

**RF TECHNIQUES FOR IEEE 802.15.4:
CIRCUIT DESIGN AND DEVICE MODELLING**

ALI ABUELMAATTI

Ph.D. THESIS

2008

RF TECHNIQUES FOR IEEE 802.15.4: CIRCUIT DESIGN AND DEVICE MODELLING

ALI ABUELMAATTI, B.SC.

A thesis submitted in fulfilment of the requirements of
the University of Glasgow for the degree of
Doctor of Philosophy

Department of Electronics and Electrical Engineering
University of Glasgow

January 2008

©

To Mum and Dad

ACKNOWLEDGEMENTS

First and foremost I would like to thank the almighty God (ALLAH) for blessing me and guiding me through all the stages of my life, up to this point. Without this guidance I would have never reached this position where I am writing this page.

Straight second to that I would like to express my deep gratitude, honour, and admiration to my supervisor Prof. Iain Thayne. I cannot ever say enough to adequately express my appreciation to him for being a supervisor of a prestige class made of him alone; a class beyond extreme excellence. With such richness in knowledge, experience and dedication, he provided the right vision and direction which led to the completion of this work; always prepared to help, always having time to listen and always open to discussion. I am forever indebted to Prof. Iain if not for anything else but for believing in me when no one else did and when hope was almost lost

Also thanks are due to the Engineering, Physics and Science Research Council, EPSRC, and the Institute for System Level Integration, ISLI, for the financial support they provided during the course of this study.

It is impossible to realise the depth of the sacrifice parents go through to provide for their children, let alone be able to reimburse their debt. When it comes to this, my parents are a distinguished elite, rising above all the rest. All the languages of the world have not produced words that can adequately describe their great contribution to my education, eventually resulting in the successful completion of this thesis. My mother, Soher Rose, she was the cause, the inspiration and the driving force without which this work would have never been started, let alone completed. It is for her that this Ph.D. was initially started. Without her dreams, her persistence and continuous prayers and her repeated words of encouragement, reaching this stage would have never materialised. My father, Prof. Muhammad Taher Abuelmaatti, I can never say enough about his immense technical, financial and inspirational support during the long years of my studies and yet to the present day. He was simply an open source which provided me with any imaginable needs and one that could happily be accessed any time, at any degree of demand. Without his various means of help, especially technical, the work of this thesis would have never been accomplished.

The support of my family and the encouragement they provide is with no equivalence. I am ever so grateful to my wife Noor Hayati for being very supportive, solely taking care of our children while I spent days and nights at home or away completing this work, serving me in my busy times, and accommodating my many moments of misery. Gratefulness also due to my daughters Sumayyah Rose and Nusaibah for making my life full of joy and happiness and for their inspirational dedication and patience. Tremendous appreciation also goes to my sister Zainab and my brothers, Omar, Osman and Abdullah, for their wide-spanned marvellous help and support during the course of this work.

A great appreciation is also due to Prof. Steven Mass who has repeatedly given me some of his precious time for technical discussion and advice. I would also like to express my thanks to my colleagues and friends in the department at Glasgow who contributed to this work through either individual or group discussions, notably, Dr. Edward Wasige, Dr. Khaled Elgaied, Ian McGregor and Griogair Whyte. Also thanks are due to the Technicians Team in the department for their support with various problems.

Ali Abuelmaatti

January 2008

ABSTRACT

The RF circuitry in the physical layer of any wireless communication node is arguably its most important part. The front-end radio is the hardware that enables communication by transmitting and receiving information. Without a robust and high performance front-end, all other higher layers of signal processing and data handling in a wireless network are irrelevant.

This thesis investigates the radio circuitry of wireless-networked nodes, and introduces several proposals for improvement. As an emerging market, analysis starts by examining available and ratified network standards suitable for low power applications. After identifying the IEEE 802.15.4 standard (commercially known as ZigBee) as the one of choice, and analysing several front-end architectures on which its transceiver circuitry can be based, an application, the Tyre Pressure Monitoring System (TPMS) is selected to examine the capabilities of the standard and its most suitable architecture in satisfying the application's requirements. From this compatibility analysis, the most significant shortcomings are identified as interference and power consumption. The work presented in this thesis focuses on the power consumption issues.

A comparison of available high frequency transistor technologies concludes Silicon CMOS to be the most appropriate solution for the implementation of low cost and low power ZigBee transceivers. Since the output power requirement of ZigBee is relatively modest, it is possible to consider the design of a single amplifier block which can act as both a Low Noise Amplifier (LNA) in the receiver chain and a Power Amplifier (PA) on the transmitter side. This work shows that by employing a suitable design methodology, a single dual-function amplifier can be realised which meets the required performance specification. In this way, power consumption and chip area can both be reduced, leading to cost savings so vital to the widespread utilisation of the ZigBee standard. Given the importance of device nonlinearity in such a design, a new transistor model based on independent representation of each of the transistor's nonlinear elements is developed with the aim of quantifying the individual contribution of each of the transistors nonlinear elements, to the total distortion.

The methodology to the design of the dual functionality (LNA/PA) amplifier starts by considering various low noise amplifier architectures and comparing them in terms of the trade-off between noise (required for LNA operation) and linearity (important for PA operation), and then examining the behaviour of the selected architecture (the common-source common-gate cascode) at higher than usual input powers. Due to the need to meet the far apart performance requirements of both the LNA and PA, a unique amplifier design methodology is developed. The design methodology is based on simultaneous graphical visualisation of the relationship between all relevant performance parameters and corresponding design parameters. A design

example is then presented to demonstrate the effectiveness of the methodology and the quality of trade-offs it allows the designer to make. The simulated performance of the final amplifier satisfies both the requirements of ZigBee's low noise and power amplification. At 2.4GHz, the amplifier is predicted to have 1.6dB Noise Figure (NF), 6dBm Input-referred 3rd-order Intercept Point (IIP₃), and 1dB compression point of -3.5dBm. In low power operation, it is predicted to have 10dB gain, consuming only 8mW. At the higher input power of 0dBm, it is predicted to achieve 24% Power-Added Efficiency (PAE) with 8dB gain and 22mW power consumption.

Finally, this thesis presents a set of future research proposals based on problems identified throughout its development.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
ABSTRACT	iv
TABLE OF CONTENTS	vi
LIST OF FIGURES	x
LIST OF TABLES	xiv
LIST OF ABBREVIATIONS	xv
CHAPTER 1 INTRODUCTION	1
1.1 Preamble	1
1.2 RF Interference	2
1.3 Low Power RF and Microwave Systems	3
1.4 Non-Linearity in Wireless Systems	3
1.5 Modulation Schemes	4
1.6 Transistor non-Linear Modelling	5
1.7 Wireless Network Standards	6
1.8 An Application: Tyre Pressure Monitoring System (TPMS)	6
1.9 Thesis Aims and Objectives	7
1.10 Summary	8
1.11 Thesis Structure	8
CHAPTER 2 TECHNICAL BACKGROUND	10
2.1 Introduction	10
2.2 Wireless Network Standards	10
2.2.1 Brief Comparison of Available Wireless Standards	11
2.2.2 Overview of the IEEE 802.15.4 Standard and ZigBee	12
2.3 RF Front-End Issues	14
2.3.1 Requirements Analysis	14
2.3.2 Topologies	16
2.4 High Frequency and Microwave Transistors	20
2.4.1 Available Technologies	20
2.4.2 High Frequency CMOS Transistors	21
2.5 Low Noise Amplifiers and Power Amplifiers	22
2.5.1 Low Noise Amplifiers	22
2.5.2 Power Amplifiers	27

2.6	Problem and Evaluation of Nonlinearity	27
2.6.1	Nonlinearity of Devices, Circuits and Systems	27
2.6.2	Nonlinearity Representation and Analysis Methods	29
2.6.3	MOSFET Nonlinearity	33
2.7	Two Port Networks	35
2.7.1	Review of S-Parameter Analysis	35
2.7.2	Reflection Coefficients and Matching	37
2.7.3	Stability.....	38
2.7.4	Gain Characterisation	39
2.7.5	Two Port Measurements in Small and Large-Signal designs	40
2.8	TPMS General Structure and Requirements.....	41
2.9	Summary.....	43
CHAPTER 3	ZIGBEE, ITS APPLICATION FOR TYRE PRESSURE MONITORING SYSTEMS (TPMS) AND ITS QUADRATURE PHASE SHIFT KEYING (QPSK) MODULATION	45
3.1	Introduction	45
3.2	TPMS Design Techniques.....	45
3.2.1	Current Approaches of TPMS.....	46
3.2.2	Network Standard-Based TPMS.....	49
3.3	ZigBee for TPMS	51
3.3.1	ZigBee Offerings vs. TPMS Requirements	51
3.3.2	The Low IF Block for ZigBee – The Image Problem	53
3.3.3	Power Consumption of Currently Available Solutions	54
3.4	QPSK Modulation and de-Modulation.....	54
3.4.1	QPSK and O-QPSK Mechanisms.....	55
3.4.2	Carrier Phase Shift Choice for Modulating Binary Bits.....	56
3.4.3	QPSK Implementation	58
3.5	Summary.....	68
CHAPTER 4	QUANTIFICATION OF CONTRIBUTION TO MOSFET DISTORTION THROUGH TRANSISTOR NON-LINEAR MODELLING	70
4.1	Introduction	70
4.2	Extraction	71
4.2.1	RF MOSFET Models and Extraction Methods	72
4.2.2	Chosen Model and Extraction Method	73
4.2.3	Automated Extraction Procedures	75
4.3	Modelling Representation Methods.....	84
4.3.1	First Method - Conductances and Capacitances (Model 1)	86
4.3.2	Second Method – Currents and Capacitances (Model 2)	91

4.4	Quantification of Sources of Distortion.....	97
4.4.1	Third Method – Currents and Charge Sources (Model 3).....	98
4.4.2	Linearising Model’s non-Linear Parameters	102
4.4.3	Quantification Results and Discussion	109
4.5	Summary.....	111
CHAPTER 5	LOW POWER, LOW NOISE AND POWER AMPLIFIERS: DESIGN AND OPTIMISATION	113
5.1	Introduction	113
5.2	Low Noise Amplifier Architectures and their Design.....	114
5.2.1	LNA Architectures	114
5.2.2	LNA Design Approaches and Optimisation Techniques	122
5.2.3	State of the Art LNAs.....	126
5.3	Amplifier non-Linearity	130
5.3.1	Circuit Factors Affecting Linearity.....	130
5.3.2	Low Power Cascode LNA Linearisation Techniques	132
5.3.3	Highly Linear Amplifiers.....	138
5.4	Frequency Components Affecting Nonlinear Performance	141
5.4.1	The Feedback Phenomena.....	141
5.4.2	Analysis and Results	143
5.5	Cascode LNAs Operating at High Input Powers - An Investigation and Design Methodology	147
5.5.1	Methodology and Design Approach	148
5.5.2	In-depth Analysis of the Cascode Architecture at Low and High Input Powers	153
5.5.3	High Dynamic Range Cascode Amplifier – A design example.....	169
5.6	Summary.....	183
CHAPTER 6	CONCLUSIONS AND FUTURE DEVELOPMENTS	185
6.1	Introduction	185
6.2	Thesis Summary	185
6.3	Contributions	187
6.4	Future Work.....	189
6.4.1	ZigBee vs. TPMS analysis	189
6.4.2	Transistor non-Linear Modelling	190
6.4.3	LNA/PA Design Approach	191
6.5	Final Remark	192
REFERENCES	193

APPENDIX A	EXPLANATION OF CODE FUNCTIONS USED IN TRANSISTOR NONLINEAR MODELLING	206
A.1	Automated Parameter Extraction.....	206
A.2.1	Extraction of Extrinsic Parameters	206
A.2.2	Extraction of Intrinsic Parameters	207
A.2.3	Finalising Intrinsic Capacitance Values	212
A.2.4	Writing Extracted Parameters Values to Data Tables	212
A.2	Generating I_{gm} and I_{gds} for DC Current Model	213
A.2.1	Accurate V_T identification	213
A.2.2	Splitting DC current to I_{gm} and I_{gds}	215
A.3	Linearising gm by Modifying its Data Table.....	217
APPENDIX B	IIP_3 DERIVATION FOR A COMMON-SOURCE MOSFET USING VOLTERRA AND TAYLOR SERIES	220
APPENDIX C	FULL RESULTS OF HARMONIC AND IMD FEEDBACK ANALYSIS.....	229
C.1	Fed-back Components From 2 nd -order Nonlinear Term	229
C.2	Fed-back Components From 3 rd -order Nonlinear Term.....	230
C.3	Fed-Back Components From 4 th -order Nonlinear Term	233
C.4	Fed-Back Components From 5 th -order Nonlinear Term	236
C.5	Fed-Back Components From 6 th -order Nonlinear Term	240

LIST OF FIGURES

Figure 2.1 Wireless network standards: Their data and distance capabilities [CyberInI 2007].....	11
Figure 2.2 IEEE 802.15.4 and ZigBee working model [Le 2004].....	13
Figure 2.3 Simple direct conversion receiver.....	16
Figure 2.4 A simple Heterodyne receiver.....	17
Figure 2.5 Simple Low-IF receiver.....	18
Figure 2.6 Offset QPSK I/Q modulator with half-sine shaping.....	19
Figure 2.7 2-Point PLL $\Sigma\Delta$ modulator with its VCO.....	19
Figure 2.8 A typical RF receiver architecture	23
Figure 2.9 A cascade system.....	23
Figure 2.10 LO leakage to the antenna through LNA.....	24
Figure 2.11 The full MOS noise Model	25
Figure 2.12 Simplified MOSFET model with input matching	30
Figure 2.13 Change in MOSFET 3 rd -order transconductance	34
Figure 2.14 The origin of S-parameters (a) Incident and reflected waves of two-port networks and (b) Determining S-parameters from these waves	36
Figure 2.15 Two port network with matching networks, reflection coefficients and associated impedances.....	37
Figure 2.16 Gains of a bilateral two port network	39
Figure 2.17 The TPMS (a) A TPMS in a 4-wheel passenger car [Siemens 2007] and (b)Diagrammatic structure of the contents of a TPMS system	42
Figure 3.1 Basic setup of a TPMS	46
Figure 3.2 Wired TPMS setup (a) Signal receiver at central module (b) Local signal receiver	47
Figure 3.3 a diagrammatic structure of a TPMS as a portable sensor network using a network protocol ..	49
Figure 3.4 Block Diagrams of QPSK Modulator and Demodulator.....	56
Figure 3.5 All possible carrier phase shifts subject to transmitted bit pairs when the system's allocation rule is (a) I reverse for 0 Q reverse for 0, (b) I reverse for 0 Q reverse for 1, (c) I reverse for 1 Q reverse for 0 and (d) I reverse for 1 , Q reverse for 1	57
Figure 3.6 Serial data bit generator (a) Top view and (b) detailed view of parallel to serial shift register	60
Figure 3.7 DeMultiplexer to split the data stream into in-phase and quadrature components.....	61
Figure 3.8 Implementation of QPSK Modulator.....	62
Figure 3.9 Implementation of QPSK Demodulator.....	63
Figure 3.10 Implementation of base band processing.....	64
Figure 3.11 The Test data signal.....	64
Figure 3.12 The testing data split to the I and Q lines (a) I line data and (b) Q line data	65
Figure 3.13 Carrier modulated by (a) I data and (b) Q data.....	66
Figure 3.14 The transmitted signal	66
Figure 3.15 Recovered data levels for (a) I line and (b) Q line	67
Figure 3.16 The received (a) I data and (b) Q data after being low-pass filtered.....	67
Figure 3.17 The received (a) I data signal and (b) Q data signal after the comparator.....	68

Figure 3.18 The received I and Q data combined in one data stream.....	68
Figure 4.1 The RF MOSFET model proposed by [Tong 2004].....	74
Figure 4.2 Plots of extracted resistance values against frequency (a) R_d , (b) R_g , (c) R_s and (d) All extrinsic resistances.....	77
Figure 4.3 ADS schematic setup for extracting the intrinsic model parameters	79
Figure 4.4 3D plot of gm variance versus drain and gate voltages	79
Figure 4.5 3D plot of gds variance versus drain and gate voltages	80
Figure 4.6 Plot of extracted C_{gs} and C_{gd} against frequency.....	81
Figure 4.7 3D plot of C_{gs} variance versus V_{GS} and V_{DS}	81
Figure 4.8 3D plot of C_{gd} variance versus V_{GS} and V_{DS}	82
Figure 4.9 All extracted intrinsic capacitances versus frequency	83
Figure 4.10 Demonstration of data representation using SDD (a) A sample data table, (b) A sample SDD and (c) its sample DAC	85
Figure 4.11 SDD representations of nonlinear intrinsic components for Model 1 (a) C_{gs} , (b) C_{gd} , (c) C_{jg} , (d) gm , (e) R_{sub} , (e) C_{ds} and (g) gds	87
Figure 4.12 complete SDD model with all intrinsic and extrinsic components	89
Figure 4.13 S-parameter test results of the transistor model and the modelled transistor.....	90
Figure 4.14 One-tone Harmonic Balance test results of the model (a) Fundamental tone and (b) Third harmonic	91
Figure 4.15 gm against V_{GS} in the linear region when $V_{DS} = 10mV$	92
Figure 4.16 Linear extrapolation of maximum gm point	93
Figure 4.17 The cross point of the linear extrapolation with V_{GS} axis	93
Figure 4.18 Extracted I_{gm} and I_{gds} at one value of V_{GS}	94
Figure 4.19 I_{gm} - V_{DS} current characteristics for a number of V_{GS} curves	95
Figure 4.20 SDDs of I_{gm} and I_{gds}	95
Figure 4.21 Complete SDD structure of Model 2	96
Figure 4.22 Z-parameters of the model and the modelled transistor for Model 2.....	96
Figure 4.23 One-tone Harmonic Balance test results of Model 2.....	97
Figure 4.24 The exact voltage applied across a capacitance in the Model.....	99
Figure 4.25 Error in integrating voltage for C_{gs}	100
Figure 4.26 Mismatch in S_{21} between model and transistor and small and large operating points (a) $V_{GG}=0.5V$, $V_{DD}=1V$, (b) $V_{GG}=1V$, $V_{DD}=4V$	101
Figure 4.27 SDD connection and table representations with (a) external and (b) internal voltages.....	101
Figure 4.28 Error in the model's fundamental output when using single- gm VCCS for linearisation.....	104
Figure 4.29 I_{gm} - V_{GS} characteristics obtained from nonlinear I_{gm} table	104
Figure 4.30 Nonlinear and linear I_{gm} - V_{GS} characteristics (in saturation).....	105
Figure 4.31 A sample V_{GS} curve in linear and nonlinear I_{gm} - V_{DS} characteristics	106
Figure 4.32 Linearised and nonlinear I_{gm} - V_{GS} characteristics.....	106
Figure 4.33 Interpolation error in linearising I_{gm} when only few sweep point are used.....	107
Figure 4.34 Fundamental output of (a) gm -linearised model and (b) original nonlinear model	108

Figure 4.35 Results of quantification of nonlinear distortion (a) gm, gds, Cgs and Cgd nonlinear. (b) gm and gds nonlinear, Cgs and Cgd linear. (c) gm, Cgs and Cgd linear, gds nonlinear. (d) gm, gds, Cgs and Cgd linear. (e) gm nonlinear, gds, Cgs and Cgd linear. (f) gm, gds, Cgs linear, Cgd nonlinear.....	110
Figure 5.1 A typical common-drain amplifier.....	115
Figure 5.2 A typical common-gate amplifier	115
Figure 5.3 A typical common-source amplifier	116
Figure 5.4 Resistive termination for input matching of common-source stage.....	117
Figure 5.5 Input/output termination using shunt and series feedback	117
Figure 5.6 Common-source with inductive degeneration for	117
Figure 5.7 The common-source common-gate cascode LNA topology	119
Figure 5.8 The current reuse LNA topology	120
Figure 5.9 LNA matching technique in [Zhenying 2003].....	128
Figure 5.10 Cascode LNA with inter stage matching network.....	128
Figure 5.11 Equivalent circuit of the cascode LNA for inter stage matching.	129
Figure 5.12 MOSFET model for derivation of IMD_3 and IIP_3 expressions using Volterra	133
Figure 5.13 g_m'' vs. V_{GS} for a common-source transistor	135
Figure 5.14 Tuned g_m'' - V_{GS} characteristics with linearisation.....	136
Figure 5.15 Circuit schematic of linearisation using the DS method	136
Figure 5.16 Tuned g_m'' - V_{GS} characteristics for auxiliary transistors (a) M_2 and (b) M_3	137
Figure 5.17 Graphical description of 2 nd -order contribution to IMD_3	141
Figure 5.18 results of the 2 nd -order nonlinearity from initial two-tone input	142
Figure 5.19 Frequency components resulting from feeding back the 2ω component.....	142
Figure 5.20 Initial optimisation circuit for the two cascode transistors.....	150
Figure 5.21 Amplifier DC current vs. CS and CG sizes at voltage supply of (a) 1.8V, (b) 1.5V, (c) 1.2V and (d) 1V	154
Figure 5.22 Amplifier mid voltage at a voltage supply of (a) 1.8V, (b) 1.5V, (c) 1.2V and (d) 1V.....	155
Figure 5.23 Common-source transistor's current versus its size and gate voltage.....	156
Figure 5.24 Minimum noise figure vs. transistor sizes	157
Figure 5.25 Minimum noise figure of a common-source transistor versus its gate voltage and size at drain voltages of (a) .3V, (b) .5V, (c) .9V, (d) 1.2V, (e) 1.5V and (f) 1.8V	158
Figure 5.26 The DC current of the amplifier drawn from the supply at an input power of (a) -20 dBm and (b) -10 dBm (c) -5 dBm and (d) 0 dBm	159
Figure 5.27 DC power consumption vs. transistor sizes at low and high input powers (a) -20 dBm and (b) -10 dBm (c) -5 dBm and (d) 0 dBm	160
Figure 5.28 Transducer gain vs. transistor sizes at low and high RF input powers (a) -20 dBm and (b) -10 dBm (c) -5 dBm and (d) 0 dBm	161
Figure 5.29 Load Power vs. transistor sizes at low and high RF input powers (a) -20 dBm and (b) -10 dBm (c) -5 dBm and (d) 0 dBm	161

Figure 5.30 PAE vs. transistor sizes at low and high RF input powers (a) -20 dBm and (b) -10 dBm (c) -5 dBm and (d) 0 dBm	162
Figure 5.31 Relative 3 rd -order intermodulation in dBc with respect to transistor sizes when the RF input power is (a) -20 dBm (b) -10dBm, (c) -5dBm and (d) 0dBm	163
Figure 5.32 Mapping between IMD ₃ and IIP ₃ calculations at -30dBm input power (a) IMD ₃ and (b) IIP ₃ vs. Transistor sizes	165
Figure 5.33 IMD ₃ vs. transistor sizes at different values of degen inductors (a) no degen (b) 0.58nH (c) 1.5nH and (d) 2.9nH when the input power is 0dBm	166
Figure 5.34 Minimum Noise Figure vs. transistor sizes at different values of degeneration inductors (a) no inductor (b) 0.58nH (c) 1nH and (d) 1.5nH	167
Figure 5.35 Effect of the on-chip load inductor on the stability and gain of the amplifier	168
Figure 5.36 Second derivative of the total transconductance in the cascode after the addition of the linearising auxiliary transistors to the CS stage.....	172
Figure 5.37 S-parameter noise circles for input matching	173
Figure 5.38 Available gain circles for low-power input matching	173
Figure 5.39 Power delivered to the load contours from the source-pull simulation	174
Figure 5.40 PAE Contours from the source-pull simulation.....	174
Figure 5.41 IMD ₃ Contours from the source-pull simulation.....	175
Figure 5.42 Power gain circles for low-power output matching.....	176
Figure 5.43 Power delivered to the load contours from the load-pull simulation	176
Figure 5.44 PAE Contours from the load-pull simulation	177
Figure 5.45 IMD ₃ Contours from the load-pull simulation.....	177
Figure 5.46 S-parameter simulation results of the matched amplifier.....	179
Figure 5.47 Noise figure vs. frequency of final amplifier.....	179
Figure 5.48 Power added efficiency and power gain vs. input power	180
Figure 5.49 DC current and DC power consumption vs. input RF power	180
Figure 5.50 One-tone harmonic balance simulation results of matched amplifier.....	181
Figure 5.51 Two-tone simulation results of matched amplifier.....	181
Figure A.1 Automated extraction of extrinsic resistances (a) Initial extraction equations and	206
Figure A.2 Extraction of g_m and g_{ds} (a) g_m , (b) g_{ds} and (c) dependency table on swept parameters..	207
Figure A.3 Extraction of C_{gs} and C_{gd}	208
Figure A.4 Equations and plots for the extraction of R_{sub}	210
Figure A.5 The extraction of C_{jd} and C_{ds}	211
Figure A.6 Taking the average of the capacitance values versus frequency	212
Figure A.7 Writing the extracted data to a text file.....	213
Figure A.8 Equations for plotting the linear extrapolation of I_{DS} vs. V_{GS} characteristics at maximum g_m	214
Figure A.9 V_T extraction using ELR method.....	215
Figure A.10 Transistor's DC characteristics	215
Figure A.11 Coded functions for the split of the transistor's current characteristics.....	216
Figure A.12 Linearising g_m by modifying the I_{gm} data table.....	219

LIST OF TABLES

Table 2.1 ZigBee specifications at its 3 frequency bands [Le 2005]	15
Table 3.1 Power consumption of available ZigBee Chips	54
Table 5.1 Results of recently reported LNAs (a) [Zhenying 2003], (b) [Zhang 2003], (c) [Asgaran 2006], (d) [Yang 2001], (e) [Kim 2004], (f) [Aparin 2005], (g) [Kim 2006a] and (h) [Huang 2006]	127
Table 5.2 Summary of 3 rd -, 5 th - and 7 th -order IMD component production	145
Table 5.3 Regions of stability and maximum gain relative to different load inductors	168
Table 5.4 Amplifier results comparison with ZigBee-compatible LNA and PA	183

LIST OF ABBREVIATIONS

3D	Three Dimensional graph
AC	Alternating Current
ADS	Advanced Design System simulation software
AM	Amplitude Modulation
ASK	Amplitude Shift Keying
BER	Bit Error Rate
BJT	Bipolar Junction Transistor
BPSK	Binary Phase Shift Keying
BSIM3v3	Berkeley Short-Channel IGFET Model 3, Version 3.0
CAD	Computer Aided Design
CCA	Clear Channel Assessment
CG	Common-Gate transistor
CMOS	Complementary Metal-Oxide Semiconductor
CS	Common-Source transistor
CS-CG	Common-Source Common-Gate cascode
CSMA-CA	Carrier Sense Multiple Access with Collision Avoidance
DAC	Data Access Component
DC	Direct Current
DS	Derivative Superposition
DSSS	Direct Sequence Spread Spectrum
DUT	Device Under Test
ED	Energy Detection
ELR	Extrapolation in the Linear Region
EVM	Error Vector Magnitude
FET	Field-Effect Transistor
FSK	Frequency shift keying
GaAs	Gallium Arsenide
HBT	Heterojunction Bipolar Transistor
HEMT	High Electron Mobility Transistor
IEEE	Institute for Electrical and Electronic Engineers
IF	Intermediate Frequency
IIP ₃	Third Order Input-referred Intercept Point
IMD ₃	Third Order Intermodulation Distortion
IP	Intercept Point
ISM	Industrial Scientific-Medical
KCL	Kirchoff's Current Law
KVL	Kirchoff's Voltage Law
LNA	Low Noise Amplifier
LNA/PA	Low Noise Amplifier/Power Amplifier
LO	Local Oscillator
LQI	Link Quality Indication
MAC	Medium Access Control Layer
MDIF	Measurement Data Interchange Format

MESFET	Metal-Semiconductor Field Effect Transistor
MGTR	Multi-Gate Transistor
MOS	Metal Oxide Semiconductor
MOSFET	Metal Oxide Semiconductor Field Effect Transistor
MSK	Minimum Shift Keying
MW	Microwave
NF	Noise Figure
NGF	Number of Gate Fingers
NMOS	Negative channel Metal- Oxide Semiconductor
NMOSFET	Negative channel Metal- Oxide Semiconductor Field-Effect Transistor
OEM	Original Equipment Manufacturer
O-QPSK	Offset Quadrature Phase Shift Keying
PA	Power Amplifier
PAE	Power-Added Efficiency
PC	Personal Computer
PDA	Personal Digital Assistant
PER	Packet Error Rate
PHY	Physical Layer
PLL	Phase Locked Loop
PM	Phase Modulation
PMOS	Positive channel Metal-Oxide Semiconductor
PSK	Phase Shift Keying
QoS	Quality of Service
QPSK	Quadrature Phase Shift Keying
RF	Radio Frequency
RFCMOS	Radio Frequency Complementary Metal-Oxide Semiconductor
RFIC	Radio Frequency Integrated Circuit
RX	Receiver
SAW	Surface Acoustic Wave
SDD	Symbolically Defined Device
SiMOSFET	Silicon Metal Oxide Semiconductor Field Effect Transistor
SNR	Signal-to-Noise Ratio
SoC	System on Chip
SOI	Silicon-on-Insulator
TPMS	Tyre Pressure Monitoring System
TX	Transmitter
VCCS	Voltage Controlled Current Source
VLSI	Very Large Scale Integration
WLAN	Wireless Local Area Network
WMAN	Wireless Metropolitan Area Network
WPAN	Wireless Personal Area Network
ZigBee	ZigBee wireless technology

Chapter 1 INTRODUCTION

1.1 Preamble

The term wireless is normally used to refer to any type of electronic operation which is accomplished without the use of a hard wired connection. The recent expansion of wireless technology has improved peoples' lives significantly. It is not difficult to find many examples of wireless communication in action in a normal day's journey. Mobile phones are now being used by hundreds of millions of people throughout the world. More exciting than that, broadband Internet wireless networks can nowadays be found in many places, not just in offices or at working sites, but also in cafes and public areas in vibrant city centres. Users are now able to link multiple wireless-capable portable devices such as digital cameras, laptop computers, mobile phones and Personal Digital Assistants (PDAs) and transfer data between them wirelessly at the touch of a button. At home, households can wirelessly link and control numerous items such as room lights, window curtains, TVs, HiFi systems and PCs all using one remote console. Also at home and with the rise of intelligent networking, household appliances such as fridges, ovens, microwaves and cookers can send information about their contents, operation and any failures to a central server which then takes the appropriate action by either informing the user by displaying this information on a screen or sending a short mobile message, or directly contacting the dealer for any required repairs. Adding to all the above, the adoption and wide use of wireless sensor networks for medical and automotive applications and in industrial plants, it becomes very difficult to count all examples of wireless technology in use by humans today in all walks of life. In short, wireless technology has introduced an irreplaceable ease and convenience for the life of humans that they cannot imagine living without anymore.

The following sections include a discussion of the major problems that come with the proliferation of wireless technology, the most important of which is interference. Next, some of the important specifications of wireless systems are introduced followed by some reflection on how meeting these requirements can improve the quality and broaden the usability of wireless technology. Attention is given to power consumption, linearity and choice of suitable

modulation schemes. Thereafter, the modelling of suitable transistors crucial to the realisation of RF integrated circuits in wireless systems is described. The importance of wireless network standards to classify the world of wireless technology is then touched upon. Following that, a specific application, a Tyre Pressure Monitoring System (TPMS) is introduced to place the work in context, in particular highlighting those aspects which will be covered in this thesis. This motivates the thesis aims and objectives which are set out before concluding the chapter with a description of the thesis structure.

1.2 RF Interference

The proliferation of wireless technology does not come without a price. It is an unavoidable fact that wireless communication systems must co-exist in extremely complicated and dense signal environments. As many of the examples mentioned above are usually operating in close proximity by either single or multiple users, these environments are comprised of multiple operating wireless networks. Taking for example a café in the centre of a busy city; some customers would be using their mobile phones, while some others are accessing the internet through the wireless network setup in the café. In the meantime two other customers would be sharing photos on their mobile phones by exchanging them through a wireless link. Another example is the networked home mentioned above; it is not difficult to imagine that all these devices may be communicating at the same time.

The number one enemy of wireless systems designer and service provider is signal interference [Anritsu 2003]. Interference hampers coverage and capacity, and limits the effectiveness of mobile networks operating in a given environment.

Interference occurs when unwanted frequency components (desired, harmonics or intermodulation) from neighbouring communication networks are picked up by a receiver node. One or more of these frequency components can be either at the same frequency as the desired signal in which case the desired signal is distorted even before it reaches the receiver; or else they can be in the pick-up band of the receiver in which case they can mix with the desired signal inside the receiver and distort it.

In Chapter 3, interference will be discussed in more details when the implementation of a TPMS using ZigBee is examined. The amplifier design in Chapter 5 will also take interference into consideration by considering several interference-reduction techniques.

1.3 Low Power RF and Microwave Systems

In order to be truly portable, wireless devices which are normally battery powered, need to be designed for high efficiency while maintaining their required performance characteristics. Demand for low power consumption network nodes becomes tougher when it is difficult to replace or recharge the nodes' battery. Examples of these are sensors deployed in hard-to-reach industrial locations or in human bodies.

The longer the battery life of a mobile device, the more the motivation for it being wireless in the first place. For instance, there is no much benefit of having a mobile phone if one has to be tied to a power source most of the time in order to use it. Also, a wireless in-body sensor is useless if the sensor has to be powered with a cable. Therefore, the development of wireless technologies must also focus on developing techniques to optimise power efficiency along with improving compact long-lasting power sources. Power consumption in wireless systems depends on many factors including; linearity, frequency used, data bandwidth and adopted modulation scheme. These issues are discussed in the following sections.

In Chapter 5, power consumption is discussed in more detail for the design and implementation of a power-efficient amplifier which can be used as both the transceiver low noise amplifier and power amplifier.

1.4 Non-Linearity in Wireless Systems

Generally, nonlinearity describes the response of a system's behaviour to its drive level. In fact, all electronic circuits display aspects of nonlinear behaviour at all levels of drive [Maas 1998]. In some circuits, for instance, amplifiers, the effect of this nonlinearity at very small signal levels is negligible and the amplifier can be assumed linear; but this is only an approximation. The effects of nonlinearity become more severe as drive level increases. As well as aspects of nonlinear behaviour being undesirable in some circuits such as amplifiers and filters, some other circuits depend on them for their operation, such as mixers, modulators and de-modulators.

The nonlinearity of an amplifier generates harmonics of the desired signal; the more nonlinear the amplifier, the larger the harmonic content. If the low noise amplifier in a receiver picks up an interfering in-band frequency component and harmonics for that were also generated, intermodulation occurs. One component of this process, the 3rd-order Intermodulation Product (IMD₃) falls very close to the desired signal and can severely distort it, leading to possible errors in the received data. The same outcome can occur in a power amplifier in a transmitter if communicating on more than one channel at the same time. Harmonics generated from every

channel can intermodulate leading to a higher probability of adjacent channel interference and also leading to a distorted transmitted signal. When this distorted signal is picked up and is affected by further interference at the receiver, a serious problem is caused where the data can be completely lost. There is a strong relationship between nonlinearity and interference. Greater nonlinearity in a transmitter system means larger harmonic and intermodulation components are generated and broadcast with the potential to be picked up by a neighbouring receiver distorting its desired signal.

There is a strong relationship between linearity and power consumption. For instance, a common way to increase the linearity of an amplifier is to optimise its design in a way which will lead to it consuming more power, for reasons which will be discussed further in Chapter 5. Hence a less linear amplifier is a less power efficient amplifier, and a transceiver with a less linear amplifier is a less power efficient transceiver, wasting battery power. Therefore, having a more linear transceiver leads to a longer lasting battery, increasing the portability and usability of a wireless system as discussed above.

In Chapter 5 the linearity of an RF transceiver will be discussed in more detail in terms of the linearity of low noise amplifiers and power amplifiers. Existing linearisation techniques will be assessed and applied in a design example to demonstrate the effectiveness of a design methodology designed to achieve high amplifier linearity.

1.5 Modulation Schemes

Modulation is the process whereby a message-bearing signal is superimposed upon a carrier signal for transmission. Carrier signals are usually high in frequency to enable the construction of small antennas.

The choice of a digital modulation scheme for a wireless transmission system significantly affects the performance and physical realisation of that system. This is usually a closed loop discussion in which two extremes are involved. These are whether a modulation scheme is to be selected to satisfy pre-set specifications of the wireless system or whether the specifications of the wireless system are to be decided by the modulation scheme being used; designers often go back and forward between these two extremes. The chosen solution is often a trade-off between all issues that are affected by the modulation scheme being used, and considering the environment where the system is to operate. Issues involved in the decision include the physical characteristics of the channel, desired levels of performance in terms of Bit Error Rate (BER), required data rate, acceptable level of latency, available bandwidth, anticipated link budget, target hardware complexity, cost and power consumption.

On the transmitter side, modulation schemes can be divided into two categories with respect to the amount of distortion that results when signals are passed through a power amplifier; those that produce a constant envelop signal, and those that produce a non constant envelop signal [Liang 1999]. If the carrier is amplitude modulated (i.e. an envelop change is introduced), nonlinearity in the power amplifier causes the carrier and the side bands to intermodulate producing undesired in-band and adjacent channel frequency components. This is why the development of modulation schemes has always focused on reducing envelop change as much as possible. Constant envelop modulation schemes do not suffer from this but they generally have larger bandwidth than non constant envelop modulation schemes for a given data rate and this can cause interference with adjacent channels if the frequency separation between channels is small. On the other side, receiver circuitry does not add any harmful effect to the received signal as a result of its modulation scheme.

In Chapter 3 the Quadrature Phase Shift Keying (QPSK) modulation scheme will be discussed in more detail and a simple, new method of implementing this sophisticated modulation scheme will be presented.

1.6 Transistor non-Linear Modelling

The ever increasing public desire for new innovations in wireless technology has led to severe competition between providers to supply to the market the best performing products on the shortest possible lead times. This has consequently put pressure on circuit designers to try to always get it right the first time, and this would not be possible without developing accurate transistor models that can be used in Computer Aided Design (CAD) simulations to speed up and verify the design before it is passed for fabrication.

Several ways of building models have been proposed over the years depending on the information the designer requires in order to correctly predict circuit performance. Examples of these are large signal models, small signal models, harmonic prediction models, power performance models and noise prediction models.

Chapter 4 of this thesis will handle transistor nonlinear modelling in extensive detail. A new nonlinear model based on representing each of the model elements independently will be proposed and verified in both small- and large- signal simulations.

1.7 Wireless Network Standards

With the increasing design and production activities that serve the desire and expansion of wireless technology, it is important to regulate the industry by setting standards. A network standard is a collection of technical and performance characteristics that any device claiming compatibility should adhere to. Wireless systems are classified into standards based on their application in terms of bandwidth, range and power consumption.

Generally, network standards are developed by many experts from different backgrounds over a long period of time and after many discussion panels. This means the standard defines all aspects of the network required to achieve certain performance characteristics. For instance, the modulation scheme adopted in a high power, high bandwidth standard is different from that adopted in a low power, low data rate one, and for each of these standards, aspects like the number of channels, receiver sensitivity, maximum output power, cross channel interference and BER which are all co-dependant on that modulation scheme, are decided. These performance parameters, determined by standard developers, are used by systems developers to form a set of rules or performance targets for the system and circuit designer which must be met when designing a transceiver or part of a transceiver that is to be compatible with the particular standard.

Having a network standard also serves another purpose; that is of enabling interoperability between different devices from various manufacturers. There should be no issues with these various devices working together if they adhere to the common standard.

In Chapter 2, network standards will be considered in more detail where a comparison between all available standards will be conducted to identify that best suited for low power and low data rate applications.

1.8 An Application: Tyre Pressure Monitoring System (TPMS)

In order to make this investigation of low power and highly linear wireless systems more tangible, an application was picked to test the capabilities of the chosen network standard in terms of required performance parameters; notably, interference and power consumption. The application of choice is the Tyre Pressure Monitoring System (TPMS). There were several reasons behind this choice. First, this is an application where the wireless transmitters attached to the pressure sensors in the tyres have to be very power efficient as it is very difficult to replace their batteries after they have been installed. Also, the vehicle environment is a very crowded one and hence interference is inevitable. That is not just because passengers in cars are

often using many forms of wireless devices that might interfere with the TPMS in their cars, but also having every private TPMS network of every vehicle operating in close vicinity to other private networks, in say a traffic queue or a car park, makes the possibility of some sort of interference occurring more likely.

Furthermore, a TPMS has never, so far, been built based on a network standard, so part of the experiment is to examine the tradeoffs of employing a network standard in its implementation and examine the capabilities of the most suitable network standard to satisfy the requirements of the system.

In Chapter 2, the performance requirements of the tyre pressure monitoring system is discussed in more detail. In Chapter 3, a compatibility analysis is conducted for examining the capabilities of ZigBee, the chosen low power standard in satisfying the requirements of the TPMS application.

1.9 Thesis Aims and Objectives

It is the aim of this thesis to review the current status of front-end RF circuits used in wireless network adapters and hence develop new techniques to contribute to improving the performance of these circuits taking into consideration the performance issues mentioned in the sections above.

The work explores the capabilities of Silicon CMOS technology as a strong competitor to traditional high frequency transistors. The design activity focuses on investigating low power techniques while maintaining other performance characteristics at high standards. Assessment of developed circuits is measured on how it would help improving the performance of the network standard-based TPMS application.

The work starts by looking at available wireless network standards and selecting the one most suitable for ultra low power applications. Available front-end architectures are then studied to select the most suitable to satisfy the requirements of that standard. Some of the problems in that architecture are then discussed and considered in examining the capabilities of the standard in satisfying the requirements of the test application. Areas of shortcomings are identified forming the basis of the development work in the thesis.

1.10 Summary

As the electronics industry continues to develop high performance wireless devices, the desire and expectations of consumers will continue to rise. Consumers will demand ever longer battery life in wireless devices, non jamming wireless communications, faster and easier wireless transfers and so on. This puts pressure on system and circuit designers to provide solutions to address the various areas outlined in this chapter which become more significant with the proliferation of wireless portable devices.

This chapter began by discussing and linking three of the most vital technical issues behind the current barriers residing in the RF front-end, which are interference, linearity and power consumption. It then showed how the choice of the modulation scheme is interdependent on many other parameters of the front-end design and how the modelling of the transistor used in the circuit design is important. An application, the tyre pressure monitoring system, on which the performance improvements proposed in this thesis will be assessed, was introduced.

Having discussed this motivation, the aims and objectives of this thesis were set. These were to contribute to overcoming the barriers mentioned in this chapter by studying the current state of the art in RF front-end design and proposing new techniques for improvements.

1.11 Thesis Structure

The rest of this thesis is structured as follows:

- Chapter 2 presents the background to the field looking at some of the technical issues involved in the design of front-end RF circuits. This includes comparison of front-end architectures and analysis of front-end performance requirements and the factors that may cause degradation to performance levels. It also includes a review of high frequency transistors with focus on Silicon CMOS as the technology of choice. The chapter also discusses low noise and power amplifiers in some extended detail and reviews some of the techniques used in their design. Also presented in this chapter is a brief overview of the ZigBee standard and the TPMS application
- Chapter 3 handles the compatibility analysis between the ZigBee and the TPMS test application by first reviewing previous methods of TPMS implementation and discussing the benefits of a ZigBee-based TPMS. Some areas where improvements are needed are then identified. The chapter also introduces a unique implementation for the QPSK modulation implemented in ZigBee based on a new method of understanding the way in which QPSK operates.

- Chapter 4 takes into account the nonlinearity of the Silicon CMOS transistor and presents a new method to implement a nonlinear model which represents each of the transistors nonlinear elements totally independently. It also proposes a superposition method by which different nonlinear elements of the model can be linearised and the distortion of the transistor re-analysed with the aim of quantifying the individual contribution to distortion from each non-linear element.
- Chapter 5 handles the investigation of the design of an amplifier for the ZigBee standard which can satisfy the performance requirements of both its low noise and power amplifiers. The investigation involves the development of a new design methodology based on graphical visualisation of how the performance parameters vary with respect to design parameters. To demonstrate the validity of the method, a design example is provided showing how different elements of the design flow should be implemented.
- Chapter 6 concludes the thesis by providing a summary of its contribution to the field and outlining some future research proposals based on issues identified during its development.

Chapter 2 TECHNICAL BACKGROUND

2.1 Introduction

After setting the motivation for this work in Chapter 1, this chapter introduces background information which will set the theme for the remainder of this thesis. First, the importance and types of wireless network standards are discussed, and a comparison between them is conducted in terms of their suitability for low power applications. As the one chosen for this work, an overview of the IEEE 802.15.4 standard is given and its relationship to ZigBee is described. Some general front-end issues are then discussed focusing on universal front-end performance considerations and common receiver and transmitter topologies. This is followed by a review of the technologies behind the transistors used in microwave circuit design, concentrating on silicon CMOS as the technology of choice. Being the main focus of this work, and having been investigated extensively in this thesis, the importance and performance requirements of the Low Noise Amplifier (LNA) and Power Amplifier (PA) circuits in a transceiver are then reviewed. This is followed by a discussion of the nonlinearity issue which has a great impact on the operation of both these crucial front-end circuits. Some of the analysis methods adopted in the design and assessment of front-end microwave circuits are then reviewed. Finally, as the assessment application of this thesis, the Tyre Pressure Monitoring System (TPMS) is introduced and its operational performance requirements discussed.

2.2 Wireless Network Standards

For the last few years, the wireless world has been bombarded regularly with news about a new generation of applications based on new or modified forms of RF technologies that is claimed to profoundly impact, if not revolutionise, the way people live and interact with their businesses and environment.

In this section, an insight into some of the wireless standards setup to regulate these evolving technologies, and their applicability, is given from the view point of their suitability for low power applications.

2.2.1 Brief Comparison of Available Wireless Standards

The Institute for Electrical and Electronic Engineers (IEEE) leads the way in developing open, leading-edge consensus standards for Wireless Networks [IEEE-WSZ 2007]. The IEEE identifies four wireless standards working groups. These working groups are; the IEEE 802.11 for Wireless Local Area Networks (Wireless LANs) [IEEE-SA 2007a], the IEEE 802.15 for Wireless Personal Area Networks (Wireless PANs) [IEEE-SA 2007b], the IEEE 802.16 for broadband Wireless Metropolitan Area Networks (Wireless MANs) [IEEE-SA 2007c], and the IEEE P1451.5 for Smart Transducer Interface for Sensors and Actuators [IEEE-SA 2007d]. Figure 2.1 shows a summary of wireless standards.

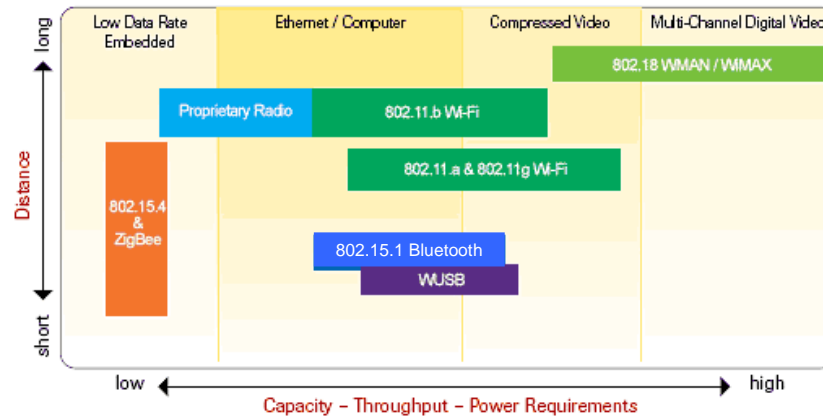


Figure 2.1 Wireless network standards: Their data and distance capabilities [CyberInI 2007]

The IEEE P1451.5 project [IEEE-SA 2007d] was established to provide a standard for wireless communication methods and data formats for transducers. It is to adopt necessary wireless interfaces and protocols to facilitate the use of technically differentiated existing wireless solutions; however it does not specify the wireless system physical design or use. The IEEE P1451.5 project is focused on the software layer and not physical front-end issues.

The IEEE 802.11 [IEEE-SA 1999] with its various standards (a, b, g and n 11-248Mbps) is commonly used for data access where a number of clients are connected to a base station providing them with access to a server or to the Internet. Clients may include computers, Personal Digital Assistants (PDAs) or compatible mobile phones. Transmitted data may be anything from very small documents to very large multimedia items over a range of about 100 metres with transmit powers up to 20dBm. This obviously makes 802.11 neither a low data rate nor a low power standard given the content transmitted and the range they are expected to reach. The broadband IEEE 802.16 [IEEE-SA 2004] is intended for even higher transmission data rates and wider range (up to 31 miles) than 802.11, and thus cannot be considered low power.

The IEEE 802.15 working group provides standards for low-complexity and low power consumption wireless connectivity. There are four main IEEE 802.15 standards; the IEEE 802.15.1 [IEEE-SA 2002], commercially known as Bluetooth (1Mb/s WPAN); the IEEE 802.15.2 Recommended Practice for Coexistence in Unlicensed Bands [IEEE-SA 2003b]; the IEEE 802.15.3 (20+ Mb/s High Rate WPAN) standard [IEEE-SA 2003c]; and finally the IEEE 802.15.4 standard [IEEE-SA 2003a] commercially known as ZigBee (250 kb/s max Low Rate WPAN).

The IEEE 802.15.3 standard [IEEE-SA 2003c] is intended for multimedia and digital imaging, leaving Bluetooth [IEEE-SA 2002, Bluetooth-SIG 2007] and ZigBee [IEEE-SA 2003a, Zigbee 2007] as the only candidates for low data rate and low power transmissions. ZigBee and Bluetooth are two solutions for two application areas hence they complement, not compete with each other. Bluetooth is designed for ad-hoc networks between capable devices with applications such as hands-free audio and small file transfers. Bluetooth is designed to maximise ad-hoc functionality. Bluetooth devices are expected to have rechargeable batteries as opposed to ZigBee devices where the battery is expected to last for a few years but with much lower duty cycle data transmission than Bluetooth. ZigBee is a much simpler standard than Bluetooth. It is intended for static rather than dynamic networks and where many devices are involved. The most important feature of ZigBee is that it is intended for ultra low power and low cost communication and hence it is the standard chosen for this work. The following section provides more information about the vision and current status of the IEEE 802.15.4 standard.

2.2.2 Overview of the IEEE 802.15.4 Standard and ZigBee

The 802.15.4 standard is defined by the IEEE for low-rate, wireless personal area networks [IEEE-SA 2003a]. The standard defines the Physical Layer (PHY) and the Medium Access Control Layer (MAC). The PHY specifications define 3 low-power spread spectrum radios operating at 2.4GHz, 915MHz and 868MHz. At 2.4GHz a basic bit rate of 250kb/s is permitted, with lower data rates at 915MHz and 868MHz. The MAC specifications define how multiple IEEE 802.15.4 radios operating in the same area will share the airwaves. The MAC supports several architectures, including a star topology (where one node acts as a network coordinator), tree topologies (where some nodes speak through other nodes in order to arrive at the network coordinator), and mesh topologies (where nodes share routing responsibilities without the need for a master coordinator).

But just defining a PHY and a MAC does not guarantee that different devices from different manufacturers will be able to communicate. This compatibility is the objective of the ZigBee

Alliance [ZigBee 2007]. The ZigBee Alliance is responsible for defining the network, security, and application layers based upon the IEEE 802.15.4 PHY and MAC layers. ZigBee also provides interoperability and conformance testing specifications. It builds upon the IEEE 802.15.4 standard to define application profiles that can be shared among different manufacturers. For instance, a ZigBee gaming profile would define all the protocols, so one can purchase a ZigBee game console from one manufacturer and a ZigBee game controller from another manufacturer and expect them to work properly together. The relationship between ZigBee and the IEEE 802.15.4 standard is shown diagrammatically in Figure 2.2 [Le 2004].

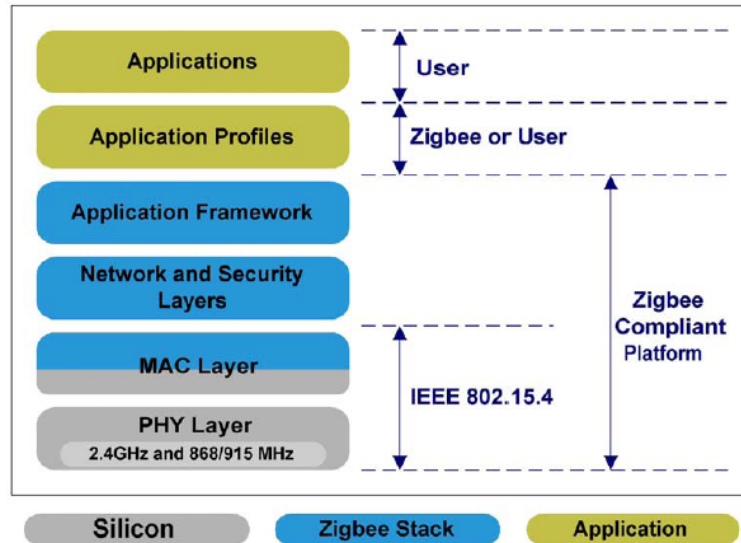


Figure 2.2 IEEE 802.15.4 and ZigBee working model [Le 2004]

ZigBee aims to provide an open standard for low-power wireless networking of monitoring and control devices based on sensors and actuators. Target applications include home and building control, automation, security, consumer electronics, PC peripherals, medical monitoring and toys. Such applications require a technology that essentially offers long battery life (measured in months or even years), small size, high reliability and in particular low system cost.

In many applications, it is not possible to reach a sensor, for instance, to change its battery. Ideally, the battery is good for the life of the sensor. The basic standard for ZigBee is fundamentally efficient in terms of battery performance. Battery lifetimes from a few months to many years are feasible as a result of a host of power-saving modes and battery-optimised network parameters. These power-saving features of the standard will be discussed in more detail in Chapter 3.

ZigBee's air interface is Direct Sequence Spread Spectrum (DSSS) using Binary Phase Shift Keying (BPSK) for 868MHz and 915MHz and Offset Quadrature Phase Shift Keying (O-QPSK) for 2.4GHz. The access method in IEEE 802.15.4-enabled networks is carrier sense

multiple access with collision avoidance (CSMA-CA). The IEEE 802.15.4 PHY includes receiver Energy Detection (ED), Link Quality Indication (LQI) and Clear Channel Assessment (CCA). Both contention-based and contention-free channel access methods are supported. The IEEE 802.15.4 standard supports IPv6 64-bit extended addresses and 16-bit short addresses. This allows it to support theoretically up to 65,536 addresses [Montenegro 2007]. The IEEE 802.15.4 standard [IEEE-SA 2003a] provides thorough details on the operation of the MAC and PHY layers; this thesis work concentrates on the RF Front-End issues.

2.3 RF Front-End Issues

In this section, the performance requirements of RF front-ends will be discussed in some detail. This is followed by a review of RF front-end receiver and transmitter architectures focusing on how they can satisfy these requirements.

2.3.1 Requirements Analysis

The front-end of a wireless receiver must meet several exacting specifications. The most important of these is its sensitivity. Sensitivity is very closely related to the noise performance of the receiver. The noise of the front-end receiver must be sufficiently low to enable a weak input signal (which may not be much larger than the thermal noise floor) to be detected. If the receiver noise is high, the magnitude of the noise may become larger than that of the weak desired signal, and it may not be detected. The sensitivity of a receiver is normally measured in terms of the lowest input power the receiver can detect while maintaining a Bit Error Rate (BER) of the received data below a pre-specified maximum percentage. There is usually a different BER specification for different standards. In a similar fashion, some standards specify the Packet Error Rate (PER) instead of BER [IEEE-SA 2003a].

Radio receivers are generally tuned to a single channel. Each channel has certain bandwidth, which describes the total frequency range that it inhabits. The receiver should accept signals which may be very weak within this channel without responding to signals on other channels, which may be much stronger. This means that the various circuits in the receiver must reject or attenuate signals at other frequencies to a degree determined by the overall performance requirements. This is achieved by a combination of filtering and linearity in the receiver amplifiers. If filtering and linearity are poor in the receiver, strong signals in neighbouring channels, may, through intermodulation with the desired signal, produce components that are within the bandwidth of this signal; thus, distorting the desired signal. Adjacent channel selectivity determines the attenuation requirement at the adjacent channel.

For the receiver to get as clear a desired signal as possible, it is important to maximise the signal-to-noise ratio (SNR) of a transmitter. This requires transmitting as much RF signal power as permitted, while pushing the noise floor as low as possible. But the maximum output level is capped by the amount of distortion introduced by undesirable output spectral components because then these can distort the signal to be received on the other end of the channel. Thus, there is a fine balance between achieving high SNR and spurious-free dynamic range while attaining maximum output power. This imposes the need for linearity in the transmitter.

The level of stringency of these design requirements reflects on the complexity, and hence size and cost, of the wireless system. For instance, a more stringent requirement on adjacent channel rejection may lead to the use of a high Q filter that is difficult to integrate; increasing the size and cost of the wireless system. These design requirements, along with other front-end specification parameters, form the prime concern when the design topology is selected, as will be seen in the next section. As an example of the design requirements and parameters of an RF front-end, the specifications of the IEEE 802.15.4 PHY are shown in Table 2.1 [Le 2005].

Table 2.1 ZigBee specifications at its 3 frequency bands [Le 2005]

Parameter	2.4 GHz PHY	868/915 MHz PHY
Sensitivity @ 1% PER	-85 dBm	-92 dBm
Receiver Maximum Input Level	-20 dBm	
Adjacent Channel Rejection	0 dB	
Alternate Channel Rejection	30 dB	
Output Power (Lowest Maximum)	-3 dBm	
Transmit Modulation Accuracy	EVM<35% for 1000 chips	
Number of Channels	16	1/10
Channel Spacing	5 MHz	single-channel/2 MHz
Transmission Rates		
Data Rate	250 kb/s	20/40 kb/s
Symbol Rate	62.5 ksymbol/s	20/40 ksymbol/s
Chip Rate	2 Mchip/s	300/600 kchip/s
Chip Modulation	O-QPSK with half-sine pulse shaping (MSK)	BPSK with raised cosine pulse shaping
RX-TX and TX-RX turnaround time	12 Symbols	

In Table 2.1, the sensitivity requirement in the 2.4GHz band is specified as -85dBm at 1% PER which means that any ZigBee compliant receiver should be able to achieve a packet error rate of 1% or less when detecting a signal with power as low as -85dBm ($\sim 3.16\mu\text{W}$). The specification of the output power as -3dBm ($\sim 0.5\text{mW}$) (Lowest Maximum) means that any compliant ZigBee transmitter may have a maximum output power not less than -3dBm, or, in other words, the transmitter must be capable of transmitting at least -3dBm of output power. EVM is the Error Vector Magnitude which is a measurement of the residual noise and distortion remaining after an ideal version of the signal has been stripped away [Le 2004]. The adjacent channels are those

immediately on either side of the wanted (tuned) channel. The alternate channels are those one channel further way.

2.3.2 Topologies

Complexity, cost, power consumption and the number of external components are usually the main criteria in selecting a front-end architecture. Practically, no optimum level can be simultaneously achieved satisfying all these criteria. Transmitter and receiver architecture selection is a trade-off issue, and compromises have to be made depending on the requirement and application of the system being designed. For ZigBee, low cost and low power consumption are the primary criteria [Le 2006]. This section will discuss possible receiver and transmitter architectures for a ZigBee transceiver implementation.

2.3.2.1 Receiver Topologies

There exist two main receiver topologies based on phase or frequency modulation, the homodyne (also called zero Intermediate Frequency (IF) or direct conversion) topology and the heterodyne (also called IF) topology. The difference between these topologies is in whether or not an intermediate frequency is used in the demodulation process. A special type of the IF topology is the image-reject (also called low IF) topology.

A simple direct conversion receiver is shown in Figure 2.3 where the incoming signal is converted directly to baseband without the use of an intermediate frequency. The main advantage of direct conversion over IF receivers is that much higher levels of integration can be achieved with direct conversion receivers than with IF receivers [Crols 1998]. This is because high Q filters are not needed, as will be explained. The direct conversion technique was invented many years ago but, even though full integration is a very important factor towards cost reduction, the use of direct conversion has been very limited in the past due to its poor performance compared to receivers based on the IF topology.

