $c_{(0,3,0)}$

The expressions for the derivatives of *long*, *mobhot*, and *clm*, with respect to V_{ds} are trivial since *long* and *mobhot* have been modelled as independent of V_{ds} and *clm* has been modelled as a linear function of V_{ds} . Nonetheless, the expressions are given here for completeness.

$$\frac{\partial long}{\partial V_{ds}} = 0 \quad (C.18)$$

$$\frac{\partial^2 long}{\partial V_{ds}^2} = 0 \quad (C.19)$$

$$\frac{\partial^3 long}{\partial V_{ds}^3} = 0 \quad (C.20)$$

$$\frac{\partial mobhot}{\partial V_{ds}} = 0 \quad (C.21)$$

$$\frac{\partial^2 mobhot}{\partial V_{ds}^2} = 0 \quad (C.22)$$

$$\frac{\partial^3 mobhot}{\partial V_{ds}^3} = 0 \quad (C.23)$$

$$\frac{\partial clm}{\partial V_{ds}} = \frac{1}{V_A} \quad (C.24)$$

$$\frac{\partial^2 clm}{\partial V_{ds}^2} = 0 \quad (C.25)$$

$$\frac{\partial^3 clm}{\partial V_{ds}^3} = 0 \quad (C.26)$$

C.3 Derivatives for $c_{(0,0,1)}$, $c_{(0,0,2)}$, and $c_{(0,0,3)}$

The following are the expressions for the derivatives with respect to V_{gs} of *long*, given by (4.44), *mobhot*, given by (4.45) and *clm*, given by (4.46), required to calculate $c_{(1,0,0)}$, $c_{(2,0,0)}$, and $c_{(3,0,0)}$. To reduce the possibilities of errors, the derivatives were calculated symbolically using Mathematica 5.0 and exported to Latex format using the "TeXForm" function.

$$\begin{aligned} V'_{th} &= \frac{\partial V_{th}}{\partial V_{sb}} \\ &= \frac{\gamma}{2\sqrt{V_{sb} + \phi}} \end{aligned} \quad (C.27)$$

$$\begin{aligned} V''_{th} &= \frac{\partial^2 V_{th}}{\partial V_{sb}^2} \\ &= \frac{-\gamma}{4(V_{sb} + \phi)^{\frac{3}{2}}} \end{aligned} \quad (C.28)$$

$$\begin{aligned} V^{(3)}_{th} &= \frac{\partial^3 V_{th}}{\partial V_{sb}^3} \\ &= \frac{3\gamma}{8(V_{sb} + \phi)^{\frac{5}{2}}} \end{aligned} \quad (C.29)$$

$$\begin{aligned} n' &= \frac{\partial n}{\partial V_{sb}} \\ &= \frac{-\gamma}{4(V_{sb} + \phi)^{\frac{3}{2}}} \end{aligned} \quad (C.30)$$

$$\begin{aligned} n'' &= \frac{\partial^2 n}{\partial V_{sb}^2} \\ &= \frac{3\gamma}{8(V_{sb} + \phi)^{\frac{5}{2}}} \end{aligned} \quad (C.31)$$

$$\begin{aligned} n^{(3)} &= \frac{\partial^3 n}{\partial V_{sb}^3} \\ &= \frac{-15\gamma}{16(V_{sb} + \phi)^{\frac{7}{2}}} \end{aligned} \quad (C.32)$$

$$\begin{aligned}
V_{DS(sat)}^{(0,1)} &= \frac{\partial V_{DS(sat)}}{\partial V_{sb}} \\
&= \frac{A_{bulk} \mathcal{E}_{sat}^2 L_{eff}^2}{(1+z) (A_{bulk} \mathcal{E}_{sat} L_{eff} + 2\phi \log(1+z) n)^2} \left(2\phi (-z \log(z) \right. \\
&\quad \left. + (1+z) \log(1+z)) n' - z V'_{th} \right) \quad (C.33)
\end{aligned}$$