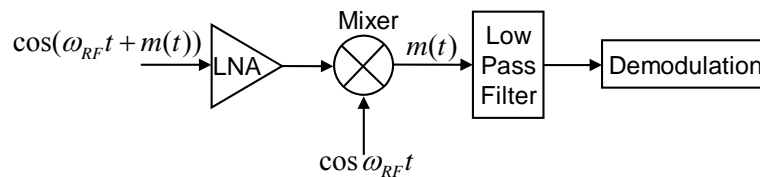


Figure 2.3 Simple direct conversion receiver

The main disadvantage of direct conversion receivers is that because the down converted band extends to the vicinity of zero frequency, extraneous offset voltages may corrupt the received signal. This is known as the DC offset problem [Razavi 1997b]. DC offset is the result of self mixing and oscillator coupling which depends on capacitive and substrate coupling. In the direct

conversion receiver the downconversion mixer is followed by a low pass filter, therefore these DC offsets can saturate the following stages in the receiver. With CMOS scaling, the dimensions of the circuit will be further reduced and this will cause stronger coupling and hence increase DC offsets.

DC offsets can be removed by employing AC coupling (i.e. high-pass filtering) in the downconverted signal path. However, AC coupling in direct conversion receivers cannot be employed if modulation schemes with high energy content at the centre of their spectrum are used since, if the corner frequency of the filter is high, the data signal may be corrupted [Razavi 1997b]. This is the case with ZigBee at 2.4GHz where an O-QPSK modulated signal with a time-bandwidth product of 0.5 has about 99% of information within 50% of its bandwidth [Choi 2003]. In addition, because direct conversion downconverts the signal directly to the baseband, flicker ($1/f$) noise effects (which increases as the frequency decreases) will be higher than in a heterodyne receiver that converts the signal to an intermediate frequency first before converting it to baseband [Razavi 1997b].

Figure 2.4 shows a simple Heterodyne (IF) receiver where some of the disadvantages of the direct conversion scheme are overcome. Because the downconversion mixing stage is followed by a bandpass filter, the DC offset effect is filtered out and is prevented from corrupting the signal or saturating further stages of the receiver.

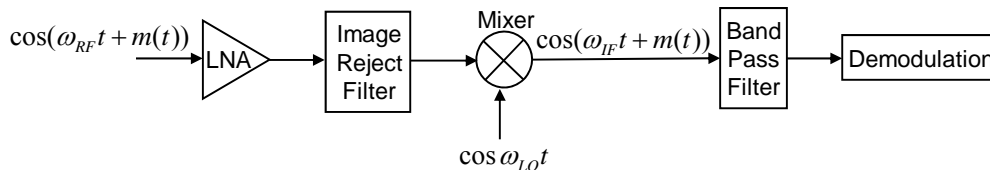


Figure 2.4 A simple Heterodyne receiver

However, because direct conversion downconverts the received RF signal directly to zero frequency, it does not suffer from the image problem as does an IF receiver. The image is an interfering signal at a frequency on the other side of the local oscillator frequency from the desired RF signal (i.e. 2IF away from RF).

If the signal being received is downconverted without eliminating the image, the desired signal and the image overlap resulting in severe corruption of the desired signal [Choi 2003]. Because the image needs to be eliminated at high frequencies before the downconversion, a very high Q and high order filter is required; such a filter may be implemented using a Surface Acoustic Wave (SAW) technique [Razavi 1997b]. However, the problem is that the integration of this filter is very difficult, resulting in a non cost-effective solution.

Low IF architectures are very similar to the heterodyne architecture except that the IF is at very low frequency (e.g. 4MHz [Choi 2003]), which makes them at the same time also very similar to the homodyne architecture. That is, they do not operate in the baseband but near the baseband. This way, low IF receivers combine advantages and eliminate disadvantages from both the other two architectures. Because the IF is very small, the image frequency is very close to the desired frequency and hence its suppression before downconversion requires an impossibly high Q filter preceding the downconversion mixer. Instead, the image can be suppressed after the downconversion using a polyphase filter. Therefore, low IF receivers do not suffer the DC offset problem as in direct conversion, or the high-Q filter integration problem as in heterodyne receivers. But while low IF architectures combine benefits and eliminate problems from both homodyne and heterodyne architectures, they do have their own considerations as will be discussed in Chapter 3. Figure 2.5 shows a simple architecture of a low IF topology.

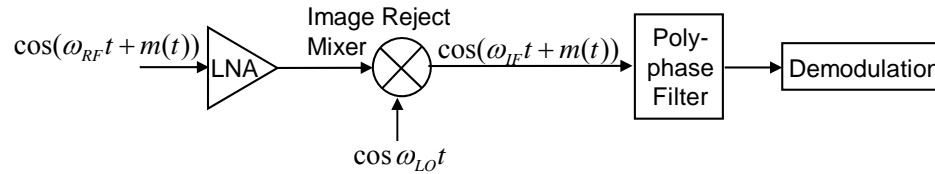


Figure 2.5 Simple Low-IF receiver

In view of the above, RF front-end designers consider the low IF as the most suitable receiver architecture for implementation of the PHY of IEEE 802.15.4 standard.

2.3.2.2 Transmitter Topologies

Transmitter topologies are not as controversial as receiver topologies, since they do not have to deal with unpredictably weak signals. Similar to their heterodyne receiver counterparts, transmitters based on multiple up-conversion and filtering stages do not comply with the low cost and low power requirements of the IEEE 802.15.4 standard [Notor 2003]. There exist two topologies that are most promising for efficient generation of the transmit signal according to the IEEE 802.15.4 PHY specifications; these are the single-step I/Q up-conversion or VCO-based modulation transmitter topologies [Notor 2003, Le 2004].

In the IQ modulator approach, the signal is up-converted using quadrature mixers resulting in a single-sideband RF output. The baseband signals drive the IQ modulator inputs, and the modulator RF output is the sum of the quadrature I and Q channels. The I and Q signals, including half-sine shaping, are processed and modulated in the digital domain and converted to analogue signals using two digital-to-analogue converters (DACs) with output data smoothing

filters to drive the quadrature mixers of the I/Q modulator. Figure 2.6 shows a block diagram of an Offset QPSK I/Q modulator with Half-Sign Shaping [Notor 2003, Le 2004].

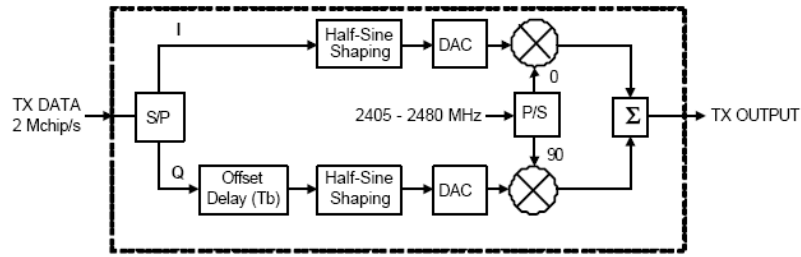


Figure 2.6 Offset QPSK I/Q modulator with half-sine shaping

The VCO-based modulation combines the LO frequency synthesis circuits with modulation circuits to achieve compact MSK modulation with a minimum number of analogue components. This can be achieved by modulating the VCO in closed-loop or open-loop operation.

In the open-loop technique the loop is opened after the Phase Locked Loop (PLL) settles to the desired channel frequency, the VCO is then directly modulated. The implementation is based on a PLL using a multi-modulus divider driven by a multi-bit Delta Sigma (DS) modulator to provide channel frequency selection and digital frequency modulation control, combined with a DAC to provide the analogue modulation control. In the closed-loop approach, the modulation is done by varying the division ratio in the PLL or a combination of varying the division ratio and modulation of the VCO directly. A block diagram of a VCO-based modulation architecture is shown in Figure 2.7.

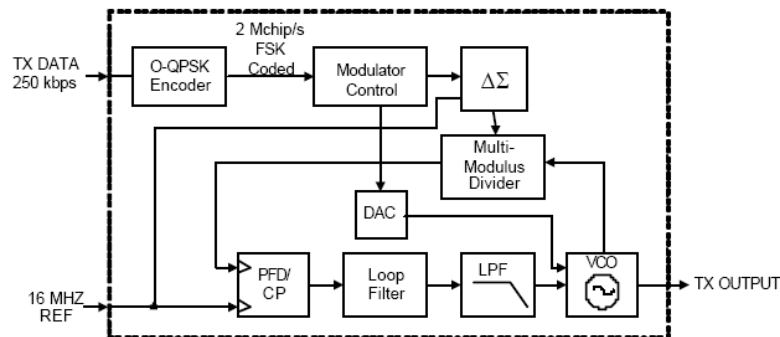


Figure 2.7 2-Point PLL $\Sigma\Delta$ modulator with its VCO.

Both open-loop and closed-loop transmitter architectures are hardware and power efficient for constant envelope modulation schemes as the modulated VCO can drive the power amplifier directly, and several analogue circuit blocks are eliminated. However, the open-loop approach has a severe drawback in terms of frequency stability, which is caused by one or a combination of several effects such as charge leakage, change in operating conditions and mechanical disturbances. The closed-loop VCO modulation does not suffer from the frequency instability,

but usually requires calibration and pre-compensation of the loop bandwidth response or signal transfer characteristic. The required calibration and compensation can be solved efficiently using digital signal processing techniques.

2.4 High Frequency and Microwave Transistors

The term microwave (MW) transistors designates transistors with operating frequencies above 1GHz [Schwierz 2000]. When discussing microwave transistors, several figures of merit which form the basis of comparison have to be considered. For example, the cut-off frequency, f_T , at which the magnitude of the short circuit current gain is 0dB, and the maximum frequency of oscillation, f_{max} , at which the transistor's unilateral power gain is unity, are often quoted. Also, noise figure (NF) and output power, which are critical parameters for RF low noise and power amplifications, respectively, are often considered. The following sections will first discuss the available microwave transistor technologies and then justify why silicon CMOS was the technology of choice in this work for the design of high frequency amplifiers. As a consequence, CMOS technology will then be discussed in more detail.

2.4.1 Available Technologies

In mainstream Very Large Scale Integration (VLSI) electronics such as memories and processors, only two types of transistors are in use, these are Metal-Oxide Semiconductor Field Effect Transistors (MOSFETs) and Bipolar Junction Transistors (BJTs). Both are made from the same semiconductor material; silicon [Schwierz 2000]. In the microwave field, a number of other semiconductors are used in addition to silicon, such as III-V semiconductors and wide band gap materials [Glover 2005]. In addition to MOSFETs and BJTs, these semiconductors enable the realisation of additional field effect and bipolar transistors such as the Metal-Semiconductor Field Effect Transistor (MESFET), High Electron Mobility Transistor (HEMT) and the Heterojunction Bipolar Transistor (HBT) [Glover 2005].

In the past, silicon MOSFETs have been considered as slow devices compared to other types of transistors. Several reasons contributed to this conviction [Liou 2003]. First, the electron mobility, a measure of how fast the free electrons can move in a semiconductor, is by nature lower in silicon than in GaAs and other compound semiconductors. Second, the inversion channel of a MOSFET is located very close to the Si/SiO₂ interface, thereby subjected to the effects of interface roughness, crystal imperfections, and interface traps. As a result, the mobility of free electrons travelling in the inverted channel is further degraded. Finally, another reason for the inferior MOSFET RF performance in the past was the relatively large gate length. In the early 1980s, the gate length of production-stage MOSFETs was in the range of 1.5 to

2.5 μm , while RF GaAs MESFETs with gate lengths between 0.25 and 0.5 μm were quite common and commercially available.

Due to aggressive feature size reduction in the past few years however, MOSFETs are now qualified as RF devices. Today, the MOSFET gate length is comparable or even smaller than that of III-V RF FETs [Liou 2003]. Currently, deep sub-micrometer CMOS processes typically reach several tens of GHz cut-off frequency and show relatively low noise figures, making them a serious alternative to the traditional III-V compound semiconductor devices. In addition, MOSFETs offer very large scale integration making it possible to realise systems-on-chip with high reliability. It should be noted that the progress in MOSFETs capability for RF operation has originated from the developments of VLSI electronics. Most of the advances achieved in increasing the switching speed of CMOS for fast logic circuits have been adopted for RF MOSFETs.

As the ability of high scale integration reduces the cost further, and with cost being a prime requirement in the design of a ZigBee compliant transceivers, CMOS was the technology of choice for this work.

2.4.2 High Frequency CMOS Transistors

The basic structure of MOSFETs, consisting of a single gate, a semiconducting substrate (frequently called body or bulk), and heavily doped source and drain regions, has not changed much in the past years; only the dimensions have been scaled down continuously to meet the demands of higher speed and increased compactness [Liou 2003]. A somewhat significant modification to the MOSFET during the last decade was the introduction of the Silicon-on-Insulator (SOI) technology where the transistor body is separated from the semiconducting wafer by an insulating layer. Advantages of SOI CMOS technology over bulk CMOS are mainly due to deep-submicron devices with low parasitics and high-Q passive components [Zencir 2005]. For System on Chip (SoC) applications, SOI is the substrate of choice since it offers good isolation, reduced crosstalk, and suppresses substrate noise [Sun 2005]. CMOS transistors on SOI also enjoy speed improvements due to the reduction of parasitic capacitances [Huang 1997, Sun 2005]. Due to this, RFIC designers have recently demonstrated interest in using SOI CMOS for low power and high speed applications; see for example [Pinel 2002] [Linten 2005], [Yao 2006], [Leong 2007] and [Gianesello 2007].

The gate length, which is directly related to the effective channel length, is a main feature controlling the MOSFET performance. As CMOS technology is scaled into the nanometre range, the transistors I_{DS} - V_{GS} characteristics in the saturation region become more linear while

the transconductance, minimum noise figure (NF_{MIN}), f_T and f_{MAX} all improve [Yao 2006]. Reducing the gate length, however, requires the scaling of other features such as the oxide thickness, drain and source junction depth, and substrate doping density [Liou 2003].

For today's 50 to 100 nm gate-length MOSFETs, the cut-off frequency can reach 200GHz. Applying a frequently used rule of thumb that the cut-off frequency should be around 10 times the operating frequency of a given application, these devices can be used to design integrated circuits operating at up to 20GHz; an operating frequency higher than that of the vast majority of modern RF electronics [Liou 2003].

It is important to note that the integration of digital baseband and RF front-ends makes RF blocks vulnerable to the coupling of the digital switching noise through the common substrate. This brings the investigation of substrate coupling noise into the forefront [Liao 2003]. Digital switching noise, thus, should be factored into the design process from the beginning. The future of single-chip CMOS RF SoC may well depend on finding an effective solution to this problem but this is beyond the focus of this thesis.

2.5 Low Noise Amplifiers and Power Amplifiers

This section discusses the main characteristics and requirements of low noise amplifiers and power amplifiers; the two transceiver components which dictate wireless systems performance, and the subject of this thesis.

2.5.1 Low Noise Amplifiers

This section first discusses the performance characteristics of the Low Noise Amplifier (LNA), and how they affect the performance of the receiver. This is followed by a review of noise sources in the MOS transistor.

2.3.2.3 Importance and Significance

The Low Noise Amplifier (LNA) is arguably the most critical building block in the receiver path of a transceiver system. In a typical receiver architecture such as the one shown in Figure 2.8, the antenna receives electromagnetic waves from free space and converts them to electric signals in transmission lines. These incoming signals which may be a combination of desired signals and unwanted interferers are then fed to the first stage of the receiver system, the LNA. It is often the case that the desired signals are weaker than the unwanted interferers. The LNA should then be able to provide adequate gain to the desired signals with minimal degradation to their signal-to-noise ratio (SNR).

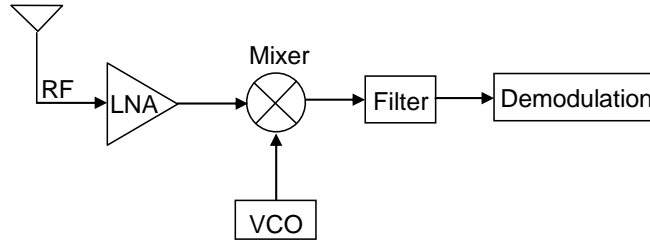


Figure 2.8 A typical RF receiver architecture

The signal-to-noise ratio is a measure of the strength of the wanted signal relative to the background noise. In other words it is the ratio between the wanted signal power to the noise power in any combination. This can be expressed mathematically by Equation (2.1)

$$SNR = \frac{P_{signal}}{P_{noise}} \text{ or in dB } SNR_{dB} = 10 \log_{10} \left(\frac{P_{signal}}{P_{noise}} \right) \quad (2.1)$$

where P_{signal} is the power of the signal and P_{noise} is the noise power. This implies that it is always desirable to have as high SNR as possible. In reality, additional noise is introduced to a signal when passing through any device or a system, the lower this additional noise, the better. The Noise Figure (NF) of a device or a system is defined as the ratio of the input SNR to the output SNR, as given by (2.2)

$$NF_{dB} = SNR_{dB_{input}} - SNR_{dB_{output}} \quad (2.2)$$

Because the output SNR will always be lower than the input SNR (due to the newly introduced noise), NF will always be higher than 1 or 0dB. The larger the noise introduced by a device or a system, the higher its NF.

In a cascade system such as the one in Figure 2.9, the noise figure at any point in the system based on the Friis equation is given by [Razavi 1997a]

$$NF_{Total} = NF_1 + \frac{NF_2 - 1}{A_1} + \frac{NF_3 - 1}{A_1 A_2} + \dots + \frac{NF_n - 1}{A_1 \dots A_{n-1}} \quad (2.3)$$



Figure 2.9 A cascade system

In Equation (2.3), A_n is the gain, and NF_n is the noise figure of the n stage in Figure 2.9. Equation (2.3) indicates that the later the block in the system, the lower its significance on the total NF (as its NF contribution is divided by its gain multiplied by all the gains of the previous stages). This means that the overall noise performance of the receiver is dominated by the noise performance of the first block, which is the LNA. This is why the amplifier at the front of the receiver path needs to be a low noise amplifier. But just as the LNA has to introduce as little

noise as possible it should also provide as high gain as possible (although this gain, which is at higher frequencies, will always be smaller than the gain that can be achieved at lower IF frequencies). By examining Equation (2.3) more closely, it is noticed that the gain of the first block appears in the denominator of every term in the equation. This means the higher the gain of the LNA, the lower the noise figure contribution from every subsequent stage in the receiver path. It is worth noting, however, that although large LNA gain is desirable for the reasons explained, it does have limits on how large it is allowed to be. A very large gain may overload the mixer and compromise dynamic range. Therefore, since the sensitivity of the receiver is defined by the lowest signal level it can detect with certain error limits, and given that the noise performance of the receiver determines the smallest signal it can handle, and that the LNA has a dominant effect on the noise performance of the receiver, it can be concluded that the LNA has a huge impact on the overall sensitivity of the receiver system.

The LNA also has to provide good matching to the antenna over all the frequency band of interest. This would guarantee that the antenna can operate with the desired frequency response and maximise power transfer from the antenna to the receiver system.

The LNA in a receiver system has to provide one additional feature, and that is to protect the antenna from Local Oscillator (LO) frequency leakage through its reverse isolation. As shown in Figure 2.10, if the reverse isolation of the LNA is not adequate, the large LO signal could leak through the mixer and the LNA to the antenna. If this happens, the RF spectrum could be contaminated and this can cause intermodulation and self mixing; establishing DC offset in the baseband.

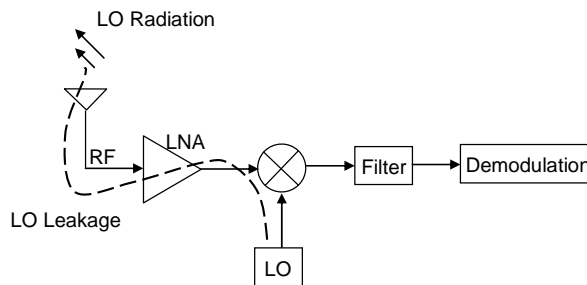


Figure 2.10 LO leakage to the antenna through LNA

2.3.2.4 Sources of Noise in MOS transistor

There are four main sources of noise in the MOS transistor; the channel thermal noise, gate resistance noise, induced gate current noise and flicker (or $1/f$) noise. The first 3 of these sources of noise are represented in the MOS transistor noise model of Figure 2.11.

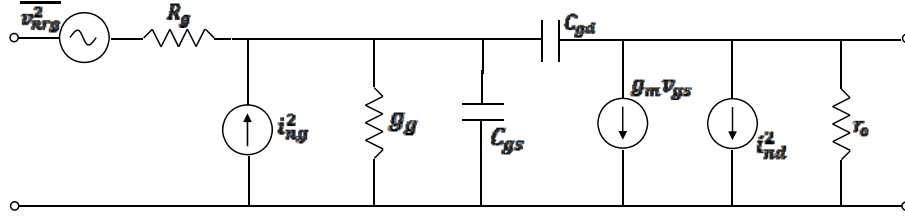


Figure 2.11 The full MOS noise Model

These noise sources can be briefly discussed as follows:

▪ Channel Thermal Noise

In the CMOS noise model of Figure 2.11, this white noise source is modelled as a shunt current source in the output circuit of the transistor (i_{nd}^2), where i_{nd}^2 is given by [Allstot 2004].

$$\overline{i_{nd}^2} = 4kT\gamma g_{d0}\Delta f \quad (2.4)$$

In Equation (2.4) k is Boltzman constant, T is absolute temperature, g_{d0} is the channel conductance at zero drain-source voltage ($g_{d0} = g_{ds}|_{V_{DS}=0}$) and γ is the coefficient of channel thermal noise which is a bias-dependant factor [Abidi 1986, Ge 1998]. As seen from (2.4), γ is proportional to the channel current noise. Thus, the larger γ , the higher the noise contribution from this source. For long channel devices, γ is usually between 2/3 and 1. The least noise (smallest γ) is achieved when the transistor is saturated (which is the normal working mode for amplifiers). If the drain-source voltage is at zero, γ is 1 and the noise reaches its maximum value. However, the high electric fields of short channel devices cause the electron temperature (T_e) to exceed the lattice temperature, this leads to the presence of hot electrons in the channel, and this causes γ to be much higher than its upper long channel boundary if the device is saturated [Abidi 1986]. This, in turn, leads to extra noise.

▪ Gate Resistance Noise

Another source of noise in the MOS device is the noise generated by distributed gate resistance [Jindal 1984]. In Figure 2.11, this noise is modelled by a series resistance in the gate circuit, R_g , and an accompanying white noise generator, $\overline{v_{nrg}^2}$, where R_g is the sheet resistance of the polysilicon. The contribution of this source of noise can be reduced to insignificant levels by interdigitating the device (i.e. reducing the gate resistance) [Shaeffer 1997, Yang 2001].

▪ Induced Gate Current Noise

When the MOS device is biased and the channel is inverted, fluctuations in the channel charge introduce a physical current in the gate due to capacitive coupling. This induced gate noise current is modelled in Figure 2.11 as a current source i_{ng}^2 . At high frequencies approaching f_T ,

and because of the distributed nature of the MOS device, the gate impedance of the device exhibits a significant phase shift from its purely capacitive (C_{gs}) value at low frequencies [Shaeffer 1997]. This effect is included in the noise model of Figure 2.11 by including a real noiseless conductance, g_g , in the gate circuit. The power spectral density of this noise source is given by [Allstot 2004] as

$$\overline{i_{ng}^2} = 4kT\delta g_g \Delta f \quad (2.5)$$

where $g_g = \omega^2 C_{gs}^2 / 5g_{d0}$ and δ is the coefficient of gate noise; classically equals to 4/3 for long-channel devices [Shaeffer 1997].

The gate noise is partially correlated with the drain noise with a correlation coefficient, $c = j0.395$ [van der Ziel 1969] where c is given by [Aparin 2005]

$$c = \overline{i_{ng} \cdot i_{nd}} / \sqrt{\overline{i_{ng}^2} \cdot \overline{i_{nd}^2}} \quad (2.6)$$

The correlated and uncorrelated parts were expressed in [Shaeffer 1997] by

$$\frac{\overline{i_g^2}}{\Delta f} = 4kT\delta g_g (1 - |c|^2) + 4kT\delta g_g |c|^2 \quad (2.7)$$

The first and the second terms of the right hand side of 2.7 are the uncorrelated and the correlated portions, respectively.

▪ **1/f Noise**

1/f noise (also called flicker noise) is inversely proportional with frequency, hence the notion “1/f”. The MOS transistor has the highest 1/f noise of all active semiconductors, due to its surface conduction mechanism. Several theories and physical models aiming to explain the 1/f noise in a MOSFET are in existence. These theories and models differ in detail but they are all based on the mobility fluctuation model expressed by the Hooge empirical relation [Hooge 1969, Hooge 1981], and the carrier density or number fluctuation model first introduced by McWhorter [McWhorter 1957].

Low frequency noise is very sensitive to the technological processes and parameters used in Integrated Circuit (IC) fabrication. Hence, it is important to take into consideration the impact of technological downscaling on the 1/f intrinsic noise sources [Valenza 2004]. Flicker noise not only degrades the noise performance of mixers and phase noise of oscillators, but also adds noise directly to the baseband [Zhang 2001].

2.5.2 Power Amplifiers

A power amplifier is characterised as an amplifier that is able to provide high output power when its load impedance is relatively small (the impedance of the antenna). Power amplifiers are normally preceded by pre-amplifiers which are voltage amplifiers designed to adjust the magnitude of the signal going into the power amplifier to a magnitude that is suitable to the power amplifier specification.

The two most important requirements of a power amplifier are high efficiency and linearity. Since the power amplifier is typically the most power-hungry component of a transceiver, its power efficiency is important as it shows how much RF power it can supply to the load for a given DC power consumption.

The importance of the linearity of the transmitter was discussed briefly in Sections 1.4 and 1.5. Additionally, as far as the power amplifier is concerned, nonlinearity results in DC power losses due to the energy in the generated unwanted harmonic and intermodulation components. For example, if the efficiency obtained from a linear amplifier was 50%, this means half the consumed DC power is converted into RF power. If the amplifier was nonlinear, part of that RF power will go to the unwanted frequency components leaving less than half the RF power to the desired frequency components; hence reducing efficiency.

Power amplifiers are divided into different classes such as A, AB, B, C and E. The classification criterion is based on the amplifier's efficiency. The later the class letter in the alphabetical order the more efficient the amplifier but also the higher its nonlinear distortion [Razavi 1997a]. The class of a power amplifier is decided by its operating point, which eventually determines its efficiency.

2.6 Problem and Evaluation of Nonlinearity

This section is dedicated to the nonlinearities of circuits and systems; their origin and assessment. First, the sources, nature and effect of nonlinearities are discussed. This is followed by a review of the main methods used to represent nonlinearity. Finally, the nonlinearities of the MOS transistor are examined more closely and the relationship between some of its several nonlinear elements discussed.

2.6.1 Nonlinearity of Devices, Circuits and Systems

Nonlinearity occurs simply because of device limitations. Analogue integrated circuits are built with passive and active silicon devices such as transistors, diodes, resistors, capacitors and

inductors. In general, each one of these devices presents a nonlinear behaviour that causes the output signal to deviate from an ideal circuit response. For instance, CMOS transistors are nonlinear active devices where the small signal behaviour is approximated by the 1st-order derivatives of the large signal performance parameters. The small-signal parameters, therefore, are only accurate for small excursions around the DC operating point. Larger excursions from the quiescent point will considerably distort the output signals due to the nonlinear characteristics of the transistors.

Any device or circuit will have nonlinear characteristics wherever it is operating. In general, as signal levels increase, the impact of nonlinearity becomes more significant. From this perspective, circuits can be divided into three categories depending on how nonlinear they are; the quasilinear, the weakly nonlinear and the strongly nonlinear [Maas 1998]. Quasilinear circuits are those who operate in a region where nonlinearities have negligible effect on the performance of the circuit. In weakly nonlinear circuits, nonlinearities do affect the circuit and have to be reduced but they do not rise to a level where the DC bias characteristics are changed. A strongly nonlinear circuit is one where this latter situation occurs.

Nonlinearity enables the generation of new frequency components relating to the excitation signal frequency. These frequencies can be in the form of harmonics which, themselves are relatively insignificant, because they are normally far from the operating frequency and so are unlikely to cause distortion. But if multiple harmonics are mixed together or with the fundamental frequency, intermodulation occurs where frequency components very near to the desired frequency may be produced. Another dangerous effect of nonlinearity is saturation where an amplifier may be saturated by a strong interferer adjacent to a weak desired signal hence not providing any amplification to the desired signal.

Linearity, or the lack of it, of a system or device can be assessed using several parameters. One of these is the Intercept Point (IP) which is a measure of how significant the unwanted intermodulation products generated are. Another is the 1dB compression point which is the point at which the gain of the nonlinear amplifier falls by 1dB below the gain that would have been achieved had this amplifier been linear.

Efficiency is also a measure of the linearity of an amplifier. Amplifier efficiency can be measured either as the drain efficiency or the power added efficiency (PAE). The drain efficiency is the ratio of the output power delivered to a load divided by the consumed DC power. However, PAE takes into account the RF input power applied at the input of the amplifier and it is defined as the difference between the RF output and input power, divided by the DC power consumption as in Equation (2.8). Hence, PAE is more representative of the

linearity of the amplifier. PAE is always less than drain efficiency. The efficiency of the power amplifier is of prime importance since the power amplifier consumes most power in the front-end of a transceiver.

$$PAE(\%) = \frac{P_{RFout} - P_{RFin}}{P_{DC}} \times 100 \quad (2.8)$$

Several nonlinearity representation and analysis methods exist, including, time domain, harmonic balance and Volterra series analysis [Maas 1998]. The applicability and limitations of these methods will be discussed in more details in the following section.

2.6.2 Nonlinearity Representation and Analysis Methods

In this section, the main methods used to represent and analyse nonlinearities are briefly reviewed. These methods are Power Series, Volterra Series, Time Domain and Harmonic Balance.

▪ Power Series

Power series can be used to represent the behaviour of any nonlinear system or circuit, an example of which is Equation (2.9) for the transconductance of an NMOSFET transistor.

$$i_D(v_{gs}) = I_{DC} + g_1 v_{gs} + g_2 v_{gs}^2 + g_3 v_{gs}^3 + \dots \quad (2.9)$$

In Equation (2.9) g_1 is the small signal transconductance and the higher-order coefficients (g_2, g_3, \dots , etc.) define the strengths of the corresponding orders of nonlinearity.

The number of terms needed to sufficiently represent the nonlinear behaviour is a reflection of the strength of nonlinearity. As long as the device or circuit does not enter into saturation, the nonlinearity is considered weak and can be represented by relatively small number of terms depending on the required accuracy. The more the device enters into deep saturation, the more terms are needed. Deep saturation is the operation region where the output changes very slowly with respect to the input, eventually becoming constant irrespective of any variation in the input. Every term of the power series represents a nonlinearity order; for instance the 3rd-order term and its coefficient represents the strength of the 3rd-order nonlinearity. Practically, this reflects the strength of the 3rd-order harmonic or intermodulation components that will result from the presence of the 3rd-order nonlinearities such as IMD_3 .

Power series are efficient in representing the nonlinearity of a system or a device irrespective of its nonlinearity. The main disadvantage of the power series however, is that it cannot represent the nonlinearity of frequency-dependent components or components with memory; a drawback overcome by Volterra series as will be shown.

▪ Volterra Series

Volterra-series was developed as a method of nonlinear system analysis by Norbert Weiner [Weiner 1980] based on the work of the Italian mathematician Vito Volterra. Weiner explained how the Volterra functional series can be used to describe nonlinear systems, as long as they were not very strongly nonlinear. Unlike conductances, capacitors and inductors are components with memory (they hold a charge) and their reactances are frequency dependant. Therefore, nonlinear effects based on the nonlinear behaviour of capacitors and inductors cannot be modelled using a time-independent power series. For this, Volterra series is used. Volterra series are time-dependent power series which can be used to describe systems where memory and frequency dependence are involved; see for example [Lambrianou 1985], [Li 2002], [Baki 2006] and [Kim 2006b].

Without going into the mathematical complications involved in the development of Volterra series, and its application for the analysis of nonlinear circuits, a simple approach to understanding the technique when applied to nonlinear circuits is provided here. This is to facilitate the understanding of equations that will be discussed in Chapter 5. This is done using the simplified NMOS transistor model with input matching impedance shown in Figure 2.12 [Aparin 2005].

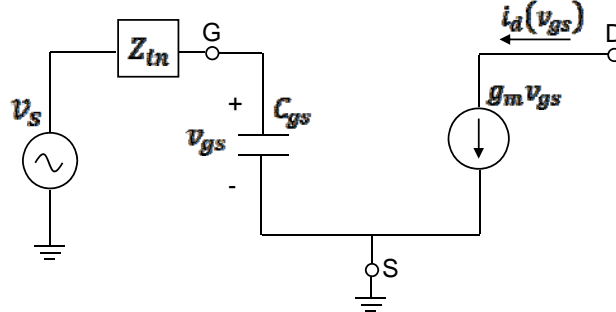


Figure 2.12 Simplified MOSFET model with input matching

The frequency independent and memory-less nonlinear part of the model (the transconductance) was expressed in a Taylor series expansion in Equation (2.9)

The first 3 coefficients in (2.9) for weakly nonlinear operation can be defined as

$$g_1 = \frac{\partial i_d}{\partial v_{gs}} \quad g_2 = \frac{1}{2!} \frac{\partial^2 i_d}{\partial v_{gs}^2} \quad g_3 = \frac{1}{3!} \frac{\partial^3 i_d}{\partial v_{gs}^3} \quad (2.10)$$

The transfer function between the excitation voltage, v_s , and v_{gs} , which includes a non linear frequency dependant and memory component (the passive input matching impedance), can be expressed in a Volterra series as

$$v_{gs} = A_1(s) \circ v_s + A_2(s_1, s_2) \circ v_s^2 + A_3(s_1, s_2, s_3) \circ v_s^3 \quad (2.11)$$

where $A_1(s)$, $A_2(s_1, s_2)$ and $A_3(s_1, s_2, s_3)$ are the Laplace transform of the 1st-, 2nd- and 3rd-order Volterra kernels, respectively. The Volterra kernel is a function used to determine the Volterra operator which is an order of the Volterra series ([Wambacq 1998] and [Koolivand 2006]).

The 1st-order in (2.11) is the linear part and the n^{th} -order terms are often called the n^{th} -order nonlinear transfer functions. $s_i = j\omega$ is the Laplace variable, where i indicates the number of the frequency component involved in determining its respective term in the Volterra series. The operator “ \circ ” means that the magnitude and phase of each spectral component of v_s^n is to be changed by the magnitude and phase of $A_n(s_1, s_2, \dots, s_n)$ [Aparin 2005].

Thereafter the relationship between the drain current and the excitation voltage can be described in another Volterra series as

$$i_d(v_s) = B_1(s) \circ v_s + B_2(s_1, s_2) \circ v_s^2 + B_3(s_1, s_2, s_3) \circ v_s^3 \quad (2.12)$$

The Volterra series in (2.12) effectively combines the Taylor series in (2.9) and the Volterra series in (2.11) in such a manner where $B_1(s)$ is determined in terms of the 1st-order terms in (2.9) and (2.11), $B_2(s_1, s_2)$ is determined in terms of the 2nd- and lower-order terms of (2.9) and (2.11) and $B_3(s_1, s_2, s_3)$ is determined in terms of the 3rd- and lower-order terms of (2.9) and (2.11). A demonstration of how this works is presented in Appendix A, where the derivation of the input 3rd-order intercept-point (IIP₃) of a simplified NMOS model is performed.

In the Volterra series of Equations (2.11) and (2.12), the number of frequencies in each order matches the number of that order. For example, 3 frequencies are involved in the determination of the 3rd-order product. This does not mean that 3 excitation frequencies are needed to determine the 3rd-order product but 3 frequency components often called *mixing products* are needed to determine it [Maas 1999a]. The frequency of each component in the series is the sum of all the frequencies in the order. For instance, the frequency of $B_3(s_1, s_2, s_3)$ is $\omega_1 + \omega_2 + \omega_3$. The example of calculating IIP₃ can be used to explain this: If the 3rd-order intercept point for a two tone excitation system, where the excitation $v_s = A[\cos(\omega_a t) + \cos(\omega_b t)]$, is to be calculated from the Taylor series of Equation (2.9), it will be [Aparin 2004]

$$A_{IP3} = \sqrt{\frac{4}{3} \left| \frac{g_1}{g_3} \right|} \quad (2.13)$$

And from Volterra series of Equation (2.12) it will be [Ha 1981]

$$A_{IP3} = \sqrt{\frac{4}{3} \left| \frac{B_1(s_a)}{B_3(s_b, s_b, -s_a)} \right|} \quad (2.14)$$

Having that the fundamental frequencies are ω_a and ω_b , and the frequency of the 3rd-order intermodulation (IM₃) term is $2\omega_b - \omega_a$, $B_3(s_b, s_b, -s_a)$ realises $2\omega_b - \omega_a$ through $\omega_b + \omega_b - \omega_a = 2\omega_b - \omega_a$.

Volterra analysis also includes phase information (ignored in a memory-less Taylor series of Equation (2.9)). This makes it suitable for simulating high frequency effects such as AM-PM conversion [Maas 1998].

The most common method of determining the Volterra kernels is the harmonic input method. The harmonic input method is based on probing the circuit with a multi-tone excitation and solving the Kirchhoff's law equations in the frequency domain at the sum of all input frequencies. The procedure starts with a single-tone excitation to determine the linear transfer function and is continued to higher-order functions by adding one more input tone at each step.

The main disadvantage of Volterra series is that it is only valid in the weak nonlinear region [Maas 1999b]. In practice, this means that the excitation level must be kept well below saturation. In a class-A amplifier, this means that Volterra analysis may not be valid after the point where the excitation pushes the amplifier to its 1dB compression level. It is important to note, however, that the 1dB compression point is not a definitive boundary between weakly and strongly nonlinear operations but is an indication to the start of compression. Volterra analysis is, therefore, not appropriate for mixers, modulators, frequency multipliers, or other strongly driven nonlinear circuits.

▪ Time Domain and Harmonic Balance

Transient analysis, in the case where there are harmonics and/or closely-spaced frequencies, is very time and memory consuming. This is because the minimum time step must be compatible with the highest frequency present while the simulation must be run for long enough time to observe one full period of the lowest frequency present.

Harmonic balance is a frequency-domain analysis technique for simulating nonlinear circuits and systems. It is well-suited for simulating analogue RF and microwave circuits, since these are most naturally handled in the frequency domain.

The harmonic balance method works in an iterative manner. It is based on the fact that for a given sinusoidal excitation there exists a steady-state solution that can be approximated, to satisfactory accuracy, by means of a finite Fourier series. Consequently, the circuit node voltages take on a set of amplitudes and phases for all frequency components. The currents flowing from nodes into linear elements, including all distributed elements, are calculated by means of a straightforward frequency-domain linear analysis. Currents from nodes into nonlinear elements are calculated in the time-domain. Generalized Fourier analysis is then used to transform from the time-domain to the frequency-domain. Hence, a frequency-domain representation of all currents at all nodes becomes available. According to Kirchhoff's Current Law (KCL), these currents should sum to zero at all nodes. The probability of obtaining this result on the first iteration is extremely small. Therefore, an error function is formulated by calculating the sum of currents at all nodes. This error function is a measure of the amount by which KCL is violated and is used to adjust the voltage amplitudes and phases. If the method converges (that is, if the error function is driven to a given small value), then the resulting voltage amplitudes and phases approximate the steady-state solution.

2.6.3 MOSFET Nonlinearity

The distortion behaviour of the MOSFET comes from its nonlinear device parameters such as the transconductance, g_m , the output conductance, g_{ds} , the gate to source capacitance, C_{gs} , and the gate to drain capacitance, C_{gd} , with the main source of distortion being the nonlinear behaviour of the transconductance. In Chapter 4, a new transistor model is developed to try to quantify the individual contribution to distortion from these nonlinearity sources.

It is important to emphasise that there are interactions between the above-mentioned nonlinear parameters in the MOSFET transistor, and of most concern are interactions that directly affect the nonlinearity of the amplifier current. The drain current in a MOSFET is dependent on three conductances; the transconductance, the output conductance and the bulk conductance, g_{mb} . Assuming that all these conductances are nonlinear and the bulk and source of the transistor are not connected, the drain current can be described as follows

$$i_d(v_{gs}, v_{ds}, v_{sb}) = \quad (2.15)$$

$$g_{m_1} v_{gs} + g_{m_2} v_{gs}^2 + g_{m_3} v_{gs}^3 + \quad (2.15-1)$$

$$g_{ds_1} v_{ds} + g_{ds_2} v_{ds}^2 + g_{ds_3} v_{ds}^3 + \quad (2.15-2)$$

$$g_{mb_1} v_{sb} + g_{mb_2} v_{sb}^2 + g_{mb_3} v_{sb}^3 + \quad (2.15-3)$$

$$g_{m_1}g_{ds_1}v_{gs}v_{ds} + g_{m_1}g_{mb_1}v_{gs}v_{sb} + g_{ds_1}g_{mb_1}v_{ds}v_{sb} + \quad (2.15-4)$$

$$g_{m_2}g_{ds_1}v_{gs}^2v_{ds} + g_{m_2}g_{mb_2}v_{gs}^2v_{sb} + \quad (2.15-5)$$

$$g_{m_1}g_{ds_2}v_{gs}v_{ds}^2 + g_{ds_2}g_{mb_1}v_{ds}^2v_{sb} + \quad (2.15-6)$$

$$g_{m_1}g_{mb_2}v_{gs}v_{sb}^2 + g_{ds_1}g_{mb_2}v_{ds}v_{sb}^2 + \quad (2.15-7)$$

$$g_{m_1}g_{ds_1}g_{mb_1}v_{gs}v_{ds}v_{sb} \quad (2.15-8)$$

In (2.15), the first three terms on the right-hand side (2.15-1) represent the small signal linear transconductance and its 2nd- and 3rd-order nonlinear derivatives. The next three terms (2.15-2) are the small signal linear output conductance and its 2nd- and 3rd order nonlinear derivatives. The following three terms (2.15-3) represent the small signal linear bulk conductance and its 2nd- and 3rd-order nonlinear derivatives. The rest of the nonlinear terms in (2.15.4) to (2.15.8) are referred to as the cross modulation terms [Pedro 1994]. These terms describe the relationship between these nonlinearities and their impact on each other. For instance, the term $g_{m_1}g_{ds_1}v_{gs}v_{ds}$ in (2.15.4) describes the dependence of g_m on V_{DS} or the dependence of g_{ds} on V_{GS} and the term $g_{m_1}g_{mb_2}v_{gs}v_{sb}^2$ in (2.15.7) describes the dependence of g_m on V_{SB} or the dependence of g_{mb} on V_{GS} .

Modern, short-channel, MOSFET devices deviate from the ideal square-law operation that normally characterises long-channel devices. Figure 2.13 shows how the 3rd-order transconductance nonlinearity (g_{m_3} or g_m'') of a common-source short-channel MOSFET transistor changes as the transistor moves from the weak inversion to the moderate inversion and then the strong inversion regions (as V_{GS} increases). The result in Figure 2.13 was produced for the 180nm NMOS transistor used for analysis throughout this thesis.

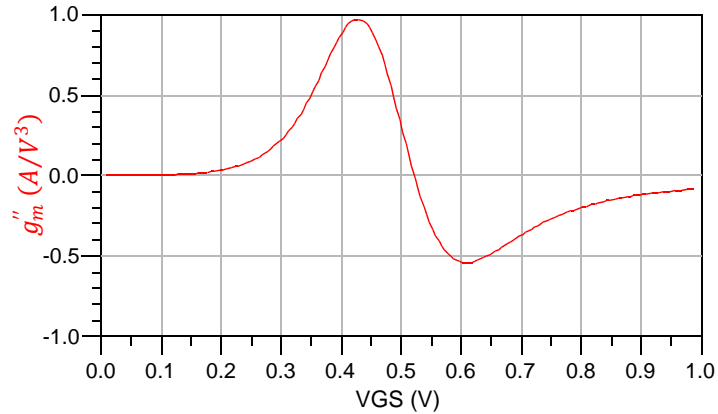


Figure 2.13 Change in MOSFET 3rd-order transconductance nonlinearity at different operation regions

In short-channel devices, g_{m_3} becomes negative in strong inversion and as a result, the device experiences gain compression [Toole 2004]. This is because all odd-order nonlinearities contribute to the fundamental (if they are mathematically broken they would have a

fundamental component within) and the significance of g_{m3} becomes higher at high input powers. Therefore, at high input powers, the fundamental component produced from g_{m3} will be deducted from the fundamental output component (due to g_{m1}) and so compression will take place. As g_{m3} passes through zero in the moderate inversion region when the operation of the transistor changes from weak inversion to strong inversion, the device acts as an ideal square-law device with no 3rd-order distortion, implying that no 3rd-order intermodulation products will result and therefore IIP_3 is expected to be infinite. This point is called the “sweet-spot” [De Carvalho 1999]. In reality, however, even though the direct source of the 3rd-order intermodulation distortion is minimised, some 3rd-order products are produced from the higher order nonlinearities such as 5th- and 7th-order, dominating the 3rd-order distortion and limiting the expected peak in IIP_3 . A more deep insight to the extent of this will be given in Chapter 5 where polynomial equations are used to assess the higher-order contribution to the 3rd-order intermodulation both through the higher-order of nonlinearity or after feedback.

Also, making use of the sweet-spot in a high frequency LNA might not be as straightforward [Baki 2005]. In a practical LNA, as the input impedance needs to be matched to the characteristic impedance of the system (typically 50Ω), matching components are added to the input of the transistor at the source and the gate. This impedance matching establishes a feedback path from the output to the input fundamental tone. This degrades the IIP_3 far below its original peak at the sweet spot, making that value of V_{GS} no longer the optimum operating point.

2.7 Two Port Networks

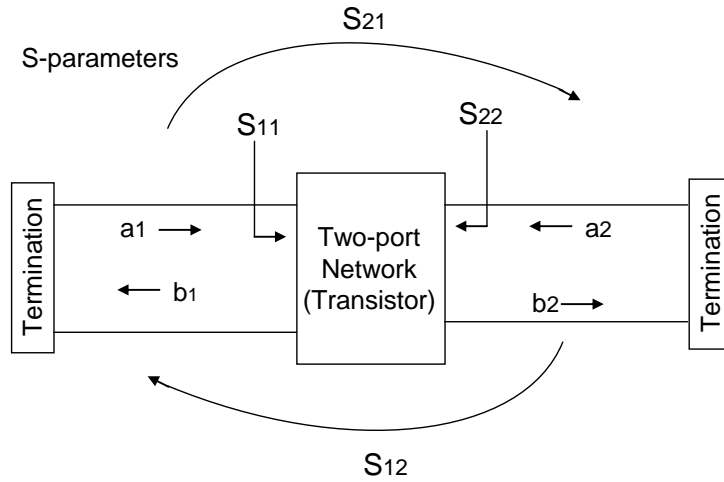
In this section, a brief review of the circuit analysis and characterisation methods used in high frequency front-end amplifier design is given. Attention is first drawn to the theory of two-port network characterisation using S-parameters. This is followed by a discussion of reflection coefficients and the importance of impedance matching. Building on that, two-port network stability assessment and gain characterisation are handled. The section concludes with a discussion of the capabilities and limitations of two-port characterisation at small- and high-power levels.

2.7.1 Review of S-Parameter Analysis

The conventional methods of low frequency circuit design and analysis, based on summing currents and voltages, is not sufficient in high frequency designs. This is because parasitics of circuit components and signal lines become more significant at high frequencies. This means conventional small-signal models used in low frequency analysis become less accurate and new

models, including all parasitics, are so complicated that they cannot be solved analytically. This leads to the use of S-parameters which relate travelling waves (power) to the reflection and transmission behaviour of a two-port network.

S-parameters (scattering parameters), belong to a group of two-port parameters used in two-port theory including Y, Z, G and H parameters. All these parameters are bias and frequency dependent but the difference between Y, Z, G or H and S-parameters, is that S-parameters relate to the travelling waves that are scattered or reflected when a network is inserted into a transmission line of certain characteristic impedance Z_0 . S-parameters are more popular in microwave circuit analysis and frequency response characterisation than Y, Z, G or H parameters.



(a)

$$s_{11} = \left. \frac{b_1}{a_1} \right|_{a_2=0} = \text{Input reflection coefficient with the output port terminated by a matched load } (Z_L = Z_0 \text{ sets } a_2=0)$$

$$s_{22} = \left. \frac{b_2}{a_2} \right|_{a_1=0} = \text{Output reflection coefficient with the input terminated by a matched load } (Z_S = Z_0 \text{ sets } a_1=0)$$

$$s_{21} = \left. \frac{b_2}{a_1} \right|_{a_2=0} = \text{Forward transmission (insertion) gain with the output port terminated in a matched load.}$$

$$s_{12} = \left. \frac{b_1}{a_2} \right|_{a_1=0} = \text{Reverse transmission (insertion) gain with the input port terminated in a matched load.}$$

(b)

Figure 2.14 The origin of S-parameters (a) Incident and reflected waves of two-port networks and (b) Determining S-parameters from these waves

Figure 2.14 details how S-parameters are defined. Figure 2.14.b details how the 4 scattering parameters which are sufficient to describe any two port microwave circuit are extracted from

the reflection coefficients and Figure 2.14.a shows graphically their origin. A two port network is placed between two terminations at its input and output. Two waves a_1 and a_2 are incident from the terminations towards the network, these result in reflected waves b_1 and b_2 . The reflected wave b_1 is the sum of part of a_1 which is reflected back from the network towards the input termination and part of a_2 which was transmitted through the network from the output to the input. Likewise, the reflected wave b_2 is the sum of part of a_2 which is reflected back from the network towards the output termination and part of a_1 which is transmitted through the network from the input to the output. Therefore, S-parameters can completely describe a two-port network by representing its reflection (S_{11} , S_{22}) or transmission (S_{12} , S_{21}). However, S-parameters, like all other two-port parameters are linear by default, so they cannot be used to describe the behaviour of a network operating at high powers when nonlinearities become more significant.

Later in this section, the application of S-parameters in the analysis and characterisation of circuit performance will be discussed.

2.7.2 Reflection Coefficients and Matching

The need for matching networks arises because amplifiers, in order to deliver maximum power to a load or to achieve other performance parameters such as noise, gain or intermodulation distortion, must be properly terminated at both the input and the output ports [Gonzalez 1996]. The choice of a matching network affects the gain, noise, stability and linearity of an amplifier.

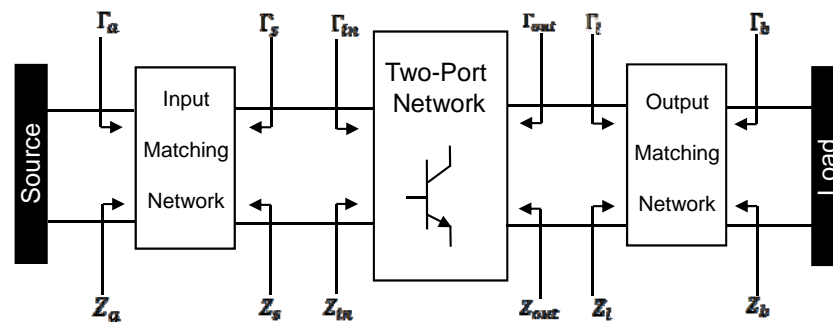


Figure 2.15 Two port network with matching networks, reflection coefficients and associated impedances.

Figure 2.15 shows a two port network terminated with source and load impedances matched at the input and the output. The input matching network is designed to transform the generator impedance to the required source impedance, Z_s , while the output matching network transforms the terminating impedance to the required load impedance, Z_l . Every reflection coefficient, Γ_i , is associated with an impedance, Z_i . A range of reflection coefficients and impedances are defined for use in various calculations when the design of the matching network is considered. Apart

from Γ_s and Γ_l , Γ_{in} , Γ_{out} , Γ_a and Γ_b are also defined in Figure 2.15 and their directions are indicated.

For maximum power transfer between the source and the network, the input matching network has to be designed such that the input is conjugately matched i.e. $\Gamma_s = \Gamma_{in}^*$ (or $Z_s = Z_{in}^*$). The same applies for the output port between Γ_l and Γ_{out} (or Z_l and Z_{out}). Apart from conjugate matching for maximum power transfer, the matching networks can be chosen to provide optimum amplifier performance in another areas such as for example the noise or linearity.

In nonlinear circuits, the voltage source may contain more than one frequency component and the source or load impedance and the impedance of the matching network may not be the same at each frequency. Matching networks could then be designed for maximum gain for the in-band frequency only and attenuate other out-of-band components. They can also be used to terminate certain frequency components which would otherwise degrade the linearity of the circuit. Chapter 5 will discuss the optimisation of matching networks in more detail and show how matching networks can be optimised to obtain a trade-off between the many target performance parameters of an amplifier.

2.7.3 Stability

The stability of an electrical network can simply be defined as its resistance to oscillate. Oscillations occur if either network ports presents a negative resistance; this happens when $|\Gamma_{in}| > 1$ or $|\Gamma_{out}| > 1$. Although negative resistances are not desired in amplifiers, they are required (but with control) in oscillators where oscillation is the main function of the circuit.

In terms of their potential stability, two port networks can be classified under two types; *unconditionally stable* and *potentially unstable*. An unconditionally stable network is one that is expected to be stable when presented with any passive termination at its input and output. A potentially unstable network is one that could be unstable when presented with some certain passive terminations at its input or output.

Potential instability can be dealt with in two ways; either by identifying the region of impedances that would cause instability and avoid them in designing Z_s and Z_l of Figure 2.15, or by stabilising an unstable network (i.e. cancel the negative resistance). Stable and unstable impedance regions can be identified using stability circles on the Smith chart. The source (or input) stability circle consists of all the source reflection coefficients, Γ_s (or source impedances Z_s), as in Figure 2.15, which would result in $\Gamma_{out} = 1$ [Pozar 2005]. This is so because Γ_{out} is determined by the impedance of the active network and Z_s , and the impedance of the active network is static. The rest of the Smith chart impedances are either outside or inside this circle.

It is then the case that either all the impedances outside the circle will result in $\Gamma_{out} < 1$ (positive real part in Z_{out}) leading to stability, or in $\Gamma_{out} > 1$ (negative real part in Z_{out}) leading to instability. Whichever the case, the opposite applies to the impedance inside the circle. Similarly, the load (or output) stability circle specifies all the output reflection coefficients that result in $\Gamma_{in} = 1$, and the rest applies accordingly.

The situation is not as straightforward in nonlinear circuits. This is because the kinds of interactions that can occur in nonlinear circuits are more complex than those taking place in linear circuits. Nonlinear circuits often exhibit transient and steady state phenomena other than sinusoidal oscillation that although bounded, are loosely classed as instability [Maas 1998]. These phenomena include parasitic oscillations, spurious outputs, and exacerbation of normal noise levels.

2.7.4 Gain Characterisation

Several gains are defined for a two port network. These gains can be classified under two types of networks, the *unilateral* and the *bilateral*. A unilateral two port network is defined as the two port network which has an S_{12} equal to 0 or so small that it can be neglected, while the bilateral case is when S_{12} is large enough not to be neglected. The unilateral case is more common in passive rather than active networks; hence only gains of the bilateral case will be discussed here. In the bilateral case, three gains are defined as shown in Figure 2.16.

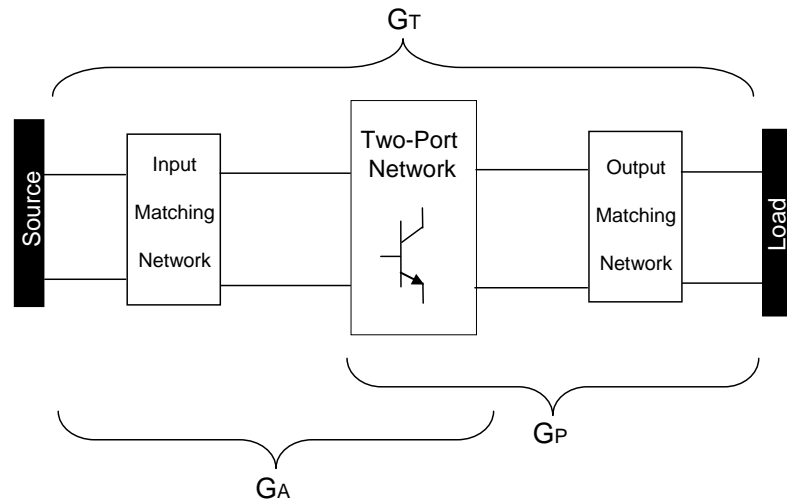


Figure 2.16 Gains of a bilateral two port network

The transducer power gain, G_T , is defined as the ratio between the power delivered to the load and the power available from the source. The power available from a source is defined as the maximum power that can be obtained from that source; which is the power delivered by the source to a conjugately matched load. The operating power gain, G_P , is defined as the ratio

between the power delivered to the load and the input power to the network. And finally, the available power gain, G_A , is defined as the ratio between the power available from the network and the power available from the source.

Under simultaneous conjugate match conditions $G_T = G_P = G_A$, it follows that $G_{T_{max}} = G_{P_{max}} = G_{A_{max}}$. Where $G_{T_{max}}$ is given by

$$G_{T_{max}} = \frac{|S_{21}|}{|S_{12}|} \left(K - \sqrt{K^2 - 1} \right) \quad (2.16)$$

where K is the stability factor.

The transducer gain is the most meaningful gain expression as it considers the overall transfer of power from the input to the output inclusive of the effect of the network gain or loss and matching losses. This phenomenon can also be used in efficiency calculations in the analysis of nonlinear circuits.

As will be seen in the next section, care should be taken when relating certain frequency components at the output to certain frequency components at the input when the circuit is nonlinear.

2.7.5 Two Port Measurements in Small and Large-Signal designs

Two port tests can supply RF designers with all the information they need to build and optimise wireless systems. Small signal two-port tests on two port networks can be performed using either S-parameter simulations in computer aided design packages on circuit or device models, or using vector network analysers on fabricated circuits. In both cases, a very small sinusoidal signal at only one frequency at a time is injected to both input and output ports terminating the circuit or device, and the reflection coefficients are calculated. From these values, scattering parameters can be calculated as has been shown; these are complex numbers that account for the signal's amplitudes and phases. The S-parameters can then be converted into impedance (Z) or admittance (Y) parameters [Everard 2001]. These values can then be used in suitable equations to obtain values for different parameters that needs to be optimised in the circuit such as gain, noise and stability of the device under test (DUT). An example for this is the expression used to calculate the transducer gain as shown in Equation (2.17) [Vizmuller 1995].

$$G_T = \frac{|S_{21}|^2 (1 - |\Gamma_L|^2) (1 - |\Gamma_S|^2)}{|1 - S_{11}\Gamma_S - S_{22}\Gamma_L + \Delta\Gamma_S\Gamma_L|^2} \quad (2.17)$$

In small signal tests, since the DUT is operating within its linear region, the input and output signal frequencies are the same as the injected frequency. In large signal operation, however,

when the device or circuit is pushed into its nonlinear region, this is no longer the case. Nonlinearities transfer energy from the injected frequency to products at new frequencies, invalidating the assumption of superposition in linear two-port network analysis. If the input signal is a large sine wave, input and output signals will be combinations of sine-wave signals, caused by the nonlinearity of the DUT. These will be either harmonics, or harmonics and intermodulation components of the injected signal depending on whether one or more frequencies are injected. Thus, under large signal operating conditions, complex travelling waves are measured at the terminating ports of a DUT not only at the stimulus frequency but also at other frequencies where energy may be generated [Jargon 2004]. These tests can be done using either large signal S-parameter simulation in computer packages on models or nonlinear vector network analysers on fabricated circuits. With these facts in mind, the major difference between a linear S-parameter test and a nonlinear large signal S-parameter test is that in linear S-parameter tests, ratios between input and output waves are taken at one frequency at a time while in large signal S-parameter test the actual input and output waves are measured simultaneously over a broad band of frequencies. These can then be used to obtain figures for nonlinear circuit design parameters such as IIP_3 .

2.8 TPMS General Structure and Requirements

As the application of choice, this section is dedicated to discuss the Tyre Pressure Monitoring System (TPMS) in some detail and reflect on the requirements of this wireless application from the perspective of the RF transceiver.

Inadequate air pressure leads to increased tyre flexing and premature tyre wear. In turn, especially at high speeds, this can lead to tyres no longer being able to withstand loads and subsequently bursting. A tyre pressure monitor warns of sudden and creeping loss of pressure and can thus protect the driver from becoming one of many victims of accidents caused by defective tyres or punctures; the majority of which are caused by inadequate tyre pressure. It can double the service life of tyres and reduce fuel consumption [Beru 2000].

An example of a tyre pressure monitoring system installed in a 4-tyre passenger car is shown in Figure 2.17.a while in Figure 2.17.b a diagrammatic structure of the contents of the system is presented. A TPMS consists of two main modules, the *Tyre Sensor Module* and the *Central Module*. The sensor module is fit in the tyre and consists of three main components, the pressure sensor, a microcontroller and the transmitter/transceiver. The central module, fitted on or near the dashboard of the car, consists of a receiver/transceiver and a microcontroller.

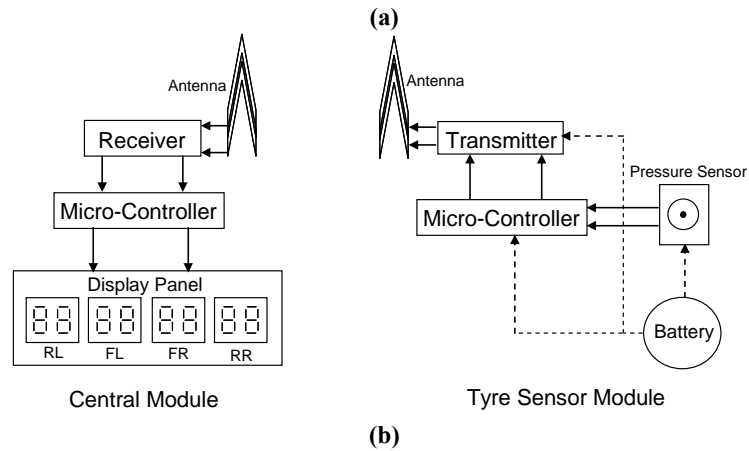


Figure 2.17 The TPMS (a) A TPMS in a 4-wheel passenger car [Siemens 2007] and (b) Diagrammatic structure of the contents of a TPMS system

The usual principle of operation can simply be explained as follows: In the sensor module, the sensor measures the tyre pressure (performing temperature compensation if needed), the transmitter transmits the signal to the central module while the micro-controller governs the relationship and interaction between both. In the central module, the receiver picks up the signal sent from the sensor modules and the microcontroller, which can be part of the overall vehicle computer system, manages the display of this information on the dashboard to the driver.

In addition to power efficiency and resilience to interference as previously discussed, some other issues are of prime concern in TPMS systems. Among these is tyre identification where the central module needs to be able to distinguish between tyres attached to it and tyres belonging to a neighbouring vehicle. It also needs to know the location of each tyre on the vehicle as tyres can be swapped. It also needs to be able to sense when a new tyre is added to the vehicle. Of equal importance are also the size and weight of the tyre module which need to be minimised in order not to affect the fitting and rotation of the tyre. The system should be able to identify if the vehicle is parked, in which case it would stop transmitting to save power, and should be able to resolve a signal congestion in the channel in case the pressure in more than one tyre changes at the same time.

Chapter 3 discusses how RF solutions can help resolve some of these issues and how the implementation of the system based on a network standard and specifically ZigBee can satisfy these conditions.

2.9 Summary

This chapter provided the necessary background information related to the topics discussed in later chapters of this thesis. After discussing the motivation for working with low-power and low-data rate applications in Chapter 1, this chapter first justified why ZigBee was the most suitable standard for this category of applications and hence chosen as the standard to base the work of this thesis on. This was done through comparison with other IEEE-ratified wireless network standards. Following that, and to give an insight into the subject of front-end circuit design, general front-end performance requirements that are commonly of concern to the RF system architect or circuit designer were reviewed and different receiver and transmitter topologies which aim to satisfy these requirements were examined. Since the transistor is the core component in any active circuit design, high frequency transistors realised with different technologies were then considered, justifying why Silicon CMOS presented the most suitable choice for ZigBee's requirements and specifications. Thereafter, some examination of high frequency CMOS transistors was provided by discussing the impact that the development of SOI technology had on the performance of the transistor and how the shrinking size of CMOS gate length changed the perception of CMOS transistors that they are now considered RF devices.

Since they are extensively discussed in this thesis, a section was dedicated to low noise and power amplifiers examining their performance requirements and how the performance of each of these two crucial circuits affects the overall performance of the receiver or the transmitter, respectively. This was followed by a detailed discussion of the nonlinearity problem which is of prime importance in both these circuits. This included a description of the nonlinearity phenomena, its causes and its effect on the performance of communication systems. Methods used for nonlinearity representation and analysis were then reviewed. Also, since it is handled in extensive detail in Chapters 4 and 5, some focus was given to the nonlinearity of the CMOS transistor including the sources of nonlinearity in the transistor and their interactions.

Following that, a review was given to the methods used in high frequency circuit analysis and characterisation; looking at: S-parameters, the theory and importance of matching in high frequency circuits, assessment and solutions of the stability requirement and gain characterisation of two-port networks. These methods will be used extensively in Chapter 5

when the amplifier design is considered. Finally, as the application of choice, the operational requirements of the Tyre Pressure Monitoring System were discussed as a foundation for Chapter 3 which is going to discuss the capability of ZigBee to satisfy these requirements and identify areas where improvements are needed.

Chapter 3 ZIGBEE, ITS APPLICATION FOR TYRE PRESSURE MONITORING SYSTEMS (TPMS) AND ITS QUADRATURE PHASE SHIFT KEYING (QPSK) MODULATION

3.1 Introduction

As an emerging standard, there are many applications that could benefit from ZigBee. These range from home automation and personal communication to industrial sensing. An application proposed here is to use ZigBee to implement a Tyre Pressure Monitoring System (TPMS). This chapter assess the capabilities of ZigBee by discussing the suitability of relevant aspects of its protocol and compliant hardware to this application. The focus is particularly on how effective the ultra low power feature of ZigBee can be in reducing the power demand of a TPMS system while maintaining or improving the level of performance. First, some of the current existing approaches to building a TPMS are studied and their problems identified, highlighting the potential for this suggested application. To further reinforce the importance of the application, this is followed by an insight of how a network standard-based TPMS, in general, would perform better than a point-to-point-based one. The focus then switches to examining ZigBee by comparing its offerings with the requirements of the TPMS application. Based on the conclusions of this analysis, the required areas of improvement are highlighted with an exploration of both the ZigBee protocol and compliant hardware.

This leads to the development of a symbolic visual method to deal with the ZigBee QPSK modulation scheme. An implementation of which demonstrates the usefulness of the visual realisation.

3.2 TPMS Design Techniques

This section is dedicated to the physical structure of the TPMS. A discussion of the current approaches for realising TPMS is first provided. Taking into account TPMS requirements

discussed in Section 2.8, limitations of these approaches are identified. Following that, a proposal to base TPMS networks on a ratified network standards is given and justified.

3.2.1 Current Approaches of TPMS

A general diagrammatic structure of a typical TPMS was presented in Figure 2.17 of Section 2.8 where the operational requirements of TPMS were also discussed. A schematic implementation of the basic TPMS is shown in Figure 3.1. This is based on point-to-point communication between transmitters at the sensor modules and the central receiver module.

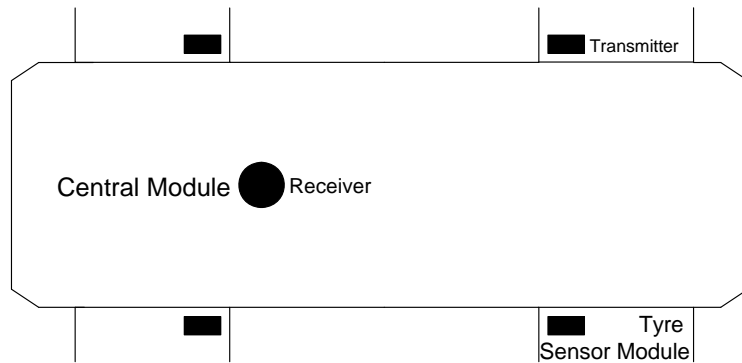


Figure 3.1 Basic setup of a TPMS

While this setup is capable of transmitting the pressure signal to the driver, the degree at which it satisfies most of the other essential TPMS requirements discussed in Section 2.8 is limited.

Clearly, there is no point in sending the pressure signal from the sensor module in the tyre to the central module if there is no driver to read it. A method must be derived to inform the sensor module in the tyre that the vehicle is not being used and therefore there is no need to transmit any information. Sending the pressure signal needlessly, unnecessarily consumes power, and this eventually means shorter life for the battery of the stand-alone sensor module. Since the module is fitted in the tyre, its battery cannot be changed without disassembling the whole tyre. The idea of constantly sending tyre pressure is, therefore, not optimal for long battery life. While many solutions to this problem have been proposed, the most popular adopted by many manufacturers is that shown in Figure 3.2. In the approach of Figure 3.2.a, a low frequency receiver is integrated in the sensor module with a low frequency transmitter fitted on the vehicle's body near the tyre, as shown. When the vehicle is parked, the low frequency transmitter sends a signal to the low frequency receiver on the sensor module instructing it to turn to its sleep mode, i.e. stop measuring the pressure and transmitting it. As long as the vehicle is parked, only the low frequency receiver will be on standby waiting for the signal indicating that the vehicle has started again (the wake-up signal). Only then, it instructs the sensor and the RF transmitter to resume normal operation. The problem with this approach is that the RF

transmitter fitted in the vehicle body near the tyre is typically hard-wired to the central module, this leads to increasing complexity and an expensive system, potentially limiting the benefit of TPMS to high-end vehicles, and making it increasingly difficult to retrofit a TPMS in a vehicle which did not have it when originally manufactured.

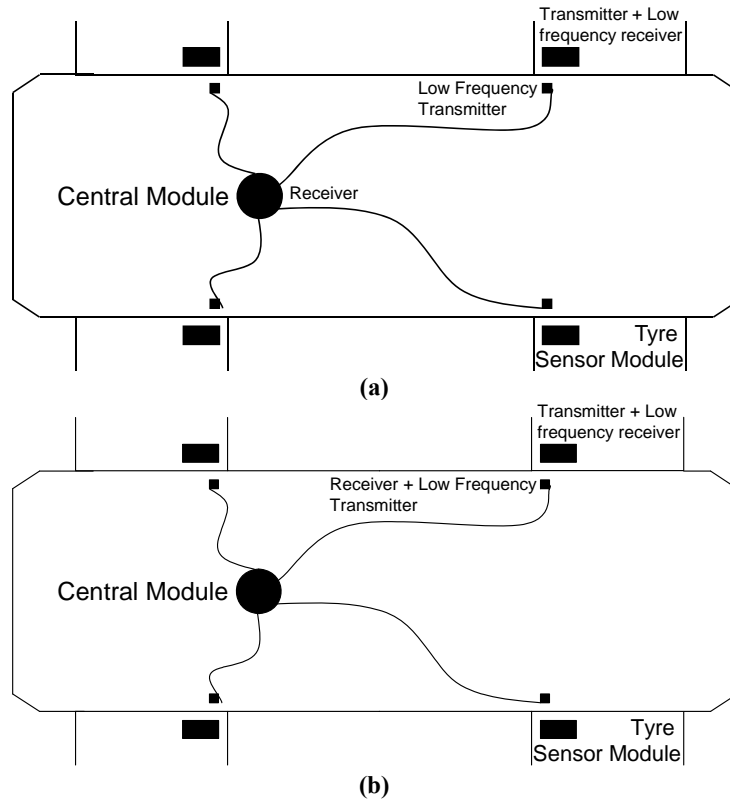


Figure 3.2 Wired TPMS setup (a) Signal receiver at central module (b) Local signal receiver.

But while this solution is adequate for saving power when the vehicle is not in use, it does not serve the issue of tyre identification. Tyre identification can be done in two ways; either manual or automatic [Mnif 2001]. Both involve giving the sensor module an ID to identify it to the central module. In the manual approach those IDs are programmed into the central module, and in case tyres are changed or rotated, the central module has to be reprogrammed. The obvious disadvantage of this approach is that this can only be done at authorised service centres and needs extra equipment and hence leads to inconvenience. In the automatic approach, the pressure signal is transmitted to a dedicated local receiver located near the tyre rather than going to the central module directly, as shown in Figure 3.2.b. In such case, every tyre deals only with the transceiver adjacent to it, providing a solution for the tyre identification issue. However, the wiring problem still exists as the local receiver is hardwired to the central module. Note that with these two additional arrangements, the communication is still based on a point-to-point configuration between the central receiver and each remote module individually.

Two more solutions have been proposed to the tyre identification issue [Mnif 2001], but they are much less popular than the one discussed above. The first is based on integrating an accelerometer with the tyre sensor module and using a sophisticated algorithm taking into account inner and outer tyre speed at turns. The disadvantage of this is that the learning procedure is long because it is based on statistical processing, and it is unreliable since it is active in road turns only. On the other hand the accelerometer can at the same time sense the speed of the tyres and switch the sensor module to sleep mode when the vehicle is parked in order to save power. But the main drawback of using an accelerometer is that it can result in a large and possibly heavy module that could violate the module size and weight restrictions as it can affect the rotation of the tyre. The second solution is based on the Amplitude Modulation (AM) generated from tyre rotation. The idea is based on the different reception levels the central module gets from each sensor module depending on the rotation of the tyre. An algorithm similar to that used with the accelerometer option then calculates the position of each tyre. This solution seems very cost effective, yet its feasibility has not been demonstrated, and it suffers from the same disadvantages as does the accelerometer-based solution.

Frequency hopping was suggested as a solution for the TPMS interference problem discussed in Section 2.8. Having the transmitter and the receiver communicating on pre-specified random frequencies can largely decrease the possibility that the system of another vehicle in range is operating on the same frequency at the same time. To enhance the benefits of this technique, more frequencies are needed, the main drawback of this solution is that a Phase Locked Loop (PLL) is required at the transmitter in order to produce all these frequencies and this leads to higher power consumption due to the extra circuitry.

To obviate the power consumption issue, several battery-less technologies have been demonstrated [Leman 2004]. Battery-less systems offer important environmental and performance benefits. However, such systems are very expensive and therefore not directly cost competitive compared to the moderate-cost battery-based systems. In addition, while a clear consolidation is ongoing among the battery-based direct TPMS OEM system providers, the technology of choice for battery-less systems is not yet clear.

As can be seen, none of the above solutions has presented the right balance between cost effectiveness and reliability. While some are very reliable, they are based on expensive technology either due to manufacturing or installation. Others offered affordable solutions but at the cost of reliability and integrity of the system.

Therefore, it is proposed that a completely wireless (i.e. entirely wire-free) solution for a TPMS will resolve most or all of the issues discussed above at low cost leading to TPMS being affordable for both low- and high-end vehicles and making it retrofittable. What is required is a completely portable network standard-based TPMS. The next section will discuss why a network standard-based TPMS can perform better than one based on point-to-point communication.

3.2.2 Network Standard-Based TPMS

To date, TPMSs have always been based on a point-to-point communication configuration which implies there are no means of constant synchronisation between the sensor module and the central module. This prompts the use of a network standard in which both the transceiver in the sensor module and the central module adhere to the same protocol of communication rules with all the benefits this brings. In other words, a portable sensor network is deployed in the vehicle, as shown in Figure 3.3, which can be easily installed in any vehicle.

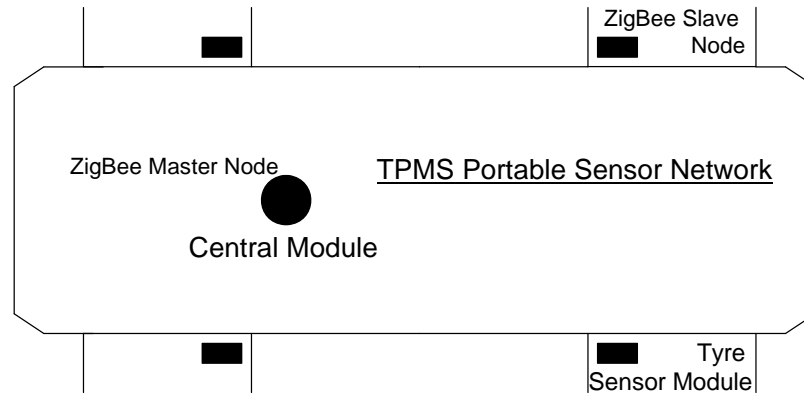


Figure 3.3 a diagrammatic structure of a TPMS as a portable sensor network using a network protocol

Initially it might be thought that employing a network standard for a TPMS does not help with the very important issue of power consumption given that power will be consumed to address the overhead needs of a network protocol. However, with careful analysis, it will be shown that this is not the case.

For instance, in a TPMS realised as a wireless sensor network, when the vehicle is stationary, the central module (master) would inform all the sensor modules (slaves) attached to it to switch to the standby mode, and this happens through the normal routines identified in the wireless network protocol with no need for extra wiring or low frequency wakeup circuits.

With a network standard-based TPMS the issue of data collision if two or more tyres tried to send their signals simultaneously is overcome as this is catered for in the protocol.

Network protocols usually employ security mechanisms where no external network can get access to information being exchanged between members of a given network. In a TPMS, this solves the problem of one central module in a neighbouring vehicle reading the sensor module signal of the other. In addition, network nodes (slaves) connected to one master communicate with each other, which could help with the issue of tyre identification after changing.

In addition, if the standard is based on an ad-hoc network with auto configuration, masters can search for slaves and automatically connect to them and add them to their network. In a TPMS this means the process of changing a tyre becomes much easier, and does not need a specialist to do it, since there is no need for reprogramming the central module as is the case with the current point-to-point based systems. With a slight adjustment, the master can connect only to a slave that is not connected to any other master. Hence a new tyre being placed on a vehicle would be recognised by only the central module of this vehicle no matter how many other vehicles are in the vicinity because it would be the only sensor module not connected to any other central module.

The sequence of operation of a TPMS realised using a wireless network standard would be as follows: When the vehicle is moving, most of the time the receiver section of the tyre sensor module would be on standby waiting for instructions from the central module (like for example that the vehicle has stopped or a pressure on demand instruction). The transmitter portion of the tyre sensor module would be OFF. The pressure sensor would be switching between ON and OFF modes on a pre-specified time intervals, controlled by the on-board microcontroller. When a tyre pressure measurement is taken, it would be compared with the previous one stored on the microcontroller, and if it is the same the transmitter would remain OFF and no signal will be transmitted. If it is different the transmitter will go to “ON mode” to transmit the signal. If a ‘pressure on demand’ instruction was received from the central controller both the sensor and the transmitter would switch ON, measure and transmit the pressure signal irrelevant of whether it was different or the same as the previous one. When the vehicle becomes stationary and the engine is off, the central module informs the tyre sensor modules of the change, the pressure sensor, microcontroller and transmitter go to “OFF mode” while the receiver remains on stand by. When the engine is turned ON again, the sensor goes to its time controlled “ON-OFF mode”, the transmitter remains OFF and the receiver goes to standby and the cycle continues. This approach is designed to run the system at levels of highest reliability and lowest power consumption.

Therefore, a network standard based TPMS can provide an entirely portable wireless solution to tyre pressure monitoring requirements, being reliable, low power consumption and with low cost and ease of installation.

3.3 ZigBee for TPMS

In this section, a compatibility analysis between ZigBee as the selected low power network standard and the test application, TPMS is conducted. This is done by first highlighting the features in the ZigBee standard which can help satisfy the requirements of TPMS discussed in Section 2.8. Following that, two weaknesses which may potentially limit the use of ZigBee in this application are introduced. It is important to note that the analysis conducted in this section focuses only on the operational features of ZigBee and TPMS, assuming the existence of a reliable channel, and disregarding the external effects (resulting from rotation, heating, breaking....etc.) which can influence the robustness of that channel.

3.3.1 ZigBee Offerings vs. TPMS Requirements

Of the three ZigBee frequency allocations, the 2.4GHz band is the most suitable for the TPMS application; for three reasons. First, the 2.4GHz band is the only ISM band that is unlicensed worldwide. The 868MHz band is unlicensed only in Europe and the 915MHz unlicensed band is used only in North America, New Zealand and Australia. Second, in the 2.4GHz band, ZigBee uses 16 channels, while it uses 10 in the 915MHz band and only 1 channel in the 868 band. Using multi frequencies helps to increase security, immunity to interference and signal mixing from other vehicles that might be operating on the same frequency within close vicinity.

The power efficiency issue was highlighted earlier as one of a prime concern in the TPMS application. Generally, some of the operational techniques that help achieve low average power consumption are reduction of the amount of data transmitted, reduction of the transceiver duty cycle and implementing strict power management mechanisms such as power-down and sleep modes.

The IEEE 802.15.4/Zigbee [IEEE-SA 2003a] technology is specified with a wide range of low-power features at physical and higher levels which can help achieve the low power consumption requirement of the TPMS application, they can be summarised as follows:

- The general operational power-saving feature of ZigBee is the low duty cycle operation. Typical applications for IEEE 802.15.4 devices are expected to run with low duty cycles (under 1%) [IEEE-SA 2003a]. It is worth noting, however, that if the warm-up time which wakes the device up from its sleep mode is high, significant power may be

unnecessarily consumed. Warm-up time can be dominated by the settling of transients in the signal path; especially the integrated active channel filters.

- Wideband techniques, such as the Direct Sequence Spread Spectrum (DSSS) employed in ZigBee, have an advantage that their wide channel filters have inherently short settling times. With their greater channel spacing, DSSS frequency synthesisers may also employ higher frequency references, reducing lock time.
- The offset-QPSK with half-sine pulse shaping modulation selected for the 2.4GHz PHY is equivalent to Minimum Shift Keying (MSK), which is a constant envelope modulation scheme. This allows the use of simple, low-cost components and constitutes low demand on the linear performance of the power amplifier, reducing the need for potential high power consumption.
- ZigBee does not specify duplex operation, reducing peak current.
- Specifications for receiver blocking (defined as the receiver ability to receive a wanted signal at its assigned channel frequency in the presence of an unwanted interferer on frequencies other than those of the adjacent channels) is reduced in ZigBee allowing lower active power consumption of the receiver front-end.
- Output power specifications are reduced in ZigBee. Compatible devices must be capable of transmitting -3dBm (as a lowest maximum), but can operate at lower power (as low as -32dBm) if acceptable by the application and the environment.

Some additional features of ZigBee that can be beneficial to the requirements of the TPMS application include:

- With its Direct Sequence Spread Spectrum (DSSS) and 16-ary quasi-orthogonal 32 chips modulation technique [Gutierrez 2003], ZigBee has a high standard of bandwidth efficiency and increased security.
- ZigBee transmission band can be up to 10 metres, and its addressing scheme can support up to 255 active nodes per network coordinator. This means a TPMS realised using ZigBee can be used in a variety of vehicles ranging from small personal cars into large trucks.
- With ZigBee, each network device employs a Carrier Sense Multiple Access-Collision Avoidance (CSMA-CA) protocol [Gutierrez 2003] to avoid wasteful collisions when multiple simultaneous transmissions occur. This exactly fits the requirement to avoid collisions if one or more sensor modules attempted to transmit to the central module simultaneously.

It can be concluded therefore that by examining ZigBee's capabilities, it has the potential to realise an efficient TPMS. However, as will be shown in the following sections, more detailed analysis of ZigBee's protocol and hardware shows that improvements are needed before such application can become a physical reality. For instance, it is evident that some of the above listed power reduction techniques employed in ZigBee can lead to severe effects on the quality of the transmission of signals. One of the goals of this thesis is to try to find a way by which to improve the signal quality of ZigBee compliant transceivers but without affecting the power consumption.

3.3.2 The Low IF Block for ZigBee – The Image Problem

Many standards and specifications for unlicensed wireless communication devices have chosen the 2.4GHz ISM band for operation, examples of these, in addition to ZigBee, are Bluetooth and IEEE 802.11 WLAN. This is due to this band's unique combination of near worldwide availability, suitable bandwidth, and technical feasibility for low cost designs [Gutierrez 2001]. However, this popular choice leads to a design challenge based on the possibility that interference between these incompatible standards may be severe and may lead to corruption of the received data. In Chapter 2 it was concluded that the Low-IF block features the best trade-off as a receiver architecture considering ZigBee requirements. But one of the most serious disadvantages of the Low-IF architecture is the image problem. The image problem becomes more severe, the more co-existing devices there are operating in the same band in close proximity.

For simplicity and low power consumption, ZigBee does not employ a Quality of Service (QoS) scheme in its protocol, that is, the protocol does not guarantee error-free reception of the signal after a link has been established. With this in mind, the impact of the image frequency, which can severely distort the received signal, becomes more serious. This is clearly a drawback since the reception of a correct pressure signal, for instance, in the TPMS application, is crucial. One way to overcome this is by repeatedly sending the signal and calculating the average at the central module, but this only adds to the power consumption. A more general solution is to try to ensure a robust and high performance ZigBee RF front-end, which should lessen the need for QoS. Apart from the image problem, other aspects need to be considered such as; adjacent channel interference, sensitivity, linearity and noise level. But image rejection is of particular importance due to the crowded 2.4GHz ISM band as discussed earlier. The image problem can be solved by enhancing the capabilities of polyphase filters used for image frequency rejection at low IF receivers.

3.3.3 Power Consumption of Currently Available Solutions

Several ZigBee compliant chips are currently available in the market from several manufacturers as shown in Table 3.1. These are the companies who make the actual ZigBee silicon chip not the module. I/O modules and development kits are made by many other providers, but discussion of these is beyond the scope of this work.

The assertion that all these chips are compliant to the requirements of the standard only means that they all satisfy the minimum performance requirements; but that does not specify a limit on the DC power consumption. ZigBee is defined as an ultra low-power standard, but it does not specify explicitly a value for low power. Table 3.1 provides a comparison between the current consumption and supply voltage of various ZigBee transceivers in the 2.4GHz band, using data obtained from their respective data sheets. As can be seen, power consumption varies significantly.

Table 3.1 Power consumption of available ZigBee Chips

	Supply Voltage	RX-mode Current (mA)	TX-Mode Current (mA)	Power Output (dBm)	RX-Power Consumption (mW)	TX-Power Consumption (mW)
Freescall (MC13191)	2.0 to 3.4 (2.7 Typ)	30	37	-27 to +4	60 to 102	74 to 125.5
Atmel (AT86RF230)	1.8 to 3.6	16	17 (Max TX 3dBm)	-17 to +3	28.8 to 57.6	30.6 to 61.2
Ember (EM 250)	1.7 to 1.9	28	28 (Max TX 3dBm)	-32 to +3	47.6 to 53.2	47.6 to 53.2
TI/Chipcon (CC2420)	2.1 to 3.6	19.7	17.4	-25 to 0	41.4 to 70.9	36.5 to 62.6

Calculations using this data show that for the current ratings of Table 3.1, it is not possible to build a TPMS system because a AAA battery would last for 2 years, assuming moderate use. Bearing in mind that some TPMS regulations require the in-tyre module battery to last 10 years and that a AAA battery is probably too heavy to be placed in a tyre, the need for reduced power consumption is clear. In Chapter 5, an amplifier is designed to reduce the power consumption in the front-end transceiver by combining the low noise and power amplifier.

3.4 QPSK Modulation and de-Modulation

In this era of vast communication expansion it is vital that modulation techniques are improved to enhance a communication system performance in terms of bandwidth efficiency, signal integrity, power consumption and transmission security.

For this purpose, many variations of the three basic digital modulation schemes (Amplitude Shift Keying (ASK), Phase Shift Keying (PSK) and Frequency shift keying (FSK)) have been developed. One of these variations is Quadrature Phase Shift Keying, QPSK.

QPSK is a form of Phase Shift Keying in which the carrier is modulated by two bits at once, selecting one of four possible carrier phase shifts. QPSK allows the transmitted signal to carry twice as much information as PSK when using the same bandwidth. QPSK is used in many applications including satellite transmission, cable modems, videoconferencing and cellular phone systems. A QPSK variant, Offset QPSK, O-QPSK, is used in ZigBee at 2.4GHz. The mechanisms of QPSK and O-QPSK are discussed below.

The work reported in this section serves two purposes. First, a comprehensive approach is proposed to aid the understanding of QPSK [Abuelma'atti 2007]. Based entirely on the analysis of phasors with reduced mathematical complexity, this focuses on how a QPSK modulator generates I and Q vectors to encode the data bits on the four possible carrier phase shifts. The second part of this work proposes a method by which modulation schemes like QPSK can be implemented using only simple first degree components (i.e. components that are not made up of other components) in Matlab Simulink, avoiding all the hurdles of mathematically-complicated blocks. Results obtained from implementations confirm the viability of the approach by showing that data received from a noiseless channel exactly matches the data sent despite a slight, unavoidable delay.

3.4.1 QPSK and O-QPSK Mechanisms

The term "quadrature" in QPSK implies that there are four possible phases in the QPSK modulation. In Binary Phase Shift Keying (BPSK) the carrier signal is modulated by changing its phase between two phases which are 180 degrees apart while in QPSK the carrier undergoes four changes in phase (four symbols) and can thus represent 2 binary bits of data per symbol [Young 1999]. QPSK is a form of phase shift keying in which two bits modulate the carrier at once, shifting its phase by one of possible four amounts (45, 135, 225, or 315). Because two binary bits are represented by one phase shift, a QPSK transmitter effectively doubles the amount of data that can be transmitted using the same bandwidth in a given time period.

The operation of QPSK can be described in a simple and effective manner as follows: The binary bits of the data to be transmitted are split into two sets, I and Q, where "I" is the "in-phase" component of the data waveform, and "Q" represents the "quadrature" component. These two, now independent data sets, are carried through two different paths, hereafter called the "I line" and the "Q line" and the data on each line called the "I data" and the "Q data",

respectively. For instance, even-numbered bits are assigned to the I line and odd-numbered bits are assigned to the Q line. The carrier signal is modulated by the data bits on the I and Q lines separately as two separate BPSK modulations. The I and Q data modulate a cosine wave and a sine wave version of the carrier frequency, respectively. This takes place simultaneously for every pair of bits on both lines. The modulated I and Q carriers are then combined producing one of the 4 phase shifts for each pair of bits. On the receiver side, the signal is demodulated by multiplying it twice, once with a cosine wave and once with a sine wave version of the carrier frequency. After some low-pass filtering and base band processing, the multiplication of the received signal by the cosine wave version of the carrier reproduces the I data which was modulated by the same cosine wave version of the carrier, and similarly, the multiplication by the sine wave reproduces the Q data which modulated the sine wave version of the carrier. In a reverse fashion to the way they were split, the received I and Q data are then combined to one line, in the correct order, to reproduce the original data. Figure 3.4 shows how a QPSK modulator and demodulator are set up.

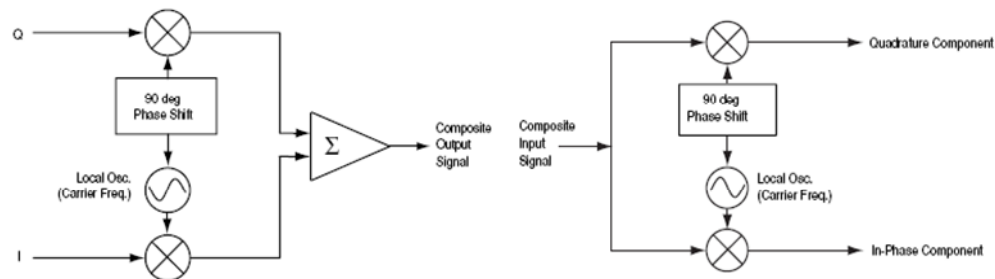


Figure 3.4 Block Diagrams of QPSK Modulator and Demodulator

The difference between O-QPSK and QPSK is that in QPSK, the phase shift representing every pair of bits can jump between the 4 specified phase shifts by as much as 180 degrees at a time, depending on the data on the I and Q lines. This can result in large amplitude fluctuations which is undesirable. In O-QPSK, this is avoided by delaying either the I or Q data lines by 1 bit period (or half a symbol). This way, the data bits on the I and Q lines can never change at the same time, preventing the 180 degree jump and limiting it to 90 degrees only.

3.4.2 Carrier Phase Shift Choice for Modulating Binary Bits

In QPSK, 4 possible phase shifts are used to represent four possible pairs of transmitted bits which are 00, 01, 10 or 11. The choice of which phase shift is to represent which pair of transmitted bits depends on the individual BPSK operation that independently take place on the I and Q lines. The deciding factor is whether the cosine or sine wave version of the carrier is modulated by reversing its own phase for a modulating binary bit of a 0 or a 1 on its respective line. Since there are two possibilities and two lines, each pair of bits can be represented by any

of the 4 phase shifts. Figure 3.5 gives a comprehensive explanation of how this works for all possible cases (16 cases made from 4 possible pairs of bits and four possible phase shifts).

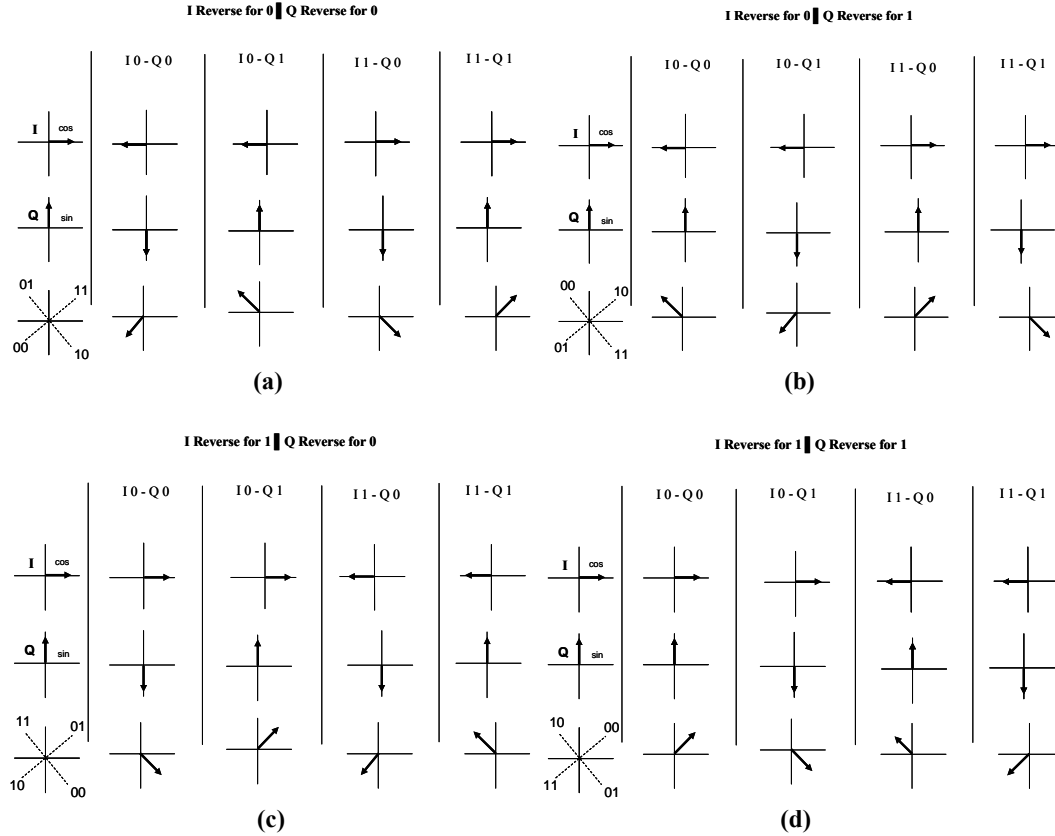


Figure 3.5 All possible carrier phase shifts subject to transmitted bit pairs when the system's allocation rule is (a) I reverse for 0 Q reverse for 0, (b) I reverse for 0 Q reverse for 1, (c) I reverse for 1 Q reverse for 0 and (d) I reverse for 1, Q reverse for 1

Taking for instance the case in Figure 3.5.b where the bit pair is [01] and it is an “I reverse for 0-Q reverse for 1” system. The cosine wave version of the carrier which would be modulated by the I data is represented by a phasor on the positive side of the X axis and the sine wave carrier version that would be modulated by the Q data is a phasor on the positive side of the Y axis. If the I data was to modulate its carrier by an I reverse for 0 system, and the I bit here is 0, then the resulting carrier phase is a 180 degrees phase-shifted cosine wave which is represented by a phasor on the negative side of the X axis, as shown. If the Q data was to modulate its carrier by a Q reverse for 1 system and the Q bit here is 1 then the resulting carrier phase is a 180 degrees phase shifted sine wave which is represented by a phasor on the negative side of the Y axis as shown. Combining these two phasors, the result is a phasor at 225 degrees. Hence, in an “I Reverse for 0-Q Reverse for 1” system, the bit pair [01] is represented by a 225 phase shift of the carrier. Performing the same analysis on the same bit pair but in an “I reverse for 1-Q reverse for 1” system (Figure 3.5.d), it can be seen that the I data does not modulate its carrier

but the Q data does, and when the two phasors are combined, a carrier phase shift of 315 degrees is obtained to represent this pair of bits to be transmitted. Similar analysis can be done to the same bit pair in the cases in Figure 3.5.a and Figure 3.5.c to see that each bit pair can be represented by any one of 4 possible phase shifts and that is decided by whether the reverse (or 180 phase shift) of the modulated carriers is passed for a 1 or a 0 bit of the modulating signals on the I and Q lines. All remaining 12 cases for the remaining 3 possible bit pairs can be analysed in the same way from Figure 3.5.

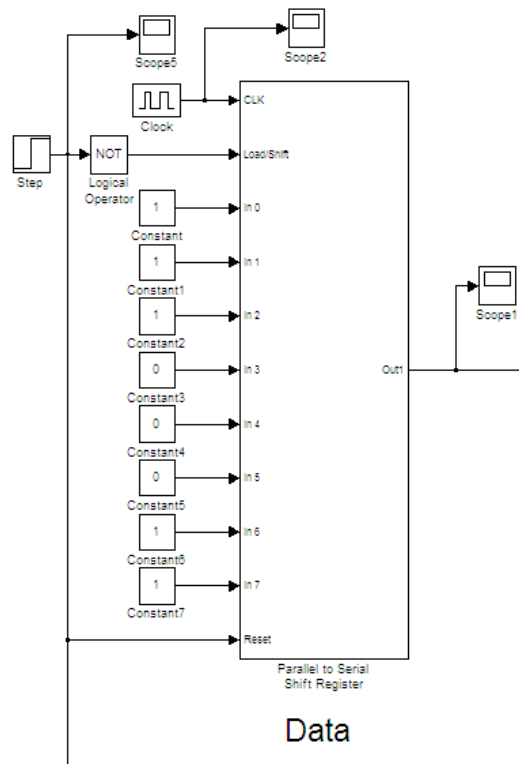
3.4.3 QPSK Implementation

It is often thought that modulation schemes like QPSK can only be modelled mathematically or implemented using mathematical expression-based blocks either in hardware or software [Maxim 2000]. For instance, Matlab's block level simulation environment, Simulink, provides a block in its tool box library which can be directly used to modulate a carrier by a data signal using QPSK. Such a block would normally be built on lengthy programming code based on mathematical analysis of how QPSK works. While a component like this is able to efficiently modulate a carrier signal using QPSK, the user has no view of what is actually happening in that block and hence repairing any fault in a system under development can be unworkable. Furthermore, the cost associated with obtaining such components either in software through licensing or hardware through off-the-shelf components (see for example [Maxim 1996]) could be considerable. The method presented here shows how simple blocks in Matlab Simulink, such as switches, flip-flops, multiplexers, adders and multipliers can be used to model a QPSK modulator and demodulator which effectively does the same job as the mathematically-complicated block. Also, the user is able to plot the signal at every point in the system on a scope. In addition to the educational purpose this serves in terms of showing the user what happens to the signal at every step of the QPSK modulation process, this implementation can also simply be extended to hardware to build a simple and cost effective QPSK modulator.

The implementation of the QPSK transceiver system is based on the description of phase shift choices for transmitted bit pairs detailed in Section 3.4.2. It demonstrates how sophisticated modulation schemes can be reliably built using simple components, cost effectively, without complication. It includes blocks for generating the data, splitting it onto I and Q lines, performing modulation, transmitting the signal, performing demodulation, some base band processing and finally recombining the received bits in the correct order. Due to size constraints, the whole setup cannot be shown here in a single figure, so the following sections will describe each part of the system individually.

3.4.3.1 The Test Data

In order to obtain some binary data to modulate the carrier using QPSK, 8 “constant” blocks are used to generate the data bits and then a parallel to series shift register is built to generate the serial stream. The value of the “constant” block can be set to any value required by the user; in this case the 8 constant blocks were set to 0s or 1s to generate an 8-bit binary data as shown in Figure 3.6.a. The parallel to series shift register was built using 8 “2-1 multiplexers” and 8 “D-flip-flops” as shown in Figure 3.6.b [Wilkinson 1992]. A step is used to clear all the flip flops before loading and shifting starts. The shift register is very adaptable in that if more or less bits were to be used, the number of 2-1 Multiplexer and D-flip-flops can be changed accordingly.



(a)

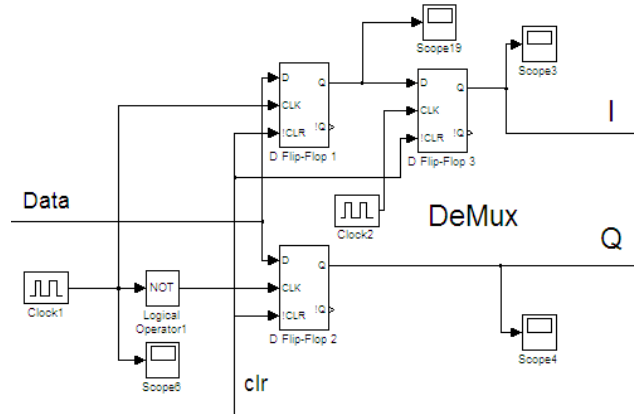


Figure 3.7 DeMultiplexer to split the data stream into in-phase and quadrature components

3.4.3.3 BPSK Modulations, and I/ Q Combination

A switch was used to implement the modulation of the sine and cosine carrier versions by the I and Q data as shown in Figure 3.8. The switch used is a three port switch with 2 input ports and one control port. The switch selects either of the inputs and passes it to the output depending on the value at its control port which is a threshold set by the user against which the two inputs are compared. For the I line, one of the inputs is connected to a cosine wave and the other to its inverse (180 degrees shifted). The threshold of the switch is set to 1, hence if the input data is 1, the cosine wave version of the carrier will be passed to the output, and the inverse of the cosine wave will be passed if the incoming data bit on the I line was 0, thus implementing a BPSK with “I reverse for 0” according to the rules set in Section 3.4.2. Similarly, the modulation of the sine wave version of the carrier by the Q data stream is made. This is also implemented here as a “Q reverse for 0” system. Finally, the signal to be transmitted is prepared by simply combining the modulated carriers in the I and Q lines using an adder. The modulated carriers and the transmitted signal are shown and discussed in Section 3.4.3.6.

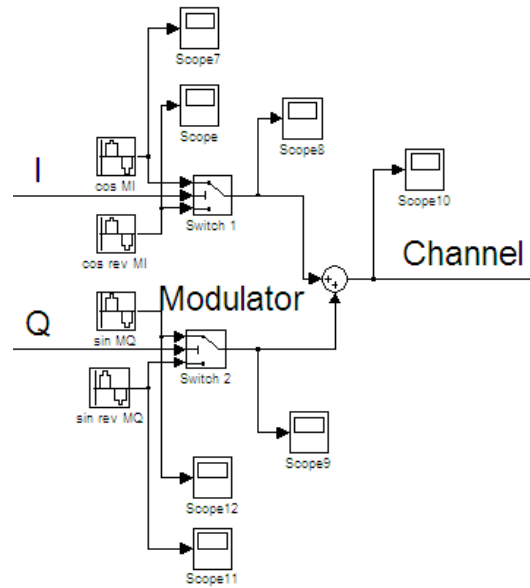


Figure 3.8 Implementation of QPSK Modulator

3.4.3.4 Demodulation

Demodulation in QPSK takes place over two stages. These are restoring the transmitted levels (0s and 1s) from the phases of the transmitted signal and removing the high frequency component. Two multipliers (“product 1” and “product 2”) are used to restore the transmitted levels from the phases of the transmitted signal as shown in Figure 3.9. The QPSK demodulation process is similar to the modulation strategy in that the QPSK modulated carrier is split into I and Q channels, each of which is BPSK demodulated. The received signal is split and multiplied by the original (not the reverse) cosine and sine wave versions of the carrier for the I and Q lines, respectively. The multiplication of the received signal by the carrier version of the I data stream (the cosine wave) generates the modulating I data and like wise for the Q data stream and its sine wave carrier. The produced signals at both lines are then low pass filtered. The filters in “Analog Filter Design 1” and “Analog Filter Design 2” are each setup as an 8th-order Butterworth low pass filter. The resultant I and Q signals after demodulation and filtering are shown and discussed in Section 3.4.3.6.

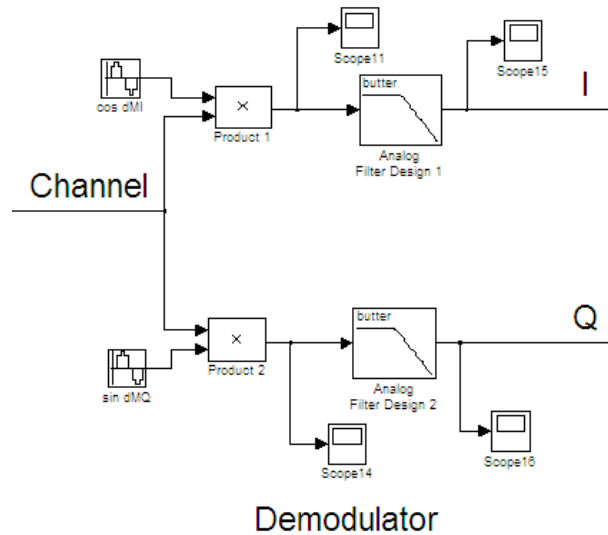


Figure 3.9 Implementation of QPSK Demodulator

3.4.3.5 Base Band Processing and Recombining

The demodulation stage described above restores the transmitted levels to “low” or “high” values as will be seen in Section 3.4.3.6. In order to recover the received signal to the correct 0 and 1 levels, two switches (switches 3 and 4) are used to implement a comparator as shown in Figure 3.10. The threshold of both the switches is set to 0 and the input ports of each switch are connected to two “constant” blocks with one set to the value of 1 and the other set to 0. If the signal on the line is higher than 0, the switch passes 1 to the output, if it is lower, a 0 is passed. In this way, the received signal can be reconstructed to a binary digital signal of 1s and 0s.

To reconstruct the received data as a one data stream like that obtained from the transmitter shift register, the data bits on the I and Q lines are combined. This is done using a switch, switch 5, as shown in Figure 3.10. The received I and Q data are connected to the input terminals of the switch. The control port is connected to a clock, clock 4, which is set to a frequency half that of the shift register clock frequency.

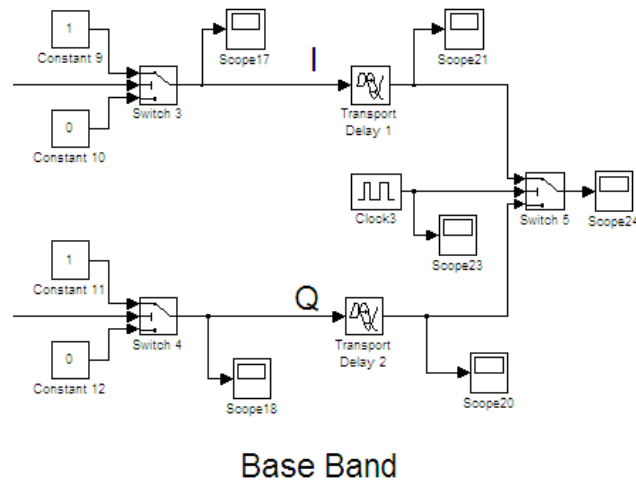


Figure 3.10 Implementation of base band processing

This is because data in the I and Q lines have double the time period of the originally generated data stream. The threshold of the switch is set to 1 so if the clock is high the data bit on the I line is passed to the switch output, and if the clock is low, the data bit on the Q line is passed. This effectively amounts to swinging between taking a data bit from the I and Q lines in an opposite fashion to the de-multiplexer used to split the original data to the I and Q lines. This results in constructing back a single data stream in the correct order. Delay blocks (Transition Delay 1 and 2) are used to synchronise the edges of the data on the I and Q lines with the edges of the clock to avoid conflicts

3.4.3.6 Implementation Results: A Step by Step QPSK

In Figure 3.11 through Figure 3.18, the implementation results are presented in the form of screen shots of scope signals at different points of the system. Figure 3.11 (Scope 1) shows the test data as generated from the shift register. The one time unit delay is due to the delayed starting point of the shift register which depends on the step and the clock as demonstrated in Figure 3.6. The Data sequence is 11100011 with one time unit for each bit.

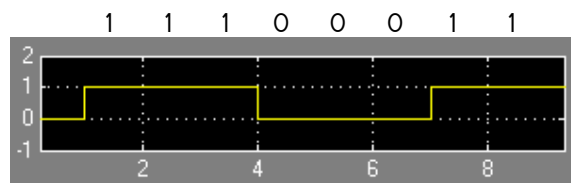


Figure 3.11 The Test data signal

The data in the I and Q lines are shown in Figure 3.12.a (Scope 3) and Figure 3.12.b (Scope 4), respectively. The extra two time unit delays are due to the use of the two flip flops in the case of the I data. In the case of the Q data, one time unit delay is due to the use of flip flop 2, the other

is because the DeMux waits one time unit to pick up a bit for the I data before working on the Q data.

Because the DeMux acted on a bit for each line while keeping the other idle, now each bit in the I and Q data lines has double the time it had on the original data stream. As expected, even numbered bits of the original data stream are now on the I line, reading 1101 as shown in Figure 3.12.a and odd-numbered bits went to the Q line which now reads 1001 as shown in Figure 3.12.b. With this sequence, only 3 phases out of the possible 4 will be transmitted. That is because in this case, only 3 out of the possible 4 bit pairs are present. These are 11, 10 and 00, as can be seen.

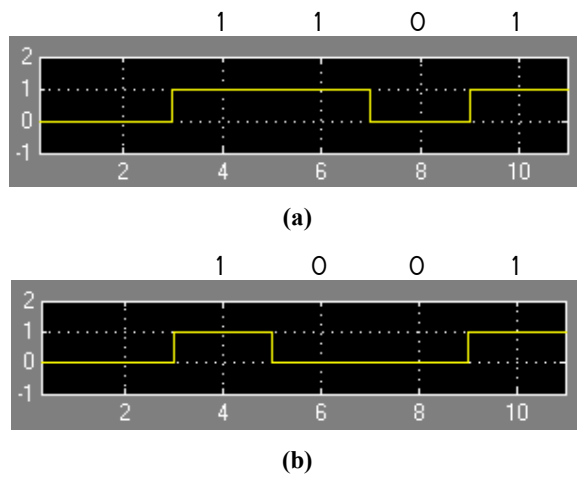


Figure 3.12 The testing data split to the I and Q lines (a) I line data and (b) Q line data

In Figure 3.13.a (Scope 8) and Figure 3.13.b (Scope 9), the carrier modulated by the I and Q data is shown. In Figure 3.13.a a cosine wave version of the carrier is presented for bits with a value of 1 and its 180 degree reverse is presented for the 0 bit. In Figure 3.13.b, a sine wave version of the carrier is presented for bits with value of 1 and its 180 degree phase shift is presented for 0 bits. Hence this implementation represents an “I reverse for 0-Q Reverse for 0” system. If the phase choice rule was different, the signals in Figure 3.13.a and b will have a different shape.

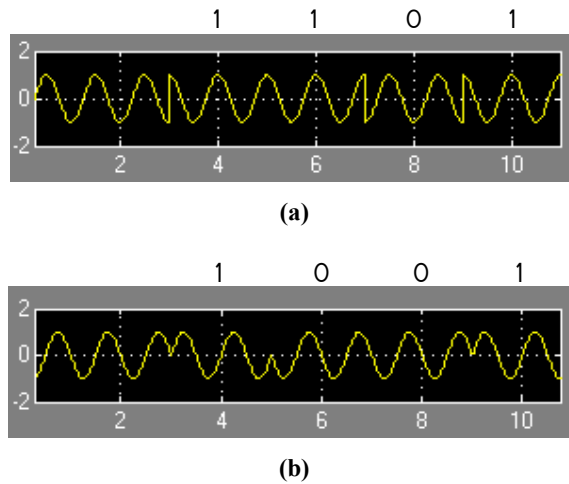


Figure 3.13 Carrier modulated by (a) I data and (b) Q data

In Figure 3.14 (Scope 10), the transmitted signal is shown. This is the signal resulting from combining the signals in Figure 3.13a and Figure 3.13b using an adder. Close inspection of the signal reveals the 3 different phases of the carrier signal being transmitted.

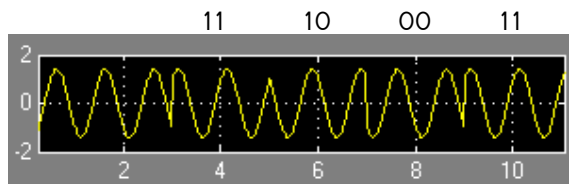


Figure 3.14 The transmitted signal

Figure 3.15 shows how the I and Q data levels are recovered in the demodulator. In Figure 3.15.a (Scope 11), multiplying the transmitted signal by a cosine wave version of the carrier recovers the transmitted levels of the I line in the form of a wave which changes its average value subject to the transmitted 1 and 0 bits. In the same fashion, the transmitted Q data is recovered and is shown in Figure 3.15.b (Scope 14). Inspection of Figure 3.15 and Figure 3.12 confirms that the level recovery is successful. The high frequency content is, however, still present.

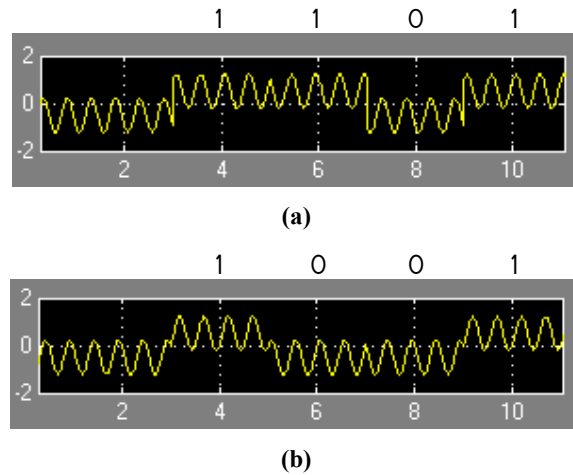


Figure 3.15 Recovered data levels for (a) I line and (b) Q line

Figure 3.16.a (Scope 15) and Figure 3.16.b (Scope 16) show the received I and Q data after the high frequency content is removed by the low-pass filter. With the amplitudes of the upper-centred and the lower-centred waves now not overlapping as in Figure 3.15, the signal can be refined to its original level by a comparator.

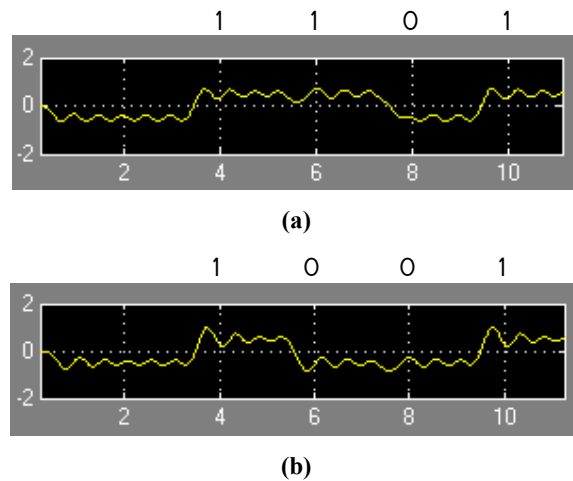


Figure 3.16 The received (a) I data and (b) Q data after being low-pass filtered

Because of the way the signals in Figure 3.16 switch between the upper and lower average value, the comparator does not produce signals with edges synchronised with the edge of the time unit (which is synchronous with the edge of the clock). The job of the time delay units is to adjust this. Figure 3.17.a (Scope 21) and Figure 3.17.b (Scope 20) show the I and Q received data bits clearly recovered to their original 1 and 0 levels by the comparator.

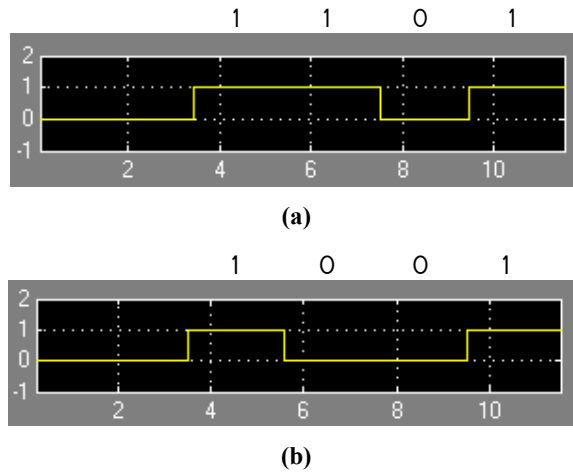


Figure 3.17 The received (a) I data signal and (b) Q data signal after the comparator

Finally, with a further delay of 1 time unit due to the data in the I line waiting for the edge of the clock controlling the combining switch, the received I and Q data are recombined to form one data stream as shown in Figure 3.18.a (Scope 24). Inspection of Figure 3.11 and Figure 3.18 show that despite the 3 time unit delay, the originally obtained data is received correctly.

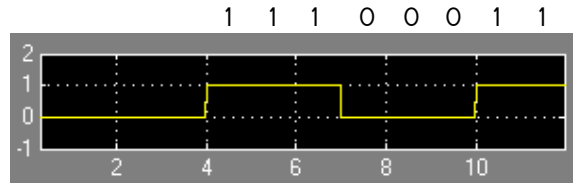


Figure 3.18 The received I and Q data combined in one data stream

3.5 Summary

In the first part of this chapter, a compatibility analysis between the selected low power standard, ZigBee, and the application used to assess its capabilities, TPMS, was conducted. This was done by first examining current approaches to realising tyre pressure monitoring systems, and then drawing on their limitations by identifying the benefits a wireless standard-based implementation can bring to the performance of TPMS. This was followed by discussing specific features of the ZigBee standard which would help satisfy TPMS requirements and also pointing to compatibility weaknesses which can limit those benefits. Two weaknesses were identified as susceptibility to interference by the image frequency due to the use of a low-IF receiver architecture for ZigBee in the TPMS's crowded environment, and power consumption. As an introduction to coming chapters of thesis, these two weaknesses were thoroughly introduced and discussed.

The second part of this chapter included the details of a QPSK implementation based on a new approach to describe the operation of this type of modulation. The approach, based on manipulation of phasors, focuses on how the modulator decides by which phase shift it represents each pair of bits. It was also shown how complicated modulation schemes like QPSK can be implemented using very simple components in a commercial modelling package avoiding mathematically complicated blocks, a method which can also be extended to hardware to realise working circuits.

This work can be utilised in enhancing the security of communication systems, and in communication node addressing. If it was possible that the demodulator can only demodulate a received signal with knowledge of which pair of bits were represented by a given phase shift at the modulator, and if this was programmable (as was demonstrated), then this feature can be used to control which receivers can receive a particular signal.

Chapter 4 QUANTIFICATION OF CONTRIBUTION TO MOSFET DISTORTION THROUGH TRANSISTOR NON-LINEAR MODELLING

4.1 Introduction

Knowledge of the small-signal equivalent circuit of a field effect transistor is very useful for device performance analysis in the design of microwave circuits and also for characterising the device technological process. It can also be used to predict the device behaviour in small and large signal conditions.

There are a number of sources of distortion in silicon CMOS transistors; among these are the transistor transconductance, nonlinear capacitances, and the output conductance. While many previous studies looked at analysing CMOS transistor nonlinearity (see for example [Park 2001], [Kwonl 2005] and [Aparin 2005]), most of these studies investigated the distortion of the transistor as a whole using mathematical representations of the transistor behaviour, but with no emphasis on how individual nonlinear parameters in the transistor might be affecting that behaviour. Accurate quantification of individual sources of distortion to the total nonlinearity of the transistor have not, to the knowledge of the author, been investigated on either CMOS or GaAs transistors in the manner presented here.

The quantification method developed in this work is based on enabling and disabling (eliminating) the nonlinearities of individual nonlinear parameters within the transistor model in a superposition manner and determining the resulting nonlinear behaviour of the transistor each time. Before applying the superposition method, the transistor must be modelled in a way which enables the switching between linear and nonlinear representations of parameters in the model. A number of previous attempts to predict the nonlinear behaviour of transistors resulted in models where nonlinear parameters were combined at high level; see for example [Schmale 1997]. In contrast, the model presented here is unique in that its representation of each nonlinear parameter is completely independent from any other elements.

The work presented here is based on the assumption that the large signal non-linear transistor performance can be determined by interpolation of extracted small signal parameters at several DC bias points.