$$\begin{aligned}
V_{DS(sat)}^{(0,2)} &= \frac{\partial^2 V_{DS(sat)}}{\partial V_{sb}^2} \\
&= \frac{A_{bulk} \mathcal{E}_{sat}^2 L_{eff}^2}{2(1+z)^2 \phi n (A_{bulk} \mathcal{E}_{sat} L_{eff} + 2\phi \log(1+z) n)^3} \\
&\quad \cdot \left((4\phi^2 (A_{bulk} \mathcal{E}_{sat} L z \log(z))^2 \right. \\
&\quad + 2\phi (4z (1+z) \log(z) \log(1+z) \\
&\quad - 2(1+z)^2 \log(1+z)^2 + z \log(z)^2 (-2z + \log(1+z))) n) n'^2 \quad (C.34) \\
&\quad + 4z\phi (A_{bulk} \mathcal{E}_{sat} L \log(z) + 2\phi (2(1+z) \log(1+z) \\
&\quad + \log(z) (-2z + \log(1+z))) n) n' V'_{th} + z (A_{bulk} \mathcal{E}_{sat} L_{eff} \\
&\quad + 2\phi (-2z + \log(1+z)) n) V_{th}'^2 + 2(1+z) \phi n (A_{bulk} \mathcal{E}_{sat} L_{eff} \\
&\quad + 2\phi \log(1+z) n) (2\phi (-z \log(z)) \\
&\quad \left. + (1+z) \log(1+z)) n'' - z V''_{th} \right)
\end{aligned}$$

$$\begin{aligned}
V_{DS(sat)}^{(0,3)} &= \frac{\partial^3 V_{DS(sat)}}{\partial V_{sb}^3} \\
&= \frac{A_{bulk} \mathcal{E}_{sat}^2 L_{eff}^2}{4(1+z)^3 \phi^2 n^2 (A_{bulk} \mathcal{E}_{sat} L_{eff} + 2\phi \log(1+z)n)^4} \\
&\quad \left(-8\phi^3 (A_{bulk}^2 \mathcal{E}_{sat}^2 L_{eff}^2 z \log(z)^2 (3+3z+\log(z)-z \log(z)) \right. \\
&\quad -4A_{bulk} \mathcal{E}_{sat} L z \phi \log(z)^2 (-6(1+z) \log(1+z) \\
&\quad +\log(z) (3z+(z-1) \log(1+z))) n \\
&\quad +4\phi^2 (-9z(1+z) \log(z)^2 (2z-\log(1+z)) \log(1+z) \\
&\quad +18z(1+z)^2 \log(z) \log(1+z)^2 -6(1+z)^3 \log(1+z)^3 \\
&\quad +z \log(z)^3 (6z^2-6z \log(1+z)-(z-1) \log(1+z)^2)) n^2) n'^3 \\
&\quad -12z\phi^2 (A_{bulk}^2 \mathcal{E}_{sat}^2 L_{eff}^2 \log(z) (2+2z+\log(z)-z \log(z)) \\
&\quad -4A_{bulk} \mathcal{E}_{sat} L \phi \log(z) (-4(1+z) \log(1+z) \\
&\quad +\log(z) (3z+(z-1) \log(1+z))) n \\
&\quad +4\phi^2 (-6(1+z) \log(z) (2z-\log(1+z)) \log(1+z) \\
&\quad +6(1+z)^2 \log(1+z)^2 +\log(z)^2 (6z^2-6z \log(1+z) \\
&\quad -(z-1) \log(1+z)^2)) n^2) n'^2 V_{th}' +z (A_{bulk}^2 \mathcal{E}_{sat}^2 L_{eff}^2 (z-1) \\
&\quad +4A_{bulk} \mathcal{E}_{sat} L \phi (3z+(z-1) \log(1+z)) n \\
&\quad -4\phi^2 (6z^2-6z \log(1+z)-(z-1) \log(1+z)^2) n^2) V_{th}'^3 \\
&\quad +6z(1+z) \phi n (A_{bulk} \mathcal{E}_{sat} L_{eff} \\
&\quad +2\phi \log(1+z)n) V_{th}' (2\phi (A_{bulk} \mathcal{E}_{sat} L \log(z) \\
&\quad +2\phi (2(1+z) \log(1+z) +\log(z) (-2z+\log(1+z))) n) n'' \\
&\quad + (A_{bulk} \mathcal{E}_{sat} L_{eff} +2\phi (-2z+\log(1+z)) n) V_{th}'' \\
&\quad +6\phi n' (- (z (A_{bulk}^2 \mathcal{E}_{sat}^2 L_{eff}^2 (1+z+\log(z)-z \log(z)) \\
&\quad +4A_{bulk} \mathcal{E}_{sat} L \phi (2(1+z) \log(1+z) +\log(z) (-3z \\
&\quad -(z-1) \log(1+z))) n +4\phi^2 (-3(1+z) (2z \\
&\quad -\log(1+z)) \log(1+z) +\log(z) (6z^2-6z \log(1+z) \\
&\quad -(z-1) \log(1+z)^2)) n^2) V_{th}'^2) \dots
\end{aligned} \tag{C.35}$$

$$\begin{aligned}
V_{DS(sat)}^{(0,3)} = & \dots + 2 (1+z) \phi n (A_{bulk} \mathcal{E}_{sat} L_{eff} \\
& + 2 \phi \log(1+z) n) (2 \phi (A_{bulk} \mathcal{E}_{sat} L_{eff} z \log(z)^2 \\
& + 2 \phi (4 z (1+z) \log(z) \log(1+z) - 2 (1+z)^2 \log(1+z)^2 \\
& + z \log(z)^2 (-2 z + \log(1+z))) n) n'' + z (A_{bulk} \mathcal{E}_{sat} L \log(z) \\
& + 2 \phi (2 (1+z) \log(1+z) + \log(z) (-2 z + \log(1+z))) n) V_{th}'') \\
& + 4 (1+z)^2 \phi^2 n^2 (A_{bulk} \mathcal{E}_{sat} L_{eff} \\
& + 2 \phi \log(1+z) n)^2 (2 \phi (- (z \log(z)) \\
& + (1+z) \log(1+z)) n^{(3)} - z V_{th}^3) \Big)
\end{aligned}$$

$$\begin{aligned}
V_A^{(0,1)} &= \frac{\partial V_A}{\partial V_{sb}} \\
&= - \left(\frac{P_{VAG} V_{ACLM} (2 \phi (z \log(z) - (1+z) \log(1+z)) n' + z V_{th}')}{\mathcal{E}_{sat} L (1+z)} \right) \quad (C.36)
\end{aligned}$$

$$\begin{aligned}
V_A^{(0,2)} &= \frac{\partial^2 V_A}{\partial V_{sb}^2} \\
&= - \frac{P_{VAG} V_{ACLM}}{2 \mathcal{E}_{sat} L (1+z)^2 \phi n} \left(- 4 z \phi^2 \log(z)^2 n'^2 - 4 z \phi \log(z) n' V_{th}' \right. \\
&\quad - z V_{th}'^2 + 2 (1+z) \phi n (2 \phi (z \log(z) - (1+z) \log(1+z)) n'' \\
&\quad \left. + z V_{th}'' \right) \Big) \quad (C.37)
\end{aligned}$$

$$\begin{aligned}
V_A^{(0,3)} &= \frac{\partial^3 V_A}{\partial V_{sb}^3} \\
&= -\frac{P_{VAG} V_{ACLM}}{4 \mathcal{E}_{sat} L (1+z)^3 \phi^2 n^2} \left(8 z \phi^3 \log(z)^2 (3 + 3 z + \log(z)) \right. \\
&\quad - z \log(z) n'^3 - 12 z \phi^2 \log(z) (-2 (1+z) \\
&\quad + (z-1) \log(z)) n'^2 V'_{th} - (z-1) z V_{th}'^3 \\
&\quad - 6 z (1+z) \phi n V'_{th} (2 \phi \log(z) n'' + V''_{th}) \\
&\quad - 6 z \phi n' \left((-1-z + (z-1) \log(z)) V_{th}'^2 \right. \\
&\quad + 2 (1+z) \phi \log(z) n (2 \phi \log(z) n'' + V''_{th})) \\
&\quad + 4 (1+z)^2 \phi^2 n^2 (2 \phi (z \log(z) \\
&\quad \left. - (1+z) \log(1+z)) n^{(3)} + z V_{th}^3) \right) \quad (C.38)
\end{aligned}$$

$$\frac{\partial long}{\partial V_{sb}} = \frac{I_{d0} \log(1+z)}{(1+z) \phi n} \left(\phi (-2 z \log(z) + (1+z) \log(1+z)) n' - z V_{th}' \right) \quad (C.39)$$