The following sections will first discuss the extraction of the small signal model parameters including the intrinsic and extrinsic components and will present the dependence of intrinsic equivalent circuit parameters on gate and drain voltages. Being able to observe the dependency of the nonlinear parameters on gate and drain voltages can prove significantly important for the RF front-end circuit designer where linearity is always a major challenge. Following that, three different models of the transistor based on different representations of the nonlinear elements will be presented and their performance in S-parameter and Harmonic balance simulations assessed and discussed. The general structure of these three models is essentially the same; Using the values of the nonlinear parameters obtained from the extraction procedures at several bias points, a Symbolically Defined Device (SDD) is used to represent each nonlinear component, and the whole model is constructed using a table based approach. Formulating a linear representation for nonlinear elements around the chosen bias point is then discussed. Finally, the superposition method to quantify the contribution of the various non-linear elements to overall transistor distortion is applied and the results are discussed. The chapter concludes with a summary of the topics presented. All the work presented in this chapter is based on a 180nm NMOS transistor from an RFCMOS mixed signal design kit sourced from the UMC foundry.

4.2 Extraction

Having that the nonlinear model developed in this work is for a design kit transistor, and given that the design kit transistor is itself a model; this work is then equivalent to reverse engineering the work done by the foundry to model their transistors. As there was no way to establish the nonlinear model on which the foundry model was based, several RF MOSFET models proposed in the literature were compared by obtaining their parameters and assessing them as will be explained in Section 4.2.2.

The small signal equivalent circuit of a FET, an example of which is shown in Figure 4.1 can be divided into, the *extrinsic* and *intrinsic* parts. The *extrinsic* elements are those whose values are independent of the bias condition, these are the elements associated with the transistor terminals such as the gate resistance, R_g , the gate inductance, L_g , the source resistance, R_s , the source inductance, L_s , the drain resistance, R_d , and the drain inductance, L_d . The *intrinsic* elements are those that vary with the voltage levels at the terminals of the transistor. These include the

transconductance, g_m , the output conductance, g_{ds} , the gate to source capacitance, C_{gs} , the gate to drain capacitance, C_{gd} , the drain to source capacitance, C_{ds} , and the drain to junction capacitance C_{jd} .

4.2.1 RF MOSFET Models and Extraction Methods

For the extraction of the nonlinear parameters, several reported models and extraction methods were assessed, among these are the works of [Tong 2004], [Lee 2002], [Jen 1999], [Voinigescu 2004], [Yang 2003], [Lee 2000], [Lovelace 1994], [Lee 1997], [Kwon 2002] and [Dambrine 1988]. The suitability of each model and extraction method to the UMC NMOS technology considered in this work was assessed by establishing the extracted parameter values, as explained in the next section.

There are noticeable differences between the small signal models presented in the above references. Examples of these references are as follows:

- 1- In [Tong 2004], the gate resistance was represented by a single resistor R_g which represents the effective lumped gate resistance consisting of both the electrode resistance and the distributed channel resistance according to [Jin 1998]. However, in [Yang 2003] the gate resistance was modelled by two resistors R_{gs} and R_{gd} representing the effective channel resistance seen by the signal flowing from gate to source via C_{gs} and the gate to drain via C_{gd} , respectively.
- 2- [Tong 2004] and [Lovelace 1994] omit the extrinsic terminal inductances which are taken into account in [Lee 1997]. In [Yang 2003], the gate inductance L_g associated with the deep submicron poly gate is included to better fit the input admittance Y_{11} at high frequencies. It has to be noted though that this may be due to this model being effective up to 110GHz while the model in [Tong 2004] and [Lovelace 1994] run up to only 20 and 10GHz respectively. [Kwon 2002] states that the effect of the parasitic inductances is very small up to 10GHz.
- 3- In [Kwon 2002] and [Tong 2004] an extra transcapacitance, C_m is included between the gate and the source. The transcapacitance is introduced due to the capacitance between the gate and the drain being split between C_{gd} and C_{dg} which are two non-reciprocal capacitances caused by the drain and gate biasing respectively where $C_m = C_{dg} - C_{gd}$. C_{gd} is the capacitive effect on the gate due to the drain while C_{dg} is the capacitive effect on the drain due to the gate. According to [Kwon 2002], the transcapacitance takes care of the different effects of the gate and drain on each other in terms of charging currents

and has to be included in order to model Y_{12} and Y_{21} accurately. [Calvo 1993] also suggested that the transcapacitance ensures charge conservation in the model.

- 4- In [Kwon 2002] and [Yang 2003], R_s and R_d were omitted for simplicity and ease of extraction as the authors claimed these will have a slight effect on R_g and g_m values, while [Tong 2004] and [Lovelace 1994] included these resistances in their models.

Extraction methods presented in these references do also differ but most are based to a great degree on the work presented in [Dambrine 1988]. In this paper, the author explained the general theme of parameter extraction which is based on extracting the non bias dependant extrinsic components using analytical equations obtained from the measured S-parameters of the device and then subtracting these extrinsic components from the measured S-parameters to obtain values for the intrinsic components. The method presented in [Lee 1997] and [Lee 2000] was complicated as it involved significant amount of mathematical manipulations and curve fitting. The extraction of the extrinsic resistances in [Lovelace 1994] was done by applying zero bias to the gate and drain, thereby eliminating the contribution of the voltage dependant current generator g_m and maximising R_{ds} , hence reducing the equivalent circuit to $R_g, R_s, R_d, C_{gs}, C_{gd}$ and C_{ds} . Analytical equations based on measured Y-parameters were then used to obtain values for the terminal resistances. The method presented in [Tong 2004] was very similar to that except that the transistor was forward biased at the gate and zero biased at the drain. This reduced the equivalent circuit to the same as in [Lovelace 1994] but with the exception of C_{gd} which was neglected. The intrinsic extraction method in [Lovelace 1994] was also adopted in [Lee 1997].

4.2.2 Chosen Model and Extraction Method

The compatibility assessment of an available parameter extraction method (and the model on which it was based) with the used UMC 180nm NMOS transistor was centred on whether the extracted values (using that model and method) for extrinsic resistances and intrinsic capacitances were constant with respect to frequency, as suggested by most of the extraction methods, and whether the S-parameters of the constructed linear model (built at one bias point using lumped components) matched those of the modelled transistor. Examination of the available models and extraction methods was performed by repeating this assessment process for every reported method and its associated model. The method found to be most suitable for the transistor investigated in this work was the one presented in [Tong 2004], which was based on a model for a similar 0.18 μ m RF SiMOSFET process. However, a slight modification was introduced to this method in relation to de-embedding the extrinsic resistors prior to extraction of the intrinsic parameters, and to the model in relation to the effect of C_m which was ignored as

tests carried out on the transistor showed that it has little impact on the results of both S-parameter and harmonic balance simulations.

The viability of the choice of this method and the introduced modification are illustrated in Section 4.2.3 where the extraction of the extrinsic and intrinsic parameters is detailed and in Section 4.3 where the resulting models are verified.

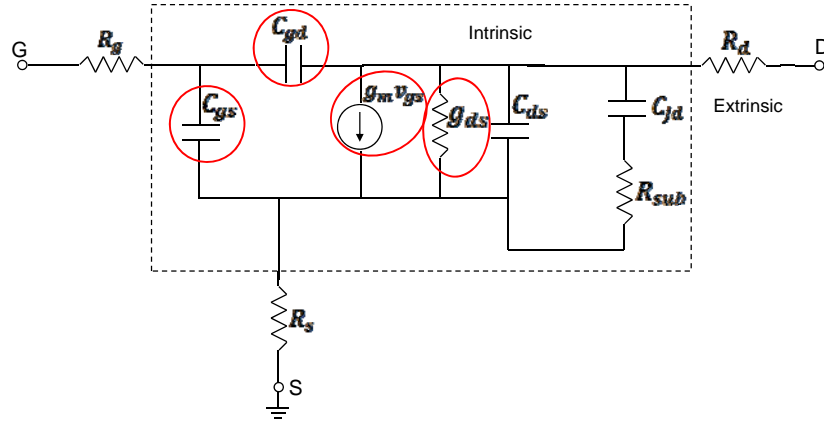


Figure 4.1 The RF MOSFET model proposed by [Tong 2004]

Figure 4.1 shows the final transistor small signal model used for this work. The red-circled elements are the nonlinear intrinsic parameters the nonlinearities of which are to be assessed.

The impact of some of the model elements in fitting the S-parameters can be summarised as follows

- The output capacitance C_{ds} resulting from capacitive coupling between fingered structures is important for the output admittance Y_{22} in the microwave regime as the transistor dimension shrinks [Yang 2003].
- C_{gd} dominates the reverse transmission, S_{12} .
- The substrate coupling effects through the drain junction and substrate resistance (C_{jd} and R_{sub}) become more significant for the fitting of the output admittance Y_{22} as the operation frequency increases [Kwon 2002]. This is because the impedance of the junction capacitance reduces with the increase in frequency.
- R_g significantly affects the input admittance Y_{11} at high frequency [Kwon 2002].
- C_{gs} represents the phase of S_{11} and its magnitude at high frequency.
- The forward transmission, S_{21} is most dominated by the transconductance and output conductance.

4.2.3 Automated Extraction Procedures

Doing this modelling for nonlinearity quantification work on a design kit transistor (model) in a simulator rather than on a fabricated transistor enables more robust verification of the model; that is by undertaking identical tests on the transistor model and the modelled transistor using the same simulation environment. Another major advantage of this is that it also makes possible easier determination of transistor parameters at many more bias points and in less time, especially when automation is used.

There are two contradictory goals of constructing transistor table models: accuracy and size of tables [Nadezhin 2003]. In order to make the nonlinear table-based model more accurate, values for parameters in the table must be obtained over smaller steps of gate and drain voltages, hence more points. More points mean larger tables and consequently, these are more time consuming in recording and manipulating entries especially when the extraction of some of the parameters depends on the extracted values of other parameters. The extended detail of the data in the table becomes more significant in harmonic balance tests where input power is involved [Van den Bosch 1999]. That is because the value of the model element selected from the table depends on the interpolation (normally performed using spline functions) between the entries of the table during simulation. The error in the 3rd-order intermodulation products, for instance, is introduced because spline functions can present discontinuous 3rd-order derivative near table entries, and 3rd-order intermodulation products are related to 3rd-order derivative [Van den Bosch 1999]. Therefore, less interpolation-introduced error is always desired and one of the ways to do that is making the tables more detailed, but the only obstruction to that is computation time.

This section, details how the extraction of the nonlinear intrinsic parameters over a range of gate and drain voltages was automated using coded functions implemented in ADS data display window equations. The evaluation of the extracted values over the whole bias range without manual intervention was enabled by careful manipulation of index numbers in the sweep lists. The core of the functions were the extraction equations presented in [Tong 2004]. Appendix A.1 provides detailed description of the coded functions implemented in the ADS data display window for automatic parameter extraction.

This automation method can also be applied when building table-based models of fabricated transistors. This can be done by placing the measured bias-dependent S-parameter data in a multi-dimensional data file (such as the Generic MDIF type in ADS) and referencing it from an equation-based data component (such as the S2P component in ADS used for 2 port S-parameters) through a Data Access Component (DAC). This way, the same sweep of gate and

drain voltages applied to the transistor can be applied to the S2P component and the relevant S-parameters to each bias level will be obtained every time. The extraction equations based on these S-parameters and the automation setup can then be applied in the same way.

4.2.3.1 Extraction of the Extrinsic Parameters

As previously stated, the extraction of the extrinsic components was performed following the method presented in [Tong 2004]. The transistor used was a 180nm RF MOSFET with total width of 100µm formed from 20 gate fingers, each 5µm wide. As the extrinsic components are independent of the bias level, no sweep of gate and drain voltages was required, and the values obtained are valid for all bias points. To extract the extrinsic resistors, the voltage at the gate was set to 1.8 Volts and the voltage at the drain was set to 0 Volt according to the extraction method in [Tong 2004]. The simulation frequency was swept from 0 to 10GHz which is the maximum operating frequency of the transistor. Based on Equations (4.1) to (4.3), the functions written in the data display window in ADS for extracting the extrinsic resistances are described in Appendix A.2.1. Figure 4.2 displays the results of the extraction.

$$R_s = \text{real}(Z_{12}) = \text{real}(Z_{21}) \quad (4.1)$$

$$R_d = \text{real}(Z_{22}) - R_s \quad (4.2)$$

$$R_g = \text{real}(Z_{11}) - R_s \quad (4.3)$$

As evident from Figure 4.2.a, b, and c, the extracted values of the extrinsic resistors vary by not more than 10% up to 10GHz. The name `Rx_plot` is used to evaluate the extracted resistances against the swept frequency range (where x is d, g, or s for the drain, gate or source resistance, respectively), reserving the name `Rx` for the final value of each resistor. In Figure 4.2.c `Rs1` and `Rs2` represent the two different approaches to evaluating R_s which are supposed to yield equal values, but as can be seen in Figure 4.2.c `Rs1` and `Rs2` slightly differ. Figure 4.2.d shows the three extracted resistances against frequency on a single plot after taking the average of `Rs1` and `Rs2` to evaluate `Rs_plot`. As explained in Appendix A.2.1, the final values of R_d and R_g are taken as their values at the centre frequency point (at 5GHz) which constitutes an average value. The final value of R_s is the average of the two centre values of `Rs1` and `Rs2`.

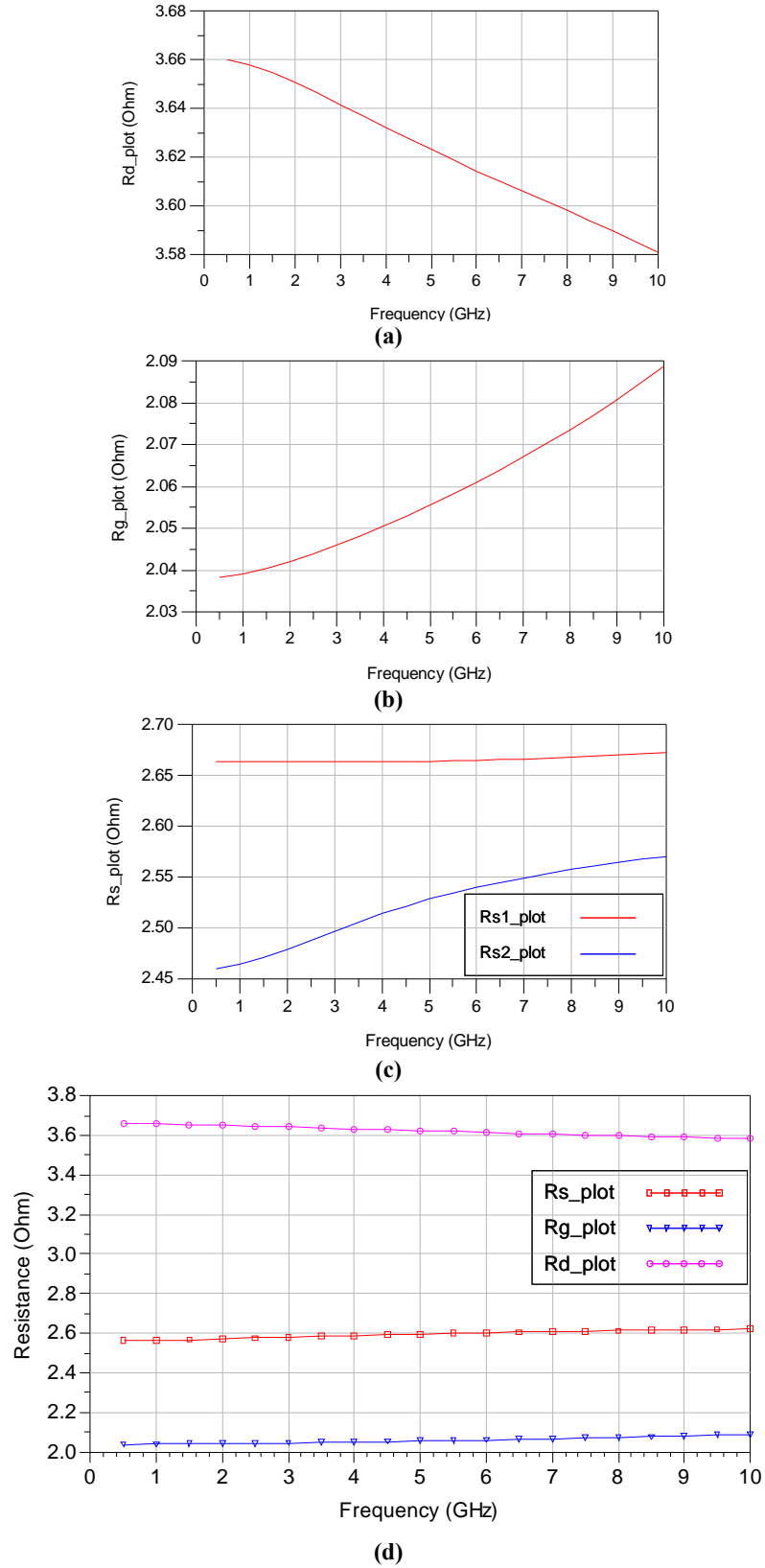


Figure 4.2 Plots of extracted resistance values against frequency
(a) R_d , (b) R_g , (c) R_s and (d) All extrinsic resistances.

In [Tong 2004], the author refers to [Enz 2000] and suggests that only R_g is to be de-embedded from the Z-parameters measured at the desired operating biasing condition of the transistor because the poles due to the terminal source and drain resistances are at a much higher frequency than the typical transit frequency and therefore, they can be neglected when calculating the Y-parameters of the equivalent circuit. However, tests on this transistor showed that not de-embedding R_s and R_d lead to a significant variation in the resulting S_{21} between the resulting model and the modelled transistor.

4.2.3.2 Extraction of the Intrinsic Parameters

To extract the nonlinear intrinsic components, the resistors mentioned above were de-embedded by connecting a negative resistance of equal value to the corresponding terminal, as shown in Figure 4.3. The figure also shows how the sweep of gate and drain voltages was made in ADS. This is an S-parameter simulation (also calculating Z and Y parameters) with the gate and drain voltages being swept over the range of voltages from 0 to 1.8V on the gate with 50mV steps and from 0 to 4V on the drain steps of 100mV. These ranges and steps are designed to improve the accuracy of the model.

When variables are swept in ADS, every sweep point is given an index number. For instance, when V_{GS} is swept from 0 to 1.8V on 0.05V steps, 37 index points are assigned. Every other parameter in the circuit, currents or voltages for instance, then has 37 values dependant on the 37 V_{GS} values. These values are accessed by the corresponding index numbers that were assigned to V_{GS} . When a double sweep is performed, such as the one in Figure 4.3, one of the sweeps has to take place inside the other. This is demonstrated in the schematic by sweeping VDS in the SP simulation controller and sweeping VGS in VDS. This means that at every swept VDS point, VGS is going to be swept over its 37 values, and the process is repeated for all 41 VDS points resulting from the VDS sweep range. Consequently, every parameter in the circuit will have a number of points equivalent to the product of multiplying the number of VGS points by the number of VDS points, and can be addressed or accessed by a double index accordingly. The [what] function can be used to test the dependencies of parameters. These features are demonstrated in Appendix A.2.2 which describes how manipulation of the sweep lists enabled the automated extraction of the model's small signal intrinsic parameters with a high level of accuracy

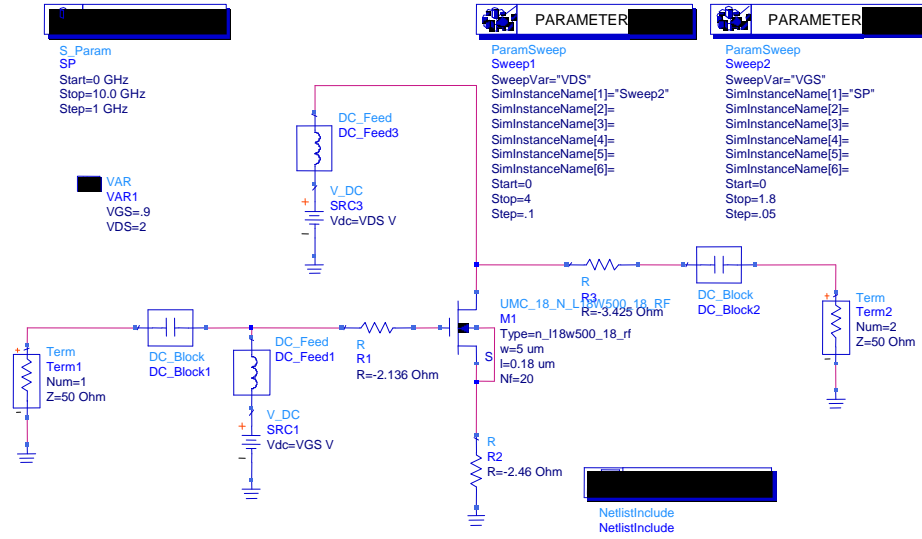


Figure 4.3 ADS schematic setup for extracting the intrinsic model parameters

Extraction of g_m and g_{ds}

The automated extraction of the transconductance, g_m and the output conductance, g_{ds} is detailed in Appendix A.1.2.1 based on Equations (4.4) and (4.5)

$$g_m = \text{real}(Y_{21})_{\omega=0} \quad (4.4)$$

$$g_{ds} = \text{real}(Y_{22})_{\omega^2=0} \quad (4.5)$$

With this setup, values for g_m and g_{ds} can be extracted for the full range of specified bias points without any manual intervention. This enabled 3D plots of these parameters with respect to V_{GS} and V_{DS} to be generated, as shown in Figure 4.4 and Figure 4.5. The current version of ADS does not support 3D plots so the data was exported to Matlab and plotted using the `[surf]` function.

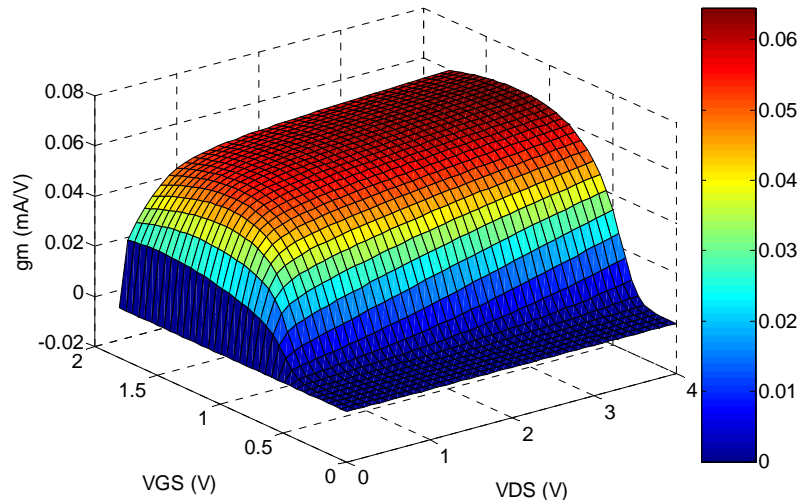


Figure 4.4 3D plot of g_m variance versus drain and gate voltages

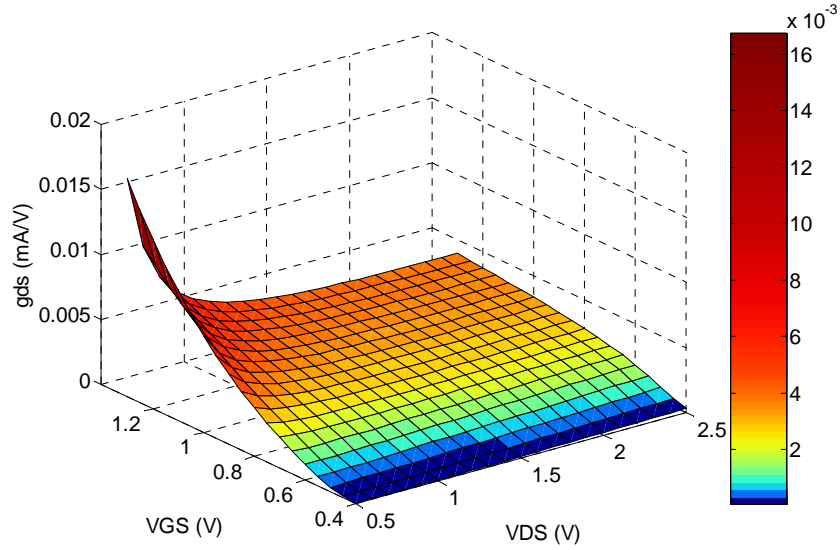


Figure 4.5 3D plot of g_{ds} variance versus drain and gate voltages

Figure 4.4 and Figure 4.5 show the nature and extent of the nonlinear behaviour of the transconductance and the output conductance in terms of how these nonlinear parameters vary with varying gate and drain voltages. Note how the transconductance increases rapidly in the gate voltage range of 0.5 to 1V as the transistor switches on, and then becomes constant as the gate voltage increases further.

▪ Extraction of C_{gs} and C_{gd}

The automated extraction of the intrinsic capacitances C_{gs} and C_{gd} , which are independent of the values of any other parameters, is detailed in Appendix A.1.2.2 utilising Equations (4.6) and (4.7) [Tong 2004]. In Figure 4.6 the extracted values of C_{gs} and C_{gd} are plotted against frequency at a chosen bias point (1V on the gate and 1.8V on the drain).

$$C_{gs} = -\frac{\text{imag}(Y_{11}) + \text{imag}(Y_{12})}{\omega} \quad (4.6)$$

$$C_{gd} = -\frac{\text{imag}(Y_{12})}{\omega} \quad (4.7)$$

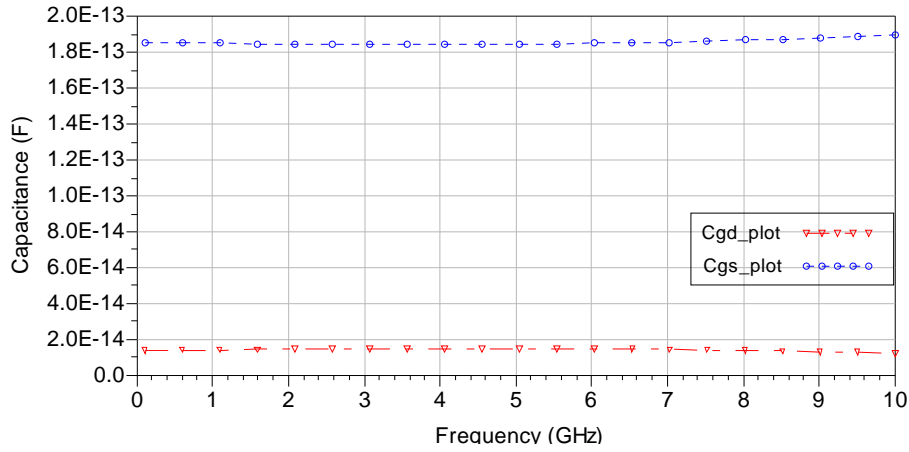


Figure 4.6 Plot of extracted C_{gs} and C_{gd} against frequency.

As can be seen from Figure 4.6, the extracted capacitances are almost constant over the simulated frequency range; this further demonstrates the validity of choosing the extraction method reported in [Tong 2004] to extract the parameters of this transistor. To show the nature and extent of the nonlinear behaviour of these two capacitances, Figure 4.7 and Figure 4.8 show, in 3D, how the extracted values of C_{gs} and C_{gd} change over an extended range of gate and drain voltages. The value of C_{gs} increases quickly from low value to near maximum as the transistor switches on around the threshold voltage. C_{gd} follows the opposite trend, decreasing as the transistor switches on and then becomes almost constant as both V_{DS} and V_{GS} increase.

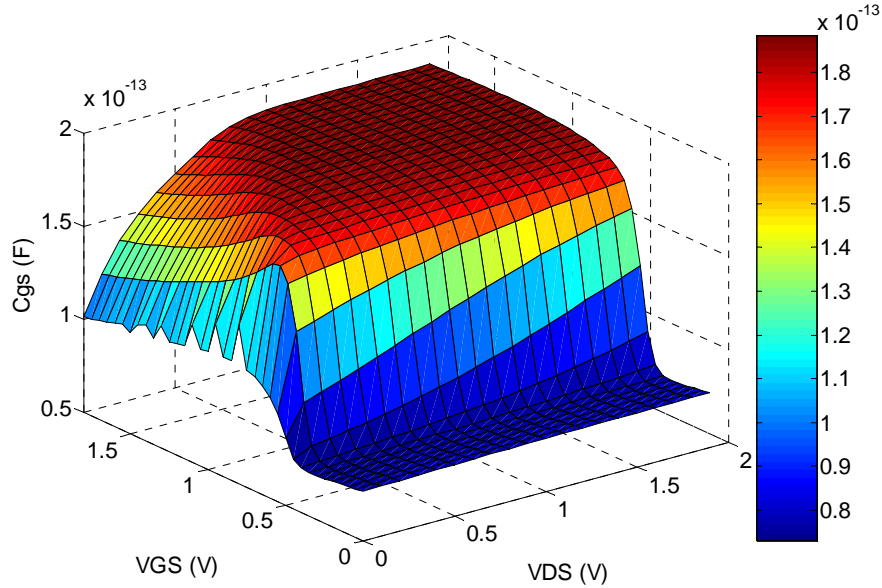


Figure 4.7 3D plot of C_{gs} variance versus V_{GS} and V_{DS}

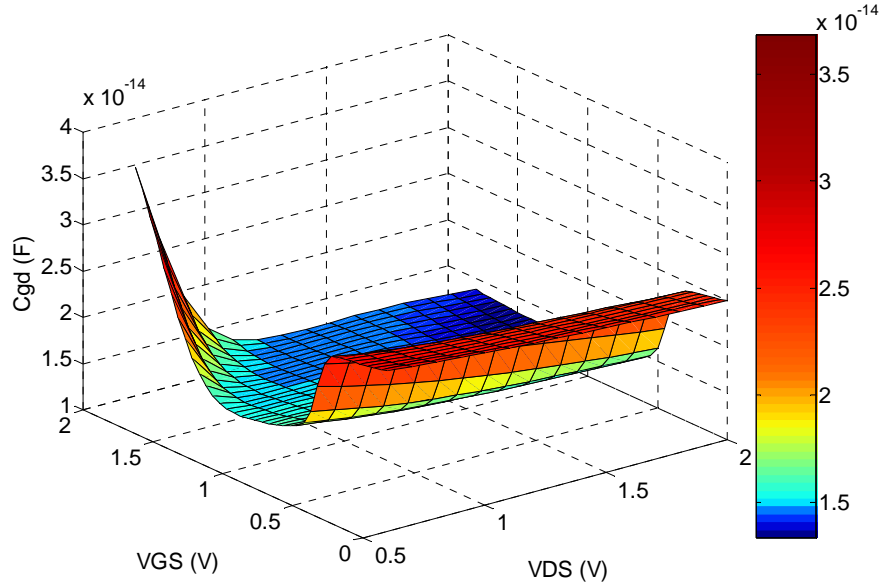


Figure 4.8 3D plot of C_{gd} variance versus V_{GS} and V_{DS}

▪ Extraction of R_{sub}

The automation of the extraction of R_{sub} is more complicated than the extraction of C_{gs} and C_{gs} , as it is not directly determined from the Z or Y parameters. According to [Tong 2004], $Y_{sub} = Y_{22} - g_{ds} - j\omega C_{ds} - j\omega C_{gd}$ and R_{sub} is the slope of the $\omega^2 / \text{real}(Y_{sub})$ versus ω^2 plot. The automated extraction of R_{sub} is described in Appendix A.1.2.3

▪ Extraction of C_{jd} and C_{ds}

Following the extraction of R_{sub} , the determination of the two remaining nonlinear parameters, C_{jd} and C_{ds} (whom depend on R_{sub}) is relatively straightforward and is performed directly from Equations (4.8) and (4.9) [Tong 2004] as described in Appendix A.1.2.4.

$$C_{jd} = \left(\frac{\omega^2 R_{sub}}{\text{real}(Y_{sub})} - \omega^2 R_{sub}^2 \right)^{-1/2} \quad (4.8)$$

$$C_{ds} = \frac{\text{imag}(Y_{22})}{\omega} - C_{gd} - \frac{C_{jd}}{1 + \omega^2 R_{sub}^2 C_{jd}^2} \quad (4.9)$$

Figure 4.9 adds to Figure 4.6 the extracted values of C_{jd} and C_{ds} versus frequency at 1.8 Volts on the drain and 1 Volts on the gate. Note that they are also almost constant with frequency, indicating the validity of the extraction of R_{sub} .

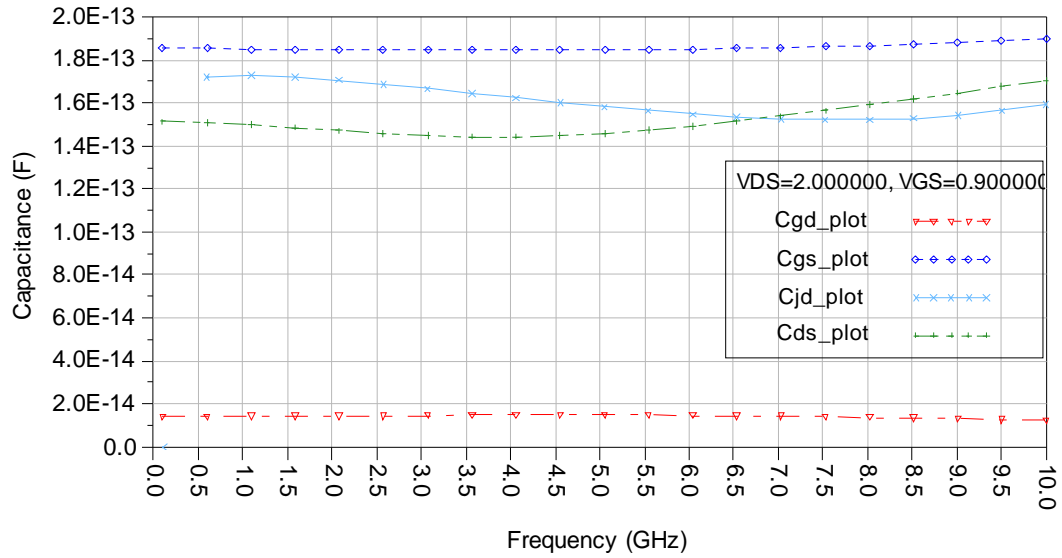


Figure 4.9 All extracted intrinsic capacitances versus frequency

As demonstrated, the intrinsic parameter extraction procedures were entirely automated by making use of the index number functions in ADS data display. Every step of the flow ensured a sustained and measured dependence on all points of the varying terminal voltages of the transistor, making this procedure completely insensitive to neither the desired range of terminal voltages nor the required accuracy. It is also very adaptable in that it can be applied to any type of transistor even those with negative gate voltages because the index numbers will always be assigned in the same way.

4.2.3.3 Finalising Extracted Parameter Values

As is the case in the work of [Tong 2004] and others, manual optimisation is often performed on the extracted values to improve the match between the model and the modelled transistor at the chosen bias point as often the extracted values do not yield the best match. In this work, such an approach is impractical due to the great accuracy desired, which translates to about 1600 different bias points. Multiplying by the 7 intrinsic nonlinear parameters to be optimised, results in more than 15,000 required manual optimisations. This is clearly an impractical route to follow due to the time required to carry out such a task.

The final values of the extracted extrinsic resistors were taken as their values at the centre of the swept frequency range, as described in Appendix A.2.1. The final values of the transconductance and the output conductance were taken as their values at $\omega = 0$ as described in Appendix A.1.2.1. Due to the extraction method in [Tong 2004], R_{sub} only has one value irrespective of frequency, as seen in Appendix A.1.2.3. To try to achieve the maximum possible accuracy in the extracted intrinsic capacitances, their average values over the frequency range was calculated. As can be seen from Figure 4.9 this is more necessary for C_{jd} and C_{ds} than for

C_{gs} and C_{gd} . Appendix A.2.3 describes how this process was automated in the ADS data display window.

4.2.3.4 Writing Extracted Intrinsic Parameters Values to Data Tables

The final concern of this work is to carry the huge number of extracted values of various intrinsic parameters to the data tables. Due to the number of points involved, performing this manually is not a realistic option. One alternative method is to export a data list from the data display window to a text file in ASCII format using menu options. This is a convenient way since it carries the whole range of data over all the bias points automatically. But the main drawback of this method is that the headers and ends of every section of the table (as shown in Figure 4.10.a) will have to be defined manually. Instead, the `[write_var]` function is used as described in Appendix A.2.4, using C_{gs} as an example.

4.3 Modelling Representation Methods

In this section, two different representations used to build the nonlinear model are introduced and their accuracies and performance in S-parameter and harmonic balance simulations are discussed. The novelty that all of these methods share, is that they are the first reported table based nonlinear models that represent each of the elements in the nonlinear model completely independent from one another.

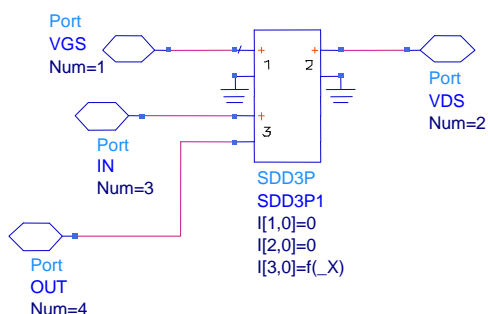
The obtained values of the nonlinear parameters at the gate and drain voltages are placed in a table in a text file in the Generic MDIF format, referenced by the voltages at which they were obtained. Figure 4.10.a shows a simplified section of a typical table for few data points of a parameter 'x', for demonstration. The data in the table is accessed by the Data Access Component (DAC) and is passed to the respective SDD representing each nonlinear parameter. Performance between measured points in the table is interpolated in the DAC to ensure a correct description of harmonics and convergence within harmonic balance simulations. Figure 4.10.b and Figure 4.10.c, show an SDD representing the nonlinear parameter 'x' and its DAC, respectively, to illustrate how this works.

```
VAR VDS(1)= .3
Begin TestData
% VGS(1) X(1)
0.00E+00 2.99E-14
1.00E-01 2.95E-14
2.00E-01 2.91E-14
3.00E-01 2.88E-14
End
```

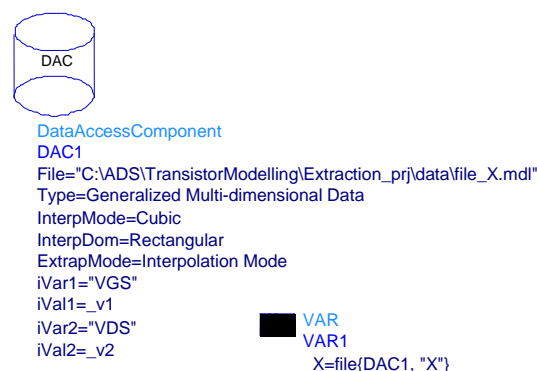
```
VAR VDS(1)= 0.4
Begin TestData
% VGS(1) X(1)
0.00E+00 2.93E-14
1.00E-01 2.90E-14
2.00E-01 2.87E-14
3.00E-01 2.84E-14
End
```

```
VAR VDS(1)= 0.5
Begin TestData
% VGS(1) X(1)
0.00E+00 2.87E-14
1.00E-01 2.85E-14
2.00E-01 2.83E-14
3.00E-01 2.81E-14
End
```

(a)



(b)



(c)

Figure 4.10 Demonstration of data representation using SDD (a) A sample data table, (b) A sample SDD and (c) its sample DAC.

In Figure 4.10.c, `_X=file{DAC1, "X"}` puts the values of `X` fetched from the table through the DAC to the on-schematic variable `_X`. The DAC specifies the location of the data file (`file_X`), the type of the data file and the type of interpolation and extrapolation to be performed on the data. It also specifies the control voltages upon which values from the table are to be selected. `iVar1=VGS` refers to the first variable on the table and associates it through `iVal1=_v1`, with `_v1` which is the voltage on Port 1 of the SDD in Figure 4.10.b. Similarly, the voltage on Port 2 of the SDD is attached to the variable `VDS` in the table through `iVar2=VDS` and `iVal2=_v2`. With the two variables in the table mapped to the two voltages on the ports of the SDD, the selected `X` value from the table will depend on the

applied VGS and VDS on Port 1 and Port 2 of the SDD, respectively. Interpolation is performed on the data in the table to fill the gaps between X points if VGS and VDS did not change enough to push the selection to the next value in the table. This is always the case when an input power is applied; hence this is fundamental for harmonic balance simulations. The output of the SDD in Figure 4.10.b is in the form of a current defined by $I[3, 0] = f(X)$. Where 3 is the number of the output port in the SDD and the second item in the square brackets indicates a weighting function. Some weighting functions are preset in the SDD, or they can also be user-defined. $f(X)$ is a function defining the current due to component X.

Based on the above description of how an element can be represented using an SDD with a data table and a DAC, every element in the nonlinear model is defined by its current, making the model a collection of the small signal currents of all its elements. The following sections will discuss how an SDD was used to represent each element in the nonlinear model and the impact of that representation on the performance of the model. Two main types of components exist, conductances and capacitances. Each of these were represented in two different ways and combinations of them were used.

4.3.1 First Method - Conductances and Capacitances (Model 1)

This section discusses the development and verification of the first method of constructing a transistor nonlinear model where each nonlinear element of the model is individually and separately represented. In this method, the transconductance and output conductance are represented as voltage-controlled conductances while model capacitances are represented as voltage-controlled capacitances.

4.3.1.1 Model Construction

Figure 4.11.a to Figure 4.11.g, show how the model's nonlinear elements were individually represented; in Figure 4.12 the whole model is constructed. Port 1 and Port 2 of all the components in Figure 4.11.a to Figure 4.11.g are for the controlling voltages, used to determine which table values are chosen, as explained in the beginning of Section 4.3 above. In the top hierarchy of Figure 4.12, all port 1 are connected to the combined DC and AC voltage on the gate, Vgs, and all port 2 are connected to the drain voltage, Vds.

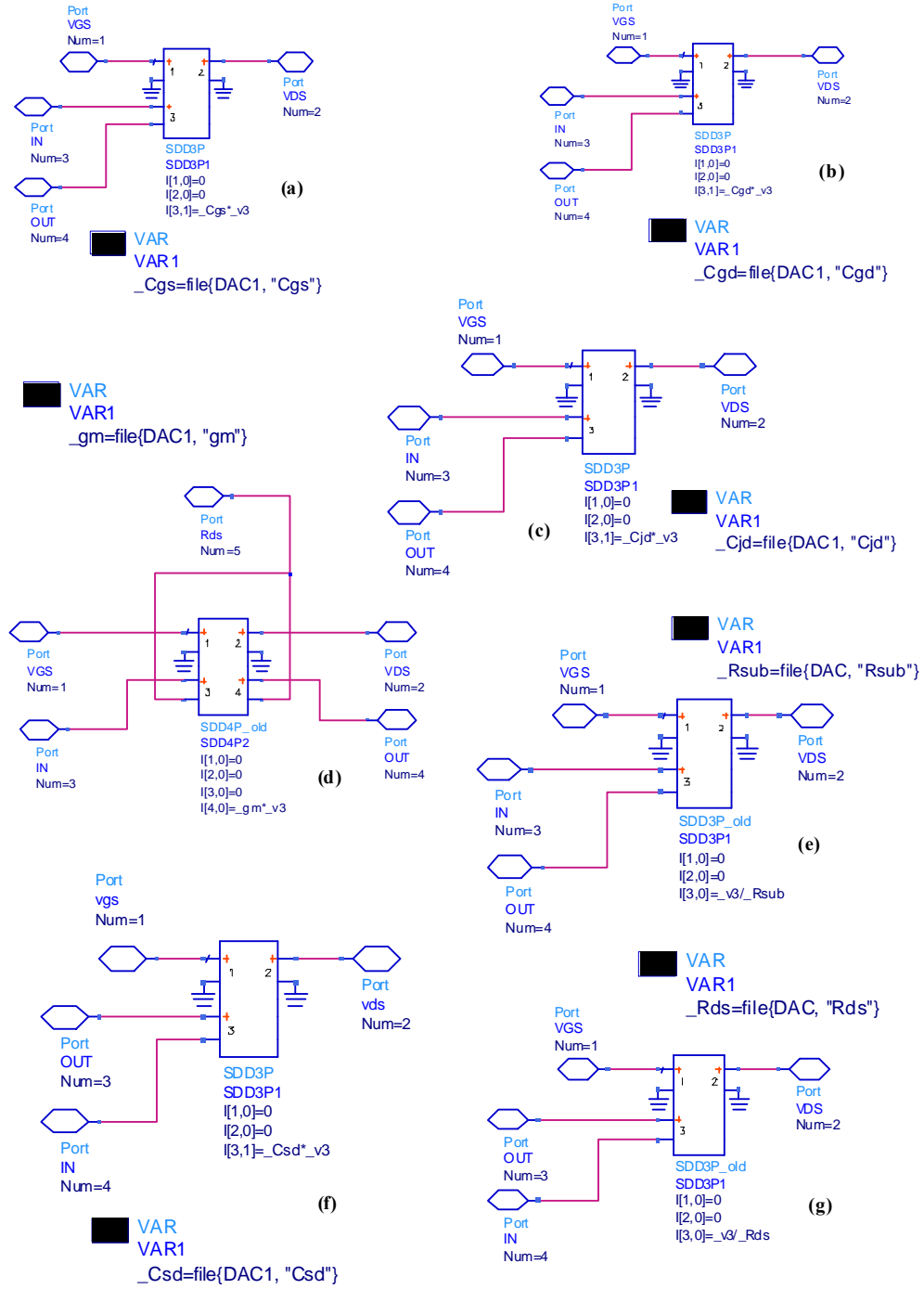


Figure 4.11 SDD representations of nonlinear intrinsic components for Model 1 (a) C_{gs} , (b) C_{gd} , (c) C_{jg} , (d) g_m , (e) R_{sub} , (e) C_{ds} and (g) g_{ds}

In Figure 4.11.d, the nonlinear transconductance, g_m , is represented as a voltage-controlled conductance (varying with V_{GS} and V_{DS}). Its small signal current is modelled as a voltage-controlled current source (varying with the small signal gate-source voltage, v_{gs}). The output current, $I[4,0]$ (that is the output current of port 4 and the 0 indicates no weighting function)

is made equal to $_gm*_v3$. gm refers to the extracted transconductance values which are fetched through the DAC from the data file. $_v3$ is the voltage between the two terminals of port 3 on the SDD, the small signal AC voltage connected in the top hierarchy in Figure 4.12 as the voltage across C_{gs} , v_{gs} . This is made possible by DC Block 3 in Figure 4.12 which prevents the DC component from reaching the input terminal making $_v3$ the AC small signal voltage only. The reference of port 3 is connected to R_s . This ensures that $_v3$ is the voltage drop across C_{gs} only excluding the voltage drop on R_s . The reference of the output (port 4) is also connected to R_s because the current $g_m v_{gs}$ goes down to R_s not the ground (therefore it is not connected to ground). This representation is equivalent to the $g_m v_{gs}$ voltage controlled current source of the model in Figure 4.1.

In Figure 4.11.g, the output conductance, g_{ds} , is represented as a voltage controlled resistance (varying with V_{GS} and V_{DS}) by converting its extracted g_{ds} values to resistance, R_{ds} . Its small signal current is modelled as a voltage-controlled current source (varying with the small signal drain-source voltage, v_{ds}). The output current, $I[3,0]$ is made equal to $_v3/R_{ds}$. Here, $_v3$ is the small signal v_{ds} and R_{ds} is the extracted value fetched from the table through the assigned DAC. As at the input, DC Block 4 in Figure 4.12 ensures that the voltage on $_v3$ is the AC signal only. This is equivalent to $g_{ds} v_{ds}$ in the original model of Figure 4.1. R_{sub} is modelled similarly using Ohms law in Figure 4.11.e.

All nonlinear capacitances in Figure 4.11.a, b, c and f are represented as voltage controlled capacitances using their extracted values. Their currents are defined using the differential equation $i_C = C(dv_C/dt)$ where the weighting function of 1 in the output current ($I[3,1]$) represents a differentiation which is predefined in the SDD. For C_{gs} in Figure 4.11.a, the differentiation of $_v3$ is performed in $I[3,1] = _Cgs*_v3$. This is the small signal voltage connected in the top hierarchy across C_{gs} . This is achieved by connecting one of the terminals in port 3 to the gate and the other to R_s , again in order to exclude the voltage drop across R_s . All other capacitors are modelled similarly. For C_{gd} in Figure 4.11.b, one of its port 3 terminals is connected to the gate and the other to the drain. Port 3 of C_{jd} in Figure 4.11.c is connected between the drain and R_{sub} and for C_{ds} in Figure 4.11.f between the drain and R_s .

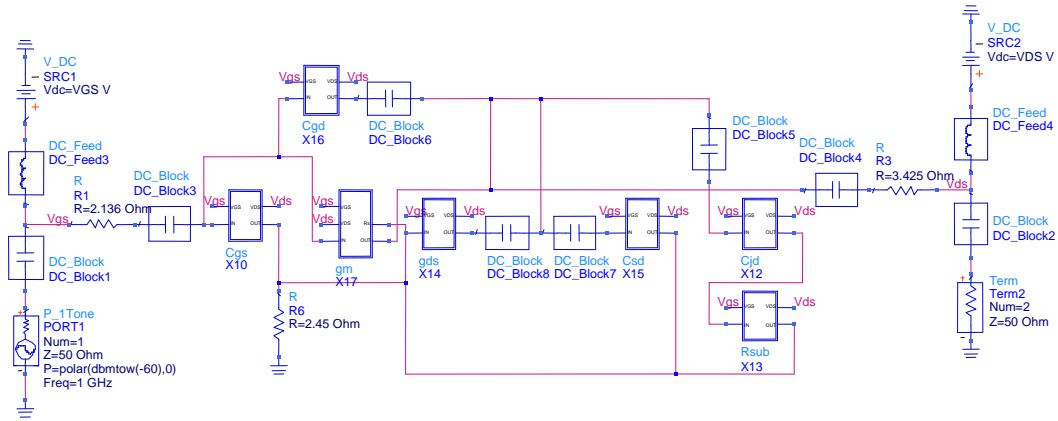


Figure 4.12 complete SDD model with all intrinsic and extrinsic components

The whole model of Figure 4.12 was reconstructed by comparison to the original small signal model of Figure 4.1. This is a straightforward process and is a further evidence to the simplicity of this method. What needs to be noted is the importance of the DC feeds (very high inductances) and the DC blocks (very high capacitances) in the model. These insure the model operates as a truly nonlinear one. In order for the model to be truly nonlinear, values of nonlinear parameters in the model should be variable subject to the input power superimposed on the DC bias as well as to the bias level itself. DC Block 1 and DC Block 2 isolate the input and output terminations from the DC supply. DC Feed 1 and DC Feed 2 prevent the AC component of the signal at the terminals from being shorted through the DC supply. The combination of the DC bias and the AC power superimposed on it is the voltage used for controlling which values of the interpolated table are used. This voltage is set on the controlling ports (port 1 and port 2) of all the SDDs representing the transistor's nonlinear parameters. In this way, the value selected from the table will depend on the input power, making the model truly nonlinear. DC Block 3 and DC Block 4 prevent the DC component from being passed to the RF input ports of all the SDDs. Finally DC Blocks 5, 6, 7 and 8 prevent the DC component produced from the differentiation in the capacitor models from going through the RF input ports of other parameters in the model.

4.3.1.2 Model Verification

Figure 4.13 shows S-parameter test results from 50MHz to 10GHz of the model and the modelled transistor at single bias point. The quality of the match is clear.

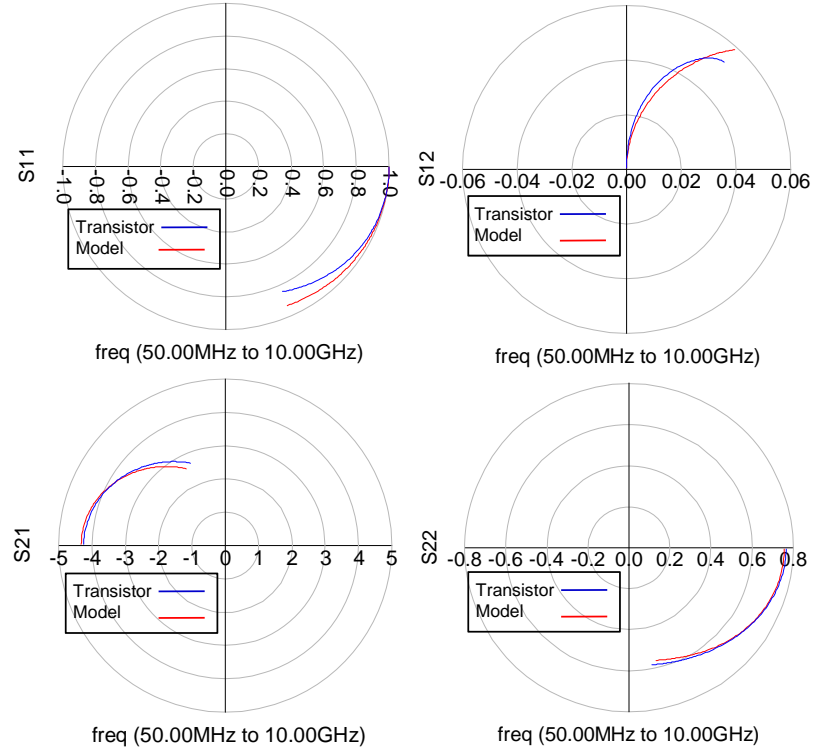
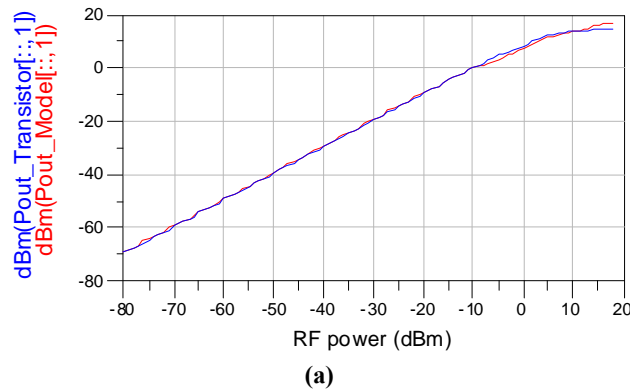


Figure 4.13 S-parameter test results of the transistor model and the modelled transistor

To verify the nonlinear behaviour of the model, a one tone test was conducted with input power swept from -80 to 18dBm. Figure 4.14 shows the fundamental and the third harmonic outputs over all the swept input power values of this test for both the modelled transistor and the transistor model. As can be seen the model output matches the transistor output even beyond compression which means its non linear behaviour and hence its nonlinear representation is accurate.



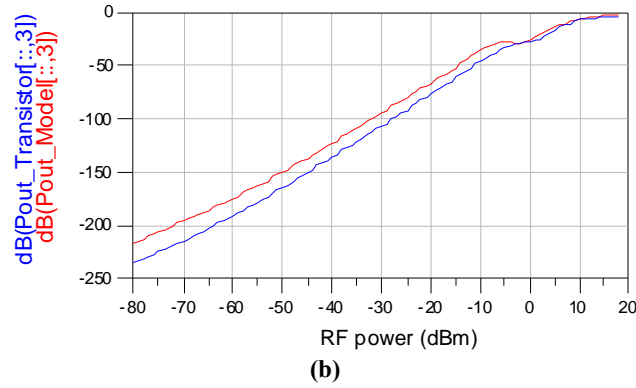


Figure 4.14 One-tone Harmonic Balance test results of the model (a) Fundamental tone and (b) Third harmonic

The main disadvantage of this representation is that the model is not very reliable in harmonic balance simulations. At some levels of bias, no convergence is achieved and matching harmonic content becomes more unreliable when the transistor is biased near the limits of the table. It was also observed that the higher the input power, the more difficult convergence becomes. This is because a high input power will introduce a bigger voltage swing in the output and may lead to saturation in the table (reaching the end point of the data provided in the table), in which case the data in the table is extrapolated by the DAC and, inevitably, errors are introduced.

4.3.2 Second Method – Currents and Capacitances (Model 2)

This representation method tries to solve the main drawback of the model in the previous section which is its performance in harmonic balance simulations, both in terms of convergence and quality of fit over a wide range of bias voltages. Instead of modelling the transconductance and output conductance using their extracted values, they are modelled from the current characteristics of the transistor. The validity of this approach is based on the non dispersive behaviour of the transistor's g_m and g_{ds} as both are taken as their extracted values at zero frequency, according to the extraction method described in [Tong 2004], with the resulting model valid at any frequency up to 10GHz.

4.3.2.1 Model Development – Generating I_{g_m} and $I_{g_{ds}}$

Representing g_m and g_{ds} from the transistor's current characteristics is comprehensive because the current characteristics describe the behaviour of the transistor in terms of the variance of its drain current with respect to its terminal voltages which is essentially what g_m and g_{ds} describe.

To separately represent each intrinsic model element (g_m and g_{ds} in this case) for the purpose of quantifying contributions to distortion from the various nonlinear elements, the individual

effect on current from g_m and g_{ds} must be obtained. This requires accurate knowledge of the transistor threshold voltage in order to determine the exact point at which it reaches its saturation region of operation. There exist numerous methods to extract the value of threshold voltage that have been proposed in the literature; many of these methods were reviewed in [Conde 2002]. The authors compared the available methods by applying them to the same transistor and discussing their performance in terms of both the obtained result and ease of use. After careful examination of available threshold voltage extraction methods, the Extrapolation in the Linear Region (ELR) method presented in [Schroeder 1998] and [Terada 2001] among others was chosen for this thesis work. It was the easiest and fastest to apply and also lead to an average result among other methods when applied to the same transistor in the work of [Conde 2002]. Appendix A.2 describes the ADS data display window coded functions designed to automate the process of generating separate g_m -related and g_{ds} -related current characteristics for the transistor.

▪ Accurate V_T Extraction

ELR works by finding the gate voltage axis intercept of the linear extrapolation of the I_{DS} - V_{GS} characteristics at its maximum first derivative (slope) point [Conde 2002], i.e. maximum g_m point when the transistor is operating in the linear region. In order to enhance accuracy and enable reusability, it was decided to do this automatically. In Figure 4.15 the transconductance, g_m , is plotted against V_{GS} when the transistor is in the linear region with V_{DS} set to 10mV. As shown in the marker readout, the maximum g_m occurs at $V_{GS} = 0.78V$. Appendix A.2.1 describes the coded functions used to automatically implement the ELR method based on this maximum g_m point.

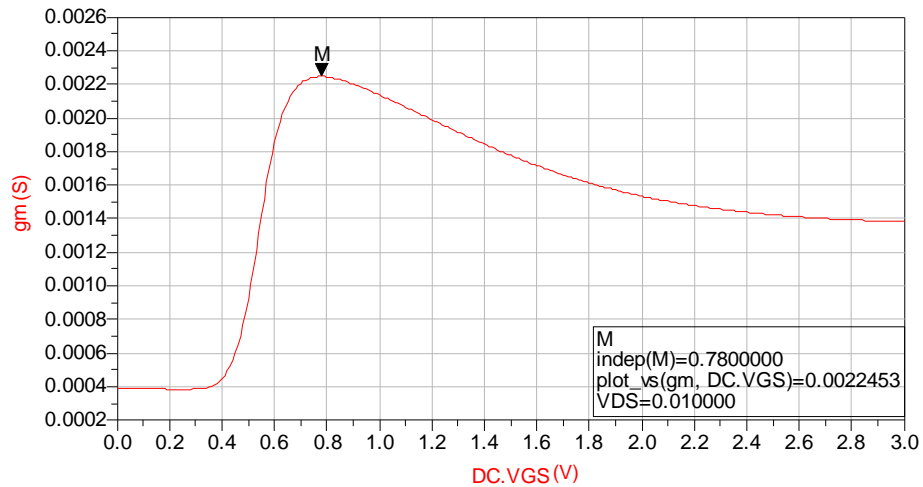


Figure 4.15 g_m against V_{GS} in the linear region when $V_{DS} = 10mV$

Figure 4.16 shows the linear extrapolation of maximum g_m point, crossing the V_{GS} axis in a straightforward implementation of the ELR V_T extraction method.

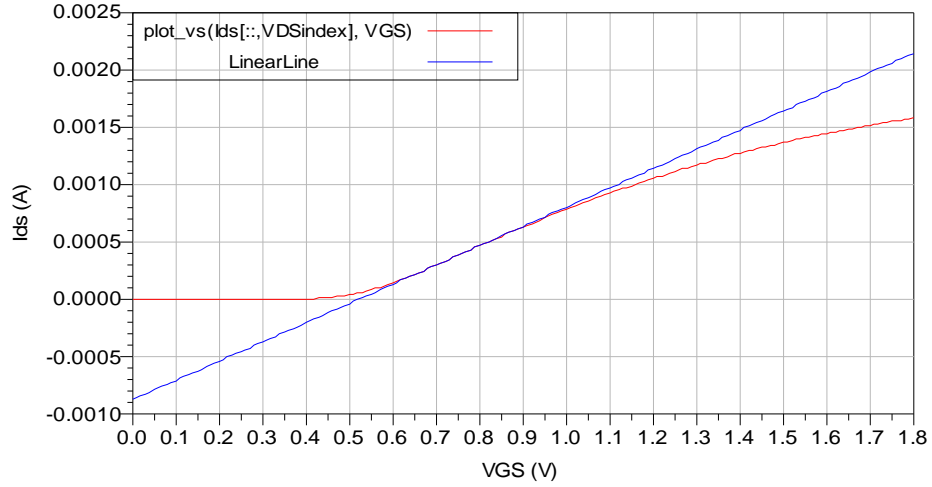


Figure 4.16 Linear extrapolation of maximum g_m point

In Figure 4.17, the graph in Figure 4.16 is zoomed around the point where the linear extrapolation crosses the V_{GS} axis showing that the transistor's threshold voltage is exactly 0.52V.