$$\begin{aligned}
\frac{\partial^2 long}{\partial V_{sb}^2} &= \frac{I_{d0}}{2 (1+z)^2 \phi^2 n^2} \left(4 z \phi^2 \log(z)^2 (z + \log(1+z)) n'^2 \right. \\
&\quad + 4 z \phi \log(z) (z + \log(1+z)) n' V'_{th} + z (z + \log(1+z)) V_{th}'^2 \\
&\quad + 2 (1+z) \phi \log(1+z) n \left[\phi (-2 z \log(z) + \right. \\
&\quad \left. (1+z) \log(1+z)) n'' - z V_{th}'' \right] \left. \right) \quad (C.40)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^3 long}{\partial V_{sb}^3} = & - \frac{I_{d0}}{4(1+z)^3 \phi^3 n^3} \left(8z \phi^3 \log(z)^2 \left[3(1+z)(z + \log(1+z)) \right. \right. \\
& + \log(z) \left(3z - (z-1) \log(1+z) \right) \left. \right] n'^3 \\
& + 12z \phi^2 \log(z) \left[2(1+z)(z + \log(1+z)) \right. \\
& + \log(z) \left(3z - (z-1) \log(1+z) \right) \left. \right] n'^2 V'_{th} \\
& - z \left(-3z + (z-1) \log(1+z) \right) V_{th}'^3 \\
& - 6z(1+z) \phi (z + \log(1+z)) n V'_{th} (2\phi \log(z) n'' + V''_{th}) \\
& - 6z \phi n' \left[- \left((1+z)(z + \log(1+z)) \right. \right. \\
& + \log(z) \left(3z - (z-1) \log(1+z) \right) \left. \right] V_{th}'^2 \\
& + 2(1+z) \phi \log(z) (z + \log(1+z)) n (2\phi \log(z) n'' + V''_{th}) \left. \right] \\
& - 4(1+z)^2 \phi^2 \log(1+z) n^2 \left[\phi (-2z \log(z) \right. \\
& + (1+z) \log(1+z)) n^{(3)} - z V_{th}^{(3)} \left. \right] \Big) \tag{C.41}
\end{aligned}$$

$$\frac{\partial mobhot}{\partial V_{sb}} = \frac{\alpha \left(2\phi [z \log(z) - (1+z) \log(1+z)] n' + z V'_{th} \right)}{(1+z) \left(1 + 2\alpha \phi \log(1+z) n \right)^2} \tag{C.42}$$

$$\begin{aligned}
\frac{\partial^2 mobhot}{\partial V_{sb}^2} = & \frac{\alpha}{2(1+z)^2 \phi n \left(1 + 2\alpha \phi \log(1+z) n\right)^3} \left(4\phi^2 \left[-z \log(z)^2 \right. \right. \\
& + 2\alpha \phi n \left[z \log(z)^2 (2z - \log(1+z)) \right. \\
& \left. \left. - 4z(1+z) \log(z) \log(1+z) + 2(1+z)^2 \log(1+z)^2 \right] \right] n'^2 \\
& + 4z\phi n \left[-\log(z) + 2\alpha \phi \left[\log(z) (2z - \log(1+z)) \right. \right. \\
& \left. \left. - 2(1+z) \log(1+z) \right] \right] n' V'_{th} \\
& + z \left(-1 + 2\alpha \phi n (2z - \log(1+z)) \right) V_{th}'^2 \\
& + 2\phi n (1+z) (1 + 2\alpha \phi n \log(1+z)) \left[2\phi \left[z \log(z) \right. \right. \\
& \left. \left. - (1+z) \log(1+z) \right] n'' + z V''_{th} \right] \Bigg) \tag{C.43}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^3 \text{mobhot}}{\partial V_{sb}^3} = & \frac{\alpha}{4(1+z)^3(1+2\alpha\phi\log(1+z)n)^4} \left(24\alpha^2 \left(2\phi \left[z \log(z) \right. \right. \right. \\
& - (1+z) \log(1+z) \left. \left. \right] n' + z V_{th}' \right)^3 \\
& + \frac{12\alpha(1+2\alpha\phi n \log(1+z))}{\phi n} \left[\left(2\phi n' \left[-z \log(z) \right. \right. \right. \\
& + (1+z) \log(1+z) \left. \left. \right] - z V_{th}' \right) \left(4z\phi^2 \log(z)^2 n'^2 \right. \\
& + 4z\phi \log(z) n' V_{th}' + z V_{th}'^2 - 2(1+z)\phi n \left[2\phi \left[z \log(z) \right. \right. \right. \\
& - (1+z) \log(1+z) \left. \left. \right] n'' + z V_{th}'' \left. \right] \right) \\
& - \frac{(1+2\alpha\phi n \log(1+z))^2}{\phi^2 n^2} \left[8z\phi^3 n'^3 \log(z)^2 \left[-3(1+z) \right. \right. \\
& + (z-1) \log(z) \left. \left. \right] + 12z\phi^2 \log(z) \left[-2(1+z) \right. \right. \\
& + (z-1) \log(z) \left. \left. \right] n'^2 V_{th}' + (z-1) z V_{th}'^3 \right. \\
& + 6z(1+z)\phi n V_{th}' (2\phi \log(z) n'' + V_{th}'') \\
& + 6z\phi n' \left[(-1-z + (z-1) \log(z)) V_{th}'^2 \right. \\
& + 2(1+z)\phi \log(z) n (2\phi \log(z) n'' + V_{th}'') \left. \right] \\
& - 4(1+z)^2 \phi^2 n^2 \left(2\phi (z \log(z) \right. \\
& \left. \left. - (1+z) \log(1+z)) n^{(3)} + z V_{th}^{(3)} \right) \right] \Big) \quad (C.44)
\end{aligned}$$

$$\frac{\partial \text{clm}}{\partial V_{sb}} = \frac{(-V_{ds} + V_{DS(sat)}) V_A^{(0,1)} - V_A V_{DS(sat)}^{(0,1)}}{V_A^2} \quad (C.45)$$

$$\begin{aligned}
\frac{\partial^2 \text{clm}}{\partial V_{sb}^2} = & \frac{1}{V_A^3} \left(2(V_{ds} - V_{DS(sat)}) V_A^{(0,1)2} + 2V_A V_A^{(0,1)} V_{DS(sat)}^{(0,1)} \right. \\
& \left. - V_A \left((V_{ds} - V_{DS(sat)}) V_A^{(0,2)} + V_A V_{DS(sat)}^{(0,2)} \right) \right) \quad (C.46)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^3 clm}{\partial V_{sb}^3} = & \frac{1}{V_A^4} \left(-6 (V_{ds} - V_{DS(sat)}) V_A^{(0,1)3} - 6 V_A V_A^{(0,1)2} V_{DS(sat)}^{(0,1)} \right. \\
& + 3 V_A V_A^{(0,1)} \left(2 (V_{ds} - V_{DS(sat)}) V_A^{(0,2)} + V_A V_{DS(sat)}^{(0,2)} \right) \\
& + V_A^2 \left(3 V_{DS(sat)}^{(0,1)} V_A^{(0,2)} + (-V_{ds} + V_{DS(sat)}) V_A^{(0,3)} \right. \\
& \left. \left. - V_A V_{DS(sat)}^{(0,3)} \right) \right)
\end{aligned} \tag{C.47}$$

References

- [1] R. Van Langevelde and F. M. Klassen, "Effect of gate-field dependent mobility degradation on distortion analysis in MOSFET's," *IEEE Transactions on Electron Devices*, vol. 44, no. 11, pp. 2044–2052, Nov. 1997.
- [2] P. H. Woerlee, M. J. Knitel, R. Van Langevelde, D. B. Klassen, L. F. Tiemeijer, A. J. Scholten, and A. T. Zegers van Duijnhoven, "RF-CMOS-performance trends," *IEEE Transactions on Electron Devices*, vol. 48, no. 8, pp. 1776–1781, Aug. 2001.
- [3] S. A. Maas and D. Neilson, "Modeling MESFET's for intermodulation analysis of mixers and amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 38, no. 12, pp. 1964–1971, Dec. 1990.
- [4] J. C. Pedro and J. Perez, "Accurate simulation of GaAs MESFET's intermodulation distortion using a new drain-source current model," *IEEE Transactions on Microwave Theory and Techniques*, vol. 42, no. 1, pp. 25–33, Jan. 1994.
- [5] R. Van Langevelde and D. B. Klassen, "Accurate drain conductance modeling for distortion analysis in MOSFETs," in *Proceedings IEDM*, 1997, pp. 313–316.
- [6] R. Van Langevelde, L. F. Tiemeijer, R. J. Havens, M. J. Knitel, R. F. Roes, P. H. Woerlee, and D. B. Klassen, "RF-distortion in deep-submicron CMOS technologies," in *Proceedings IEDM*, 2000, pp. 807–811.
- [7] P. Wambacq and W. Sansen, *Distortion Analysis of Analog Integrated Circuits*, Kluwer Academic Publishers, Boston, 1998.
- [8] P. Wambacq, G.. Gielen, P. Kinget, and W. Sansen, "High-frequency distortion analysis of analog integrated circuits," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 46, no. 3, pp. 335–345, Mar. 1999.