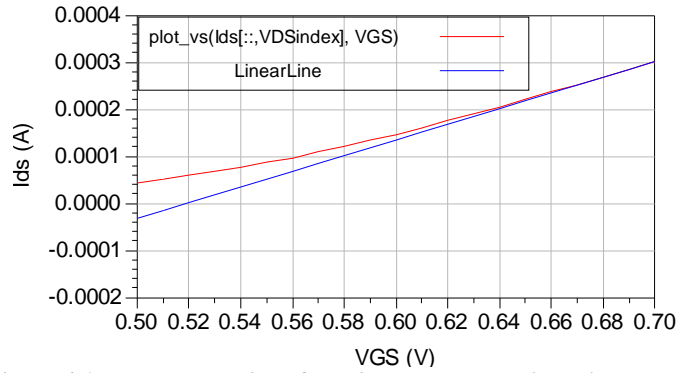


Figure 4.17 The cross point of the linear extrapolation with V_{GS} axis

▪ Splitting Current Characteristics Into I_{g_m} and $I_{g_{ds}}$

The procedures of splitting the transistor's current characteristics shown in Figure A.10 to g_m -related current, I_{g_m} , and g_{ds} -related current, $I_{g_{ds}}$, starts with identifying the knee point on each VGS curve at which the transistor switches between the linear and saturation regions of operation.

In the saturation region, the current is split but in the linear region, the current is added to the I_{g_m} current and $I_{g_{ds}}$ current is set to 0. It is important to note that this split procedure is therefore only valid in the saturation region. This is because in the linear region, V_{DS} does have an influence on I_{DS} so $I_{g_{ds}}$ in that region should not be zero and I_{g_m} should not be equal to I_{DS} .

However, as it is not possible to split the two contributions in the linear region, the split is done in this way so that when these two currents are added in the model, the current characteristics of the transistor are reconstructed. Also, this work aims to operate the transistor in the saturation region in which the split is valid.

As before, the obtained I_{gm} and I_{gds} are stored in data tables accessed through a DAC where the DAC also performs interpolation on the data when required. It was found, as with the extracted values, that the more detailed the data in the table, the more accurate the model, especially in harmonic balance tests where variable input power is involved. That is because less interpolation error is introduced. Manually calculating the saturation point of every V_{GS} curve and subsequent g_m - and g_{ds} -related currents would be very time consuming especially when higher accuracy and wider range are required, and can also easily lead to errors, hence impractical. As with the parameter extraction, this process was programmed to take place almost automatically using coded functions implemented in Data Display Window equations. Appendix A.2.2 fully describes the automation of this process.

Figure 4.18 shows the obtained I_{gm} and I_{gds} currents. As can be seen, I_{gm} is equal to I_{DS} up to the saturation point and then becomes constant with respect to V_{DS} . I_{gds} is zero up to the knee point and then gradually increases as V_{DS} increases.

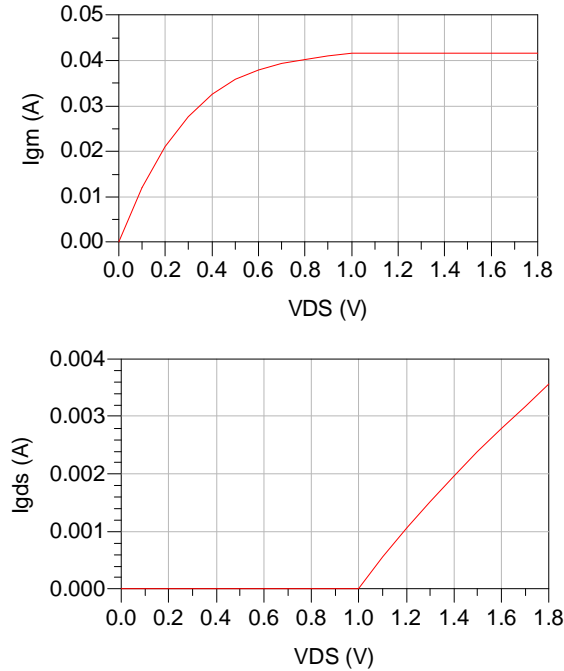


Figure 4.18 Extracted I_{gm} and I_{gds} at one value of V_{GS}

Figure 4.19 shows selected curves of the generated I_{gm} - V_{DS} characteristics which will represent the transconductance, g_m , in the transistor model.

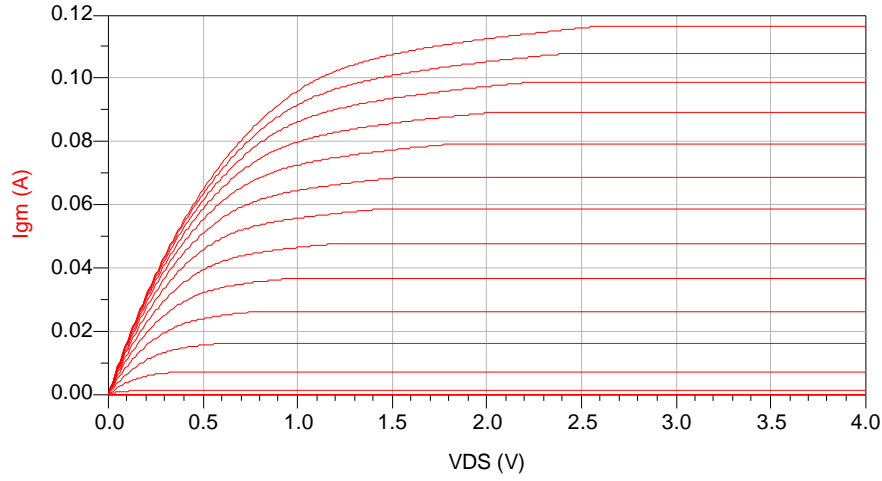


Figure 4.19 I_{gm} - V_{DS} current characteristics for a number of V_{GS} curves

Because the code described in Appendix A.2.2 takes care at every step of retaining the dependency on the swept values of V_{GS} and V_{DS} , it is highly adaptable. This method can be used with any transistor for any range of terminal voltages and with any degree of accuracy. The produced I_{gm} and I_{gds} current data are automatically transferred to data tables using data export equations in a similar fashion to the way the extracted model's parameters were exported.

4.3.2.2 Model Construction and Verification

Figure 4.20 shows the SDDs of I_{gm} and I_{gds} . As can be seen, they are modelled directly as current sources from the table. In Figure 4.21, the whole model is constructed.

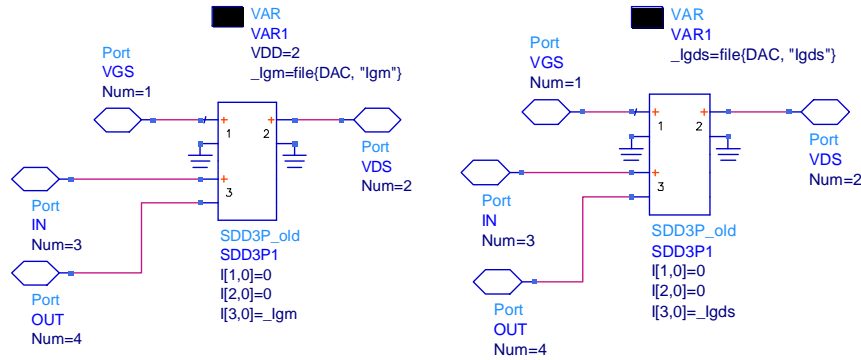


Figure 4.20 SDDs of I_{gm} and I_{gds}

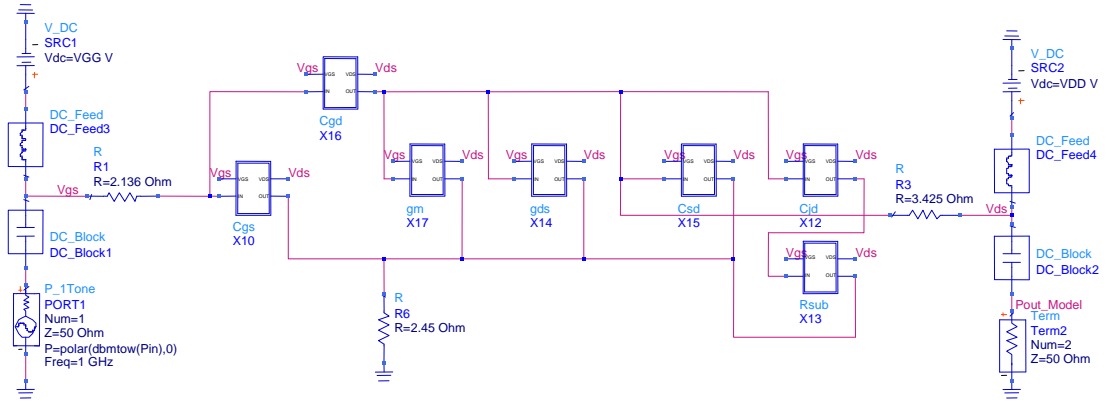


Figure 4.21 Complete SDD structure of Model 2

As with the first model, this new model was verified in both S-parameter and harmonic balance simulations. The S-parameter simulation obtained the same good matching results of the first model, but rather than presenting them again, the Z-parameters are presented instead. As shown in Figure 4.22, there is an excellent match between the Z-parameters of the model and the modelled transistor.

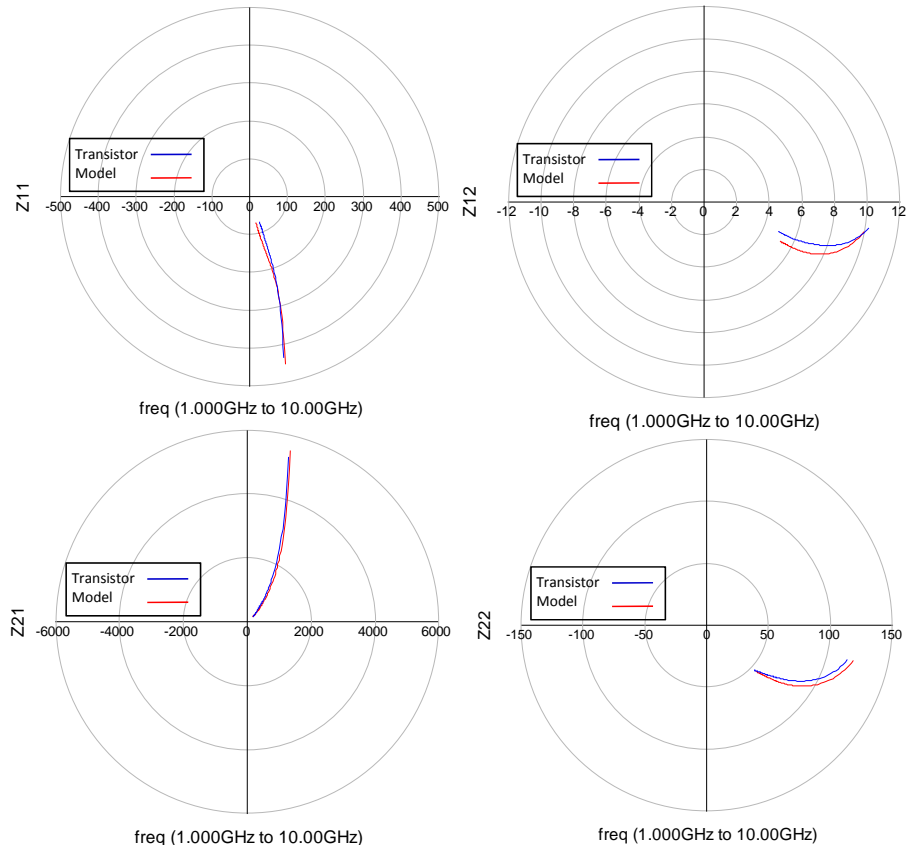


Figure 4.22 Z-parameters of the model and the modelled transistor for Model 2

In harmonic balance, this model performs much better than the first model. The simulation converges at much wider range of bias voltages and the accuracy of the match with the modelled transistor is much better. Figure 4.23 shows an example where the harmonic balance test was performed over input powers ranging only from -10 to 15dBm to study more closely the behaviour of the model around compression where the transistor enters its nonlinear operation region. As can be seen, there is a good match between the model and the modelled transistor in both the fundamental and third harmonic components produced over the range of input powers.

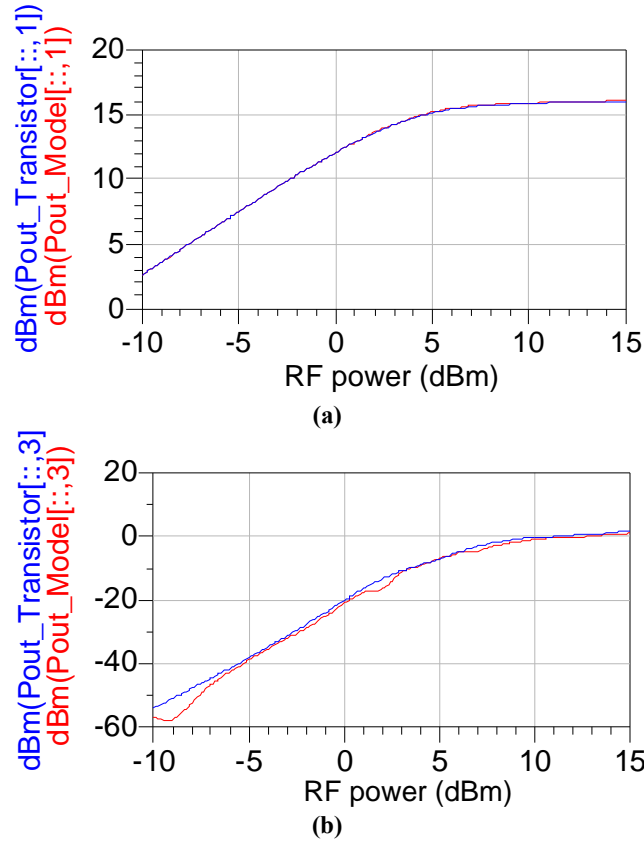


Figure 4.23 One-tone Harmonic Balance test results of Model 2

4.4 Quantification of Sources of Distortion

This section handles the quantification of nonlinear distortion from the model's nonlinear elements using the superposition method proposed in Section 4.1. The investigation begins by improving the nonlinear model in an effort to achieve better convergence and fitting in harmonic balance simulations, hence increasing the robustness of the quantification results. The linearisation of nonlinear elements is then discussed, followed by presentation and discussion of the quantification results.

4.4.1 Third Method – Currents and Charge Sources (Model 3)

In this third representation, transistor capacitances are modelled as voltage-controlled charge sources rather than as voltage-controlled capacitances as in the previous two models.

4.4.1.1 Model Development – Generating Charge Sources

When a DC voltage is applied across the terminals of a capacitor, the capacitor takes time to charge. At the steady state, when the capacitor is fully charged, the charge on the capacitor can be obtained by

$$Q_C = \int_0^{V_{DC}} C dV \quad (4.10)$$

where V_{DC} is the DC voltage applied across the capacitor. The current due to the charge can then be found from $i = dQ_C / dt$, hence the 1 (differentiation) in $\mathbb{I}[3, 1] = Q_{GS}$ of the SDDs in Figure 4.27, as previously explained in Section 4.3.1.1.

The integrations for Q_{gs} , Q_{gd} , Q_{ds} and Q_{jd} are performed as detailed in Equations (4.11), (4.12), (4.13) and (4.14), respectively.

$$Q_{gs} = \int_{0, V_{DS(1)}}^{V_{GS((1):(m))}, V_{DS(n)}} C_{gs}(V_{DS}, V_{GS}) dV_{GS} \quad (4.11)$$

$$Q_{gd} = \int_{0, V_{GS(1)}}^{V_{GD((1):(m))}, V_{GS(n)}} C_{gd}(V_{GS}, V_{DS}) dV_{GD} \quad (4.12)$$

$$Q_{ds} = \int_{0, V_{GS(1)}}^{V_{DS((1):(m))}, V_{GS(n)}} C_{ds}(V_{GS}, V_{DS}) dV_{DS} \quad (4.13)$$

$$Q_{jd} = \int_{0, V_{GS(1)}}^{V_{DS((1):(m))}, V_{GS(n)}} C_{jd}(V_{GS}, V_{DS}) dV_{DS} \quad (4.14)$$

In Equations (4.11) to (4.14), all intrinsic capacitances are nonlinear, depending on V_{GS} and V_{DS} . n and m represent the maximum simulated voltages; their definition is specific for each equation, as explained below

For C_{gs} in Equation (4.11), the full range of simulated drain voltages can be written as $V_{DS(x)}$ where $x = 1:n$ and the full range of simulated gate voltages can be written as $V_{GS(y)}$ where $y = 1:m$. The full range of extracted C_{gs} values can then be written as $C_{gs(x,y)}$. To determine Q_{gs} in Equation (4.11), C_{gs} is integrated with respect to V_{GS} , the voltage across its terminals. Starting at $V_{DS(1)}$ (the first simulated value of V_{DS}), a function is first generated to define the relationship between $C_{gs(1,y)}$ and $V_{GS(y)}$. The integration is then performed from 0 to $V_{GS(y)}$, starting with $V_{GS(1)}$ (the first simulated value of V_{GS}) to determine $Q_{gs(1,1)}$ and repeated until $V_{GS(m)}$ (the last simulated value of V_{GS}) to determine $Q_{gs(1,m)}$. All this is then repeated for the rest of the simulated values of V_{DS} until $V_{DS(n)}$. Similar analysis applies to the rest of the charge sources in Equations (4.12), (4.13) and (4.14).

Looking at Figure 4.24 of the transistor model, it is important to establish the exact potential across a given capacitor for integration, as explained below.

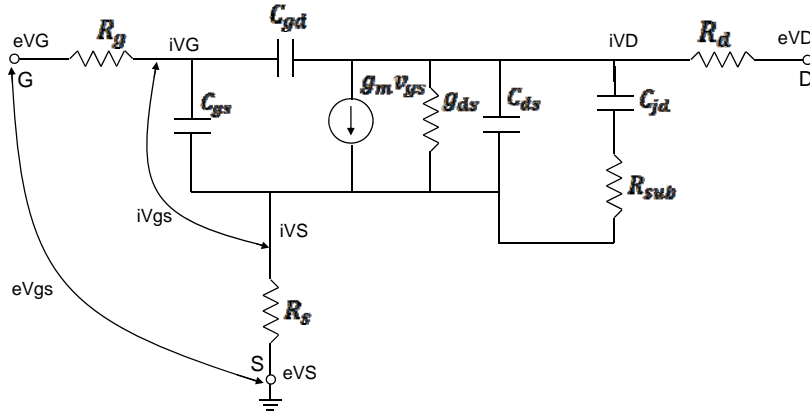


Figure 4.24 The exact voltage applied across a capacitance in the Model

Taking for example C_{gs} , the external V_{GS} ($eVgs$) is not equal to the internal V_{GS} ($iVgs$). At the steady state, C_{gs} and C_{gd} are fully charged and are open circuits so there is no DC current in R_g and eVG equals iVG . But there is current through R_s , hence iVS is not equal to eVS . Therefore, to integrate with respect to the voltage across C_{gs} , the voltage drop across R_s has to be taken into consideration. Similarly, the voltage drop across R_d has to be taken into account to accurately calculate V_{GD} , the voltage drop across C_{gd} . For C_{ds} , the voltage drop across both R_d and R_s must be taken into account. It is the same case for C_{jd} since there is no DC current in R_{sub} as C_{jd} is open circuit at the steady state.

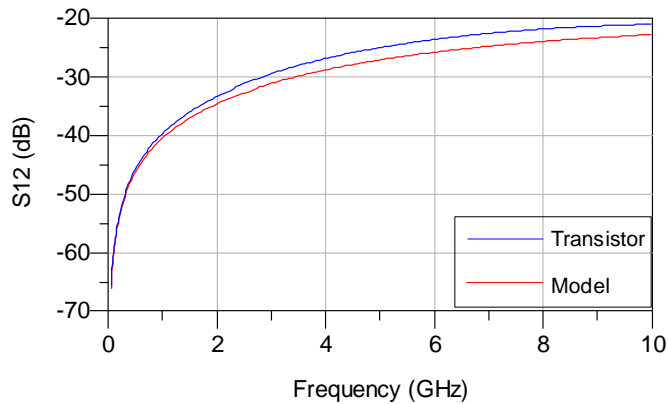
As an example of how significant the difference between the internal and external terminal voltages can be, Figure 4.25 shows a sample of the results of a simulation performed to

calculate the exact voltage across the capacitances by taking into account the current in the terminal resistances. V_{GG} is the DC voltage applied to the gate (equivalent to eV_{gs} in Figure 4.24) and V_S is the voltage drop across the source resistance. As expected, V_S increases as the gate voltage increases, leading to a larger difference between V_{GG} and V_{GS} (equivalent to iV_{gs} in Figure 4.24).

VGG	VS	VGS
0.750	0.033	0.717
0.800	0.039	0.761
0.850	0.045	0.805
0.900	0.051	0.849
0.950	0.057	0.893
1.000	0.064	0.936
1.050	0.070	0.980
1.100	0.077	1.023
1.150	0.083	1.067
1.200	0.090	1.110
1.250	0.097	1.153
1.300	0.103	1.197
1.350	0.110	1.240
1.400	0.117	1.283
1.450	0.123	1.327
1.500	0.130	1.370
1.550	0.137	1.413
1.600	0.143	1.457
1.650	0.150	1.500
1.700	0.156	1.544
1.750	0.163	1.587
1.800	0.169	1.631

Figure 4.25 Error in integrating voltage for C_{gs}

As an example to demonstrate the significance in the mismatch between the model and the transistor this difference in the integrating voltages can cause, Figure 4.26 shows two simulation results for S_{12} at two different bias points. This simulation was performed when the charge sources of the capacitances were produced by integrating the capacitances with respect to the external terminal voltages. As shown in Figure 4.26, the higher the operating point the larger the mismatch. This is because a higher operating point means larger current, leading to more voltage drop across the terminal resistances, hence larger difference between the internal and external terminal voltages.



(a)

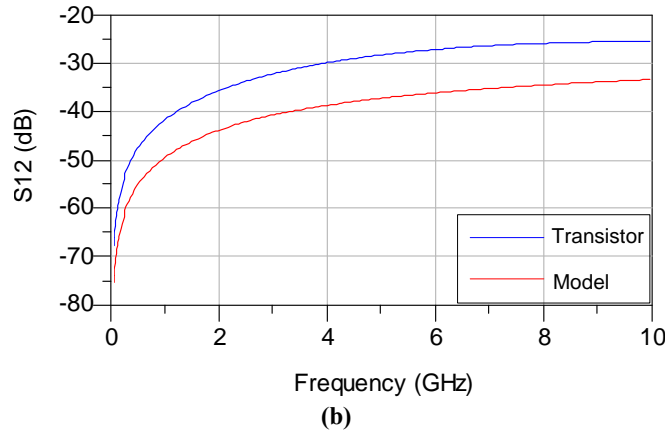


Figure 4.26 Mismatch in S_{21} between model and transistor and small and large operating points (a) $V_{GG}=0.5V$, $V_{DD}=1V$, (b) $V_{GG}=1V$, $V_{DD}=4V$

It is also important to observe the relevance between listing the charge source values (produced from integration with respect to the internal terminal voltages) against the external or internal terminal voltages in the data table, and connecting the SDD's voltage control ports to the external or internal terminal voltages of the transistor. This is demonstrated in Figure 4.27. If the table lists the charges against the external terminal voltages, the SDD connections should follow, as shown in Figure 4.27.a. However, if the charges in the table are listed against values of internal terminal voltages, the controlling ports of the SDD should be connected after the terminal resistors and not directly to the voltage sources as shown in Figure 4.27.b. Otherwise, a serious error may be introduced.

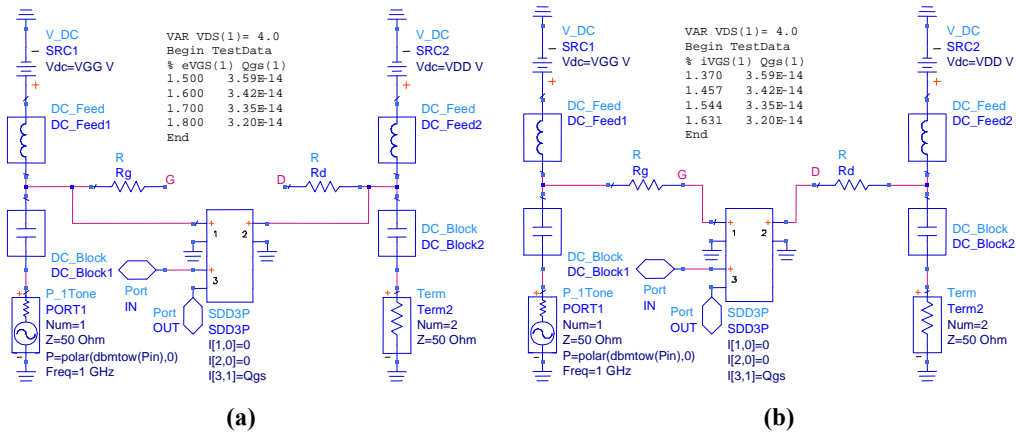


Figure 4.27 SDD connection and table representations with (a) external and (b) internal voltages

4.4.1.2 Assessment of Model Performance

Implementation of this model, which was constructed from the SDDs representing each of the model's elements in the same manner as performed in models 1 and 2, yielded comparable fit in S-parameter simulation as that of those models. However, the representation of C_{ds} and C_{jd} was

left as in Model 2. A yet unexplainable problem caused severe fitting problems when these two components were represented by their voltage-controlled charge sources. But this is not going to significantly affect the quantification procedures since quantification will not involve these two capacitances, as discussed earlier in Section 4.2.2.

Convergence in harmonic balance simulations was much easier to obtain and extended to a much wider range of bias voltages with this model. This can be related to the number of operations the harmonic balance simulator has to perform in the representation of the capacitances. In models 1 and 2, the current in the capacitances was represented by $i = C(dV_c/dt)$ which meant the harmonic balance simulator had to constantly perform two operations, differentiation and multiplication, as the input power changed. With this representation, the integration described above was performed externally, in Matlab, and the charge source values were put in data tables for the model, which meant only one operation, differentiation, had to be constantly performed.

4.4.2 Linearising Model's non-Linear Parameters

This section discusses the methods used to obtain linear representations of the nonlinear parameters in the transistor model. Linearising nonlinear elements of the transistor implies that the resulting model's characteristics are not a real representation of the original transistor but of an assumed transistor where the nonlinear parameter is made linear. As will be shown, every effort is taken to try to ensure that the linearisation mechanism will have minimal effect on the fundamental output of the transistor but will affect more its nonlinear performance. These linear representations of the model parameters will be used in the quantification process by switching the representation of the model's parameters between their linear and nonlinear representations.

4.4.2.1 Linearising the Transconductance, g_m

The transconductance, g_m , is essentially a voltage-controlled current source which describes the change in the small signal drain current with respect to the small signal gate-to-source voltage. Therefore, when considering the linearisation of the transconductance, two issues must be kept in mind.

- In order to correctly model g_m , linearly or nonlinearly, variance has to be present. Therefore, replacing the SDD representing the current characteristics with a current source at single value (that of the respective operating bias point) or inserting the respective current value in the SDD as a single number are not valid methods of linearisation, because no change is presented, and hence g_m is not modelled.

- Also, it is well known that a voltage-controlled current source produces harmonics only if its current nonlinearly varies with its controlling voltage. If the current linearly varies with the controlling voltage, no harmonics are produced. The situation is different with nonlinear capacitances, as will be discussed.

In light of these two realities, and bearing in mind that the linearisation should not affect the fundamental output of the transistor but only its nonlinear distortion, there are two options for linearising the model's transconductance: Either by using a voltage controlled current source (such as the VCCS component from the ADS library) and setting its g_m to its value at the respective operating bias point, or by linearising the current characteristics' data table. These two options can be discussed as follows:

▪ **Single- g_m VCCS**

This method satisfies the variance condition in modelling the transconductance since g_m itself describes the change in i_{DS} with respect to v_{GS} . Also, since there is only one value of g_m at any input power, this change is linear. However, while a good S-parameter match can be obtained with this (since this value of g_m will be valid around close proximity of the bias point due to the very small signal used in the S-parameter test), the problem with this method is at high input power. When the input power significantly changes, the value of g_m set in the VCCS will no longer be valid and hence the fundamental output of the model is not expected to match that of the original transistor. This is because the total input voltage will expect to see another g_m value from the table which no longer exists. This is demonstrated in Figure 4.28 which was produced by replacing the I_{g_m} table representing the transconductance in the model with a VCCS (from the ADS component library) set to the correct g_m value at the corresponding bias point. As can be seen, there is a good match between the fundamental output of the model and the transistor at low input power, but as the input power increases, the fundamental output of the model deviates further from the transistor's fundamental output.

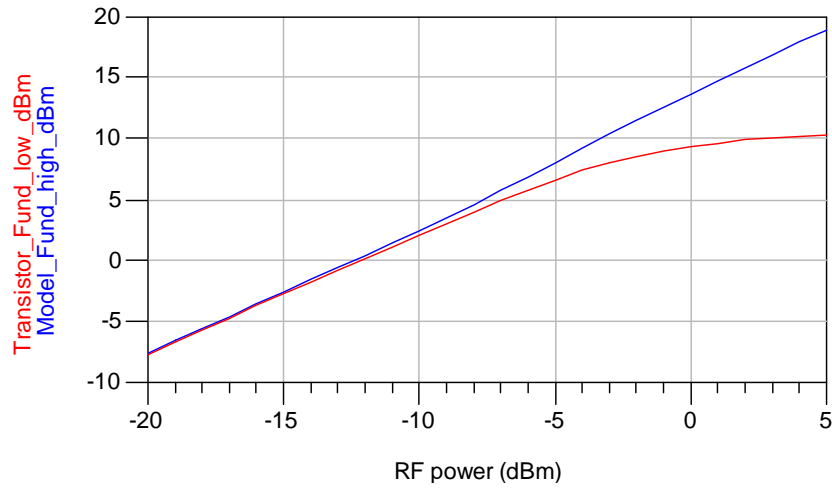


Figure 4.28 Error in the model's fundamental output when using single- g_m VCCS for linearisation

▪ Linearising the I_{g_m} current table

Linearising by changing the data in the I_{g_m} table works by changing the current level in each V_{GS} curve of the current characteristics in Figure 4.19 when the transistor is in saturation. This ensures variance of i_{DS} with v_{GS} and also, if properly done, ensures that this change is linear. This method overcomes the drawback of the VCCS method in that several characteristic curves at different levels will be presented from which g_m is to be extracted, hence g_m no longer has a single value. The strategy, as will be described in detail below, is to use the original nonlinear I_{g_m} data table to plot I_{g_m} vs. V_{GS} at the highest value of V_{DS} in the table, as shown in Figure 4.29, choose a V_{GS} bias point and plot a tangent of the $I_{g_m} - V_{GS}$ characteristics at this point. The new values of the current in saturation are then taken from the linear tangent line rather than from the original characteristic line, as will be seen.

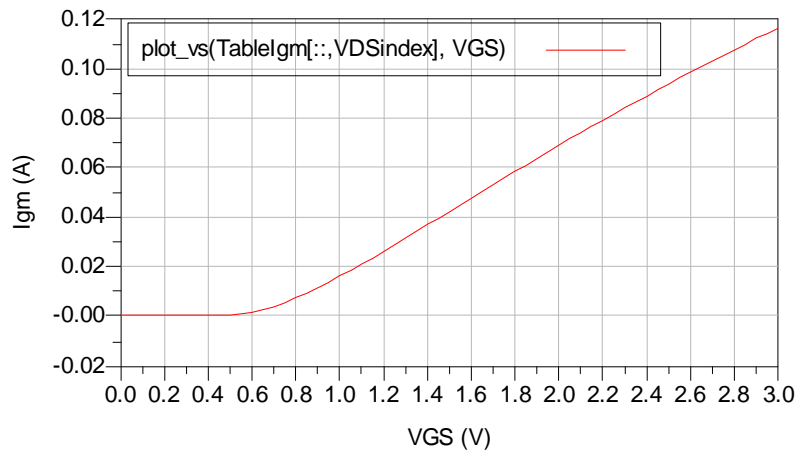


Figure 4.29 $I_{g_m} - V_{GS}$ characteristics obtained from nonlinear I_{g_m} table

Similar to the extraction of transistor parameters, this linearisation method was designed with full automation, adaptability and reusability in mind. The user can simply change two variables (the gate and drain bias points) and the new linearised current characteristics are automatically recalculated. Appendix A.3 details the data display window coding used to perform this procedure.

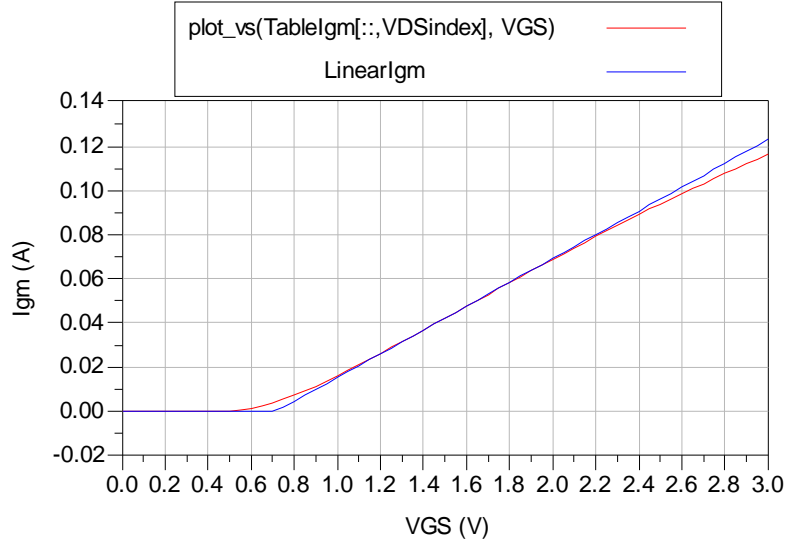


Figure 4.30 Nonlinear and linear I_{gm} - V_{GS} characteristics (in saturation)

The new I_{gm} - V_{GS} curve shown in Figure 4.31 assumes a different characteristics for a transistor that switches on suddenly at its exact V_T , not in the normal fashion where the transistor switches on gradually and subsequently the current through it increases. However, this will not affect the results of the quantification since I_{gm} levels are changed in every V_{GS} curve only after saturation which is far beyond this point. In Figure 4.31, a single V_{GS} curve is selected to demonstrate how the current is changed. With this value of V_{GS} , the transistor saturates at V_{DS} of 2.3V. From this point forward, the current is replaced by its new value, as explained above, and the V_{GS} curve is modified. These new V_{GS} curves form the new linearised I_{gm} data table used for quantification.

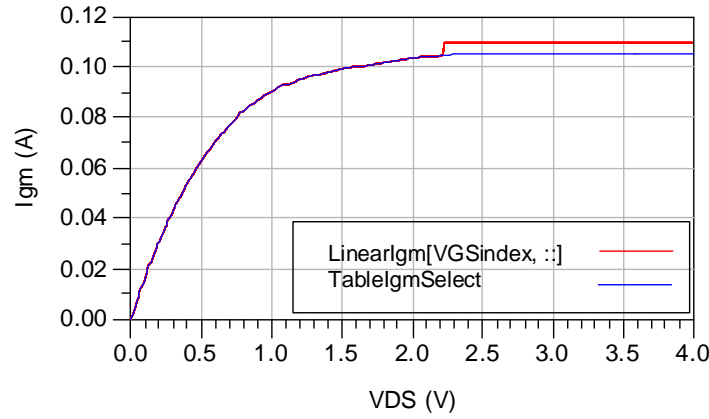


Figure 4.31 A sample V_{GS} curve in linear and nonlinear I_{gm} - V_{DS} characteristics

As demonstrated in Appendix A.3, the coded functions used in linearising the I_{gm} current ensure at every step the retaining of the dependency on the swept values of V_{GS} and V_{DS} , making the procedure accurately adaptable to any device and easily applicable by the designer for any accuracy required. The new linearised I_{gm} - V_{GS} characteristics, along with the original nonlinear characteristics are shown in Figure 4.32 for few V_{GS} curves to demonstrate the changes. Note that the actual work was done on V_{GS} steps of 0.05V for increased accuracy. Now the relationship between the current levels in saturation and their respective V_{GS} curves is linear.

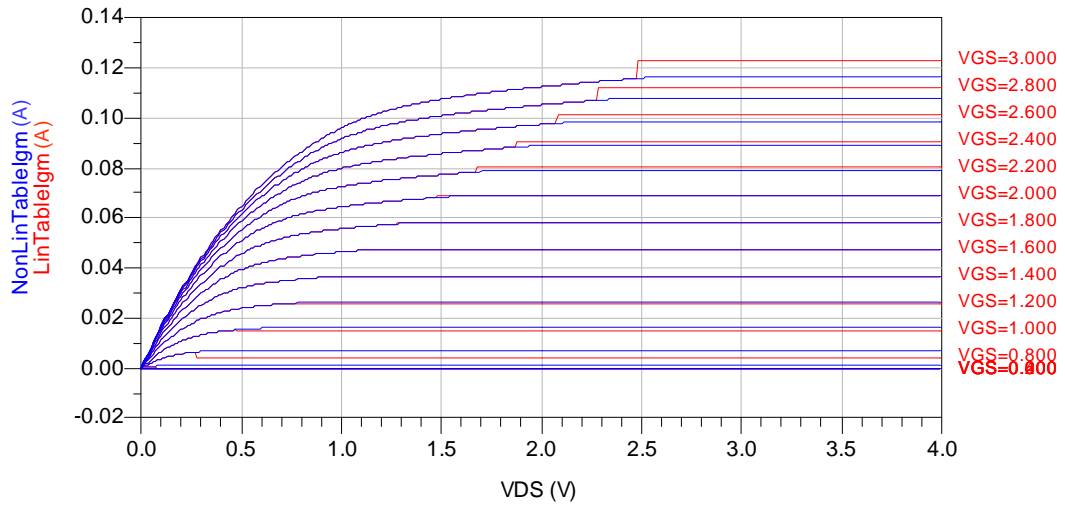


Figure 4.32 Linearised and nonlinear I_{gm} - V_{GS} characteristics

Figure 4.33, demonstrates the need for automating the linearisation process. This is a zoom of a linearised I_{gm} - V_{GS} characteristics similar to that of Figure 4.32 but produced when fewer V_{GS} curves are used (at steps of 0.25V), and with $V_T=0.4V$ (to ease observation of values in the figure). Thorough observation of the curves in Figure 4.33 reveals a serious problem. The above description of the linearisation procedure made clear that the current level of each V_{GS} curve is

to be changed from its saturation point forward, but only once. However, observing Figure 4.33 reveals that this is not the case. Taking the curve of $V_{GS}=2.9V$ as a simple example, the current level first changes when the transistor saturates at $V_{DS}=2.35V$, but then it changes again at $V_{DS}=2.6V$ which is the same point at which the upper curve ($V_{GS}=3V$) saturates and changes. The lower curves ($V_{GS}=2.8, 2.7$ and $2.6V$) follow the same route, each with all the curves above it. But this does not happen to the V_{GS} curve at $V_{GS}=2.5V$. This phenomenon is due to the interpolation. Since the simulated step is $0.25V$, the curves at $V_{GS}=2.5, 2.75$ and $3.0V$ have their data available in the table. What is between them is interpolated, and the observed error is due to the way the interpolation works. The straightforward solution to minimising this error is to use many more data points. As discussed earlier in this chapter, this can be very time consuming and may lead to errors, hence automation is needed.

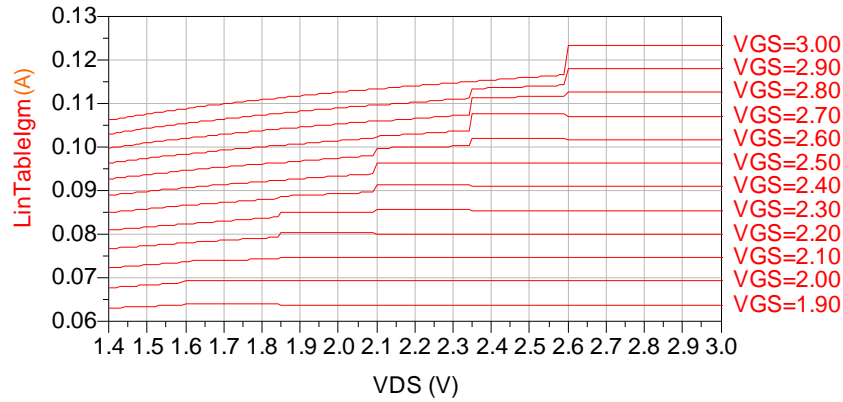


Figure 4.33 Interpolation error in linearising I_{gm} when only few sweep point are used

For most accurate results, the produced model with the linear I_{gm} should be biased at the same point at which the linear tangent line was drawn. With this, an S-parameter simulation produces a good fit to the original transistor. That is because the S-parameters' test signal is very small and hence will not change the operating point from that at which the linear tangent line was drawn and where the model is biased. However, when a large input signal spread over the linear line is applied, the fundamental is expected to be different because the linear line does not represent the original characteristics. Nevertheless, Figure 4.34 shows that this error is very inconsiderable. It shows that the error in the matching between the fundamental output component of the model and the transistor does not greatly differ when the linearised or the original I_{gm} table is used to model the transconductance. This is because the difference between the linear line and the original characteristics is insignificant as shown in Figure 4.30. But, as will be demonstrated in the next section this linearisation has an important effect on the 3rd-order intermodulation distortion (IMD_3) components of the model.

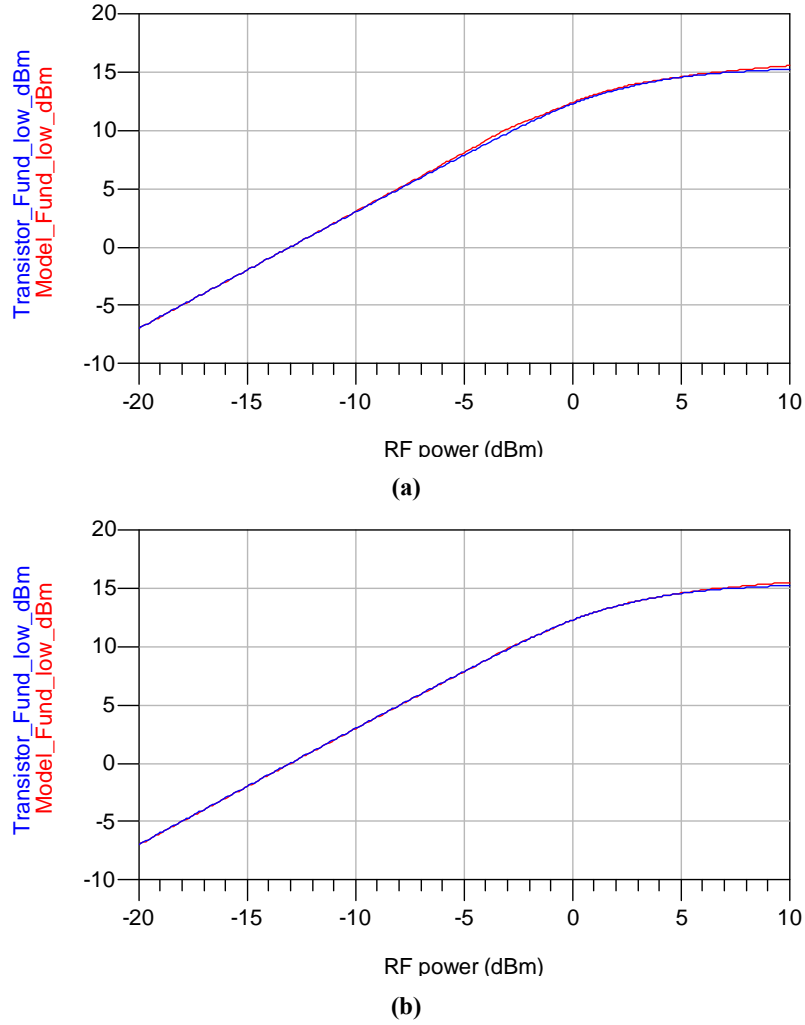


Figure 4.34 Fundamental output of (a) g_m -linearised model and (b) original nonlinear model

4.4.2.2 Linearising the Output Conductance

The output conductance, g_{ds} , is linearised by changing its representation to a voltage controlled resistor as described for Model 1 in Section 4.3.1.1, and using the corresponding value of g_{ds} extracted at the respective bias point where the quantification test is being performed.

4.4.2.3 Linearising the Capacitances

While the current of a voltage-controlled current source has to be nonlinearly variant with its controlling voltage in order for harmonics to be produced (no harmonics are produced if it is linearly variant), a voltage-controlled capacitance produces harmonics just if it is variant with its controlling voltage, linearly or nonlinearly. In light of this, to linearise a nonlinear capacitor, it has to have a single value at any applied external voltage. For the purpose of quantification in this work, this can be achieved in two ways:

- 1- By replacing the SDD representing the nonlinear capacitance with a single value capacitor from the library. This is an easy and straightforward option, but it lacks the benefits of representing capacitance as charge sources as discussed earlier.
- 2- If the capacitance is to be represented as a charge, non-variance can be achieved by changing the SDD's table, copying the charge value at the operating bias point to all other values of V_{GS} and V_{DS} in the table.

The first option was adopted in this work due to its simplicity and straightforwardness.

4.4.3 Quantification Results and Discussion

The bias point and the range of input powers used in quantifying the nonlinear distortion from the transistor's nonlinear elements are crucial to the validity of the quantification process, and its results. Broadly, it is recommended that very high input powers should not be used, that is due to three reasons. First, it is important that the transistor operates in the area where the I_{gm} - I_{gds} separation discussed in Section 4.3.2.1 is valid, i.e. the input power signal swing should not take the operation of the transistor to the linear region. Otherwise, it will enter the region where the representation of the transconductance and output conductance is combined. The choice of the bias point also helps avoiding this by biasing the model far from the linear region. Also, the operation of the transistor needs to be confined to the area where the interpolation error of the linearised I_{gm} (discussed in Section 4.4.2.1) does not come into effect. This can lead to inaccuracy and introduce errors. Finally, very high input powers will make the single values used for linearising the output conductance and capacitances no longer valid, as it will push the transistor far from its bias point and can corrupt the result. Also, the input power should not be very small since that makes the effect of interpolation (present in any data table as required by harmonic balance simulation) more significant. Bearing these points in mind; some analysis was done on the linearised model to study its reaction to very large and very small input powers. This analysis led to choosing to carry the quantification process at the moderate input power range of -25 to -5dBm

In Figure 4.35, some quantification results are presented to demonstrate the viability of the idea of quantifying the contribution to distortion from nonlinear elements through nonlinear modelling. These are based on switching on and off single or combinations of nonlinear parameters. An s-parameter test was done every time to verify the small signal operation of the modified model.

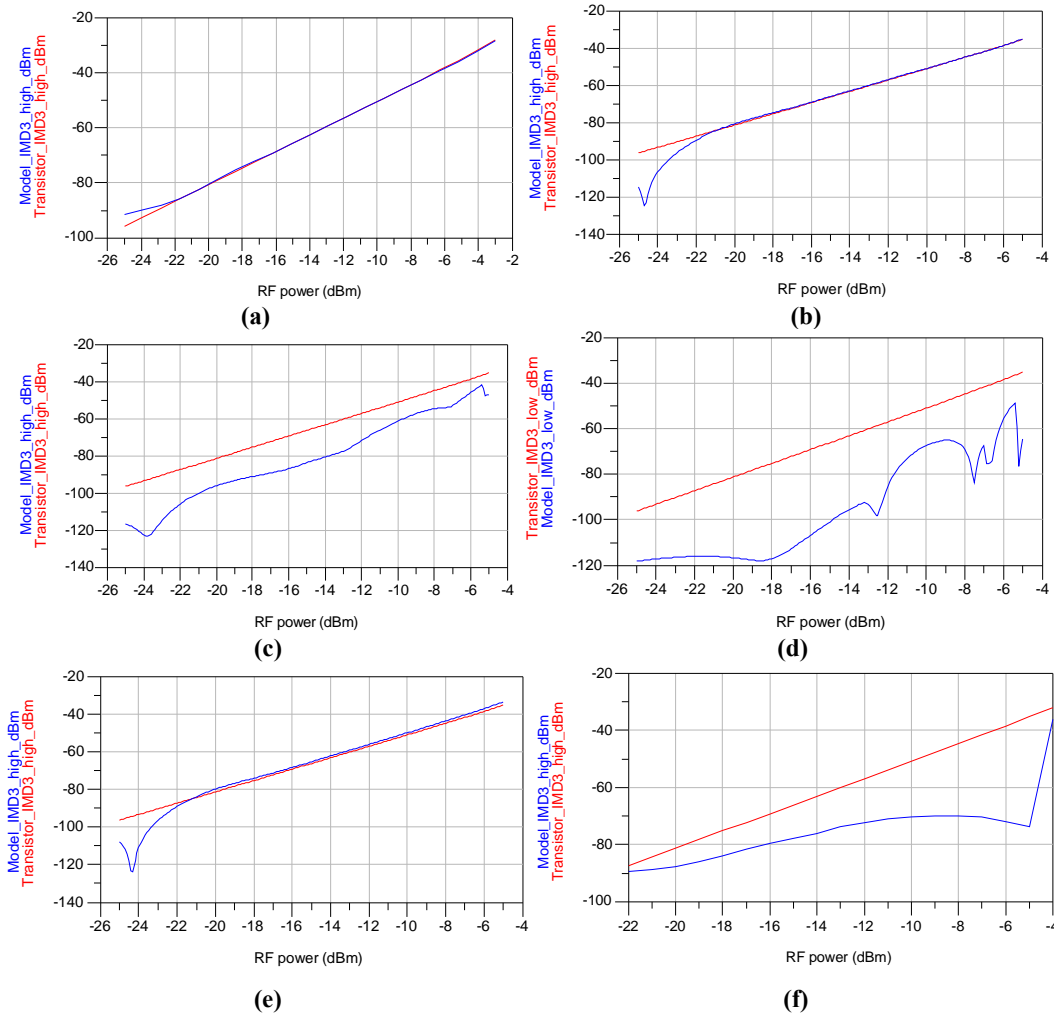


Figure 4.35 Results of quantification of nonlinear distortion (a) g_m , g_{ds} , C_{gs} and C_{gd} nonlinear. (b) g_m and g_{ds} nonlinear, C_{gs} and C_{gd} linear. (c) g_m , C_{gs} and C_{gd} linear, g_{ds} nonlinear. (d) g_m , g_{ds} , C_{gs} and C_{gd} linear. (e) g_m nonlinear, g_{ds} , C_{gs} and C_{gd} linear. (f) g_m , g_{ds} , C_{gs} linear, C_{gd} nonlinear.

In Figure 4.35.a, all elements are nonlinear. g_m and g_{ds} are represented by their nonlinear data tables, while C_{gs} and C_{gd} are represented by their charge source tables. As expected, almost identical IMD₃ product over the simulated input power range is produced from the transistor and the model. In Figure 4.35.b, g_m and g_{ds} are nonlinear but C_{gs} and C_{gd} are linearised by replacing each table with a capacitor, setting it to the corresponding value extracted at that bias point, following the method proposed in Section 4.4.2.3. Note that there is almost no difference between the IMD₃ product of the transistor and the model indicating that the contribution from these capacitances to the nonlinear distortion is minimal. Therefore, in the results of Figure 4.35.c to Figure 4.35.e, the representation of the capacitances is left linear to study the effect of a linear or nonlinear g_m and g_{ds} on the total IMD₃ distortion. In Figure 4.35.c, g_m is linearised using the technique presented in Section 4.4.2.1 but g_{ds} is still represented by its nonlinear current table. A reduction in IMD₃ of about 15dBm can be observed. In Figure 4.35.d, both g_m

and g_{ds} are linearised. g_{ds} is linearised according to the method described in Section 4.4.2.2. Note that in Figure 4.35.d the reduction in IMD_3 is much greater than in Figure 4.35.c reaching 45Bm at maximum. However, in Figure 4.35.e when g_m is nonlinear and g_{ds} is linear, no reduction in IMD_3 is observed. The results in Figure 4.35.c and Figure 4.35.d in comparison with Figure 4.35.e present an interesting observation. While the contribution of g_{ds} nonlinearity is obvious when comparing Figure 4.35.c and Figure 4.35.d, Figure 4.35.e suggests that this contribution is only significant when g_m is linear. When g_m is nonlinear, its nonlinearity dominates and the effect of a linear g_{ds} on reducing IMD_3 is insignificant. A similar observation can be made from the result in Figure 4.35.f. In this test, all nonlinear elements are set as they were for the result in Figure 4.35.d except C_{gd} which is made nonlinear using its voltage-controlled charge sources table. Note that even though Figure 4.35.b shows no reduction in IMD_3 as a result of linearising C_{gd} , comparing the results in Figure 4.35.f with that of Figure 4.35.d reveals that a notable increase in IMD_3 resulted from making C_{gd} nonlinear.

These quantification results reveal the validity of the process of quantifying nonlinear distortion through transistor nonlinear modelling where each nonlinear element is independently represented. They also represent the effectiveness of the linearisation techniques proposed. However, these results also show that there is a strong correlation between the nonlinearities in the transistor. This raises the necessity for improving this method by maintaining separate representation of each nonlinear element while also accounting for the correlation between these nonlinearities.

4.5 Summary

This chapter presented a new concept in table-based nonlinear modelling where each nonlinear parameter was individually represented. It also introduced a set of automation procedures aimed at reducing the time taken for extracting the model parameters and organising them in relevant tables. The chapter also discussed and presented results for quantifying the contribution to distortion from various nonlinear elements in the model, based on a proposed superposition method.

The chapter first started by discussing and examining several CMOS transistor models proposed in the literature for suitability to be used with the UMC 180nm RFCMOS transistor. It then presented a set of procedures aimed at automating the extraction of the model parameters at several bias points, which was necessary for forming the nonlinear model. The automation was based on embedding parameter extraction equations in coded functions in the ADS data display

window and organising them using sweep index manipulations to maintain dependence on the swept gate and drain voltages, in such a way that all parameters are automatically recalculated as the terminal voltages change.

Following that, three different models were presented. The difference between them was in how the transistor nonlinear parameters were modelled. In the first model, the transconductance was represented as a voltage controlled conductance. The output conductance was converted and represented as a voltage controlled resistance. The current in both the transconductance and output conductance was modelled as a voltage-controlled current source. Model capacitances were represented as a voltage-controlled capacitance. Their current was modelled using their current differential equation. This model produced good fit in S-parameter simulations but was difficult to converge in harmonic balance simulations. The second model represented the capacitances as in the first model, but changed the representation of the transconductance and output conductance to be extracted from the transistor current characteristics. For this, a method was developed to separately define the current due to the transconductance and that due to the output conductance in the saturation region. That is in order to make the model suitable for nonlinear quantification from these two different sources of nonlinearity. This method was also automated in order to enhance accuracy by enabling the use of many data points. The resultant model converged better in harmonic balance simulations than the first model. In order to enhance convergence further, and make the model more robust in harmonic balance simulations, a third model was developed in which capacitances were converted to charge sources. The significance of identifying the exact potential across the capacitance for this purpose was demonstrated.

Subsequently, quantification of the nonlinear contribution from various nonlinear elements was discussed by first discussing methods of linearising the characteristics of the nonlinear elements. An automated method used to linearise the characteristics of the transconductance current, the biggest distortion contributor was presented and discussed. After applying the super position method, some quantification results were presented to demonstrate the validity of quantifying nonlinear distortion through nonlinear modelling. The discussion of the results highlighted how this method may be improved by including the correlation between various nonlinearities.

Chapter 5 LOW POWER, LOW NOISE AND POWER AMPLIFIERS: DESIGN AND OPTIMISATION

5.1 Introduction

In Chapter 2, the importance of the Low Noise Amplifier (LNA) as the first component in the receiver chain was discussed. This chapter looks more closely at low power LNAs in terms of their design and optimisation. It proposes a new design methodology for power-constrained LNAs and investigates the implications of increasing the input power of a commonly used LNA architecture in an attempt to make the amplifier capable of performing Power Amplifier (PA) functionality for low power standards as well as the low noise amplification.

First, a review of the available LNA topologies is presented and reasons are stated for choosing one for this dual functionality LNA/PA design. A review of current LNA design approaches is then provided and the limitations of these approaches are highlighted. This is followed by examination of some results of recently published LNA designs and discussion of the techniques employed to reach these performance levels. These topics are covered in Section 5.2.

Due to their complexity, no variable gain techniques have been considered for the amplifier design, instead, the proposed design methodology presents a route through which a designer can obtain the best linear performance through optimisation and employing the most suitable linearisation techniques available for a given technology under the constraint of power consumption. For this, the linearity of amplifiers is discussed in Section 5.3 with particular emphasis on the nonlinear performance and linearisation techniques suitable for an LNA intended for low power applications.

Section 5.4 presents a unique and comprehensive investigation of how amplifier negative feedback affects the production of distortion components. The analysis, which is based on mathematical substitutions of harmonic and intermodulation products of several orders of nonlinearity, looks at how the negative feedback of particular components produced from single-order nonlinearities can produce intermodulation products if mixed with the fundamental

frequency and placed under another order of nonlinearity. Results reveal that previously unknown components can cause significant contributions to 3rd-, 5th- and 7th-order intermodulation products.

The proposed LNA design methodology is presented in Section 5.5. This begins with an outline of the novelty of the methodology, and how it can aid the designer through the design process. Following that, individual aspects of the design methodology are described in more detail. The section concludes by demonstrating the viability of the approach using a design example.

5.2 Low Noise Amplifier Architectures and their Design

The design strategy of the LNA/PA amplifier is to select a suitable LNA topology on which the design would be based, and then study ways of increasing the dynamic range of the amplifier through linearisation and improved design methodology. This section provides a targeted review of LNA architectures and LNA design approaches to give clear insight into the issues involved in such an approach.

5.2.1 LNA Architectures

This section discusses amplifier topologies for consideration in the LNA/PA design. Available choices are examined first, followed by discussion of the reasons for choosing the cascode inductor-degenerated architecture.

5.2.1.1 Topology Choices

In order to consider existing LNA topologies, the three basic single-transistor amplifier stages; the common-drain, the common-gate and the common-source, are first discussed as possible candidates of the first stage of the amplifier. In addition to the gain that each stage can provide, there are two other important issues to consider; namely, how the stage can be matched to the antenna, and the likely noise performance.

▪ Common Drain

A simplified common-drain amplifier is shown in Figure 5.1. A common-drain amplifier has its input from its gate and its output taken from its source. The common-drain amplifier provides less than unity voltage gain with low output impedance. Because the output current (at the source) is very large in comparison to the input current, the common-drain amplifier can achieve high power gain, sufficient to reduce the significance of the noise figure of subsequent stages in the receiver system as discussed in Section 2.5.1. However, the severe drawback of this topology is that a large input voltage swing is required to turn on the transistor, hence a

push pull stage would typically be used for both half cycles. The requirement of high voltage swing at the input is an option that does not exist for an LNA since the incoming signal is usually very weak. Therefore, the common-drain configuration is eliminated as a possible option for the first stage of an LNA design.

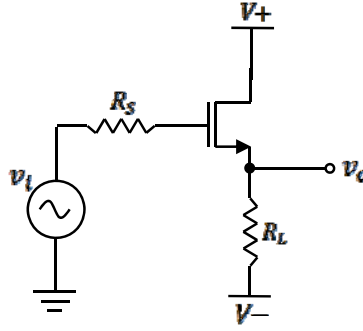


Figure 5.1 A typical common-drain amplifier

Common Gate

A simple common-gate amplifier is shown in Figure 5.2. A common-gate amplifier has its input from the source and its output taken from the drain.