- [9] T. Soorapanth and T. Lee, "RF linearity of short-channel MOSFETS," in *First International Workshop on Design of Mixed-Mode Integrated Circuits and Applications, Cancun, Mexico*, Jul. 1997, pp. 81–84.
- [10] Q. Huang, F. Piazza, P. Orsatti, and T. Ohguro, "The impact of scaling down to deep submicron on CMOS RF circuits," *IEEE J. of Solid-State Circuits*, vol. 33, no. 7, pp. 1023–1037, Jul. 1998.
- [11] M. T. Terrovitis and R. G. Meyer, "Intermodulation distortion in current-commutating CMOS mixers," *IEEE J. of Solid-State Circuits*, vol. 35, no. 10, pp. 1461–1473, Oct. 2000.
- [12] S. Tanaka, F. Behbahani, and A. A. Abidi, "A linearization technique for CMOS RF power amplifiers," in *1997 Symposium on VLSI Circuits Digest of Technical Papers*, 1997, pp. 93 – 94.
- [13] B. Kim, J. S. Ko, and K. Lee, "A new linearization technique for MOSFET RF amplifier using multiple gated transistors," *IEEE Microwave and Guided Wave Letters*, vol. 10, no. 9, pp. 371–373, Sept. 2000.
- [14] S. S. Weng, L. K. Chong, C. K. Vai, C. W. Wa, K. W. Tam, and R. P. Martins, "An analytical linearization method for CMOS MMIC power amplifier using multiple gated transistors," in *Proceeding of 4th International Conference on ASIC*, 2001, pp. 670–672.
- [15] B. Kim, J. Ko, and K. Lee, "Highly linear CMOS RF MMIC amplifier using multiple gated transistors and its volterra series analysis," in *IEEE MTT-S Symposium Digest*, 2001, pp. 515–518.
- [16] S. Kang, B. Choi, and B. Kim, "Linearity analysis of CMOS for RF application," *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, no. 3, pp. 972 – 977, Mar. 2003.
- [17] Y. Tsividis, K. Suyama, and K. Vavelidis, "A simple 'reconciliation' MOSFET model valid in all regions," *Electronics Letters*, vol. 31, pp. 506–508, Mar. 1995.
- [18] B. Razavi, *RF Microelectronics*, Prentice Hall, Englewood Cliffs, New Jersey, 1998.

- [19] J. Rogers and C. Plett, *Radio Frequency Integrated Circuit Design*, Artech House, Boston, MA, 2003.
- [20] V. Volterra, *Theory of Functionals and of Integral and Integro-differential Equations*, Dover Publication, New York, Unabridged republication of the first English translation edition, 1959.
- [21] S. Narayanan, "Transistor distortion analysis using Volterra series representation," *The Bell System Technical Journal*, vol. 46, pp. 991–1024, May/June 1967.
- [22] S. Narayanan, "Application of Volterra series to intermodulation distortion analysis of transistor feedback amplifiers," *IEEE Transaction on Circuit Theory*, vol. CT-17, no. 4, pp. 518–527, Nov. 1970.
- [23] R. Meyer, M. Shensa, and R. Eschenbach, "Cross modulation and intermodulation in amplifiers at high frequencies," *IEEE J. of Solid-State Circuits*, vol. SC-7, no. 1, pp. 16 – 23, Feb. 1972.
- [24] Y.L Kuo, "Distortion analysis of bipolar transistor circuits," *IEEE Transaction on Circuit Theory*, vol. CT-20, no. 6, pp. 709–716, Nov. 1973.
- [25] A. M. Khadr and R. H. Johnston, "Distortion in high-frequency FET amplifiers," *IEEE J. of Solid-State Circuits*, , no. 4, pp. 180 – 189, Aug 1974.
- [26] R. A. Minasian, "Intermodulation distortion analysis of MESFET amplifiers using the Volterra series representation," *IEEE Trans. on Microwave Theory and Techniques*, vol. MTT-28, no. 1, pp. 1 – 8, Jan. 1980.
- [27] D. D. Weiner and J. Spina, *Sinusoidal Analysis and Modeling of Weakly Non-linear circuits*, Van Nostrand Reinhold Company, New York, 1980.
- [28] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*, John Wiley and Sons, New York, 1980.
- [29] R. Meyer, "Intermodulation in high frequency bipolar transistor integrated-circuit mixers," *IEEE J. of Solid-State Circuits*, vol. SC-21, no. 4, pp. 534 – 537, Aug. 1986.

- [30] J. Bussgang, L. Ehrman, and J. Graham, "Analysis of nonlinear systems with multiple inputs," *Proceedings IEEE*, vol. 62, no. 8, pp. 1088 – 1118, Aug. 1974.
- [31] L. Chua and C Ng, "Frequency domain analysis of nonlinear systems: formulation of transfer functions," *IEE Journal of Electronic Circuits and Systems*, vol. 3, no. 6, pp. 257 – 269, Nov. 1979.
- [32] Y. Cheng and C. Hu, *MOSFET Modeling and BSIM3 User's Guide*, Kluwer Academic Publishers, Boston, 1999.
- [33] D. Foty, *MOSFET modeling with SPICE: principles and practice*, Prentice-Hall, 1996.
- [34] R. M. Velghe, D. B. Klassen, and F. M Klassen, "Compact MOS modeling for analog circuit simulation," in *Technical Digest of IEEE IEDM 1993*, Dec. 1993, pp. 485 – 488.
- [35] A. Boothroyd, S. Taraswicz, and C. Slaby, "MISNAN-a physically based continuous MOSFET model for CAD applications," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 10, no. 12, pp. 1512 – 1529 15, Dec. 1991 Pages:1512 - 1529 Volume: 10 , Issue: 12 , Dec. 1991 Pages:1512 - 1529 Volume: 10 , Issue: 12 , Dec. 1991 Pages:1512 - 1529 1991.
- [36] N. D. Arora, R. Rios, C. Huang, and K. Raol, "PCIM: A physically based continuous short-channel IGFET model for circuit simulation," *IEEE Trans. on Electron Devices*, vol. 41, no. 6, pp. 988–997, Jun. 1994.
- [37] C. Enz, F. Krummenacher, and E. Vittoz, "An analytical MOS model valid in all regions of operation and dedicated to low voltage and low power applications," *Analog integrated circuits and signal processing*, vol. 8, no. 1, pp. 83–114, July 1995.
- [38] B. Sheu, D. Sharfetter, P. Ko, and M. Jeng, "BSIM: Berkeley short-channel IGFET model for MOS transistors," *IEEE Journal of Solid-State Circuits*, vol. SC-22, no. 4, pp. 558 – 566, Aug 1987.
- [39] W. Liu, *MOSFET Models for SPICE Simulation, including BSIM3v3 and BSIM4*, John Wiley and Sons, Inc, 2001.

- [40] Y. Tsividis, *Operation and Modeling of the MOS Transistor*, McGraw-Hill, Boston, 1999.
- [41] Y. Tsividis and K. Suyama, "MOSFET modeling for analog circuit CAD: Problems and prospects," *IEEE J. of Solid State Circuits*, vol. 29, no. 3, pp. 210–216, Mar. 1994.
- [42] Y. Tsividis and D. L. Fraser, "Harmonic distortion in single-channel MOS integrated circuits," *IEEE J. of Solid State Circuits*, vol. SC-16, no. 6, pp. 694–702, Dec. 1981.
- [43] N. D. Arora and L. M. Richardson, "MOSFET modeling for circuit simulation," in *Advanced MOS Device Physics*, N. G. Einspruch and G. Gildenblat, Eds., San Diego, California, 1989, vol. 18, pp. 237–276, Academic Press, Inc.
- [44] P. K. Ko, "Approaches to scaling," in *Advanced MOS Device Physics*, N. G. Einspruch and G. Gildenblat, Eds., San Diego, California, 1989, vol. 18, pp. 1–37, Academic Press, Inc.
- [45] P. R. Gray and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits, Third Edition*, John Wiley and Sons, New York, 1993.
- [46] T. Grotjohn and B. Hoefflinger, "A parametric short-channel MOS transistor and model for subthreshold and strong inversion current," *IEEE Transactions on Electron Devices*, vol. ED-31, no. 2, pp. 234–246, Feb. 1984.
- [47] A. I. Cunha, O. C. Gouveia-Filho, M. C. Schneider, and C. Galup-Montoro, "A current-based model for the MOS transistor," in *Proceeding 1997 International Symposium on Circuits and Systems*, Hong Kong, June 1997, pp. 1608 – 1611.
- [48] W. H. Press, *Numerical recipes in C : the art of scientific computing*, Cambridge University Press, Cambridge, New York, 2nd ed. edition, 1992.
- [49] G. Baccarani, M. R. Wordeman, and R. H. Dennard, "Generalized scaling theory and its application to 1/4 micron MOSFET design," *IEEE Trans. on Electron Devices*, vol. ED-31, pp. 452–462, 1984.