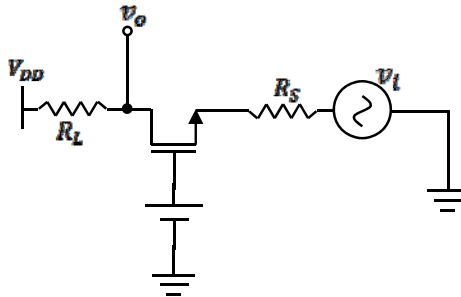


Figure 5.2 A typical common-gate amplifier

A known disadvantage of a common-gate amplifier is its low input impedance. The input impedance of a common-gate stage is typically $1/g_m$ where g_m is the transconductance of the MOSFET. For impedance matching, this is required to be equal to the source impedance which is normally 50 Ohms for an LNA following a typical antenna. This matching can be very difficult to attain. Additionally, while g_m of a MOSFET depends on its dimension and biasing, the noise figure (NF) for this stage can be expressed as [Ge 1998].

$$NF = 1 + \frac{\gamma g_{do} + \delta g_g}{g_m} \quad (5.1)$$

g_{do} , g_g , γ and δ have been defined in Section 2.5.2.

If g_m is fixed for a given power consumption (i.e. biasing + dimensions), then the noise factor cannot be minimised. However, since $g_m = 2K(V_{GS} - V_T)$, V_{GS} can be varied to alter g_m and

optimise the noise figure. But because lower noise figure needs higher g_m according to (5.1) and higher g_m requires higher V_{GS} , this will lead to an increase in power consumption.

Furthermore, assuming matched conditions, a common-gate stage with its $1/g_m$ input termination results in a minimum noise factor for the MOS device given by (5.2) [Shaeffer 1997]

$$NF = 1 + \frac{\gamma}{\alpha} \quad (5.2)$$

where $\alpha = g_m/g_{d0}$. In short channel devices, due to short channel effects $\alpha \leq 1$ and due to excess thermal noise from hot electrons $\gamma \geq 2/3$ ($\alpha = 1$ and $\gamma = 2/3$ for long channel devices). This yields a minimum theoretically achievable noise figure of around 3dB, which will be higher in practice. This noise factor is high for an LNA if the sensitivity of the receiver is required to be -80dBm as is the case in the ZigBee standard. Thus, implementing a common-gate configuration as the first stage of an LNA is not recommended.

▪ Common Source

A simplified common-source amplifier is shown in Figure 5.3. A common-source amplifier has its input at the gate with its output taken from the drain. The common-source amplifier has high voltage gain and high input impedance. There are several techniques to match a common-source amplifier to the input impedance.

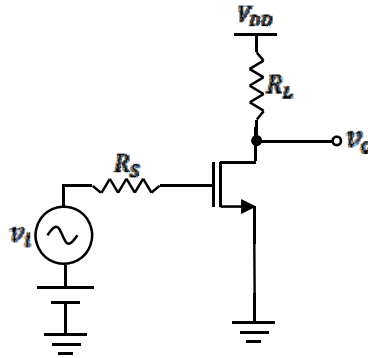


Figure 5.3 A typical common-source amplifier

The first employs a resistive termination of the input port as shown in Figure 5.4. Although this is a convenient and easy way to provide 50Ω real impedance, the use of real resistors in this fashion has a severe effect on the amplifier's noise figure. In [Shaeffer 1997], it was found that terminating the input of the LNA in this way can double the noise figure of the amplifier. This was due to the thermal noise contribution of the resistor as well as the effect of the input signal being attenuated by the resistor.

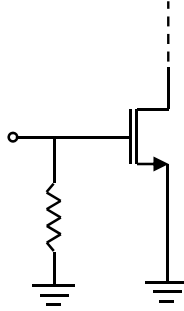


Figure 5.4 Resistive termination for input matching of common-source stage

Another approach for obtaining a 50Ω input termination to match a common-source transistor is shown in Figure 5.5. Here, a shunt and a series feedback are used to set the input and output impedances of the LNA. This approach was implemented in [Benton 1992], [Sheng 1991] and [Kobayashi 1994]. From the comparison in [Shaeffer 1997], it is shown that amplifiers using this technique are usually broadband amplifiers and their power consumption is much higher than other amplifiers with similar performance.

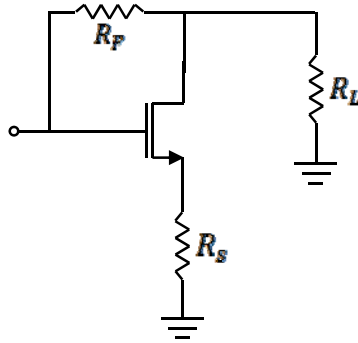


Figure 5.5 Input/output termination using shunt and series feedback

A more convenient technique for matching common-source transistors that has widely been used in LNA circuits in recent years is to use inductive source degeneration as shown in Figure 5.6 (see for example [Chiu 2005] and [Liao 2003]).

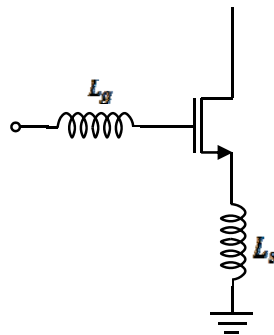


Figure 5.6 Common-source with inductive degeneration for

In this configuration, the input impedance, Z_{in} is given by (5.3) [Zhang 2003]

$$Z_{in} = j\omega(L_s + L_g) + \frac{1}{j\omega C_{gs}} + R_g + \frac{g_m}{C_{gs}} L_s \quad (5.3)$$

where C_{gs} is the gate to source capacitance of the transistor, L_g and L_s are the inductance at the gate and the source, respectively, and R_g is the effective gate resistance including the series parasitic resistance of L_g .

At resonance, $\omega_o = 1/(\sqrt{(L_g + L_s)C_{gs}})$ and Z_{in} becomes a real impedance given by Equation (5.4)

$$Z_{in} = (g_m/C_{gs})L_s \approx \omega_T L_s \quad (5.4)$$

where ω_T the unity gain frequency of the transistor which is equal to g_m/C_{gs} . With this arrangement, L_s can be adjusted so that $\omega_T L_s$ equals 50Ω , $L_g + L_s$ then resonate with C_{gs} at the operating frequency. In this way, a 50Ω input impedance can be produced without the need for any real resistance, hence avoiding any deterioration in the amplifier noise figure. In reality however, since the Q factor of the integrated spiral inductor (if indeed an integrated spiral inductor was used) is normally low due to technology limitations, there will be some effect on the noise figure due to the series resistance of the inductors. For this reason, L_g is required to be as small as possible. Under power matched conditions at the operating frequency, the input reactance is zero while the input resistance equals the source resistance R_s , then L_s and L_g can be determined by [Zhang 2003].

$$L_s = \frac{(R_s - R_g)C_{gs}}{g_m} \text{ and } L_g = \frac{1}{\omega_o^2 C_{gs}} - L_s \quad (5.5)$$

From Equation (5.5), if L_g is to be small, C_{gs} has to be increased and to achieve higher C_{gs} the width of the transistor has to be increased (adding more capacitances in parallel). But although this will result in lower noise figure when the transistor is matched, increasing the width will lead to higher power consumption, hence a trade-off has to be found. Nevertheless, a common-source stage is probably the most appropriate for consideration as the first stage in low-power LNA design. Having said that, there are important issues of concern with this topology that have to be addressed; as will be seen in Section 5.5.2.

Bearing in mind the properties of the single-stage amplifiers studied above in terms of matching and noise performance, there exist two main low noise amplifier architectures in use today.

▪ The Common-Source Common-Gate Cascode

The common-source common-gate low noise amplifier topology has been, and still is by far, the most common LNA topology in use (see for example [Chen 2006a], [Vidojkovic 2004], [Ock 2002]). It can be shown that the gain of a single-stage common-source amplifier is $A = -g_m Z_{load}$ which is the same as for a common-gate stage but with the 180° phase shift. However, in a conventional common-source stage, C_{gd} provides a feedback path between the input and output which degrades reverse isolation and therefore affects the amplifier stability and reduces the bandwidth. In a common-gate amplifier, this problem does not exist since C_{gd} is connected to ground.

To exploit the advantages of both the common-source and common-gate topologies, a cascode design with a common-source first stage and a common-gate second stage shown in Figure 5.7 is worth investigating. In this configuration, C_{gd} of transistor $M1$ is no longer a feedback path since it is no longer located between the amplifier input and output, therefore reverse isolation and stability can be improved.

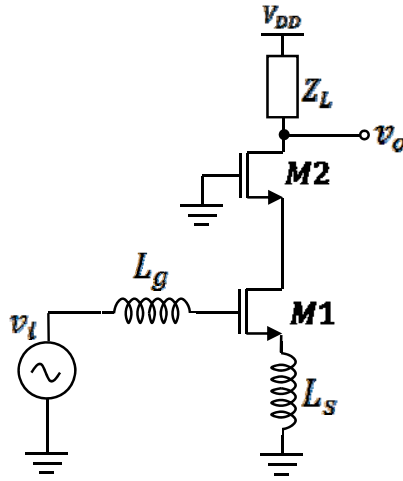


Figure 5.7 The common-source common-gate cascode LNA topology (with inductive degeneration)

It can be shown by a simple mathematical analysis of the circuit in Figure 5.7 that if the transconductance of the common-source and the common-gate transistors were equal, then the voltage gain of this amplifier topology will be that of the common-source stage alone. Also in that case, the gain of the common-source stage will be unity (due to the low input impedance of the common-gate stage). Therefore, the common-source common-gate cascode can also help to eliminate the large capacitance appearing at the gate of a single common-source stage (according to the Miller theorem) due to the parasitic capacitance between the gate and drain, C_{gd} .

An important advantage of this topology is that it allows independent optimisation of noise and linearity by optimising the common-source and common-gate transistors, respectively. This will be discussed in more detail in the next section. One disadvantage of this topology is that the transconductance of the cascode amplifier is limited by the transconductance of the common-source stage only [Amor 2006]. Another disadvantage is that the high frequency noise and gain can be significantly degraded by several parasitics at the node between the two cascode transistors [Cha 2000].

▪ The Common-Source Common-Source Cascade

Another topology often used in LNA design is the common-source common-source topology shown in Figure 5.8 (see for example [Chen 2006b], [Cha 2003] and [Cha 2000]). In this topology, the gain is boosted through the multiplication of the transconductance of the two cascading stages [Amor 2006]. An alternative name for this topology is the “current reuse topology”. This comes from the fact that both the two stacked common-source transistors are using the same current in order to reduce power consumption. In the topology in Figure 5.8, C_1 is a coupling capacitor to pass the output of $M1$ at its drain to the input of $M2$ at its gate and L_1 is an RF choke to block the AC coupling between the drain of $M1$ and the source of $M2$ (which is bypassed by C_2) while passing the DC current between the transistors.

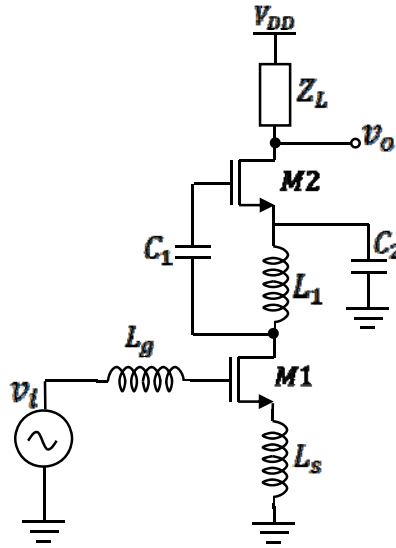


Figure 5.8 The current reuse LNA topology

One form of the current reuse topology is the inductively degenerated NMOS and PMOS pair in shunt configuration (see for example [Fouad 2002] and [Gatta 2001a]), but the disadvantage of this method is that the linearity is of concern since the mobility of the NMOS and the PMOS devices is different [Liao 2003]. Transistor mobility has a significant effect on its nonlinearity since mobility degradation is dependant on the gate to source voltage, V_{GS} .

▪ Other topologies

For the sake of completeness, it is important to mention some of the other, less popular LNA topologies which have been reported. Fully differential LNAs have been proposed (see for example [Alam 2006] and [Zencir 2003]). The advantage of this topology is that it can mitigate the effects of common mode noise and clock feed-through [Liao 2003], and reduce even-order nonlinearities. The disadvantage of this topology however, is the increased power consumption and complexity.

5.2.1.2 Topology Selection

The most important factor for the selection of an amplifier topology in this work is the fact that the design strategy is to start with an LNA and then increase its power handling capability to enable its use as a power amplifier. The feasibility of the common-source common-gate cascode in which the noise and linearity can be independently optimised was a strong motivating factor in the final choice of this topology for this work.

In [Guo 2002], it was found that the noise figure contribution from the second stage (transistor M_2 in Figure 5.7) can be approximated to

$$NF_{M_2} = NF_{M_1} \left(1 + \frac{\omega_T L_s}{R_s} \right)^2 \left(\frac{\omega_o}{\omega_T} \right) \quad (5.6)$$

where R_s is the source resistance and L_s is the degeneration inductance. Because in practice $\omega_o \ll \omega_T$, the noise contribution of M_2 is much less than the noise contribution of M_1 .

Linearity of the cascode was analysed in [Guo 2002] by calculating the input-referred 3rd-order intercept point (IIP_3) of the cascode topology using two 3rd-order Taylor series expansions to represent the nonlinearities of each of the transistors. Using this approach, the IIP_3 of the cascode amplifier was expressed as

$$\frac{1}{IIP_3^2} = \frac{1}{IIP_{3_1}^2} + \frac{3\alpha_2\beta_2}{2\beta_1} + \frac{\alpha_1^2}{IIP_{3_2}^2} \quad (5.7)$$

where IIP_3 is the overall input-referred 3rd-order intercept point, IIP_{3_1} and IIP_{3_2} are the input-referred third-order intercept points of the first and second stage, respectively. α_1 and α_2 are the first and second coefficients of Taylor series describing the nonlinearity of the first stage. β_1 and β_2 are defined similarly for the second stage. [Guo 2002] explained that inspection of Equation (5.7) shows that if α_1 is greater than unity then the second stage plays a more important role in determining the amplifier IIP_3 than the first stage.

The authors of [Guo 2002] concluded that this means that with the common-source common-gate topology, the low noise and high linearity performance can be achieved by the separate optimisation of $M1$ and $M2$ with almost no tradeoffs. This conclusion was also supported by work reported in [Park 2001] and [Razavi 1997a]. The extent of this result will be discussed and studied in more details in Section 5.5 where it will be shown that although this result is valid and very useful, alternative non mathematical expression-based analysis methods reveal that it has serious limitations.

Nevertheless, this result represents an extremely important feature for this work. If the intended amplifier is to work as both an LNA and a PA, then $M1$ can be optimised for the noise performance required for the LNA (as discussed in Section 2.5.1) while $M2$ is separately optimised for the linearity required by the PA (as discussed in Section 2.5.2).

In conclusion, and due to the reasons above, the common-source (CS), common-gate (CG) cascode topology with inductive degeneration was the one chosen for this work. More analysis of this amplifier architecture will be shown in Section 5.5 where the design of the amplifier is explained. Some modifications on this architecture have been proposed in the literature. These and some of the recent results obtained will be discussed in Section 5.2.3.

5.2.2 LNA Design Approaches and Optimisation Techniques

RF designers usually need a methodology to handle the complex relationship between design and performance parameters that often exists in analogue circuits. This section handles LNA design approaches and optimisation techniques by first looking at the nature and reasons for diversity of design approaches and how they can be assessed, then reviewing some of the recently proposed LNA design approaches, both general and application-specific. The section concludes by a discussion of the limitations of the currently available design approaches.

5.2.2.1 Diversity and Nature of Design Approaches

Several design approaches for high performance and low power LNAs have been proposed in the literature, (see for example [Amor 2006], [Egels 2004], [Lee 2004] and [Rafla 1999]). Design approaches usually have a starting point, and a path to completion. Commonly, the starting point greatly differs from one approach to another. This diversity can be due to two main reasons:

- The first is that different applications have different priorities on the performance requirements of the LNA. If the requirement on one performance parameter is more stringent than others, the designer tries to guarantee achieving that first.

- If there is no exceptionally stringent requirement on one performance parameter over others, the design approach becomes broader. This motivates the second reason for diversity, the judgement of the designer on the parameters which negatively affect the least number of performance parameters; usually these are set first to reduce the scope of possibilities. This comes from the fact that in RF design, in general, there is usually a very complex dependency of virtually every design parameter on almost every performance metric.

Once the starting point has been identified, the design procedure depends on an assessment of how influential various design parameters and their combinations are on performance parameters. Although there are general rules for every circuit to govern these relationships, there is no absolute right or absolute wrong in making these judgements. Instead, design approaches can be assessed on their nature, where there are three points:

- 1- The accuracy of the design methodology. Considerations here include the parameters that were neglected to simplify design equations or the level of complexity of the component model used. For example, [Lee 2004] used a simplified MOSFET model eliminating C_{gd} , while [Rafla 1999] used a simplified inductor model containing only the inductance, series resistance and capacitance to substrate.
- 2- The flexibility and reuse of the design methodology. This is where the design methodology is flexible enough to allow the designer to go back to any point of the design flow and choose other design parameters if different performance parameters are required for another design target. The ultimate design approach then is the one that can analyse all parameters simultaneously [Tulunay 2004].
- 3- The degree to which the methodology informs the designer. That is, how much does the approach inform the designer of available design options and their relative performance metrics and tradeoffs.

5.2.2.2 Design Approaches of Inductively Degenerated Cascode LNAs

Shaffer and Lee [Shaeffer 1997] reported a detailed examination of the inductively degenerated cascode LNA, which to date represents the most comprehensive analysis of this architecture in the literature. Also in this paper, they proposed an optimisation strategy based on adjustment of the gate-source overdrive voltage to achieve a suitable trade-off between noise, gain and power consumption.

However, Shaffer and Lee ignored the effect of the series resistance of the gate inductance on the performance of the amplifier by assuming a high Q off-chip inductor. This later became the focus of many of the optimisation methods which were based on circuits adopting integrated gate inductors (see for example [Andreani 2001], [Rafla 1999] and [Sun 2006]). The influence of an integrated gate inductance with low Q is particularly important at high frequencies as the loss due to its parasitic resistance increases, thus contributing to an increased noise figure [Lee 2004]. In these works, the analysis is commonly centred on the relationship between the achievable noise figure, gain and current consumption with circuit elements at the input of the amplifier including, in addition to L_g and its series resistance, the transistor width, its C_{gs} , and the degeneration inductance, L_s . The basis of this comes from the result of Equation (5.5) which introduces an important trade-off challenge arising from the fact that in order to reduce L_g (hence reducing its parasitic resistance), C_{gs} has to be increased and because C_{gs} can only be increased by increasing the width of the transistor, the current consumption will increase along with the device gain.

5.2.2.3 Purpose-Specific Optimisation Methods

Some published literature has discussed the optimisation of cascode low noise amplifiers but with focus on the relationship between two or more specific performance parameters, most commonly, noise and linearity. This is of particular interest to this work since the successful optimisation of noise and linearity is the single most important factor that is going to make the combined low noise and power amplifications possible.

In [Guo 2002], the authors studied the simultaneous optimisation of noise and linearity in a cascode LNA. In this work, the noise was analysed using a cascoded noise model for the two transistors in the architecture. This study, however, was based on the Taylor series expansion which only takes into consideration the transconductance nonlinearity of the transistors. But a more serious deficiency of this work is that the simulation carried out to assess the theoretical results only analysed the noise and IIP₃ performance subject to the gate widths of the two devices. No consideration or mention was given to the current and the gate overdrive present at each transistor, which, as will be seen in Section 5.5, has a great effect on both noise and nonlinear performance, and indeed the power consumption.

The work in [Park 2001] proposed a different route to the simultaneous noise and linearity optimisation. The approach was based on optimising the first stage for noise performance and the second stage for linearity, but separately. The design was carried out by selecting a suitable gate voltage for the first stage to constrain the current and then finding the optimum transistor size for noise performance. Separately, the IIP₃ of differently sized transistors at a range of gate

voltages for the second stage was studied. The problem with this approach is that it fails to take into account the cross-dependency of the two stages. For instance, the voltage at the node between the two stages will be influenced by the bias conditions in each stage, and this will affect the noise performance separately optimised at the first stage and the nonlinearity separately optimised at the second stage. Also, although there is overwhelming dominance of the CS transistor in controlling the DC current of the amplifier, there are regions, as will be seen in Section 5.5, where the effect of CG cannot be ignored. Having that, guaranteeing that the amplifier will perform as expected when the two separately-optimised stages are cascoded will either be impossible or will involve performing many iterations back and forward between the two separate optimisations of the first and second stage.

5.2.2.4 Limitations of Current Design Approaches

In addition to the limitation of separately optimising the CS and CG stages which often requires repeated iterations in the design process, as in [Roy 2006], there are three limiting factors that are common between all the design approaches discussed in the previous sections. These can be summarised as follows:

- 1- They are all based on theoretical mathematical expression-based modelling of the devices used and the circuit topologies. However, this approach is not conclusive, for a number of reasons. First, the theoretical mathematical expression-based analysis often involves several assumptions to both simplify the models by ignoring supposedly insignificant parasitics (see for example [Govind 2004], [Chen 2002] and [Andreani 2001]) or simplification of the equations by removing supposed insignificant terms (see for example the design methods in [Gatta 2001b], [Goo 2002], [Zhenying 2003], [Nguyen 2004], [Sivonen 2005], [Sun 2006] and [Asgaran 2007]). While this step is necessary since manipulation can be impossible if the full models and equations are considered, the insignificance judgement is left to the designer and hence is very variable and can be unreliable. This reduces the accuracy of the results obtained with the design methodology [Ragheb 2007]. Secondly, the modelling equations often have a narrow spectrum, meaning, although they show how parameters are related, they are difficult to interpret in terms of how this relationship will change if one of the parameters was confined to certain limits of values. This reduces the amount of information that the design approach can give, and can lead to an incorrect identification of the starting point.

- 2- Because of the limited scope of the mathematical equations, most of these mathematical expression-based design optimisation methods lead to only a starting point for each design parameter. Designers then have to run an automatic optimisation routine in a commercial simulator specifying performance requirements and letting the simulator find the ultimate optimum point (see for example [Li 1999], [Lavasani 2003] and [Lapuyade 2006]). But this can lead to inaccuracies since, in addition to the possibility that the starting point resulting from the simplified equation is not accurate, the optimiser which is itself based on mathematical methods might not be able to find the optimum solution starting from the initial point with the number of iterations specified. Increasing the number of iterations requires more memory power and can be very time consuming, adding pressure on time-to-market demands.
- 3- Finally, but most importantly, in these design approaches, the designer will not be made aware of what performance parameters were traded in the “optimisation” and therefore cannot make a quick decision on how to make amendments to the design to change its performance if needed. This reduces the flexibility of the design method.

The design approach presented in Section 5.5 takes into account these limitations. In addition to considering power added efficiency for the power amplifier functionality, the noise and linearity under power constraint optimisation is more conclusive in that it simultaneously analyses graphically the effect on all performance parameters from all related design parameters. Thus, increasing flexibility and reducing the need for iterations. Moreover, it is based on simulations of near-real foundry design kit components rather than simplified mathematical expressions and simplified models.

5.2.3 State of the Art LNAs

A problem with amplifier design using CMOS technology is that it has low transconductance at high frequencies [Amor 2006]. The consequent high power dissipation required for high-gain and low-noise implementation of LNAs intended for portable applications is therefore highly undesirable. Here, some of the useful modifications that have been introduced to the cascode architecture to improve the performance of the LNA at high frequency while constraining power consumption are analysed. This section presents and discusses performance results of some of the recently published LNA designs based on the common-source common-gate cascode architecture, hereafter called the CS-CG architecture, with focus on CMOS, narrow-band implementations. These performance results are listed in Table 5.1.

Table 5.1 Results of recently reported LNAs (a) [Zhenying 2003], (b) [Zhang 2003], (c) [Asgaran 2006], (d) [Yang 2001], (e) [Kim 2004], (f) [Aparin 2005], (g) [Kim 2006a] and (h) [Huang 2006]

	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
Freq (GHz)	2.4		5.7		0.9	0.9	2	2.14
NF (dB)	3.8	2.5	3.4	2.4	2.85	1.65	1.4	0.66
IIP ₃ (dBm)	-9.1	2		-3.4	15.6	+22	13.3	+18
1-dB (dBm)		-12	-8	-21				
Power gain (dB)	23	19.9	11.5	20	10	15.5	12.8	18.6
input return loss (dB)	-16.8		-14	-19				
Output return loss (dB)	-10.2		-17	-21				
Reverse Isolation (dB)	-51.7	-47.8		-35				
Stability factor				3.6				
Power dissipation (mW)		14.7	4	7.2				11.8
Process (μm)		0.35	0.18	0.25				
Supply voltage (V)		2	1.8	3.3	2.7	2.6	1.8	
Current (mA)			2.2	2.4	7.45	9.3	8	

In [Zhenying 2003] the authors covered the fact that it is not always possible to tune the input matching impedance to 50Ω with limited values of the inductors L_g and L_s . This is often the case when using foundry design libraries where inductor values can only be selected from pre-set values. To solve this they suggested connecting an extra capacitor from the gate of transistor $M1$ to ground to give extra freedom in tuning Z_{in} , as shown in Figure 5.9. For the output circuitry, they extensively discussed the effect of the load inductance on the drain of $M2$ on the stability, and showed why the value of this inductance has to be carefully chosen to prevent the effective output resistance from going negative. The output matching was achieved using an extra transistor stage and its load inductor as shown in Figure 5.9. This is beneficial in isolating the effect of the output match on the input match of the amplifier.

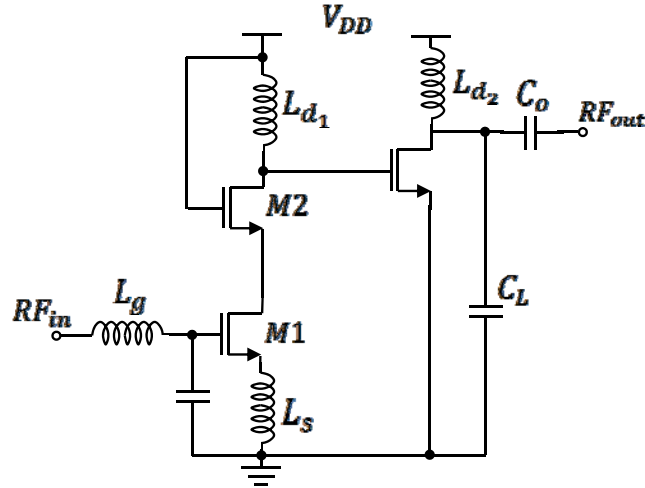


Figure 5.9 LNA matching technique in [Zhenying 2003]

One of the most advantageous modifications to the ordinary cascode architecture has been the inclusion of an inter-stage matching inductor between the two stages of the amplifier as shown in Figure 5.10.

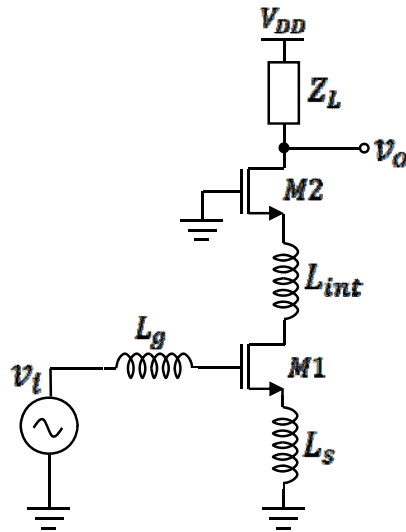


Figure 5.10 Cascode LNA with inter stage matching network

The idea behind this technique is that any power loss in the LNA signal path means more noise contribution of the following blocks. This technique was first presented in [Li 1999]. The output impedance of the first stage and the input impedance to the second stage are given in Equations (5.8) and (5.9), respectively [Zhang 2003]. These are extracted from the small signal equivalent circuits of the common-source and the common-gate stages.

$$Z_{out1}(S) = \frac{C_{gd1} s g_m R_L}{g_{m1} R_L + 1} \quad (5.8)$$

$$Z_{in2}(S) = \frac{1}{C_{gs2} s + g_{m2}} \quad (5.9)$$

As both the reactances of the output impedance of the first stage and the input impedance of the second stage are capacitive, the cascode LNA can be simplified to Figure 5.11 for the purpose of the inter-stage matching network. In Figure 5.11 C_1 and C_2 are the equivalent output capacitance of the first stage and the equivalent input capacitance of the second stage, respectively. R_1 and R_2 represent the respective real part of these impedances.

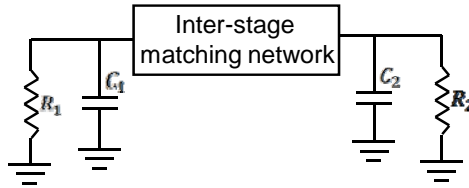


Figure 5.11 Equivalent circuit of the cascode LNA for inter stage matching.

For maximum power transfer, conjugate matching requires the output impedance of the CS stage to have a value equal to the complex conjugate of the input impedance of the common-source stage, hence requiring an inductor to be connected between these two stages. As will be shown later, the inter-stage matching impedance also has an effect on the linearity of the amplifier and can be used to optimise the IIP_3 . With the inter-stage matching network, however, the gain of the first stage is improved by adding the additional inductor (as the load impedance of the first stage is no longer the relatively low $1/g_m$ input impedance of the CG transistor). This causes the capacitance appearing at the gate of the CS transistor (due to the Miller effect of C_{gd1}) to become more significant, which may invalidate the input matching criteria [Kim 1999]. Usually, optimisation using computer simulation, can then be done to obtain new values for L_g , L_s and the inter-stage matching inductor to achieve the required optimum performance.

[Asgaran 2006] proposed eliminating the source degeneration inductance arguing that at high frequency, the degeneration inductance requires an increase in the DC current to meet a given gain requirement. However as will be shown later in Section 5.5, the degeneration inductance has another important role in addition to facilitating input matching, that is to provide negative feedback which improves the linearity of the amplifier.

The work in [Kim 2004], [Aparin 2005], [Kim 2006a] and [Huang 2006] are proposals of linearisation techniques aimed at improving the LNA nonlinear performance. These techniques will be discussed in more details in the next section when amplifier linearity and linearisation techniques are considered.

5.3 Amplifier non-Linearity

This section explores nonlinearities in devices and circuits with emphasis on the nonlinearities and linearisation techniques of low power LNAs. First, some consideration is given to circuit design issues which affect the nonlinear performance of amplifiers. A discussion of some of the linearisation techniques proposed in the literature that are suitable for low-power LNAs is then given. The section concludes by considering how these various techniques achieve highly linear LNAs with reference to recently published results.

5.3.1 Circuit Factors Affecting Linearity

In this section, circuit design factors that affect the nonlinear performance of amplifiers will be discussed with emphasis on bias and termination.

5.3.1.1 Choice of Bias Network

The effect of the bias level of a transistor on its linearity is well documented. For instance, a more linear performance can be obtained if the transistor is biased away from the nonlinear region of its input/output transfer characteristics. It is not just the bias level, however, that has an impact on linearity but the choice of the bias network as well. For instance, current-derived bias results in less IIP_3 variations from part to part than a voltage-derived bias [Aparin 2005].

In class-A amplifiers, low IMD_3 cannot be simultaneously obtained with good efficiency. If the amplifier is driven into saturation, efficiency will improve as the output power increases at the same DC power consumption, but IMD_3 distortion also increases as the amplifier is driven into the more nonlinear region of its characteristics. At low drive, on the other hand, the amplifier will be within its linear region achieving low IMD_3 but the efficiency decreases as the output power decreases while the same level of power consumption is maintained. Adaptive biasing (or dynamic bias modulation) can be used to solve this problem and increase the dynamic range of the amplifier. In this technique, the output signal is rectified and the result is used to adjust the bias such that the bias level is tuned in order for the amplifier to achieve high efficiency over a wider range of input powers. This technique was used in [Miers 1992], [Taniguchi 2001] and [Singh 2006].

5.3.1.2 Source and Load Termination

Source and load terminations are very critical in determining the nonlinear performance of an amplifier. At the output, the output matching network determines the amount of power delivered to the load; a better matching network delivers more power to the load, hence increasing the efficiency of the amplifier. Also, since the gain of the amplifier is dependant on its load impedance (this is true for both common-gate and common-source stages where $|G| = g_m Z_L$),

the point at which the gain compresses is also dependant on output termination which determines the load impedance seen by the active device.

Matching impedances also affect resultant IIP_3 and other intercept points of the amplifier. At the load, a power-series analysis of a simplified MOSFET model including only the nonlinear transconductance shows that the second and 3rd-order intermodulation powers can be given by [Maas 1998]

$$P_{IM2} = \frac{1}{2} a_2^2 V_s^4 |H(\omega)|^4 R_L \quad (5.10)$$

$$P_{IM3} = \frac{3}{4} a_3^3 V_s^6 |H(\omega)|^6 R_L \quad (5.11)$$

where a_2 and a_3 are the 2nd- and 3rd-order nonlinearity coefficients in the Taylor series representing the nonlinear transconductance, respectively, V_s is the source voltage and $H(\omega)$ is the linear transfer function. Equations (5.10) and (5.11) show that the 2nd- and 3rd-order intercept points will depend on the selection of the load termination network which determines the load impedance seen by the transistor as well as, of course, the power-series coefficients of the controlled current source. In a more realistic MOSFET when other nonlinearities are considered, the situation becomes more complicated, but the concept that the load impedance and the linearity of the controlled current source primarily establish the amplifier output intercept points is still fundamentally valid [Maas 1998].

Also, different matching networks have varying responses to different frequency components. For instance, if the output matching network at the load had a terminating response to the 3rd-order intermodulation distortion frequency, a considerable magnitude of the IMD_3 product maybe attenuated before it reaches the load, leading to an increase in IIP_3 . Also, an input matching network that was designed for optimum noise performance at the source, for instance, may have a frequency response to the 2nd harmonic component that terminates this frequency leading to IIP_3 improvement.

At the source, the input termination network has a significant influence of the magnitude and number of frequency components being fed back to the input. In addition to affecting IIP_3 , termination of the second harmonic component can also help improve efficiency [Snider 1967]. This is since this frequency component can change the shape of the drain current waveform and thus affect the power consumption. In [Gao 2005], the authors measured 76% and 42% saturated power-added efficiency, for the best and worst case second harmonic input terminations, respectively. Similar findings were made in [Colantonio 2004] and [Maeda 1995]. In general, efficiency of power amplifiers can be improved by providing proper terminations at harmonic frequencies for both load and source.

The optimum source and load terminations for low power operation can be found using techniques such as gain and noise circles as discussed in Chapter 2. For high power operation, optimum source and load terminations are determined by carrying out a simulation-based source-pull and load-pull investigation. These techniques will be jointly employed in Section 5.5.3 when the matching of the dual-function LNA/PA amplifier is discussed.

5.3.2 Low Power Cascode LNA Linearisation Techniques

Several system-level linearisation methods have been proposed in the literature, such as pre-distortion [Nojima 1985], feed forward error-cancellation [Seidel 1971] and Cartesian feedback [Minowa 1990]. But those methods are not suitable for low power applications such as ZigBee because their complexity increases cost and power consumption.

Two main linearisation methods suitable for low-power applications exist, transconductance linearisation by Transistor-Multi Gating [Webster 1996] and out-of-band harmonic termination [Aparin 1999].

In [Aparin 1999] the author used Volterra series analysis to find expressions for the 3rd-order intermodulation component (IMD₃) and the 3rd-order intercept point (IIP₃) of a common-emitter Bipolar Junction Transistor (BJT). If these equations were to be adapted for a common-source MOSFET transistor, two observations must be made, as follows:

- 1- In the MOSFET transistor, at RF, the gate current is usually very small compared to the drain current. Thus, the current gain (i_d/i_g) is very large. In order to simplify the analysis, this ratio can be assumed infinity. This corresponds to $\beta = \infty$ in the BJT transistor equations.
- 2- The junction-emitter capacitance in BJT, C_{jE} corresponds to the gate-source capacitance in MOSFET, C_{gs} . However, in BJT, C_{DE} is the base-emitter diffusion capacitance (resulting from the direct contact between the base and emitter material), which has no correspondent in MOSFET due to the silicon dioxide insulation between the gate and the source. Therefore, there is no diffusion capacitance current in MOSFET (defined in BJT as $i_{C_{DE}} = \tau di_c/dt$) where τ is the forward transit time. This corresponds to $\tau = 0$ in the BJT transistor equations.

Replacing β with ∞ and τ with 0 in the IMD₃ and IIP₃ equations for the BJT in [Aparin 1999] yields Equations (5.12) to (5.19) for the MOSFET model in Figure 5.12 [Kim 2001]. These equations have emerged to be the basis of a significant portion of linearisation concepts proposed by researchers in recent times.

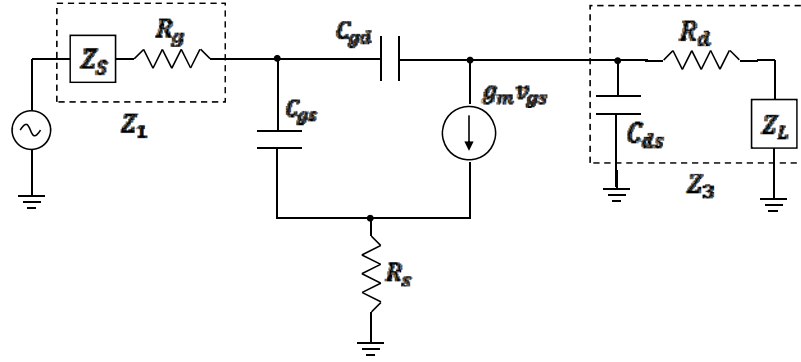


Figure 5.12 MOSFET model for derivation of IMD_3 and IIP_3 expressions using Volterra

The mathematical derivation of these equations for the full MOSFET small signal model is very lengthy and tedious. However, the procedures of obtaining them is important to the understanding of how the Volterra and Taylor series can jointly describe nonlinearities when memory components are involved, as discussed in Section 2.6.2. As an example to demonstrate how the complex substitutions take place, in Appendix A, the expression for IIP_3 is derived for a simplified version of the transistor model which includes only the transconductance and C_{gs} nonlinearities with all other capacitances considered zero and the body effect and channel length modulation neglected.

$$\text{IM}(2\omega_a - \omega_b) = \frac{3}{4} \cdot |H(\omega)| \cdot |A_1(\omega)|^3 \cdot |\varepsilon(\Delta\omega, 2\omega)| \cdot V^2 \quad (5.12)$$

$$\text{IIP}_3(2\omega_a - \omega_b) = \frac{1}{6\text{Re}[Z_s(\omega)] \cdot |H(\omega)| \cdot |A_1(\omega)|^3 |\varepsilon(\Delta\omega, 2\omega)|} \quad (5.13)$$

Where

$$H(\omega) = \frac{1 + j\omega C_{gs}[Z_1(\omega) + R_p] + j\omega C_{gd}Z_1(\omega)}{g_1 - j\omega C_{gd}[1 + R_p(g_1 + j\omega C_{gs})]} \quad (5.14)$$

$$A_1(\omega) = \frac{1}{g_1 + g(\omega)} \cdot \frac{1 + j\omega C_{gd}Z_3(\omega)}{Z_x(\omega)} \quad (5.15)$$

$$\varepsilon(\Delta\omega, 2\omega) = g_3 - g_{OB} \quad (5.16)$$

$$g_{OB} = \frac{2g_2^2}{3} \left[\frac{2}{g_1 + g(\Delta\omega)} + \frac{1}{g_1 + g(2\omega)} \right] \quad (5.17)$$

$$g(\omega) = \frac{1 + j\omega C_{gd}[Z_1(\omega) + Z_3(\omega)] + j\omega C_{gs}[Z_1(\omega) + Z_x(\omega)]}{Z_x(\omega)} \quad (5.18)$$

$$Z_x(\omega) = R_p + j\omega C_{gd}[Z_1(\omega)R_p + Z_1(\omega)Z_3(\omega) + R_pZ_3(\omega)] \quad (5.19)$$

Z_s represents the source impedance, $g(\Delta\omega)$ and $g(2\omega)$ are the conductance functions defined at the sub-harmonic $(f_1 - f_2)$ and the second harmonic $(2f)$ frequencies, respectively. $H(\omega)$ relates the equivalent IMD_3 voltage to the IMD_3 response of the drain current nonlinear term

and $A_1(\omega)$ is the linear transfer function for the input voltage of v_{gs} . $\varepsilon(\Delta\omega, 2\omega)$ shows how the drain current nonlinearities contribute to its IMD_3 response [Kim 2004].

From (5.12) it can be seen that a lower IMD_3 can be achieved when $|H(\omega)|$, $|A_1(\omega)|$ and $|\varepsilon(\Delta\omega, 2\omega)|$ are reduced. Equations (5.14) and (5.15) indicate that $|H(\omega)|$ and $|A_1(\omega)|$ depend on the in-band source and load impedances which are usually optimised to provide the required gain, noise figure or input/output match performance at the operating frequency. Equations (5.16) and (5.17) show that $\varepsilon(\Delta\omega, 2\omega)$ represents the contribution to IMD_3 from the 3rd- and 2nd-order nonlinearities where g_3 represents the direct contribution from the 3rd-order nonlinearity while g_{OB} represents the indirect contribution from the 2nd-order non linearity. In order to improve the nonlinear performance, both these nonlinearities have to be minimised. The indirect contribution from the 2nd-order nonlinearity (the sub harmonic and second harmonic frequencies) to IMD_3 will be explained in the following sub sections.

5.3.2.1 Linearisation by Harmonic Termination

As can be seen from Equation (5.16), if g_3 is negligibly small, $\varepsilon(\Delta\omega, 2\omega)$ becomes dominated by g_{OB} which is proportional to the square of g_2 that represents the strength of the 2nd-order nonlinearity. The contribution from the 2nd-order and other orders of nonlinearity to IMD_3 will be handled in more detail in Section 5.4. The basic principle is as follows: The 2nd-order nonlinearity produces the second harmonic ($2f_1$ or $2f_2$) and the sub harmonic ($f_2 - f_1$ or $f_1 - f_2$) frequency components. These components can be fed back to the input through multiple paths in the circuit and mix with the fundamental. When the total re-enters the system and is exposed to the 2nd-order nonlinearity, IMD_3 components are produced and these can be in phase with the IMD_3 components generated directly from the 3rd-order nonlinearity and hence add up.

One of the main feedback paths is the gate to drain capacitance, the other is the degeneration inductance at the source frequently used in LNA design to generate a positive real part at the input for matching purposes. As the inductance represents a non-zero impedance at the second harmonic frequency, the second harmonic currents generated in the source of the MOSFET establish non-zero second harmonic responses in the gate-source voltage of the MOSFET, effectively feeding back the second harmonic components to the input. It is possible to reduce these effects by out-of-band termination.

The authors of [Kim 2001] used equations (5.12) to (5.19) along with an NMOS BSIM3V3 simulation model to demonstrate the effect of the 2nd harmonic components on the linearity of a CMOS amplifier. They argued that $Z_s(\Delta\omega)$ and $Z_l(\Delta\omega)$ are sufficiently small and do not affect $\varepsilon(\Delta\omega, 2\omega)$. Upon that they demonstrated graphically how the value of g_{OB} dominates

$\varepsilon(\Delta\omega, 2\omega)$ when the 3rd-order nonlinearity is cancelled by two auxiliary transistors. To show that this is due to the 2nd harmonic frequency, they substituted for $Z_s(2\omega)$ and $Z_l(2\omega)$ by zero, and they then showed that in this case, the already low $\varepsilon(\Delta\omega, 2\omega)$ has a comparable value to g_3 as the value of g_{OB} becomes very insignificant. They hence used this to demonstrate that the improvements in IIP₃ and IMD₃ that can be achieved in a multi-gated and second harmonic-terminated amplifier is much more than in a multi-gated only amplifier.

5.3.2.2 Linearisation by Multi-Gating

Figure 5.13 shows the variation of the 3rd-order derivative of a MOSFET DC current (g_m'') with its gate to source voltage over a range from 0 to 1 volts (this is produced for transistor M1 in Figure 5.15).

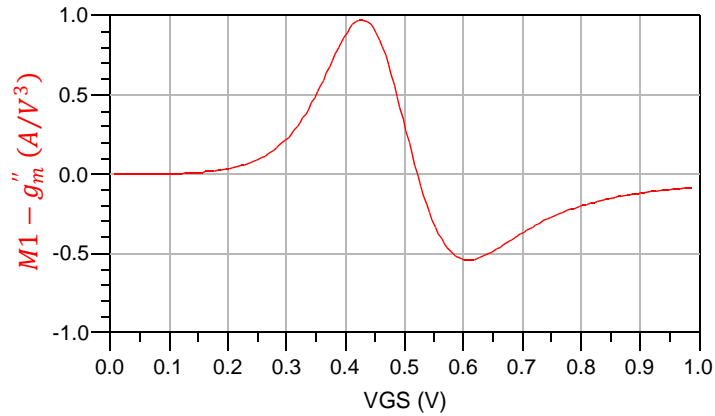


Figure 5.13 g_m'' vs. V_{GS} for a common-source transistor

A positive or negative g_m'' means 3rd-order intermodulation components will be produced and will degrade the linearity of the amplifier. Therefore, a MOSFET can be linearised by biasing it at the point where g_m'' is zero. Theoretically, this means that at this point no 3rd-order intermodulation components should be produced and hence IIP₃ should peak (or can be infinite if the effect on IMD₃ from other orders of nonlinearity is neglected). In a practical amplifier, however, there are two important issues to consider. First, this peak will happen at a very narrow point at the exact value of gate to source voltage when g_m'' is zero as shown in Figure 5.13, hence not giving the designer enough room to optimise the bias point for other performance requirements such as noise or gain, and also making the design very sensitive to bias variations which can be caused by unavoidable changes in temperature. Second, as shown in Figure 5.13, the nature of dependence of $g_3(V_{GS})$ (or g_m'') on the gate to source voltage is such that g_m'' is positive in the weak and moderate inversion region and transitions to negative as a higher V_{GS} takes the transistor into the strong inversion region [Aparin 2004]. This means that the bias point at which g_m'' is zero will always be far off from the point usually chosen for optimum low power performance which is typically 0.2-0.4V above threshold. Instead, g_m'' will

have a negative peak at that point leading to higher IMD_3 distortion. A desirable g_m'' characteristic, therefore, will be something similar to Figure 5.14 where the zero g_m'' region is shifted and expanded (this is total g_m'' of the multi-gated transistor in Figure 5.15).

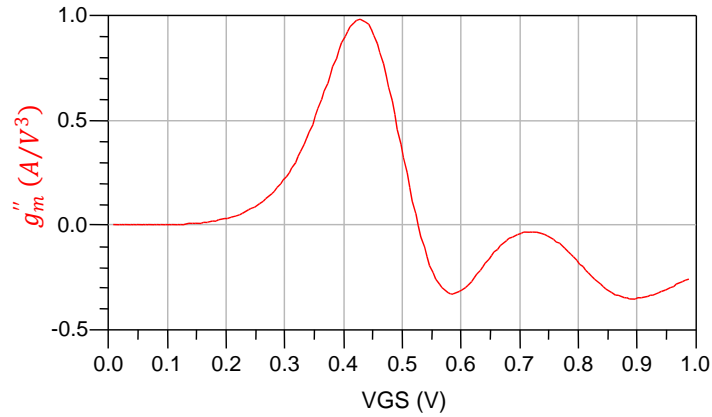


Figure 5.14 Tuned g_m'' - V_{GS} characteristics with linearisation

This can be achieved using the Derivative Superposition (DS) method which was first proposed in [Webster 1996]. In this method, two or more MOSFETs of different widths and gate biases are connected in parallel as shown in Figure 5.15 to achieve a combined DC transfer characteristic where the range of bias voltages at which g_m'' is zero or close to zero is shifted and extended. This method is sometimes referred to as Multiple Gated Transistor (MGTR) [Kim 2000].

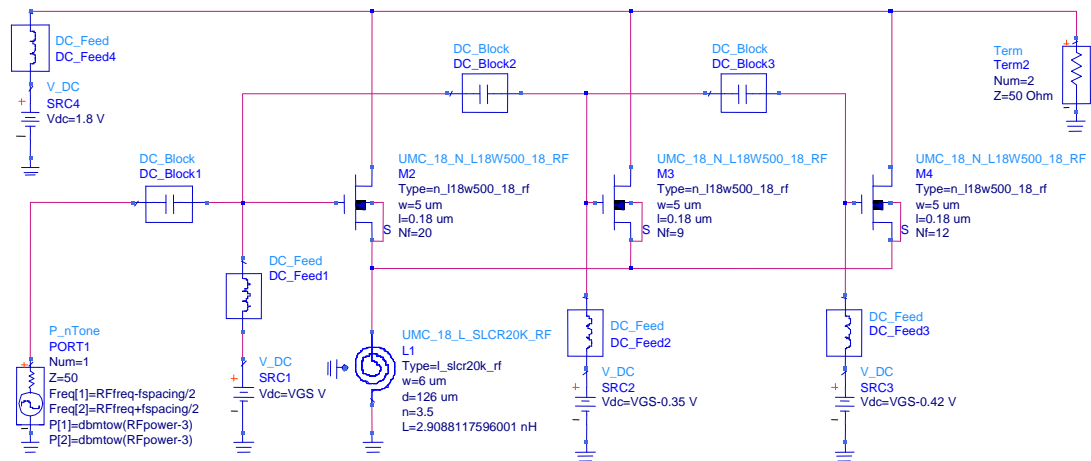


Figure 5.15 Circuit schematic of linearisation using the DS method

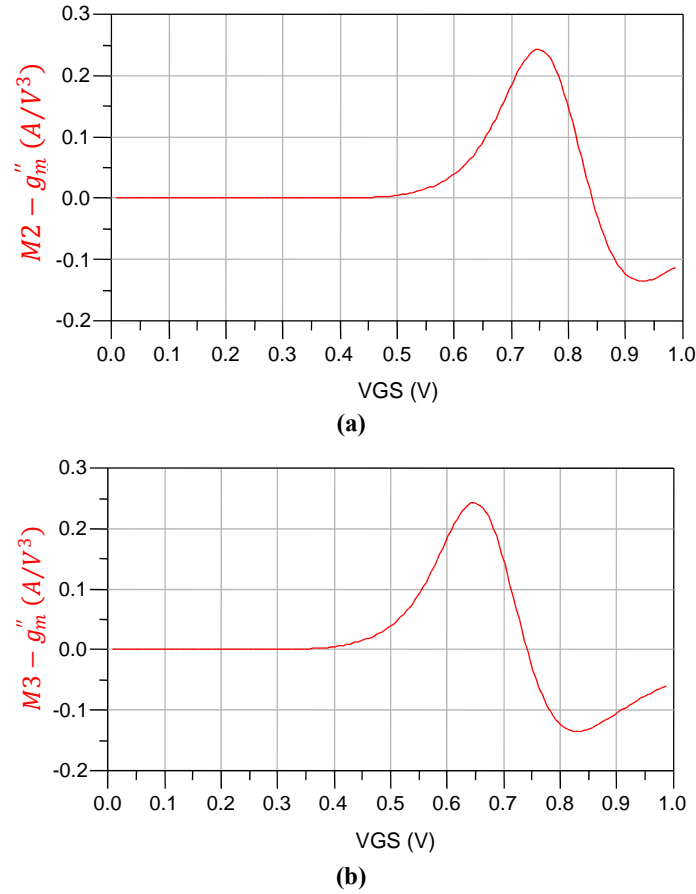


Figure 5.16 Tuned g''_m - V_{GS} characteristics for auxiliary transistors (a) $M2$ and (b) $M3$

As shown in Figure 5.16, the idea of the DS (or MGTR) method is that by adjusting the width and bias of the auxiliary transistors, the region (on the V_{GS} scale) where g''_m of the main transistor is negative can be aligned with the region at which g''_m of the auxiliary transistors is positive, thereby yielding a composite g''_m that is zero or close to zero at the desirable region at which a transistor would normally be biased, as in Figure 5.14. The more auxiliary transistors that are added, the more flat the characteristic can be around the desired region. The gate voltage of the auxiliary transistor is used to shift its g''_m along the V_{GS} scale while transistor size scaling controls the magnitude of the positive or negative parts, that is since scaling the transistor controls the current through it.

The beauty of the DS linearisation method is that because the auxiliary transistor is biased in the subthreshold regime, it does not lead to any extra current consumption, making it suitable for low power applications. One of the disadvantages of this technique, however, is the degradation of the noise figure. In [Aparin 2005], the authors studied the effect of the DS method on the noise figure and argued that the assumption that NF of the composite MOSFET is dominant by the transistor in strong inversion since it draws more than 20 times the current as the transistor

in the weak inversion, does not hold in measured data as oppose to simulations using BSIM3v3 models.

They used Equations (2.4), (2.5) and (2.6) and extended the van der Ziel model [van der Ziel 1986] to find values for γ , δ and c for the a transistor in weak inversion to be 1/2, 5/4 and $j0.707$, respectively. They then argued that since the drain current in weak inversion is due to diffusion, $g_{d0} = I_{D_{SAT}}/\phi_t$ where ϕ_t is the thermal voltage ($\phi_t = KT/q$), and substituting for g_{d0} in (2.5) yields

$$\overline{i_{ng}^2} = 4kT\delta\Delta f \frac{\omega^2 C_{gs}^2 \phi_t}{5I_{D_{SAT}}} \quad (5.20)$$

From this they concluded that while the MOSFET in weak inversion draws a negligible drain current, its induced gate noise is inversely proportional to the drain current and therefore can be quite significant as it adds to the induced gate noise current of the main transistor in strong inversion, degrading the overall Noise Figure in the DS method.

5.3.2.3 Other Low-Power Linearisation Techniques

By ignoring the gate-drain capacitance, the authors of [Yang 2003] simplified Equation 5.13 to

$$H(\omega) = \frac{1 + j\omega C_{gs}[Z_g(\omega) + Z_s(\omega)]}{g_m} \quad (5.21)$$

Using Equations (5.12) and (5.21), they extracted the following dependence of IMD_3 of the common-source transconductance stage:

$$|IM_3(2\omega_1 - \omega_2)| \propto |1 + j\omega C_{gs}Z_s(\omega) + j\omega C_{gs}Z_g(\omega)| \quad (5.22)$$

In an inductively degenerated common-source stage, $Z_s(\omega)$ becomes $j\omega L_s$. As (5.22) shows, multiplying $j\omega C_{gs}$ with an inductive $Z_s(\omega)$ results in a negative real part, partially cancelling the “1” term and hence reducing the magnitude of IMD_3 . As a result, the authors then showed that linearity can be increased by increasing either the degeneration inductance or C_{gs} . Upon that they proposed adding an external capacitance between the gate and the source of the common-source stage since increasing C_{gs} by increasing the size of the transistor will lead to higher power consumption, and increasing the degeneration inductance will increase the noise figure.

5.3.3 Highly Linear Amplifiers

This section aims to discuss recently reported results of highly linear amplifiers achieved by employing the above mechanisms with focus on CMOS low noise amplifier designs of the common-source common-gate architecture. Some of these results are listed in Table 5.1.

In the original paper for BJT [Aparin 1999], the authors achieved IIP_3 improvement by considering the sub-harmonic and second harmonic frequencies which generate IMD_3 components when fed back to the input. Based on the original common-emitter Volterra equations, they explained that for most narrow-band RF circuits, the load impedance, Z_L , is much smaller than the impedance of base-collector capacitance, C_μ , in the medium and high frequency bands so the feedback through C_μ at $\Delta\omega = \omega_2 - \omega_1$ and the effect of $Z_L(\Delta\omega)$ can be neglected. They then used micro-strip lines and passive components to simultaneously in-band match the circuit and optimally terminate the 2ω components at the input and output, and the $\Delta\omega$ component at the input. The design was such that the in-band and out-of-band terminations did not affect each other and hence the LNA's in-band performance was not affected. The IIP_3 improvement achieved was 9.6 dBm in the 1.95 to 2.05GHz frequency range.

In [Kim 2000], the authors used the DS method alone to achieve a 3dB improvement in IIP_3 . They compared the performance of single-gated and double gated amplifiers and demonstrated that the DS method alone helps improve IIP_3 without considerable effect on other performance parameters such as gain and power consumption.

In [Kim 2004] the authors combined the DS method with out-of-band termination which resulted in an IIP_3 improvement of 10dB. They stated that as g_m'' is effectively cancelled by MGTR, the value of g_m' which dominates g_{OB} in (5.17) is still appreciable and has to be decreased. From (5.17), they noted that one of the best ways to do this is to increase both $g(\Delta\omega)$ and $g(2\omega)$. They argued that $g(\Delta\omega)$ is much larger compared with $g(2\omega)$ and hence they focused on the 2nd harmonic frequency and derived an approximate expression for $g(2\omega)$ as

$$g(2\omega) \approx g_m \times \frac{1 + 2j\omega C_{gs}Z_1 + 2j\omega C_{gd}Z_2}{1 + \omega_T C_{gd}Z_2} \quad (5.23)$$

where Z_1 and Z_2 are the impedance looking into the source and load of the common-source stage, respectively, and $\omega_T = g_m/C_{gs}$. They noted that as the magnitude of Z_1 in a typical common-source NMOS is of the order $1/(\omega C_{gs})$ then the second term in the numerator of Equation (5.23) ($2j\omega C_{gs}Z_1$) is comparable to 1. Furthermore, since $2\omega \ll \omega_T$, then the second term in the denominator ($\omega_T C_{gd}Z_2$) is the most dominant factor and must be decreased in order to increase $g(2\omega)$. To decrease this term, they proposed reducing Z_2 by using a cascode stage at the drain of the main transistor, as the input impedance of which will be $1/g_m$. This method was better than the harmonic tuning method proposed in [Aparin 1999] because it didn't require the use of large passive components. However, the disadvantage of this solution is that with a very small degeneration inductance, it is difficult to simultaneously achieve a good VSWR and

NF [Aparin 2005] which may explain the relatively low (10dB) gain and high (2.85dB) NF achieved in this work.

The authors of [Aparin 2005] focused on the relationship between the performance of the DS method and the operating frequency in relation to the degeneration inductance. They explained that the performance of the DS method degrades at higher frequencies due to the effect of circuit reactances, limiting the improvement in IIP_3 . They used a simplified MOSFET model including only a nonlinear transconductance, a linear C_{gs} and the degeneration inductance and expressed its IIP_3 as

$$IIP_3 = \frac{4g_1^2\omega^2LC_{gs}}{3|\varepsilon|} \quad (5.24)$$

Where

$$\varepsilon = g_3 - \frac{2g_2^2/3}{g_1 + \frac{1}{j2\omega L} + j2\omega C_{gs} + Z_1(2\omega)\frac{C_{gs}}{L}} \quad (5.25)$$

Based on the expression for IIP_3 in (5.24) and (5.25), it was shown that the zero composite g_3 would not result in an infinite IIP_3 at higher frequencies as it does at low frequencies. The reason for this can be extracted from the second term in (5.25) which represents the contribution of the 2nd-order nonlinearity to the 3rd-order intermodulation (IMD_3) and as can be seen, this contribution depends on the degeneration inductance as previously explained in Section 5.3.2.1. This 2nd-order nonlinearity-contributed IMD_3 will still exist even if g_3 was completely cancelled by DS, and it increases as the operating frequency increases. This is since the impedance of the degeneration inductance ($= j\omega L_s$) increases as the frequency increases.

To solve this, the authors proposed a modified version of the DS method in which two series inductors were used for degeneration instead of one. The idea was, instead of terminating the 2nd harmonic components, the two inductors are used to adjust the IMD_3 contribution from the second harmonic to be equal in magnitude and opposite in phase to the direct IMD_3 generated from the 3rd-order nonlinearities, and hence obtain IIP_3 improvements by making these components cancel each other.

The work in [Kim 2006a] achieved linearisation using a slightly modified variation of the DS method in which the 3rd-order transconductance non-linearity of the common-source transistor was compensated by placing a folded PMOS transistor in parallel and adjusting its size and bias to sink to ground the IMD_3 current component in the cascode channel. In [Huang 2006], a similar idea was proposed but with the IMD_3 current sinker being an NMOS transistor in parallel with the common-gate stage transistor.

5.4 Frequency Components Affecting Nonlinear Performance

The concept of generating intermodulation components by feeding back frequency components produced from different orders of nonlinearity was briefly introduced in Section 5.3.2. Here, this phenomena is further investigated and analysed in more detail. While most of the work in the literature reviewed in Section 5.3 focused on two components, the second harmonic and the difference frequency in contributing to the 3rd-order intermodulation, here this analysis is widely extended by studying the potential effect of all components produced from all single-orders of nonlinearity in contributing to 3rd-, 5th- and 7th-order intermodulation distortion when exposed to any other single-order of nonlinearity up to the 7th-order. The analysis is based entirely on mathematical substitutions of polynomial terms, carried out in Mathematica software (Wolfram Research, Inc.) [Wolfram 2007]. In Section 5.4.1 the feedback phenomenon is explained further and the method in which mathematical substitutions were done in Mathematica is explained. In Section 5.4.2, a summary of all the results is presented and discussed.

5.4.1 The Feedback Phenomena

Figure 5.17 explains graphically how 2nd-order nonlinearity can contribute to 3rd-order intermodulation distortion as an example of the more general frequency component feedback concept.

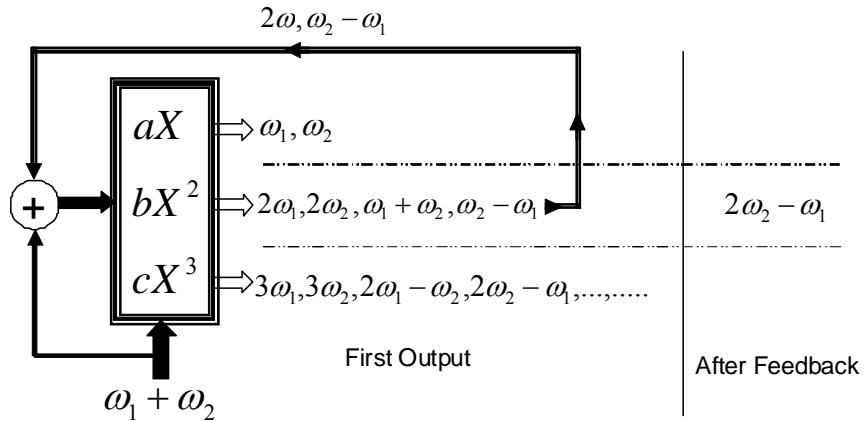


Figure 5.17 Graphical description of 2nd-order contribution to IMD_3

The original two-tone input signal is $A \sin \omega_1 t + A \sin \omega_2 t$ (shown in short in Figure 5.17 as $\omega_1 + \omega_2$ to resemble the frequency content only for simplification). The system has up to 3rd-order significant nonlinear terms. The two-tone input is exposed to the 2nd- and 3rd-order nonlinearity separately. Frequency components produced from the 2nd-order nonlinearity are as shown in Figure 5.17. The Mathematica snapshot of Figure 5.18 shows the exact magnitudes and coefficients of these components.

$$\mathbf{x} = \mathbf{A} \sin[\omega_1] + \mathbf{A} \sin[\omega_2];$$

$$\mathbf{y} = \mathbf{b} \mathbf{x}^2;$$

$$\text{TrigReduce}[\text{Expand}[\mathbf{y}]]$$

$$\frac{1}{2} (2 \mathbf{A}^2 \mathbf{b} - \mathbf{A}^2 \mathbf{b} \cos[2 \omega_1] + 2 \mathbf{A}^2 \mathbf{b} \cos[\omega_1 - \omega_2] - \mathbf{A}^2 \mathbf{b} \cos[2 \omega_2] - 2 \mathbf{A}^2 \mathbf{b} \cos[\omega_1 + \omega_2])$$

Figure 5.18 results of the 2nd-order nonlinearity from initial two-tone input

All frequency components generated from the 3rd-order system can be fed-back to the input, but focus here is on the components produced from the 2nd-order nonlinearity which can contribute to 3rd-order intermodulation distortion when fed-back. These are 2ω and $\omega_1 - \omega_2$ as indicated on the arrow of the feedback path in Figure 5.17. These components mix with the fundamental and re-enter the 3rd-order nonlinear system. The new total input is therefore exposed to all orders of nonlinearity in a similar fashion to the original two-tone input. The new input to the system becomes as shown in top of Figure 5.19 which is a snapshot of the substitution in Mathematica for the example when the 2nd-order component 2ω is fed-back to the input. The 2nd-order nonlinear term produces the sum and difference of all the frequency components that it can see. As $2\omega_1$ (fed-back) and ω_2 (input tone) are present at the input, when the 2nd-order nonlinearity calculates the difference, the IMD₃ component $2\omega_1 - \omega_2$ results as shown in Figure 5.19. In another example for Figure 5.17, with the difference frequency $\omega_1 - \omega_2$ (fed-back) and ω_1 (input tone) present at the input, if the 2nd-order nonlinearity calculates the sum, the IMD₃ component $2\omega_1 - \omega_2$ will result.

$$\mathbf{x} = \mathbf{A} \sin[\omega_1] + \mathbf{A} \sin[\omega_2] + \frac{1}{2} \mathbf{b} \mathbf{A}^2 \cos[2 \omega_1] + \frac{1}{2} \mathbf{b} \mathbf{A}^2 \cos[2 \omega_2];$$

$$\mathbf{y} = \mathbf{a} \mathbf{x} + \mathbf{b} \mathbf{x}^2 + \mathbf{c} \mathbf{x}^3;$$

$$\text{TrigReduce}[\text{Expand}[\mathbf{y}]]$$

$$\begin{aligned} & \mathbf{A}^2 \mathbf{b} + \frac{\mathbf{A}^4 \mathbf{b}^3}{4} - \frac{3}{4} \mathbf{A}^4 \mathbf{b} \mathbf{c} - \frac{1}{2} \mathbf{A}^2 \mathbf{b} \cos[2 \omega_1] + \frac{1}{2} \mathbf{A}^2 \mathbf{b} \cos[2 \omega_2] + \frac{3}{2} \mathbf{A}^4 \mathbf{b} \mathbf{c} \cos[2 \omega_1] + \frac{9}{32} \mathbf{A}^6 \mathbf{b}^3 \mathbf{c} \cos[2 \omega_1] + \frac{1}{8} \mathbf{A}^4 \mathbf{b}^3 \cos[4 \omega_1] - \\ & \frac{3}{8} \mathbf{A}^4 \mathbf{b} \mathbf{c} \cos[4 \omega_1] + \frac{1}{32} \mathbf{A}^6 \mathbf{b}^3 \mathbf{c} \cos[6 \omega_1] + \frac{3}{32} \mathbf{A}^6 \mathbf{b}^3 \mathbf{c} \cos[2 \omega_1 - 4 \omega_2] + \frac{3}{4} \mathbf{A}^4 \mathbf{b} \mathbf{c} \cos[\omega_1 - 3 \omega_2] + \frac{1}{4} \mathbf{A}^4 \mathbf{b}^3 \cos[2 \omega_1 - 2 \omega_2] - \\ & \frac{3}{4} \mathbf{A}^4 \mathbf{b} \mathbf{c} \cos[2 \omega_1 - 2 \omega_2] + \frac{3}{32} \mathbf{A}^6 \mathbf{b}^3 \mathbf{c} \cos[4 \omega_1 - 2 \omega_2] + \mathbf{A}^2 \mathbf{b} \cos[\omega_1 - \omega_2] - \frac{3}{2} \mathbf{A}^4 \mathbf{b} \mathbf{c} \cos[\omega_1 - \omega_2] + \frac{3}{4} \mathbf{A}^4 \mathbf{b} \mathbf{c} \cos[3 \omega_1 - \omega_2] - \frac{1}{2} \mathbf{A}^2 \mathbf{b} \cos[2 \omega_2] + \\ & \frac{1}{2} \mathbf{A}^2 \mathbf{b} \cos[2 \omega_2] + \frac{3}{2} \mathbf{A}^4 \mathbf{b} \mathbf{c} \cos[2 \omega_2] + \frac{9}{32} \mathbf{A}^6 \mathbf{b}^3 \mathbf{c} \cos[2 \omega_2] + \frac{1}{8} \mathbf{A}^4 \mathbf{b}^3 \cos[4 \omega_2] - \frac{3}{8} \mathbf{A}^4 \mathbf{b} \mathbf{c} \cos[4 \omega_2] + \frac{1}{32} \mathbf{A}^6 \mathbf{b}^3 \mathbf{c} \cos[6 \omega_2] - \\ & \mathbf{A}^2 \mathbf{b} \cos[\omega_1 + \omega_2] + \frac{3}{2} \mathbf{A}^4 \mathbf{b} \mathbf{c} \cos[\omega_1 + \omega_2] - \frac{3}{4} \mathbf{A}^4 \mathbf{b} \mathbf{c} \cos[3 \omega_1 + \omega_2] + \frac{1}{4} \mathbf{A}^4 \mathbf{b}^3 \cos[2 \omega_1 + 2 \omega_2] - \frac{3}{4} \mathbf{A}^4 \mathbf{b} \mathbf{c} \cos[2 \omega_1 + 2 \omega_2] + \\ & \frac{3}{32} \mathbf{A}^6 \mathbf{b}^3 \mathbf{c} \cos[4 \omega_1 + 2 \omega_2] - \frac{3}{4} \mathbf{A}^4 \mathbf{b} \mathbf{c} \cos[\omega_1 + 3 \omega_2] + \frac{3}{32} \mathbf{A}^6 \mathbf{b}^3 \mathbf{c} \cos[2 \omega_1 + 4 \omega_2] + \mathbf{A} \sin[\omega_1] - \frac{1}{2} \mathbf{A}^3 \mathbf{b}^2 \sin[\omega_1] + \frac{9}{4} \mathbf{A}^3 \mathbf{c} \sin[\omega_1] + \\ & \frac{3}{4} \mathbf{A}^5 \mathbf{b}^2 \mathbf{c} \sin[\omega_1] + \frac{1}{2} \mathbf{A}^3 \mathbf{b}^2 \sin[3 \omega_1] - \frac{1}{4} \mathbf{A}^3 \mathbf{c} \sin[3 \omega_1] - \frac{3}{16} \mathbf{A}^5 \mathbf{b}^2 \mathbf{c} \sin[3 \omega_1] + \frac{3}{16} \mathbf{A}^5 \mathbf{b}^2 \mathbf{c} \sin[5 \omega_1] + \frac{3}{16} \mathbf{A}^5 \mathbf{b}^2 \mathbf{c} \sin[\omega_1 - 4 \omega_2] - \\ & \frac{3}{8} \mathbf{A}^5 \mathbf{b}^2 \mathbf{c} \sin[2 \omega_1 - 3 \omega_2] + \frac{1}{2} \mathbf{A}^3 \mathbf{b}^2 \sin[\omega_1 - 2 \omega_2] - \frac{3}{4} \mathbf{A}^3 \mathbf{c} \sin[\omega_1 - 2 \omega_2] - \frac{3}{8} \mathbf{A}^5 \mathbf{b}^2 \mathbf{c} \sin[\omega_1 - 2 \omega_2] + \frac{3}{8} \mathbf{A}^5 \mathbf{b}^2 \mathbf{c} \sin[3 \omega_1 - 2 \omega_2] - \\ & \frac{3}{8} \mathbf{A}^5 \mathbf{b}^2 \mathbf{c} \sin[2 \omega_1 - \omega_2] - \frac{3}{4} \mathbf{A}^3 \mathbf{c} \sin[2 \omega_1 - \omega_2] + \frac{3}{8} \mathbf{A}^5 \mathbf{b}^2 \mathbf{c} \sin[2 \omega_1 - \omega_2] - \frac{3}{16} \mathbf{A}^5 \mathbf{b}^2 \mathbf{c} \sin[4 \omega_1 - \omega_2] + \mathbf{A} \sin[\omega_2] - \frac{1}{2} \mathbf{A}^3 \mathbf{b}^2 \sin[\omega_2] + \\ & \frac{9}{4} \mathbf{A}^3 \mathbf{c} \sin[\omega_2] + \frac{3}{4} \mathbf{A}^5 \mathbf{b}^2 \mathbf{c} \sin[\omega_2] + \frac{1}{2} \mathbf{A}^3 \mathbf{b}^2 \sin[3 \omega_2] - \frac{1}{4} \mathbf{A}^3 \mathbf{c} \sin[3 \omega_2] - \frac{3}{16} \mathbf{A}^5 \mathbf{b}^2 \mathbf{c} \sin[3 \omega_2] + \frac{3}{16} \mathbf{A}^5 \mathbf{b}^2 \mathbf{c} \sin[5 \omega_2] + \\ & \frac{1}{2} \mathbf{A}^3 \mathbf{b}^2 \sin[2 \omega_1 + \omega_2] - \frac{3}{4} \mathbf{A}^3 \mathbf{c} \sin[2 \omega_1 + \omega_2] - \frac{3}{8} \mathbf{A}^5 \mathbf{b}^2 \mathbf{c} \sin[2 \omega_1 + \omega_2] + \frac{3}{16} \mathbf{A}^5 \mathbf{b}^2 \mathbf{c} \sin[4 \omega_1 + \omega_2] + \frac{1}{2} \mathbf{A}^3 \mathbf{b}^2 \sin[\omega_1 + 2 \omega_2] - \\ & \frac{3}{4} \mathbf{A}^3 \mathbf{c} \sin[\omega_1 + 2 \omega_2] - \frac{3}{8} \mathbf{A}^5 \mathbf{b}^2 \mathbf{c} \sin[\omega_1 + 2 \omega_2] + \frac{3}{8} \mathbf{A}^5 \mathbf{b}^2 \mathbf{c} \sin[3 \omega_1 + 2 \omega_2] + \frac{3}{8} \mathbf{A}^5 \mathbf{b}^2 \mathbf{c} \sin[2 \omega_1 + 3 \omega_2] + \frac{3}{16} \mathbf{A}^5 \mathbf{b}^2 \mathbf{c} \sin[\omega_1 + 4 \omega_2] \end{aligned}$$

Figure 5.19 Frequency components resulting from feeding back the 2ω component

In Figure 5.19, the circled components are resulting 3rd-order intermodulation products when the 2ω component produced from 2nd-order nonlinearity is fed-back. The IMD₃ product circled with the solid line has the coefficient c only, which indicates that it is produced from mixing the two original input tones by the 3rd-order nonlinearity. The IMD₃ product circled with the dotted line has the coefficient b only which indicates that it is the result of mixing the fed-back components and the original input tones by the 2nd-order nonlinearity. The IMD₃ product circled with the dashed line has the coefficients b and c which indicates that it is the result of mixing the 2nd-order fed-back components and the original input tones in the 3rd-order nonlinearity.

5.4.2 Analysis and Results

Specifically designed power amplifiers employ linearisation techniques that are often very effective but are also often complicated and require higher power consumption. This effectiveness cannot be matched by simple low power linearisation techniques. Therefore, higher power operation of an amplifier based on a common low-power topology, such as the cascode architecture, might lead to additional distortion from higher-orders of nonlinearity that would otherwise be ineffective if the amplifier was working in its low power regime [Aparin 2005].

In order to analyse these possibilities, a mathematical analysis was performed using a 7th-order polynomial to represent the nonlinearity of an assumed amplifier where up to 7th-order nonlinearity is significant. The procedures work as follow: A two-tone signal of the form $A \sin \omega_1 t + A \sin \omega_2 t$ is assumed as an initial input to the nonlinear amplifier; this is thereafter referred to as the original fundamental two tone input. The original fundamental input is processed by each nonlinearity order of the amplifier individually. Several harmonic and intermodulation components are generated from this process; these are thereafter referred to as the initially generated frequency components. Since filtering takes place only at the output of the amplifier, then all the initially generated frequency components will return to the input due to the degeneration inductor in the source of the common-source stage (as that of Figure 5.7), which resembles a negative feedback. Therefore, initially generated frequency components are also thereafter referred to as fed back components. Each fed back component will be added to the fundamental at the input and the total new input processed through all orders of nonlinearity up to the 7th-order. The process will result in the generation of new frequency components from each order of nonlinearity that are not generated when only the original fundamental two-tone input is processed through that respective order. Focus is on the 3rd-, 5th- or 7th-order intermodulation components generated from this process, these are thereafter referred to as the feedback-generated intermodulation components. Any fed-back components that result in the generation of 3rd-, 5th- or 7th-order intermodulation distortion components are then noted. To

generalise the analysis and make it valid for as many amplifiers of various standards (which may have different operating powers) as possible, the following assumptions are made:

- All components will return with equal amplitudes and zero phase; it is difficult to make up for the limited bandwidth of the return path.
- All the resulting intermodulation components will add up algebraically; it is difficult to make up for the phase-frequency characteristics. This will result in a worst case analysis.

Despite these simplifying assumptions, this analysis is useful to provide an insight into the likely significance of generated intermodulation components from mixing all the fed-back components and the original fundamental input tones. Table 5.2 provides a summary of the results showing which fed-back components resulted in feedback-generated intermodulation components. Table 5.2 should be read as follows: The frequency components on the left are those which result from the mixing of the original two-tone input in their respective order of nonlinearity alone. These components are mixed with the fundamental two-tone input into each order of nonlinearity from 2nd to 7th individually. ^x✓ indicates that an intermodulation distortion component of order x was produced and ^x× means no component has been produced. Appendix C presents a detailed summary of these results showing the magnitudes of these components.

Some important observations can be made from studying the summary results in Table 5.2 and the more detailed results in Appendix C, these are as follows:

- The feeding back of frequency components not only introduces intermodulation distortion components from orders of nonlinearity that would normally not produce them, but adds to the intermodulation components which are normally produced by their respective order of nonlinearity. For example, the feeding back of the $\omega_1 - \omega_2$ component from 2nd-order nonlinearity to the input produces the 3rd-order intermodulation component $2A^3b^2\sin(2\omega_1 - \omega_2)$ when mixed with the fundamental in the 2nd-order nonlinearity, but also, when mixed with the fundamental in the 3rd-order nonlinearity it produces the 3rd-order intermodulation component $\frac{3}{4}A^5b^2c\sin(2\omega_1 - \omega_2)$ which adds to the 3rd-order intermodulation component $\frac{3}{4}A^3c\sin(2\omega_1 - \omega_2)$ that is normally produced by the 3rd-order nonlinearity from the original two-tone input.

- All fed back components mixed with the fundamental in any odd-order of nonlinearity produce at least two respective intermodulation components of that order; one is due to the exposure of the fundamental components to that nonlinearity order and the other is due to mixing with the feedback components. Usually, the magnitude of the non-mixed term is higher. For example $2\omega_1 + \omega_2$ from 3rd into 5th produces $\frac{5}{8}A^5e\sin(3\omega_1 - 2\omega_2)$ and $\frac{15}{16}A^7ce\sin(3\omega_1 - 2\omega_2)$ where the coefficient e of the first indicates it is from the fundamental and ce in the second indicates it is from mixing with the 3rd-order term fed back.
- Generally, if a component of an order of intermodulation distortion is produced from any even-order of nonlinearity; all lower-order intermodulation products are also produced. For example $2\omega_1 - 4\omega_2$ from 6th into 6th produces 7th-, 5th-, and 3rd-order components. Also 2ω from 2nd into 3rd produces 5th- and 3rd-order intermodulation components. Very few exceptions to this observation exist, for example $4\omega_1 - 2\omega_2$ from 6th into 2nd produces 5th- and 7th-order intermodulation components but not 3rd. Also, 4ω from 4th into 3rd produces 3rd- and 7th-order intermodulation components but not 5th.
- Several previously unnoticed frequency components produce 3rd-order intermodulation distortion, but with smaller magnitudes than those resulting from the feeding back of the 2nd-order nonlinearity components. For example $2\omega_1 + \omega_2$ from 3rd into 3rd produces $\frac{9}{16}A^5c^2\sin(2\omega_1 - \omega_2)$ and $2\omega_1 - 2\omega_2$ from 4th into 2nd produces $24A^5bd\sin(2\omega_1 - \omega_2)$. However, these components can be of concern if the input power was sufficiently high to cause their magnitudes to increase to significant levels where they can cause noticeable distortion. This is specific for every particular circuit and there is no general rule to define its boundaries.
- In some cases, frequency components produced from some low-orders of nonlinearity mix with the fundamental in other low-orders of nonlinearity to produce higher-order intermodulation distortion. For example 2ω from 2nd into 3rd produces 5th-order components, $2\omega_1 - 2\omega_2$ from 4th into 3rd produces 7th-order components and $\omega_1 - \omega_2$ from 2nd into 5th produces 7th-order components.
- Perhaps the most important observation is that any fed back frequency component from any order of nonlinearity, mixed with the fundamental in any odd-order of nonlinearity produces feedback-related intermodulation distortion components of that respective

order and all lower odd-orders. For example $2\omega_1$ from 2nd into 5th produces $\frac{25}{8}A^5e\sin(2\omega_1 - \omega_2)$ which is a 3rd-order intermodulation term produced from the fundamental tones due to 5th-order nonlinearity, but also produces $\frac{175}{32}A^7b^2e\sin(2\omega_1 - \omega_2)$ and $\frac{15}{32}A^9b^4e\sin(2\omega_1 - \omega_2)$ which are 3rd-order intermodulation terms produced from the mixing of the fundamental and feedback component in the 5th-order nonlinearity. This means that not only 5th- and 7th-order nonlinearity directly contributes to 3rd-order intermodulation distortion as previously discussed, but also components fed back and mixed with the fundamental into these higher-orders of non linearity contributes further to 3rd-order intermodulation distortion.

5.5 Cascode LNAs Operating at High Input Powers - An Investigation and Design Methodology

In this age of fast time to market requirements, RF front-end designers are increasingly dependent on Computer Aided Design (CAD) tools for fast and first-time-correct designs. Furthermore, with the existence of different transistor technologies from different foundries, the need for CAD tools becomes even more significant as theoretical design methods may need to differ from one technology to another, or from a technology made by one foundry to another. But the effective use of CAD tools has to be combined with a globally (circuit-specific) valid design methodology in order to deliver these robust results. RF circuit design is similar to any other analogue circuit design in that achieving acceptable performance parameters (including requirements and constraints) not only relies on the ability of the designer to strike valuable and convenient trade-offs using design parameters but also on the ability to extract the full potential of the available components in serving the needed purpose. This process becomes more complicated as the number of concerned performance parameters and/or the number of circuit components are increased. Such is the case with the target design of this work which requires the amplifier to achieve acceptable performance when operating as both an LNA and a PA for a low power ZigBee front-end.

In Sections 5.2.3 and 5.3.3, some of the recently realised and reported CMOS low noise amplifiers meeting challenging requirements were discussed. The design approaches which led to these performances and their limitations were discussed in Section 5.2.2. The lack of accurate, flexible and informative simultaneous analysis of performance requirements were highlighted as the most significant limitations.

The optimal design methodology is one that would simultaneously consider all performance requirements against all design parameters. But as the functional requirements of the circuit under design and the number of components involved increase, this becomes increasingly difficult to attain. Instead, increasing knowledge of how circuit elements operate and the significance of the influence certain design parameters have on certain performance parameters can facilitate a design approach where effective early eliminations can be made leading to as many performance parameters as possible being considered simultaneously against a limited number of design parameters.

The design methodology proposed here is built around this course of action. The representation is based on visual observation, in 3D graphs, of how performance parameters vary with related design parameters taking into account the multiple dependencies of performance parameters on design parameters. Components are simulated between their minimum and maximum values to explore their full potential and uncover areas of their operation where better results may be obtained. After some initial eliminations and considerations, most of the performance parameters are simulated with respect to the same design parameters. This allows the designer to simultaneously analyse the effect of these design parameters on all those performance parameters and hence make satisfactory trade-off decisions. This also gives the designer the ability to reuse the results of this analysis and change trade-off choices if another design with different constraints is needed. The design procedure is also based on simultaneously optimising the two stages of the cascode architecture, reducing the need for iterations in later stages of the design flow as has been the case with some of the work reviewed in previous sections.

The design methodology is laid out and discussed in Section 5.5.1. Based on this proposed plan, Section 5.5.2 provides an in-depth analysis of the cascode architecture taking into account all performance parameters of concern. In Section 5.5.3, a design example is presented to discuss how the results of the analysis in Section 5.5.2 should be applied. In addition the validity of the design approach is examined. All analysis, optimisation and design work presented in this section was carried out using components from a 180nm RFCMOS mixed signal design kit from UMC. This is the design kit with the smallest available CMOS transistors compatible with the design software, ADS.

5.5.1 Methodology and Design Approach

This section outlines the main aspects of the proposed design approach and determines the key steps of the design flow.

The design approach can be split into three main steps: The optimisation of the core circuit components (including bias voltages); the addition and optimisation of linearisation techniques; and the optimisation of input and output matching networks.

5.5.1.1 Optimisation of the Core Circuit Components

The core amplifier circuit (i.e. excluding the matching) to be optimised in this work is shown in Figure 5.20. It includes the two cascode transistors, the degeneration inductance, and the load inductance.

The optimisation procedures are developed with bearing in mind that the inductors of the cascode architecture will only insignificantly affect the DC characteristics of the amplifier due to their relatively small internal resistances. Therefore, the DC analysis can safely be done on the two cascode transistors only without including the inductors, and the results will still be valid when inductors are connected. Also, although the effect of these inductors on RF performance is well understood, tests show that in the cascode architecture, a compromised configuration (bias point and device width) for the cascode transistors produced from analysis of the circuit conducted without the inductors cannot be different than if the analysis was undertaken with the existence of the inductors, and hence the transistors cannot have a better configuration. In other words, although including the inductors will change the resulting circuit performance, if the optimisation (without the inductors) to select a suitable configuration for the two cascode transistors would have been performed with the inductors in place, then the chosen configuration for the transistors (based on the same judgement procedures) would have been the same. However, it is important to emphasise that this assertion is only valid for the Cascode LNA architecture due to the limited range of values of inductors that are commonly considered for use as either a degeneration inductance in the source of the CS stage or as a load inductance in the drain of the CG stage. Analogy can be made between this and operating on a small section of a nonlinear curve where the small section can be considered linear. In Section 5.5.2.5, two results are shown to demonstrate the validity of this assertion.

With appreciation of the above assertion, the first step in optimising circuit components is to tune the sizes and biases of the two cascode transistors; this is done on the circuit in Figure 5.20. This is the simple cascode architecture without the inductive degeneration.

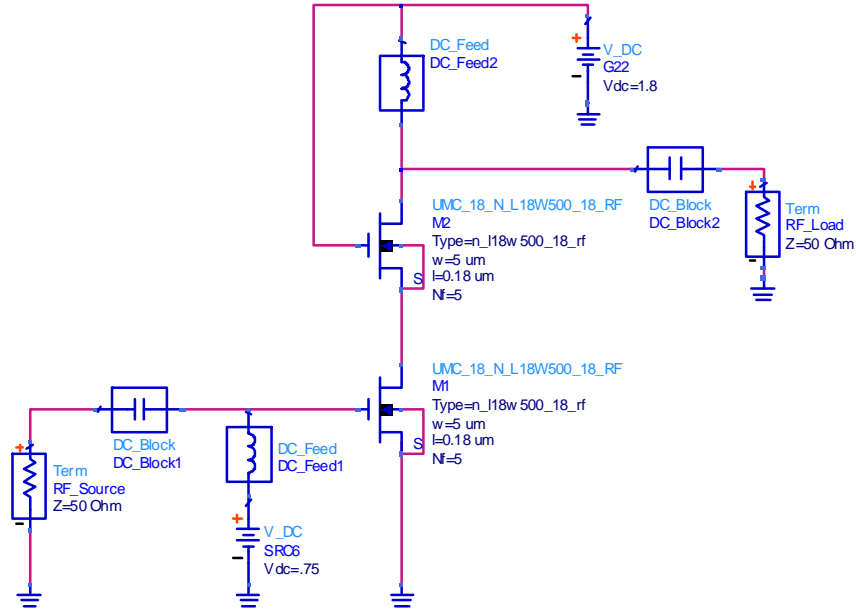


Figure 5.20 Initial optimisation circuit for the two cascode transistors

The general approach is to examine all the low and high-power performance parameters concerned against a sweep of sizes of the CS and CG transistors, after appropriately setting the bias voltages at the gates of each stage. In Figure 5.20, the gate of the common-gate transistor is connected to the amplifier's V_{DD} . This choice is motivated by consideration of Equation (5.26) [Zhenying 2003]

$$IIP3_2 = \frac{4}{3} \left| \frac{V_{GS_2} (2 + \theta V_{GS_2}) (1 + \theta V_{GS_2})}{\theta} \right| \quad (5.26)$$

Equation (5.26) is an expression for the input-referred 3rd-order intercept point of $M2$ where θ is the normal field mobility degradation factor and V_{GS_2} is the gate to source voltage of the common-gate transistor. As can be seen, this expression suggests that IIP_3 of the second stage can be enhanced by increasing V_{GS_2} . But V_{GS} of transistor $M2$ is also determined by the voltage between the two stages, hereafter called the “mid Voltage”, which is affected by V_{DD} (the bias voltage on the drain of $M2$) and the sizes of both transistors, and it can affect current consumption even though this is predominantly controlled by the common-source stage, as will be seen.

Hence, to increase flexibility and reduce the scope of possibilities, the gate of $M2$ is assigned a maximum voltage by connecting it to V_{DD} . This does not constitute a constraint on V_{GS_2} since full control over it will still be available through controlling the sizes of both transistors, as will be demonstrated in the DC analysis of Section 5.5.2.1. The bias voltage of the CS transistor is adjusted with respect to its noise performance, as will be seen. The RF choke placed on the

drain of the CG transistor isolates the supply voltage from the amplifier RF output, and keeps the amplifier independent, for now, of any specific load impedance for the purpose of optimisation.

In Section 5.5.2.1, a comprehensive DC analysis of the common-source common-gate architecture is carried out. This is undertaken to investigate the nature of variations of the DC current and the mid Voltage subject to varying transistor sizes.

The noise performance of the cascode architecture is investigated in Section 5.5.2.2 with emphasis on the bias and size of the CS transistor as the dominant contributor to noise in the amplifier. It is the intention that the most suitable CS gate voltage produced from this analysis will be used as a fixed value for the rest of the optimisation process. This does not imply any limit or constraint on any other performance parameters since the current of the CS transistor (which as will be shown controls the current of the circuit) is equally dependant on its gate voltage and size in the simulated range.

Power, gain and nonlinear analysis are performed in Sections 5.5.2.3 and 5.5.2.4 by investigating the effect of varying the sizes of the CS and CG transistors on several performance parameters at low and high input powers.

The second step in setting the core circuit components is to include the degeneration inductance and study its effect on both the low power and high power operation. The degeneration inductance affects low power operation in terms of noise for the LNA functionality and affects the nonlinear performance since it is classified as a negative feedback element. The results of these simulations are presented in Section 5.5.2.5.

In Section 5.5.2.6, the final step in setting core circuit parameters is performed by tuning the load inductor. This step is carried out after the transistor sizes and the degenerating inductor have been chosen. The load inductor has a considerable effect on two performance parameters; power gain and stability. The stability of the final amplifier circuit will depend on the impedances presented to the amplifier at the input and the output, as discussed in Section 2.7.3. In the best case, the core amplifier circuit (before matching) should present unconditional stability. This gives the designer more flexibility in matching the circuit for other performance parameters such as noise, gain and harmonic termination. However, trying to peruse this goal at the beginning of the design flow will significantly complicate the design process, add to design time and may result in unnecessary constraints on other desired performance parameters. Instead this step is left as the last step in the design of the core amplifier, and the value of the inductor can be decided as a simple trade-off of how it affects stability and gain only.

5.5.1.2 Employing Linearisation

The two linearisation techniques discussed in Section 5.3.2, the linearisation by multi-gating and linearisation by harmonic termination are employed in this design.

The auxiliary transistors for multi-gate linearisation cannot be added before deciding the final sizes and biases of the core transistors. The reason for this is that since the bias and width of the auxiliary transistors are tuned to cancel the g_3 nonlinearity of the main transistor, their configuration will not be effective if the configuration of the main transistor changed. If this step is performed as the size of the CS transistor is being swept for optimisation, it might give a better early indication on the nature of the expected nonlinear performance. But this would involve too many iterations as the configuration of the auxiliary transistors would have to be changed every time the CS transistor configuration is changed; hugely increasing design time. It is also not possible to write a mathematic routine in the simulator to automatically re-adjust the sizes and biases of the auxiliary transistors for g_3 cancellation whenever the core transistor is changed. Studying Equation (5.27) [Toole 2004] reveals why this is the case.

$$g_3 = \frac{K\xi(1 + \lambda V_{DS})}{24(n\phi_t)^2(1 + \xi)^3(1 + \alpha X)^4} (24n^2\phi_t^2\xi^2\alpha - 12n\phi_t\xi(1 + \alpha X) + X(2 + \alpha X)(1 + \alpha X)^2) \quad (5.27)$$

where

$$X = 2n\phi_t \ln \left(1 + e^{\frac{(V_{gs}-V_{th})}{(2n\phi_t)}} \right), \quad \alpha = \theta + \frac{\mu_0}{2nv_{sat}L}, \quad K = \frac{\mu_0 C_{ox} W}{2nL} \quad \text{and} \quad \xi = e^{\frac{(V_{GD}-V_{th})}{2n\phi_t}}$$

Definition of various parameters in equation (5.27) can be obtained by reference to the original paper ([Toole 2004]). Equation (5.27) represents the relationship between the 3rd-order transconductance nonlinearity of a common-source transistor in moderate inversion with its dimensions and bias. Inspection of Equation (5.27) shows that because of the exponential term (in X and ξ), no analytical mathematical solution can be found. This problem has to be solved numerically in order to readjust the bias and size of the auxiliary transistors. Alternatively, the auxiliary transistors widths and biases can be adjusted manually in the simulator.

In light of this, a better approach for optimising the auxiliary transistors is to choose the width of the core transistors based on the various amplifier performance parameters (such as noise, gain, PAE, etc.) and then linearise that by MGTR. Since it does not significantly affect the DC current in the cascode, adding the auxiliary transistors will not affect any other performance parameters set by optimising the main transistor, but will only improve its linearity.

The linearisation by MGTR method will still be valid at relatively high power because although the instantaneous g_m will change with higher powers, the average g_m (around the DC bias) will remain the same.

The harmonic termination will be incorporated within the source- and load-pull analysis which will be extensively extended (by increasing the number of simulated-for impedances) to take into account all possible terminations. Any impedance representing considerable termination to any frequency component which contributes to IMD_3 will be reflected in the input or output contours and can then be picked by the designer for matching.

5.5.1.3 Matching the Amplifier

In this design methodology, matching the amplifier comes as the last step after setting the core circuit parameters and applying linearisation; this is because these two processes contribute to the impedances seen at the input and the output of the amplifier core circuit. The importance of matching for low and high power operation was discussed in Sections 2.7 and 5.3.1.2, respectively. In high power operation, the matching techniques discussed in Section 2.7 are no longer valid; because they are primarily based on S-parameters which is essentially a small signal analysis. For example, the load of the amplifier cannot be matched for maximum power transfer with the conjugate matching requirements obtained from S-parameter measurements when the amplifier is operating at its 1dB compression point which is a large-signal operating condition. Instead, source and load pull simulations have to be performed. Source and load pull simulations consider a number of different matching impedances spread across the Smith chart (pre-specified by the designer) and investigates the response of the circuit in terms of obtainable performance parameters when these impedances are presented to the input and the output of the core amplifier circuit, respectively.

In Section 5.5.3.2 the matching of the amplifier will be discussed by simultaneously considering low and high power matching techniques, giving the designer the flexibility of selecting the desired trade-off.

5.5.2 In-depth Analysis of the Cascode Architecture at Low and High Input Powers

This section presents the circuit analysis of the cascode architecture following the proposed design methodology outlined in Section 5.5.1.1, and discusses its results. Based on these results, component values of the core amplifier circuit will be set in the design example of Section 5.5.3.

5.5.2.1 DC Analysis of the Cascode Architecture

In this section, the DC current and DC voltage of the cascode architecture are analysed based on the circuit in Figure 5.20. The analysis is performed over a range of sizes of the common-source and common-gate transistors. Figure 5.21 and Figure 5.22 show how the DC current and the DC voltage at the mid point between the two stages vary with respect to different combinations of sizes of the two transistors, at several supply voltages (voltage at the drain of the CG transistor). CS NGF is the number of gate fingers of the CS transistor and CG NGF is the number of gate fingers of the CG transistor, each gate finger is $5\mu\text{m}$ wide.

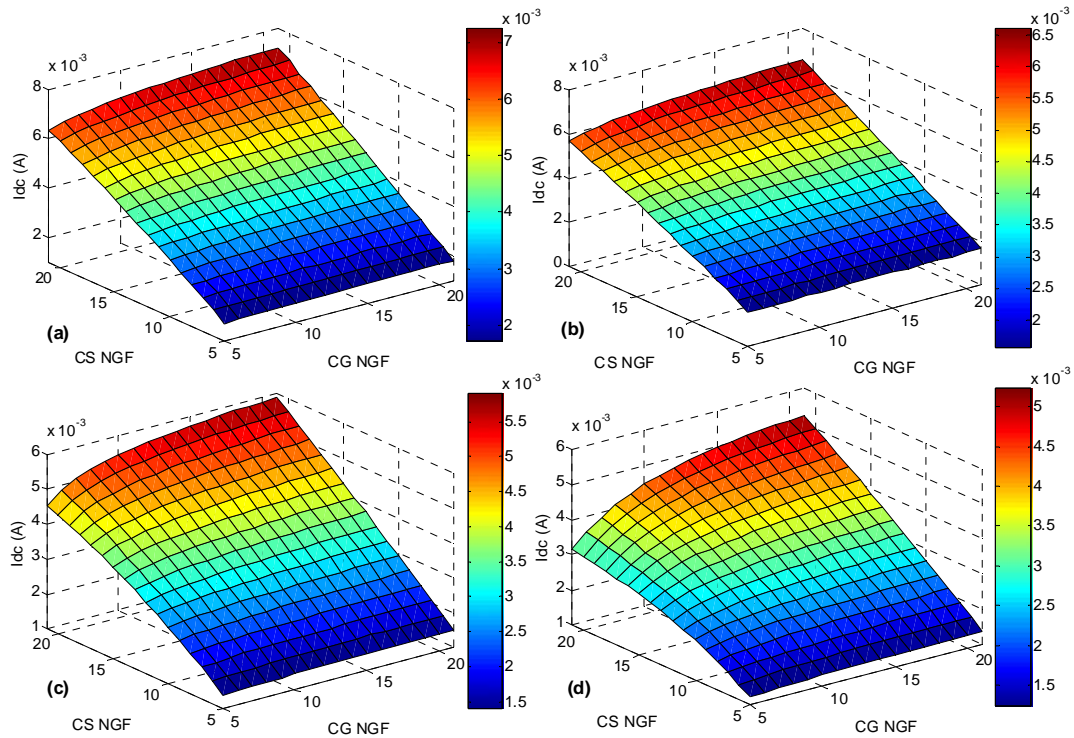


Figure 5.21 Amplifier DC current vs. CS and CG sizes at voltage supply of (a) 1.8V, (b) 1.5V, (c) 1.2V and (d) 1V

Figure 5.21 clearly indicates a higher dependence of the DC current on the size of the CS transistor than on the size of the CG transistor. This is because the gate-to-source voltage of the common-source transistor is fixed, unlike the V_{GS} of the common-gate device. As from Figure 5.22, because the voltage at the gate of the CG transistor is fixed, and since the voltage on its source is the mid voltage, its V_{GS} changes every time the sizes of the transistors change. In this sense, the CS transistor acts as a current source, the current of which depends on its size and gate voltage. The CG transistor absorbs that current and the only change it observes is the distribution of that current beneath each of its gate fingers, but the total current generally does not change as a result of changing the size of the CG transistor.

An important observation to note from Figure 5.21 is that the dependence of the DC current on the size of the CG transistor changes as the supply voltage varies. As shown in Figure 5.21, this dependence increases at smaller sizes of CG device as the supply voltage drops when the CS transistor is large. This can be explained as follows: when the CS transistor size is increased, it demands more current, and since the two transistors are cascoded, this current has to be passed through by the CG transistor. At higher values of supply voltage, the CG device does that by changing its overdrive voltage. This is evident from Figure 5.22.a; the mid voltage decreases as the size of the CS transistor increases, hence increasing V_{GS} of the CG transistor, facilitating the current increase. However, Figure 5.22 shows that as the CG transistor size decreases, the mid voltage decreases. This affects the current on two fronts: First, since the decrease in the mid voltage is not proportional to the decrease in the supply voltage, the CG transistor can no longer attain high V_{GS} at lower supply voltages as it does when the supply voltage is high. Second, the mid voltage may be reduced to a level effectively taking the CS device out of its saturation region. When the CS transistor is in the linear region, the current it tries to draw decreases and the more the CG device size reduces (when CS transistor is large) the more the mid voltage reduces and hence the CS transistor is pushed further away from saturation. These two notions explain the increased dependence of the DC current on CG transistor size at lower supply voltages. Note that this observation was only possible to make as a result of this unique 3D graphical representation of the current changes, insight that would be very hard to acquire by studying model equations describing the DC current in the cascode configuration alone.

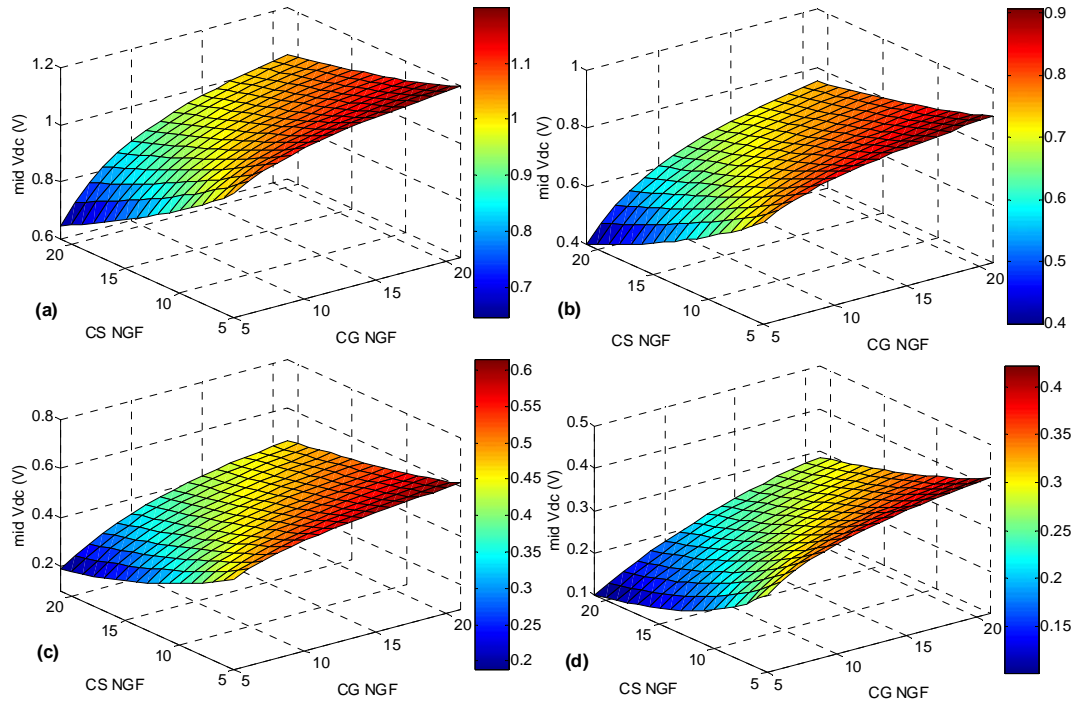


Figure 5.22 Amplifier mid voltage at a voltage supply of (a) 1.8V, (b) 1.5V, (c) 1.2V and (d) 1V

Figure 5.22 is also useful in guiding the designer on how the sizes of the CS and CG transistors might affect their areas of operation taking them between the linear and saturation regions. The CG transistor will always be in saturation since the DC voltage on its gate is the same as that on its drain. This means, with any voltage at its source (the amplifier mid voltage), its V_{DS} is always going to be larger than $V_{GS} - V_T$. For the CS transistor, however, the situation is different. If the DC voltage on its gate is set for example to 0.75 volts, Figure 5.22.d shows that when the supply voltage is set to 1V, there is a range of CS and CG sizes that is going to take the CS transistor out of saturation. The same applies to Figure 5.22.c but as the supply voltage increases to 1.2V, the range of sizes taking the CS transistor out of saturation reduces. It can also be observed from Figure 5.22 that the mid voltage is highest when CS device is very small and CG transistor is very large, this is because in this case, the effective resistance of the CS transistor is at its highest while the effective resistance of the CG transistor is at its lowest, and therefore there is more voltage drop on the CS transistor than on the CG transistor. Opposite analysis explains why the mid voltage is at its lowest when the CS transistor is at its largest size and the CG transistor is at its smallest size.

Since the DC current in the cascode is largely more dependant on the CS transistor rather than the CG transistor as explained above, it is useful to study how this current is affected by both the size of the CS device and gate voltage. Figure 5.23 presents this relationship.

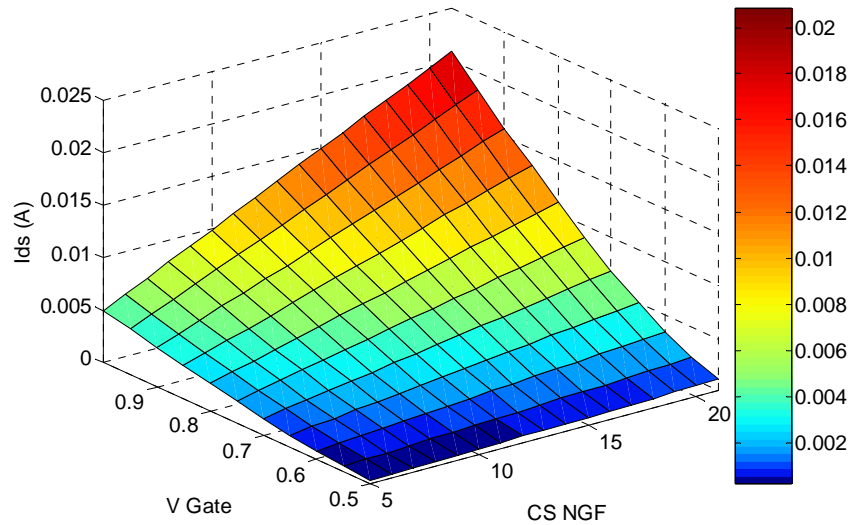


Figure 5.23 Common-source transistor's current versus its size and gate voltage

As shown in Figure 5.23, the DC current in the CS transistor is equally dependant on its gate voltage and size for the span simulated, which is the maximum span for the transistor size and the maximum realistic (for low power operation) span of the gate voltage. This feature will be useful in taking design decisions at later stages of the design flow.

5.5.2.2 Noise Analysis of the Cascode Architecture

Figure 5.24 shows how the minimum Noise Figure (NF_{min}) changes with respect to different combinations of sizes of the CS and CG devices. This graphical representation of the relationship generally agrees with the mathematical equation-based conclusions of [Park 2001] and [Guo 2002] that the noise figure is almost independent of the size of the common-gate transistor. However, because this graphical representation is more comprehensive, it shows that this general conclusion starts to become questionable when the CS transistor is significantly bigger than the CG transistor. The reason for this can be extracted from the DC analysis in Figure 5.21 and Figure 5.22 as follows: At larger sizes of the CS transistor, and as the size of the CG transistor reduces, both the total current and the mid voltage decrease. This means the CS transistor is retreating from its relatively high gain operation region to its linear region, inevitably increasing the minimum obtainable noise figure.

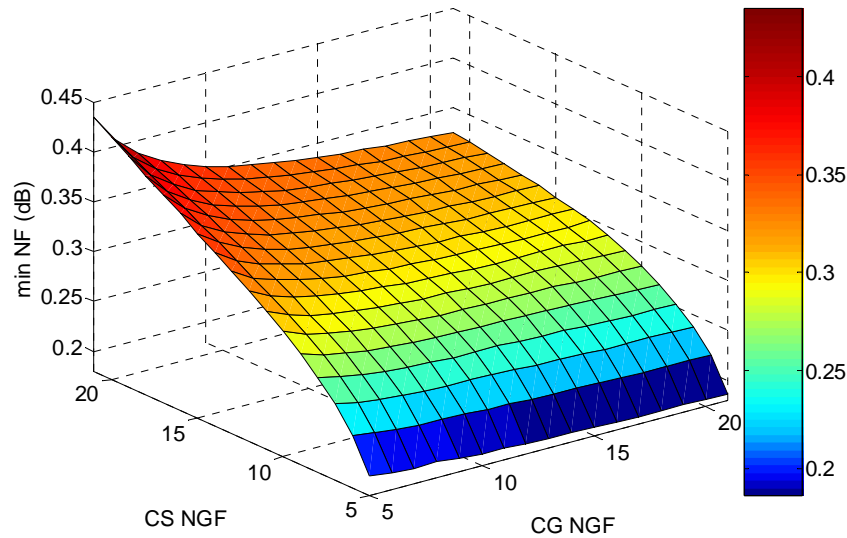


Figure 5.24 Minimum noise figure vs. transistor sizes

Based on the results of Figure 5.24, and given that it is very unlikely for the size of the CS device to be chosen less than $50\mu\text{m}$ (since this will mean very small current and insufficient gain), it follows that it is safe to optimise the noise performance of the cascode by considering the noise performance of the CS transistor alone, so long as the fact that the bias condition of the CG transistor will affect the voltage on the drain of the CS transistor is taken into consideration. This implies that the CS transistor noise analysis has to be repeated for a range of possible values of mid voltage generated in Figure 5.22. The voltage on the drain of the CS device is important because it will affect the current through it and hence the noise. Therefore, in order to take into account the three parameters of the CS transistor that affect its noise performance (its size, gate and drain voltages) the minimum noise figure is simulated with respect to the gate voltage for different transistor sizes, and this simulation is then repeated for

several drain voltages. The results of these simulations are presented in the 3D graphs of Figure 5.25.

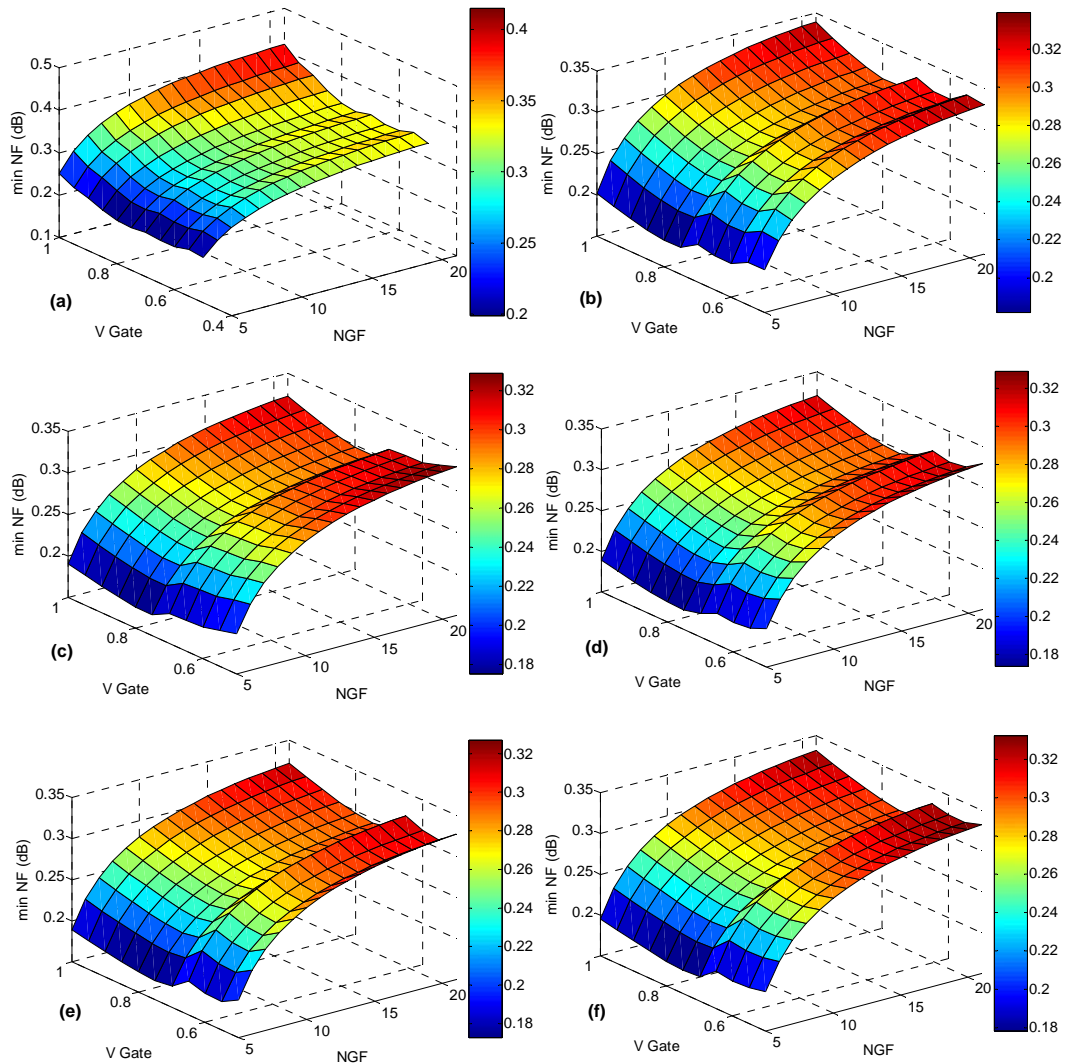


Figure 5.25 Minimum noise figure of a common-source transistor versus its gate voltage and size at drain voltages of (a) .3V, (b) .5V, (c) .9V, (d) 1.2V, (e) 1.5V and (f) 1.8V

Figure 5.25 presents an interesting and very useful result which is that there is a persistent reduction in the minimum noise figure at gate voltage in the proximity of 0.75V and that is maintained for all sizes of transistor, at all drain voltages. Figure 5.25 also shows that the noise figure generally decreases as the supply voltage on the drain of the common-source transistor increases, which agrees with the conclusions of [Wu 2006].

In Section 5.5.2.5, the effect of the degeneration inductor on noise will be analysed when the optimisation of this component is discussed.

5.5.2.3 Power and Gain Analysis of the Cascode Architecture

In this section, the influence of varying the sizes of the two transistors in the cascode architecture on power consumption, gain, and power-added efficiency is analysed. The analysis is performed by simulating for the various parameters at different input powers to examine the effect on the performance of the amplifier in low and high power regimes of operation. Figure 5.26 shows how the DC current drawn from the supply changes as the transistor sizes are varied, and at different input powers.

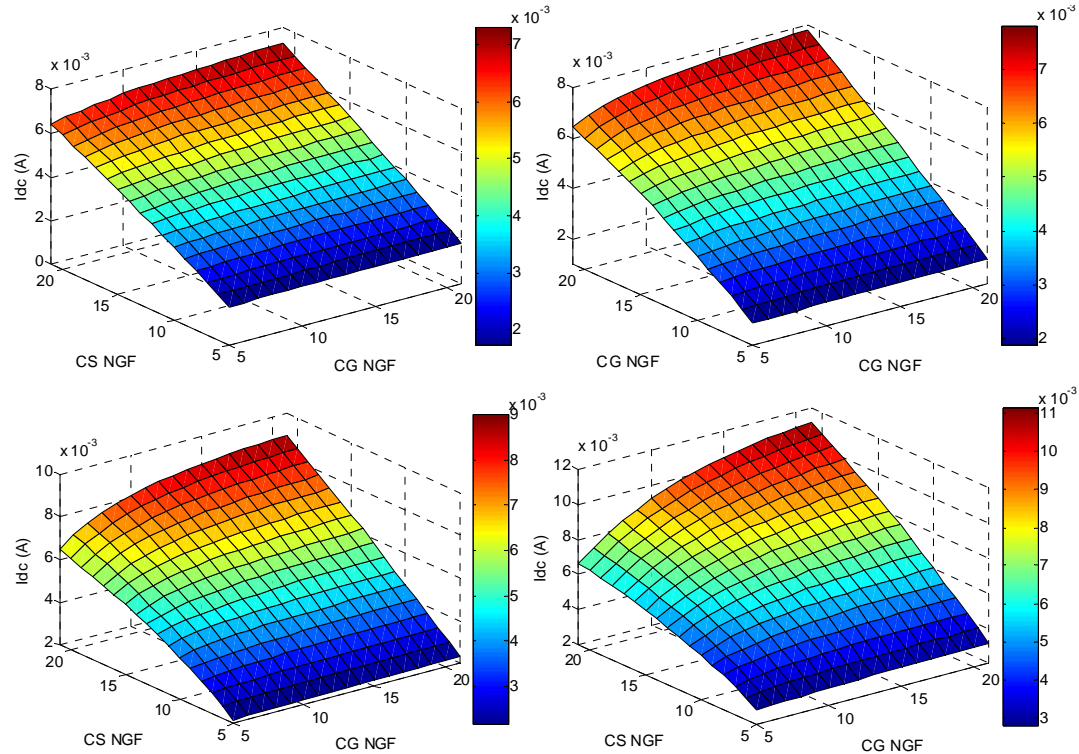


Figure 5.26 The DC current of the amplifier drawn from the supply at an input power of (a) -20 dBm and (b) -10 dBm (c) -5 dBm and (d) 0 dBm

It is observed that there is no noticeable difference between Figure 5.21.a and Figure 5.26.a, that is, the low power operation does not change the DC power characteristic from the DC analysis. However, by inspecting the graphs in Figure 5.26.b, c, and d, and comparing with Figure 5.26.a, an interesting observation can be made. That is, in the low power regime, the current is more influenced by the size of the CS transistor than the CG transistor, but as the input power increases, an extended dependency on the size of the CG transistor is introduced. In fact, when the CG transistor is small, i.e. its transconductance is small and does not dominate the overall performance, the DC current is the same at all levels of input powers. But at higher input powers, and as the size of the CG transistor increases, the current drawn from the supply starts to increase. The justification of this phenomenon can be related to non linearity since as the input power increases, the two active devices are pushed more into their non linear regions. But

the fact that the increase in DC current at high powers comes as the size of the CG transistor become larger means that the nonlinearities (particularly for this case the even-order nonlinearities which produce DC components) are produced more from the CG transistor than from the CS transistor and hence the CG transistor dominates the nonlinear performance. This result generally agrees with the conclusions of [Park 2001] and [Guo 2002], which were based on mathematical equations, but further gives a visual presentation of the extent of the effect. In Section 5.5.2.4, the magnitude of the dominance of the CG transistor on the nonlinear performance will be analysed in more details demonstrating the limitations that apply to this verdict.

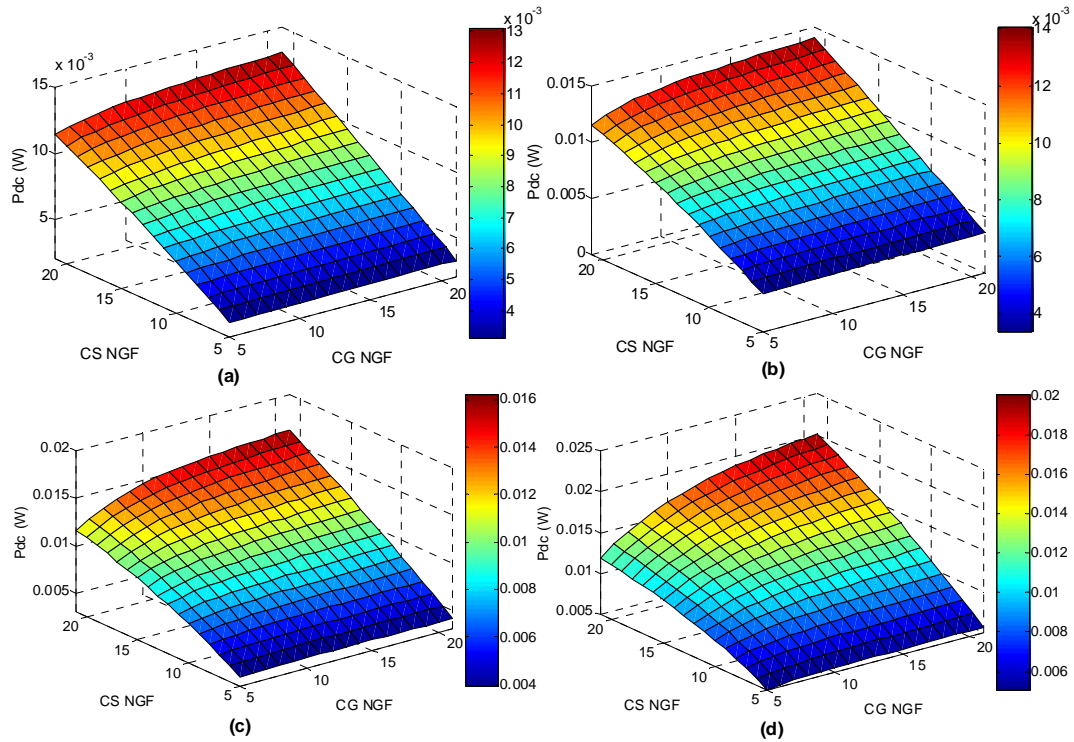


Figure 5.27 DC power consumption vs. transistor sizes at low and high input powers (a) -20 dBm and (b) -10 dBm (c) -5 dBm and (d) 0 dBm

Figure 5.27 shows how the DC power consumption changes with respect to varying transistor sizes at different input powers. As expected, there is a direct relationship between the DC current and the DC power consumption. Nevertheless, this graph is important as it informs the designer on the expected DC power consumption for given transistor sizes.