- [50] Y. Taur, D. Buchanan, W. Chen, D. Frank, K. Ismail, S-H. Lo, G. Sai-Halasz, R. Viswanathan, H-J. Wann, S. Wind, and H-S. Wong, "CMOS scaling into the nanometer regime," *Proc. of the IEEE*, vol. 85, no. 4, pp. 486–503, Apr. 1997.
- [51] S. Thompson, P. Packan, and M. Bohr, "MOS scaling: Transistor challenges for the 21st century," *Intel Technology Journal*, , no. 3, pp. 1–19, 1998, Also available at <http://www.intel.com/technology/itj/q31998.htm>.
- [52] H. P. Wong, D. J. Frank, P. M. Solomon, C. H. J. Wann, and J. J. Welser, "Nanoscale CMOS," *Proc. of the IEEE*, vol. 87, no. 4, pp. 537–569, Apr. 1999.
- [53] B. Toole, C. Plett, and M. Cloutier, "RF circuit implications of moderate inversion enhanced linear region in MOSFETS," *IEEE Transactions on Circuit and Systems - I*, vol. 51, no. 2, pp. 319 – 327, Feb. 2004.
- [54] C. C. McAndrew, G. Zaneski, P. A. Layman, and S. Ayyar, "Accurate characterization of MOSFET overlap/fringing capacitance for circuit design," in *Proceedings IEEE 1994 Int. Conference on Microelectronic Test Structures*, Mar. 1994, pp. 15 – 20.
- [55] B. Toole and C. Plett, "A low voltage, low power RF CMOS LNA for Bluetooth applications using transmission line transformers," in *Proceedings of the 27th European Solid-State Circuits Conference*, 2001.
- [56] T. Manku, "Microwave CMOS - device physics and design," *IEEE Journal of Solid-State Circuits*, vol. 34, no. 3, pp. 277 – 286, March 1999.
- [57] T.P. Liu and E. Westerwick, "5-GHz CMOS radio transceiver front-end chipset," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 12, pp. 1927 – 1933, Dec. 2000.
- [58] M. Harada, T. Tsukahara, J. Kodate, A. Yamagishi, and J. Yamada, "2-GHz RF front-end circuits in CMOS/SIMOX operating at an extremely low voltage of 0.5V," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 12, Dec. 2000.
- [59] F. Behbahani, J.C. Leete, Y. Kishigami, A. Roithmeier, K. Hoshino, and A.A. Abidi, "A 2.4-GHz Low-IF receiver for wideband WLAN in 0.6 – μm CMOS - architecture and front-end," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 12, pp. 1908 – 1915, Dec. 2000.

- [60] J. Rogin, I. Kouchev, G. Brenna, D. Tschopp, and Q. Huang, "A 1.5 V 45-mW direct-conversion WCDMA receiver IC in $0.13\mu\text{m}$ CMOS," *IEEE Journal of Solid-State Circuits*, vol. 38, no. 12, pp. 2239 – 2248, Dec. 2003.
- [61] F. Gatta, D. Manstretta, P. Rossi, and F. Svelto, "A fully integrated $0.18\mu\text{m}$ CMOS direct conversion receiver front-end with on-chip LO for UMTS," *IEEE Journal of Solid-State Circuits*, vol. 39, no. 1, pp. 15 – 23, Jan. 2004.
- [62] T. Kim, B. Kim, and K. Lee, "Highly linear receiver front-end adopting MOS-FET transconductance linearization by multiple gated transistors," *IEEE Journal of Solid-State Circuits*, vol. 39, no. 1, pp. 223 – 229, Jan. 2004.
- [63] W. Hioe, K. Maio, T. Oshima, Y. Shibahara, T. Doi, K. Ozaki, and S. Arayashiki, " $0.18\mu\text{m}$ CMOS bluetooth analog receiver with -88 dbm sensitivity," *IEEE Journal of Solid-State Circuits*, vol. 39, no. 2, pp. 374 – 377, Feb. 2004.