Figure 5.28, Figure 5.29, and Figure 5.30, respectively, relate the gain, power delivered to the load, and PAE to the CS and CG device widths. These plots show that all these performance parameters merely depend on the current in the cascode. These graphs can be used simultaneously by the designer to make trade-off decisions when choosing the transistor sizes as will be demonstrated in the design example of Section 5.5.3.

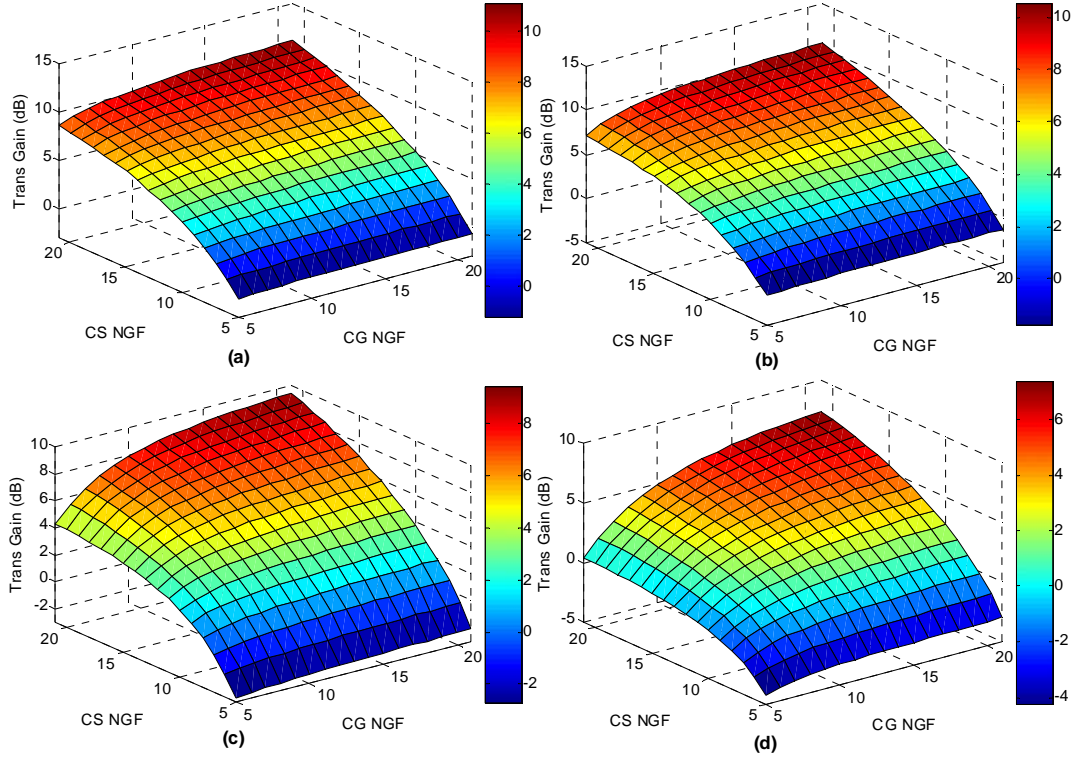


Figure 5.28 Transducer gain vs. transistor sizes at low and high RF input powers (a) -20 dBm and (b) -10 dBm (c) -5 dBm and (d) 0 dBm

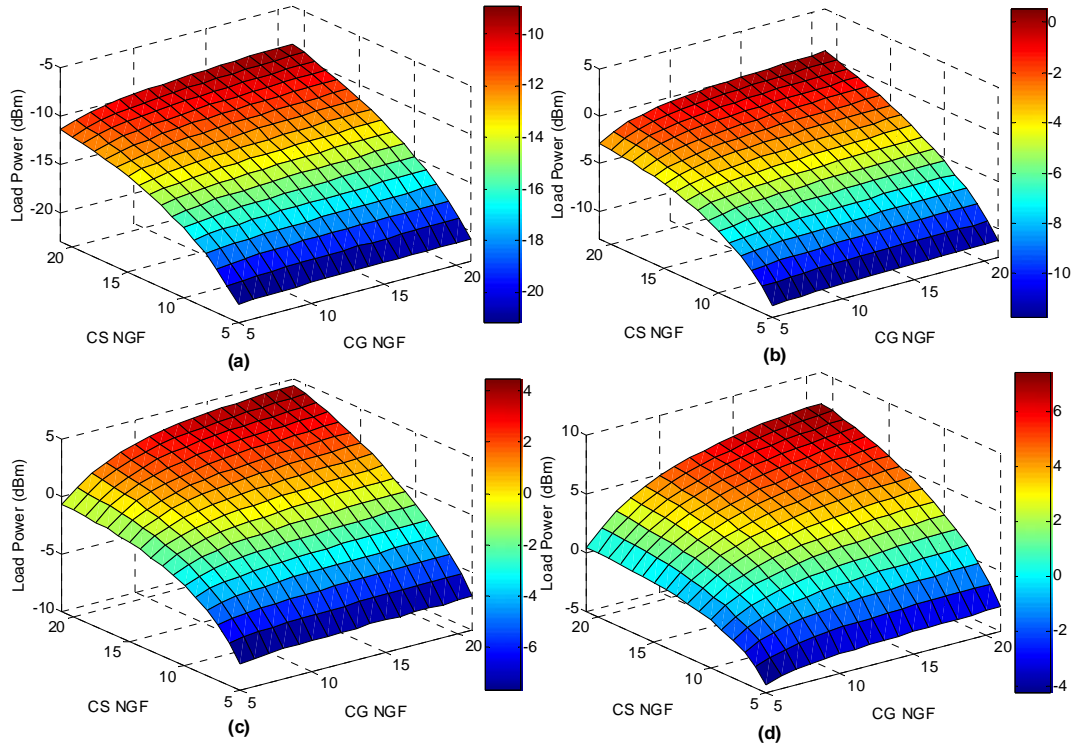


Figure 5.29 Load Power vs. transistor sizes at low and high RF input powers (a) -20 dBm and (b) -10 dBm (c) -5 dBm and (d) 0 dBm

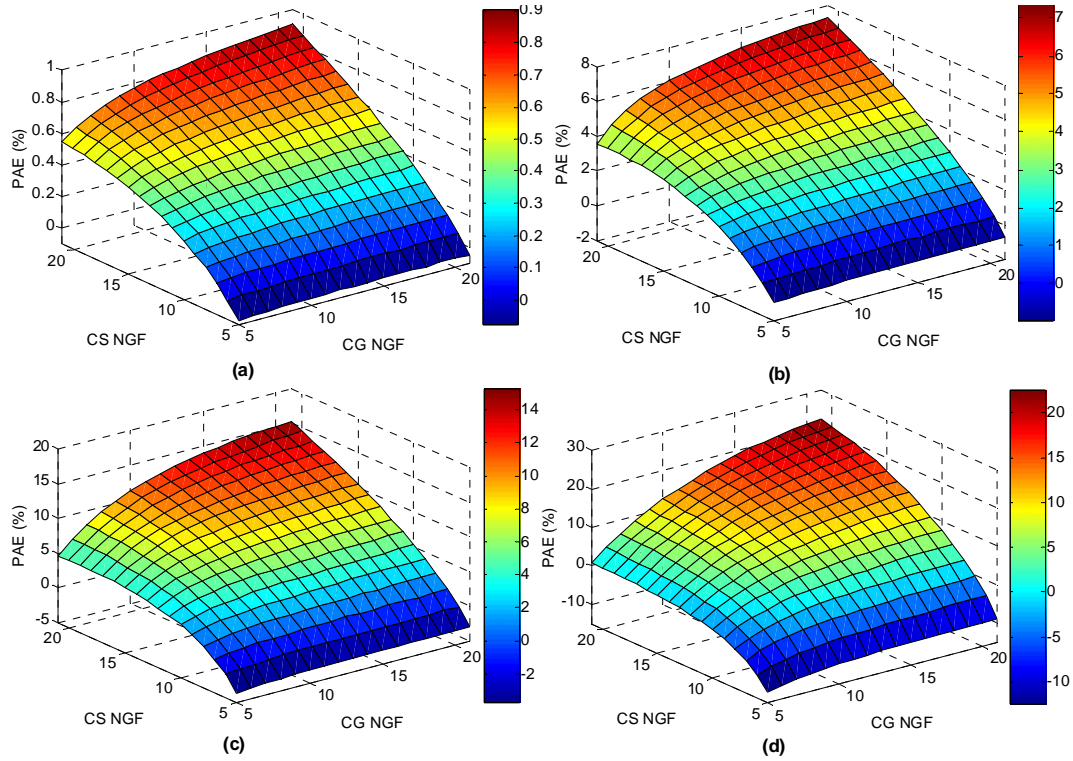


Figure 5.30 PAE vs. transistor sizes at low and high RF input powers (a) -20 dBm and (b) -10 dBm (c) -5 dBm and (d) 0 dBm

It is important to emphasise that these results present grounds for comparison on which the designer can make early trade-off decisions but are not definitive final results of the amplifier. This is because, particularly with the dual functionality amplifier intended in this thesis, trade-off decisions on the matching conditions of the amplifier are yet to be made.

5.5.2.4 Detailed non-Linear Analysis of the Cascode Architecture

The non-linear behaviour of the cascode architecture is studied by analysing how the relative 3rd-order intermodulation distortion (the magnitude of the 3rd-order intermodulation component relative to the magnitude of the fundamental component) varies when the sizes of the common-source and common-gate transistors are varied. Frequency separation between the two input tones in this two-tone test is set to 5MHz, which is the frequency separation between any two channels in the 2.4GHz ZigBee band. Figure 5.31 shows in 3D graphs how the relative 3rd-order intermodulation changes with respect to the sizes of the CS and CG transistors, at the low and high power operation regimes.

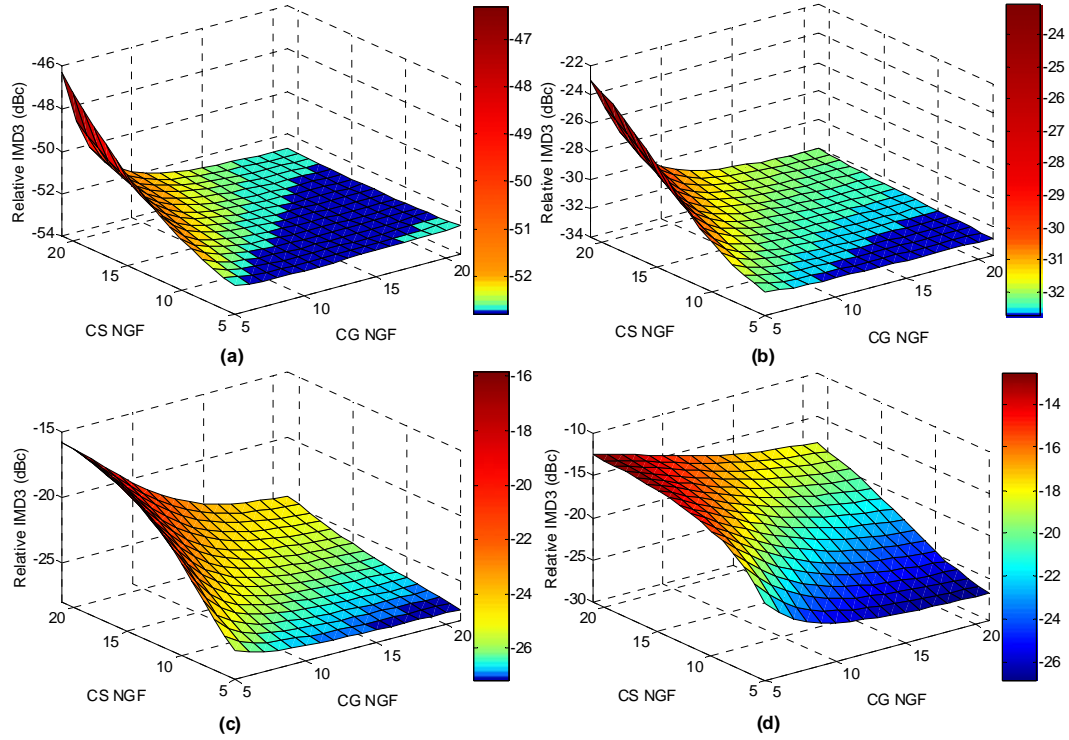


Figure 5.31 Relative 3rd-order intermodulation in dBc with respect to transistor sizes when the RF input power is (a) -20 dBm (b) -10dBm, (c) -5dBm and (d) 0dBm

Figure 5.31 demonstrates the extent to which the assertion in the literature discussed in Section 5.2.1.2 that the CG transistor dominates the nonlinear performance in the cascode architecture is valid. In Figure 5.31.a, with the input power at -20dBm, it can be seen that when the CG transistor is large ($CG\ NGF > 13$) there is almost no effect from the change in its size on the relative IMD_3 , and that is true for all sizes of the CS device. Outside this region ($13 > CG\ NGF > 5$) the effect of the CG transistor size on the IMD_3 is larger than that of the CS device. This is noted by observing the change in IMD_3 relative to 1 unit size up or down of both transistors. This observation is manifested more at larger sizes of the CS device and starts to become untrue as the CS device becomes smaller.

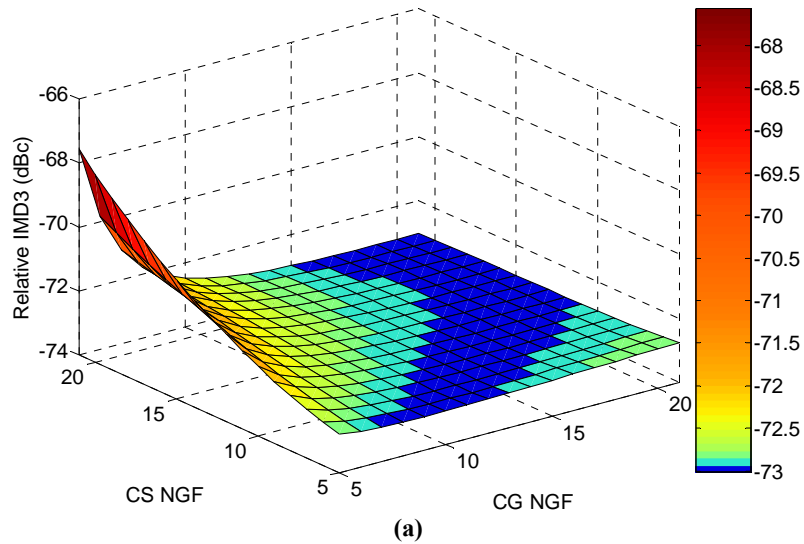
Therefore, a conclusion can be drawn from Figure 5.31.a that, when operating at small input powers, the effect of the CG device size on IMD_3 increases as the size of the CS transistor increases. This is another phenomenon that demonstrates the advantage of this design method which is based on 3D visualisation of parameter variations. It can be explained by the fact that when the CS device and hence its transconductance are large, it delivers a larger signal to the CG transistor driving it more into the nonlinear region, and hence the nonlinearities of the CG device are more apparent.

The observation drawn from Figure 5.31.a is still true in Figure 5.31.b when the input power is increased to -10dBm. However, it can be observed that when the CS transistor is large, the

range where the change in CG device size has no effect on IMD_3 starts to decrease. In Figure 5.31.c, when the input power is increased to -5dBm, this range becomes minimal and in Figure 5.31.d when the input power is 0dBm it becomes non existent. It is also observed that with the 0dBm input power in Figure 5.31.d, the effect of the CG size on IMD_3 is no longer dependant on the CS size. This is because this input power is already very large, and irrespective of whether the CS amplifies the input signal significantly or not, both transistors are driven well into their nonlinear regions and their effect on IMD_3 is almost identical.

Therefore, the analysis of Figure 5.31 concludes that the assertion in the literature ([Park 2001 and Guo 2002]) that the CG transistor dominates nonlinear performance is only partially valid at small input powers, and becomes an even weaker approximation as the operating input power increases.

Figure 5.32.a and b show the 3rd-order intermodulation and the IIP_3 relative to changes in the CS and CG device sizes. Note the complete opposite colour-coded mapping between these two figures which shows that there is a direct relationship between the relative 3rd-order intermodulation and IIP_3 . This feature will be used in considering amplifier optimisation in the design example of Section 5.5.3.



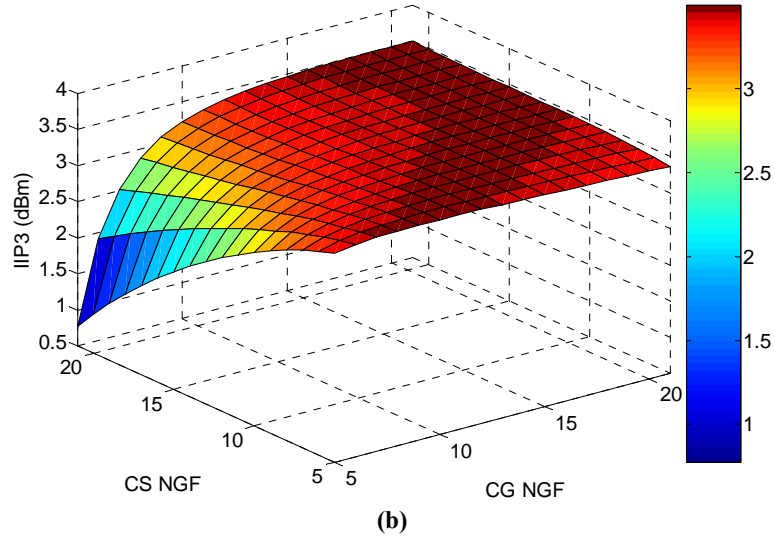


Figure 5.32 Mapping between IMD_3 and IIP_3 calculations at -30dBm input power
(a) IMD_3 and (b) IIP_3 vs. Transistor sizes

5.5.2.5 Effect of the Degeneration Inductance

The importance of the degeneration inductance in the cascode architecture has been discussed in Section 5.2.1. The feedback provided by the degeneration inductance has two effects on the nonlinear performance of the amplifier. Since it is a negative feedback, it reduces the nonlinearity, and may lead to a smaller IMD_3 . However, some frequency components of that feedback contribute to increasing IMD_3 as explained in Section 5.4. Therefore, careful analysis has to be conducted to evaluate the extent of these two effects on the nonlinear performance and establish whether one dominates the other. Figure 5.33 shows how the relative 3rd-order intermodulation distortion varies with changing transistor sizes when different values of degeneration inductances are used. This analysis is carried out at 0dBm since with this relatively high input power the nonlinearities of the circuit are manifested more than at low power operations.

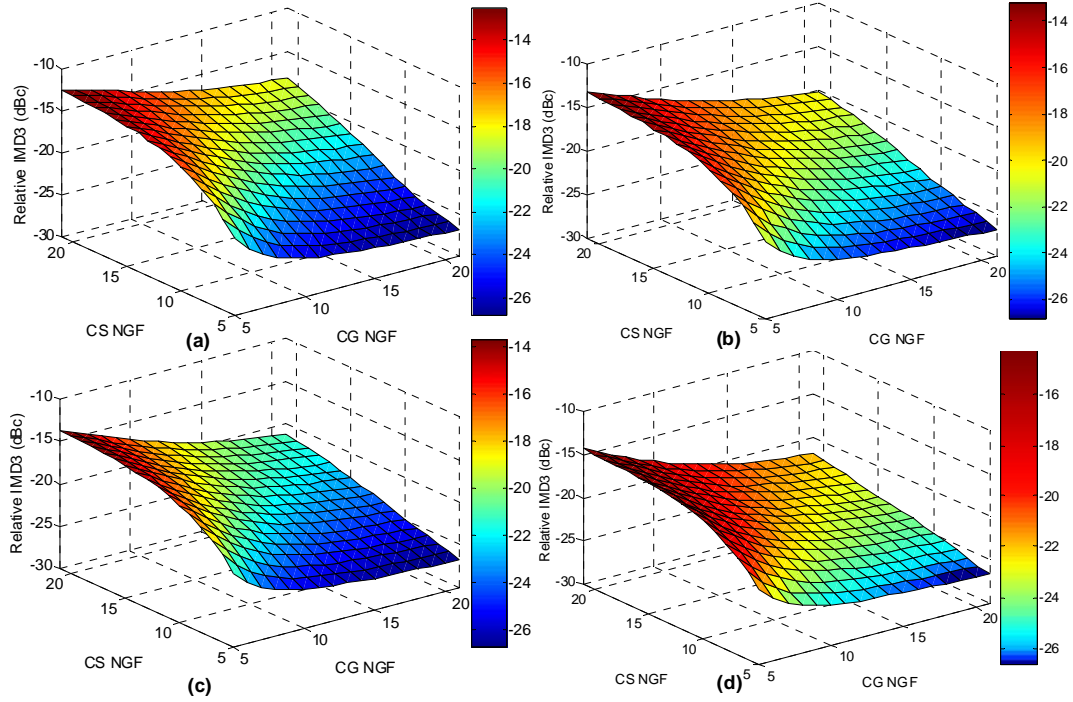


Figure 5.33 IMD_3 vs. transistor sizes at different values of degen inductors (a) no degen (b) 0.58nH (c) 1.5nH and (d) 2.9nH when the input power is 0dBm

An important observation from Figure 5.33 is that the manner of variation of IMD_3 with transistor sizes does not change as the value of the degeneration inductance changes, only the value of IMD_3 changes. This confirms the initial assertion in Section 5.5.1.1. that within a certain range of inductor values, the existence of the inductor does not change the nature of variation of RF performance parameters with respect to transistor sizes in the cascode and hence initial analysis for these performance parameters with respect to transistor sizes can safely be done without the inclusion of the degeneration inductance.

Figure 5.33 also shows that as the value of the degeneration inductance increases, IMD_3 decreases. This indicates that the effect of nonlinearity improvement through negative feedback exceeds the increase in IMD_3 contributed by the mixing of certain feedback components with the fundamental in various orders of nonlinearity as explained in Section 5.4.

As previously discussed in Section 5.2.1.1, the degeneration inductance also affects the noise figure of the amplifier. Figure 5.34 shows how the minimum noise figure varies with changing transistor sizes when different values of degeneration inductances are used. As with all the 3D graphs presented in this analysis, these results will be used in the design example of Section 5.5.3 to select a suitable value for the degeneration inductance.

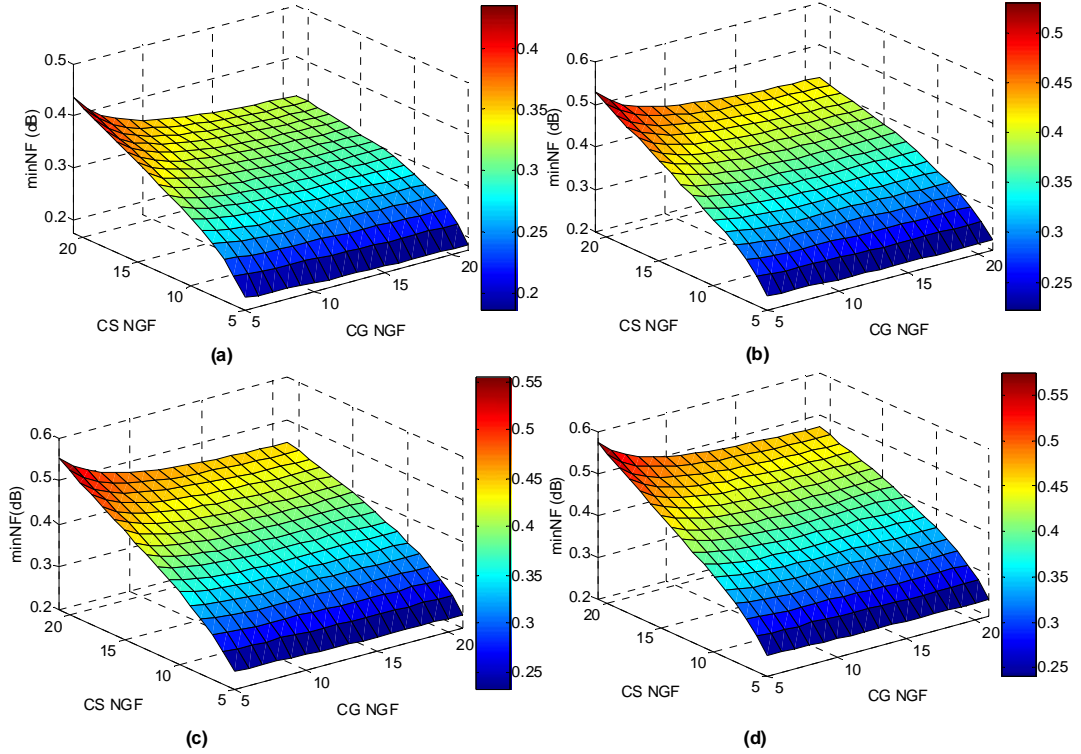
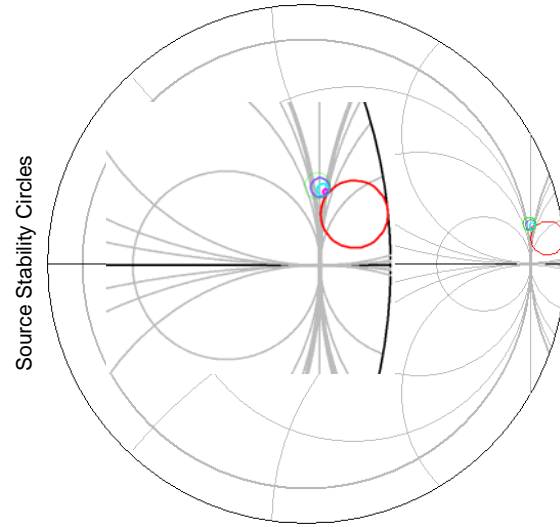


Figure 5.34 Minimum Noise Figure vs. transistor sizes at different values of degeneration inductors (a) no inductor (b) 0.58nH (c) 1nH and (d) 1.5nH

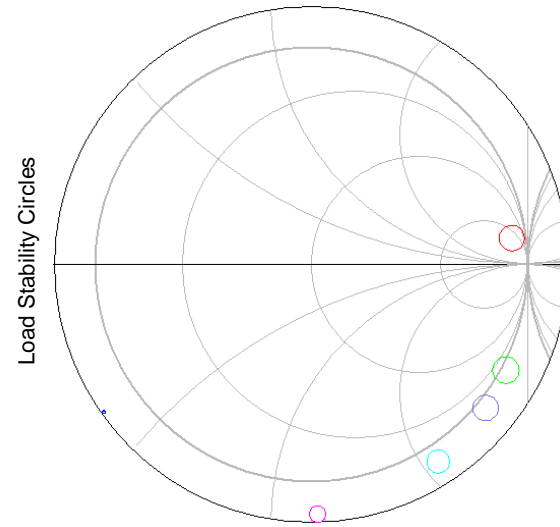
Figure 5.34 is another example to demonstrate the validity of the assertion in Section 5.5.1.1. It is clear that the manner of variation in the minimum noise figure with respect to transistor sizes does not significantly change as the value of the degeneration inductance changes. Figure 5.34 also shows that the minimum obtainable noise figure will increase as the value of the degeneration inductance increases.

5.5.2.6 Choice of the Load Inductor

Section 2.7.3 showed how stability circles can be used to determine the stability of an amplifier. The load inductor influences two performance parameters; the stability (because it acts as a resistive load in the output), and gain (since it is the amplifier load). The influence of the load inductor was investigated for different on-chip inductor values whilst observing the location of the stability circles and the effect on power gain. Figure 5.35 shows how the value of load inductor produces different stability circles at the input and output. Table 5.3 lists the stability regions of these circles and the maximum obtainable output power for each on-chip inductor at low and high input powers; obtained from load-pull simulations.



(a)



(b)

Figure 5.35 Effect of the on-chip load inductor on the stability and gain of the amplifier

Table 5.3 Regions of stability and maximum gain relative to different load inductors

	Input Stability region	Output stability region	Unconditionally stable	Maximum output power at -30 dBm	Maximum output power at 0dBm
— Choke	Inside	Outside	no	-13.7dBm	9.55dBm
— 1nH	Outside	Outside	yes	-24.7dBm	0.03dBm
— 3nH	Outside	Outside	yes	-17.5dBm	6.8dBm
— 5nH	Outside	Outside	yes	-15.5dBm	8.3dBm
— 7nH	Outside	Outside	no	-14.7dBm	8.8dBm
— 9nH	Outside	Outside	no	-14.4dBm	9dBm

Figure 5.35.a and b show how the stability situation dramatically changes when an ideal inductor (choke) is replaced by on-chip inductors; confirming that the variation is due to the series resistance of the inductors which result in resistive loading at the output. Also noticeable is the dependence of location of input and output stability circles on inductor value. Two further observations can be made from the result in Figure 5.35 and Table 5.3. The first is that a higher value of load inductor does not necessarily mean increased stability as the amplifier becomes conditionally unstable at the two highest values 7 and 9nH. The second observation is that as the value of the load inductor increases, the power gain increases. This is because the total load impedance is increased, as discussed in Section 5.3.1.2. It is worth noting that when performing this analysis, it is important to recall that increasing the on-chip inductor value increases the chip area, which is not desirable for a low cost product.

5.5.3 High Dynamic Range Cascode Amplifier – A design example

This section builds on the analysis in Section 5.5.2 of the common-source common-gate cascode architecture operating at low and high powers and implements the proposals in Section 5.5.1 to complete a design example demonstrating the validity and usefulness of the design approach. An amplifier for the ZigBee front-end that can be used as both an LNA and a PA with acceptable performance requirements is next described.

5.5.3.1 Setting Up The Core Amplifier Circuit

This sub Section is dedicated to discussing the choices made in setting up the core circuit components of the amplifier which includes the degeneration and load inductors, both cascode transistors and their bias voltages.

▪ Bias Voltages

Three bias voltages need to be determined in the cascode architecture; the main supply voltage (drain of the CG transistor), the gate voltage of CG device, and the gate voltage of the CS transistor. The analysis presented in Section 5.5.2 is adaptable and can be repeated for any value of supply voltage if the design strategy is to investigate the effect of reducing the supply voltage and hence power consumption on the other remaining performance parameters. Due to size constraints, this option was not taken here, and the supply voltage is set at 1.8V which is flexible enough to allow optimisation of other circuit elements.

The choice of the CG gate voltage was explained in Section 5.5.1.1. In the noise analysis of Section 5.5.2.2, Figure 5.25 shows that a broad minimum in noise performance is obtained for a gate voltage of around 0.75V, independent of device width. Since the amplifier noise is strongly related to current and there is an equal influence on current from the gate voltage and the

transistor size in a common-source transistor as depicted by Figure 5.23 in the DC analysis of Section 5.5.2.2, the gate of the CS transistor was set to 0.75V. This does not represent a constraint on any other performance parameter since full control over the DC current is still available through varying the size of the CS transistor.

▪ Transistor sizes

The choice of transistor sizes highlights the power of the design methodology developed in this work; the ability to simultaneously analyse the effect of the two most important design parameters (the sizes of both transistors) on various performance parameters. Simultaneously using the results presented in Sections 5.5.2.2, 5.5.2.3 and 5.5.2.4 and considering the required dual functionality of this amplifier, it can be concluded that, in general, a large CG transistor and a relatively smaller CS transistor will achieve a good compromise. The noise analysis in Figure 5.25 indicates that as the size of the CS transistor decreases, the minimum noise figure decreases. From Figure 5.31, the lowest IMD₃ at low powers can be achieved only when the CS device is very small but over a wide range of large CG transistor sizes. At high power, the same level of minimal IMD₃ can only be achieved when the CG device is at its maximum size as long as the CS transistor remains as small as possible.

Figure 5.26 clearly indicates that if CG transistor was to be large, the size of the CS transistor alone controls the DC current, and as a consequence, DC power consumption, gain, delivered power, and PAE as shown in Figure 5.27, Figure 5.28, Figure 5.29 and Figure 5.30, respectively. Figure 5.28, Figure 5.29 and Figure 5.30 show that the optimum gain, power delivered to the load and PAE will result when both transistors are as large as possible but Figure 5.27 suggests that in this case the amplifier will consume 13mW and 21mW at -20dBm and 0dBm, respectively, which are relatively high and therefore not acceptable levels of DC power consumption.

With all dependencies simultaneously presented, designers realising amplifiers to meet varying specifications can choose device widths to meet their requirements. For the purpose of demonstrating this example, and based on the discussion above, the number of CG transistor gate fingers is set to 20 (100µm in width) and the number of gate fingers of the CS transistor is set to 12 (60µm in width). Based on the results presented in Section 5.5.2, this choice is believed to represent a good compromise between all low and high power performance parameters and satisfy the operational requirements for an LNA and a PA for a low power ZigBee front-end.

▪ The degeneration inductor and the load inductor

The effect of the degeneration inductance on the noise and nonlinear performance was discussed in Section 5.5.2.5 with aid of Figure 5.33 and Figure 5.34. Figure 5.33 indicates that IMD_3 reduces as the degeneration inductance increases, but the largest reduction is observed up to an inductance value of 0.58nH, and thereafter only 1dBc improvement is obtained by increasing the inductance value to 1.5nH. In Figure 5.34, the minimum noise figure increases as the degeneration inductance increases, with almost equal increase between each simulated inductance value and the one below. Considering these two results and bearing in mind that noise is more of a concern at the input stage of the amplifier, a degeneration inductor of 0.58nH was chosen. With this value, a real part will be generated in the input impedance (similarly to any other inductor value, but opposite to the choice of eliminating the degeneration inductance), easing the choice of matching circuitry. Further, the linearity of the amplifier will be improved with only a slight increase in noise figure.

From the analysis in Section 5.5.2.6, choosing a load inductor of 5nH yields an appropriate trade-off between stability and power gain at both low and high powers. This value results in unconditional stability and a reasonably-sized on-chip inductor ($0.05\mu\text{m}^2$). It also provides good maximum output power for both low and high drive power. It is important at this stage to ensure that maximum obtainable output power is achieved since subsequent matching to optimise other performance metrics such as PAE and IMD_3 will degrade the final output power.

▪ Linearisation by multi-gating

It is important to ensure that the current in the CS transistor used to improve linearity by cancelling g_3 matches the current in the CS transistor of the cascode amplifier, and also that the impact of the cascode CG transistor on the current is taken into account. These requirements can only be guaranteed by tuning the auxiliary transistors in parallel to the CS transistor while it is connected in the cascode. The optimisation strategy was to include three auxiliary transistors to try to achieve as wide as possible the region on the V_{GS} scale where $g_3 = 0$. Because this optimisation must be performed manually, as explained in Section 5.5.1.2, the widths of the three auxiliary transistors were set to the three smallest number of gate fingers (5, 6 and 7). With this arrangement, as well as restricting additional current drawn by the auxiliary transistors, only the gate voltages of the auxiliary transistors need be optimised, simplifying the design procedure.

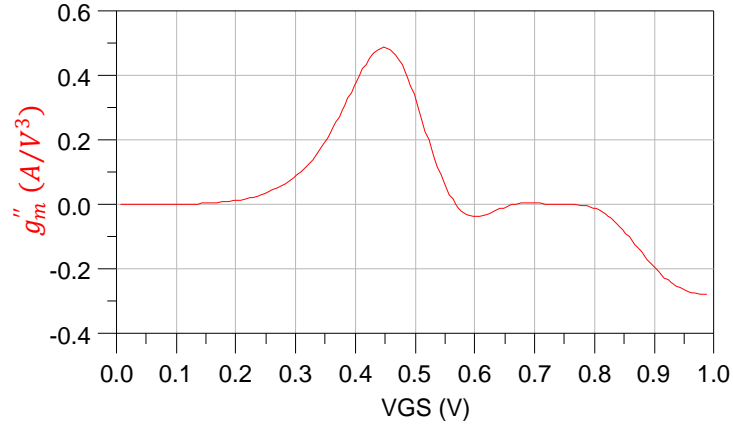


Figure 5.36 Second derivative of the total transconductance in the cascode after the addition of the linearising auxiliary transistors to the CS stage

Figure 5.36 shows the 2nd derivative of the cascode current with respect to the CS transistor gate to source voltage. Around $V_{GS} = 0.75\text{V}$ where the transistor is biased, $g_3 = 0$ is achieved between $0.68 < V_{GS} < 0.78\text{V}$. When testing the effect of linearisation, 7dBm IMD_3 improvement was observed at low input powers in the non-matched amplifier, reducing to 4dBm with an input power of 0dBm. This is because at low power, IMD_3 comes from the 3rd-order nonlinearity (g_3) only, and the contribution to IMD_3 from higher-orders of nonlinearity is negligible. As the operating power increases however, the contribution of the higher-orders of the nonlinearity becomes more significant and has to be considered.

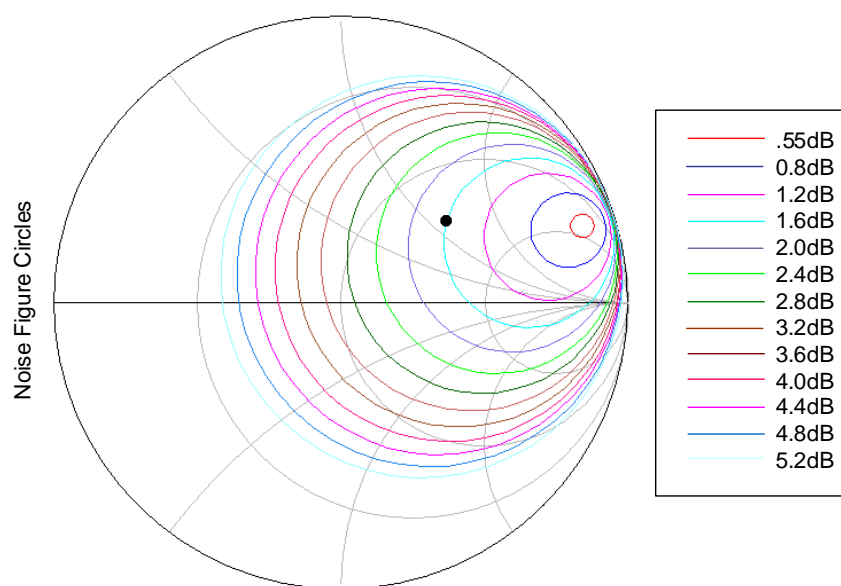
5.5.3.2 Matching the Amplifier

Following selection of the core elements, the amplifier is matched, as explained in Section 5.5.1. Matching at the input and the output will take into account the amplifier response to low and high power inputs simultaneously to make suitable tradeoffs which guarantee an operation satisfactory to meet requirements of both modes of operation.

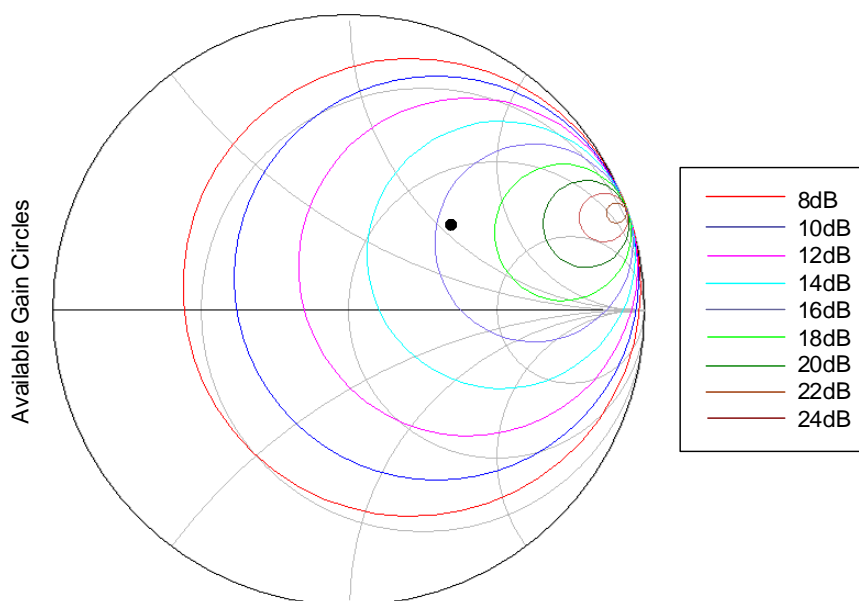
▪ Matching the input

Figure 5.37 shows the small signal noise circles generated from an S-parameter simulation for the input matching. Similar to stability circles, each of the noise circles represents a group of impedances which, if presented to the input of the amplifier, will yield the corresponding noise figure. It is important to note that these noise figure values are calculated in the simulator with the assumption that the output of the amplifier is conjugately matched, which is the usual practice in low power LNA design. This is not the case in this amplifier design however, since the output will be matched with respect to few other parameters (such as PAE, power delivered to the load and IMD_3) due to its dual functionality requirement. Thus, final noise figure value will be expected to change slightly. In Section 2.7.4, the gain characterisation of a two-port

network with S-parameters was discussed, the available gain was defined and its relationship with the input matching network was highlighted. Figure 5.38 shows the available gain circles for the core amplifier circuit generated from an S-parameter simulation. Figure 5.37 and Figure 5.38 indicate that little trade-off is required between noise and available gain for small signal operations as the location of the optimum impedance for both are at approximately the same location on the Smith chart.



**Figure 5.37 S-parameter noise circles for input matching
(with a dark spot indicating chosen impedance)**



**Figure 5.38 Available gain circles for low-power input matching
(with a dark spot indicating chosen impedance)**

For matching at high power operation, a source-pull simulation was performed at 0dBm input power to generate contours indicating the relationship between the impedance presented to the input and the power delivered to the load, power added efficiency and relative 3rd-order intermodulation distortion. These contours are presented in Figure 5.39, Figure 5.40 and Figure 5.41, respectively. It is important to emphasise that these contours (and the contours produced for matching the output in Figure 5.43, Figure 5.44 and Figure 5.45) were produced by simulating as many impedance points on the smith chart as possible in order to increase the conclusively of the results.

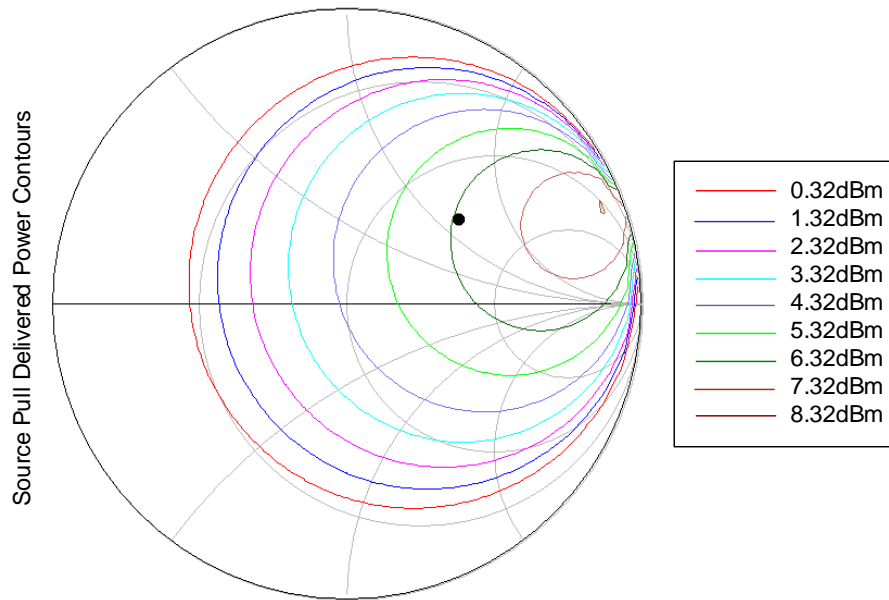


Figure 5.39 Power delivered to the load contours from the source-pull simulation (with a dark spot indicating chosen impedance)

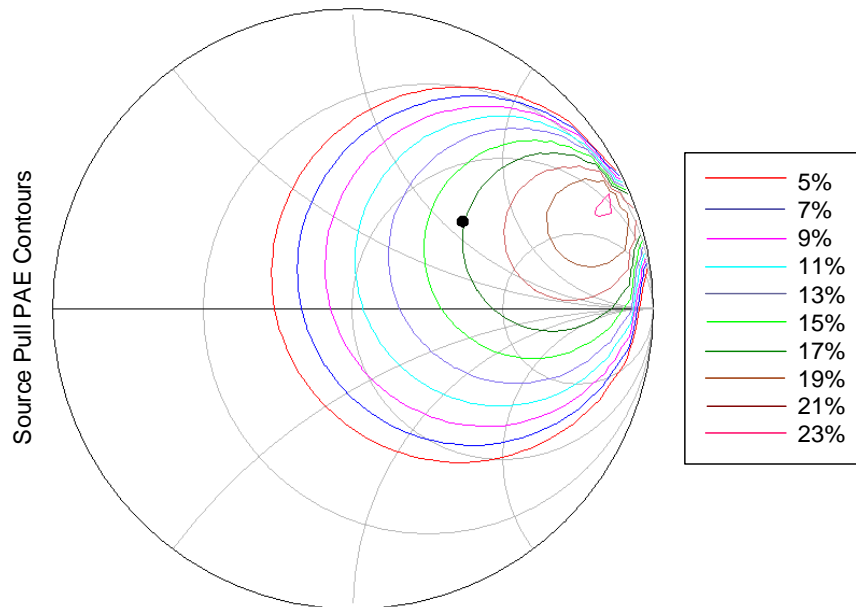
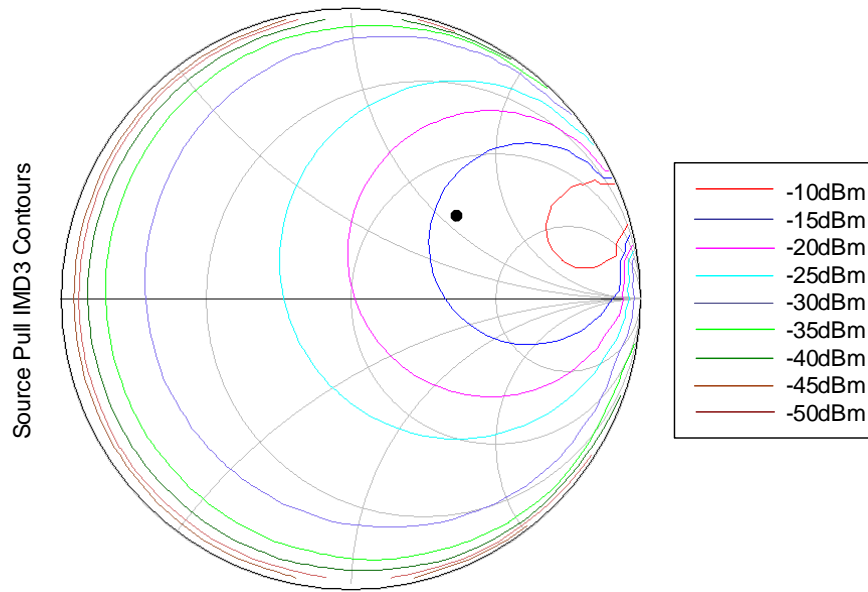


Figure 5.40 PAE Contours from the source-pull simulation (with a dark spot indicating chosen impedance)



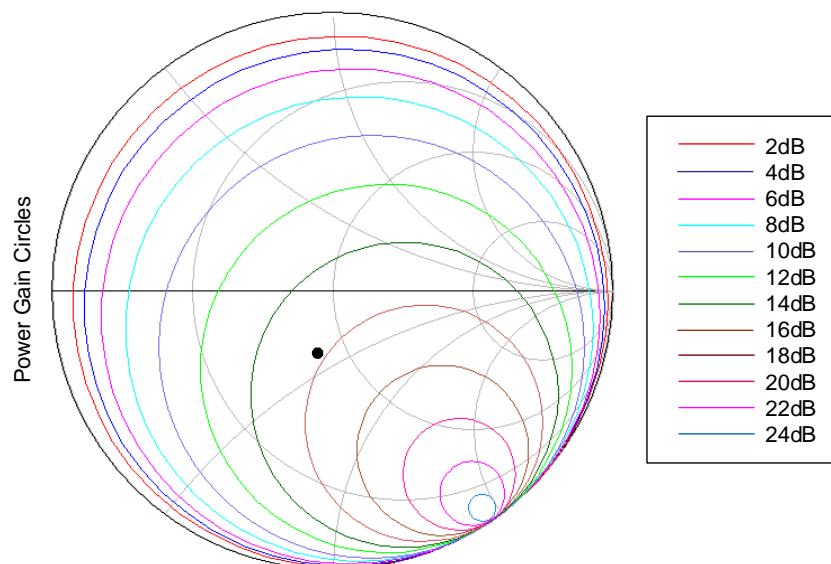
**Figure 5.41 IMD₃ Contours from the source-pull simulation
(with a dark spot indicating chosen impedance)**

Comparing Figure 5.37, Figure 5.39 and Figure 5.40 reveals that little compromise is required between the power and gain characteristics for high and low power operation. However, the result in Figure 5.41 presents an interesting challenge. The area of the Smith chart which represents the optimum impedance for power, gain and noise requirements also represents the worst case results for IMD₃. Also, in the direction where IMD₃ becomes better (smaller), the rest of the performance parameters worsen. This represents the ultimate trade-off challenge for matching the amplifier input taking into consideration noise, gain and nonlinear performance.

Given that the noise figure is influenced by the input match much more than the output match and that the output match has a significant effect on all the other performance parameters, the input matching impedance of $84.35 + j55.75 \Omega$ shown on each Smith chart (Figure 5.37 to Figure 5.41) as a dark spot was chosen. With this impedance, the noise figure is expected to be around 1.6dB and a good compromise is made between gain and linearity at the input. It is important to emphasize that these results present an initial prediction and are not final since the output matching is going to influence these performance parameters.

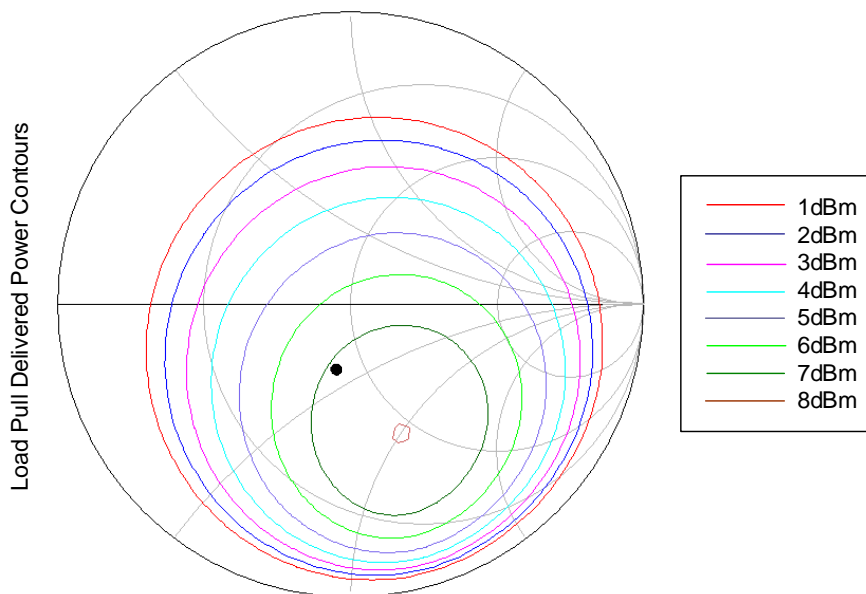
▪ Matching the output

Matching the output of the amplifier was considered with the input matched to the impedance value mentioned above. As well as available gain, Section 2.7.4 also defined power gain and highlighted its relationship with the output matching of the two-port network. Figure 5.42 shows the power gain circles of the amplifier generated from an S-parameter simulation.

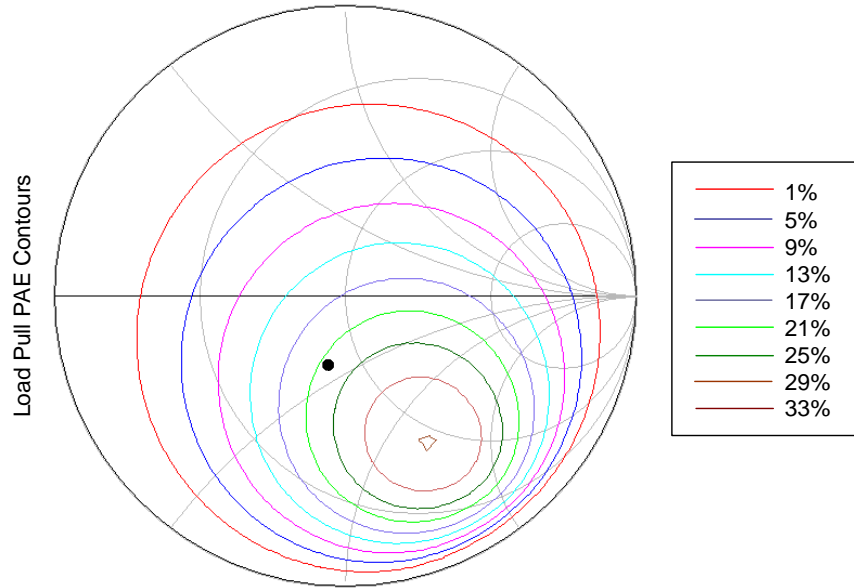


**Figure 5.42 Power gain circles for low-power output matching
(with a dark spot indicating chosen impedance)**

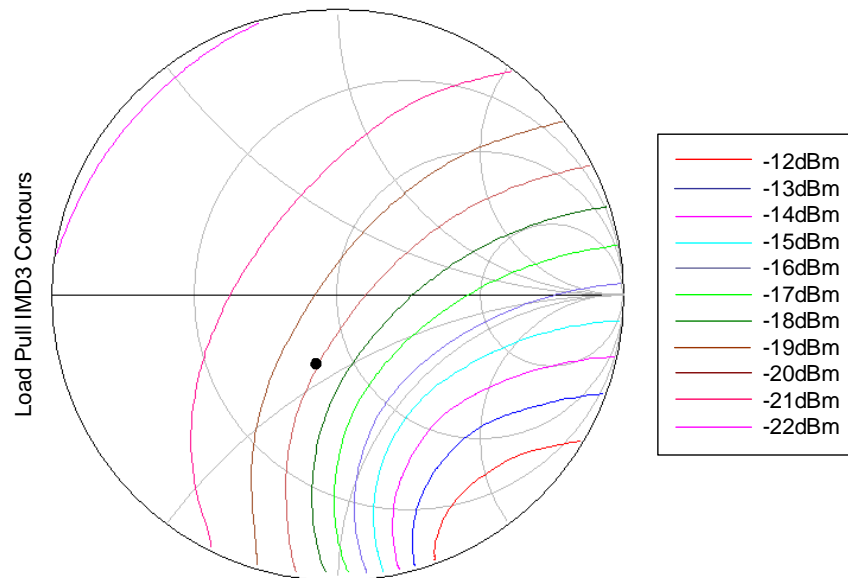
As with the input case, a load-pull simulation at 0dBm input power was performed at the output to investigate the matching condition for the high power operation. Figure 5.43, Figure 5.44 and Figure 5.45 show the generated contours for the power delivered to the load, power-added efficiency and relative 3rd-order intermodulation distortion, respectively.



**Figure 5.43 Power delivered to the load contours from the load-pull simulation
(with a dark spot indicating chosen impedance)**



**Figure 5.44 PAE Contours from the load-pull simulation
(with a dark spot indicating chosen impedance)**



**Figure 5.45 IMD₃ Contours from the load-pull simulation
(with a dark spot indicating chosen impedance)**

Inspection of Figure 5.42, Figure 5.43 and Figure 5.44 yields a similar result to that of the input match, that little trade-off is required for output matching for gain at low and high power levels. Also similar to the situation at the input, Figure 5.45 shows that the ultimate trade-off is required between matching for gain and nonlinear performance at the output. This is since, again, the impedances which yield the optimum gain and delivered power performance are shown to result in the worst intermodulation performance, and vice versa.

Furthermore, because the output match does not change DC power consumption at any RF input power, then a matching impedance that is going to result in a higher PAE (at the simulated 0dBm input power) is also going to result in a higher 1dB compression point (as long as it is less than 0dBm, which is normally the case). This is because a higher PAE (at the same input power and same DC power consumption) imply that more power is delivered to the load, indicting a later compression. Since there is a direct relationship between IIP_3 and IMD_3 as Figure 5.32 indicates, analysing the contours in Figure 5.44 and Figure 5.45 reveals that also a compromise need to be made between the 1dB compression point and IIP_3 when considering the output match of the amplifier.

After simultaneous consideration of the results in Figure 5.42 to Figure 5.45 it was decided that an output matching impedance of $42.4-j15.8\Omega$ will present an acceptable compromise. This point is shown on each Smith chart of Figure 5.42 to Figure 5.45 as a dark spot.

Note that the matching impedances at the input and output were implemented with ideal lumped off-chip components from the ADS library and not using the inductors and capacitors provided in the UMC 180nm design kit. This is because the quality factors of these lumped components were very high that they invalidated the designed-for matching impedances and made the matching conditions presented in Figure 5.37 to Figure 5.45 practically irrelevant. This is also bearing in mind that the work being presented here emphasises more on presenting a unique and useful design methodology for LNA/PA design rather than the actual final results that can be achieved with this particular foundry design kit.

5.5.3.3 LNA/PA Performance Evaluation

This Section presents the final performance results of the amplifier at low and high input powers and discusses how these satisfy the required specifications of the ZigBee front-end radio transceiver at the 2.4GHz band.

Figure 5.45 presents the S-parameter simulation of the matched amplifier over a range of frequencies from 1 to 4GHz. It is shown that in the 2.40 to 2.48GHz region (which covers ZigBee's 16 channels in this band) S_{21} is 10dB. S_{12} is -42dB which presents an excellent reverse isolation at the input. However results for S_{11} and S_{22} represent a poor input and output return loss, respectively. The reason for this can be referred to the matching conditions discussed in Section 5.5.3.2 where the worst case compromise was required between gain and intermodulation distortion matching.

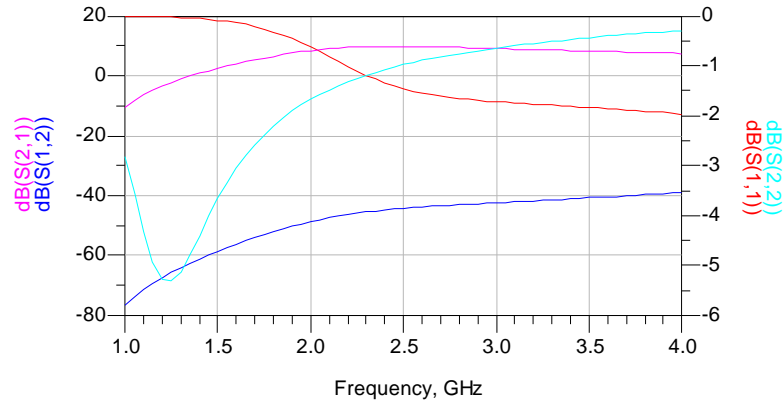


Figure 5.46 S-parameter simulation results of the matched amplifier

Figure 5.46 presents the noise performance of the amplifier over the range from 1 to 4GHz. A minimum is observed in the band of interest between 2.4 and 2.48GHz, with a noise figure of 1.65dB, slightly higher, as expected, than that predicted by the noise circles in Figure 5.37.

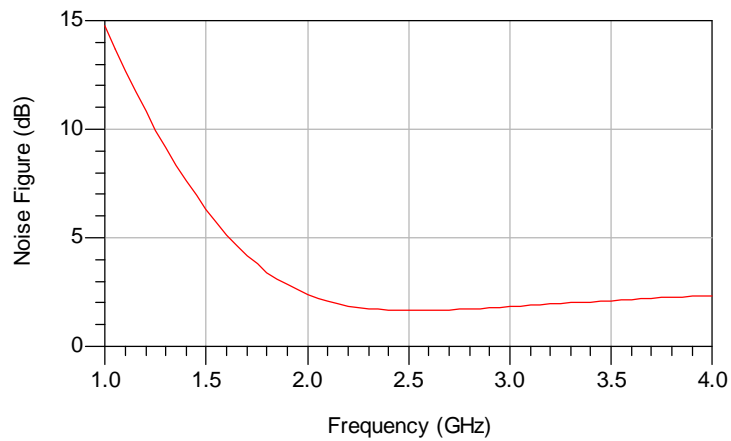


Figure 5.47 Noise figure vs. frequency of final amplifier

Power added efficiency and power gain results from a one-tone simulation are shown in Figure 5.48. The amplifier has its highest PAE of 20% when its input power is about 1dBm and achieves about 7dB of gain at that point. This more than satisfies the specification of ZigBee front-end transmitter which requires an output power of at least -3dBm, but as discussed in Chapter 3, a higher output power is desirable to increase the integrity of the signal and reduce the effect of interference.

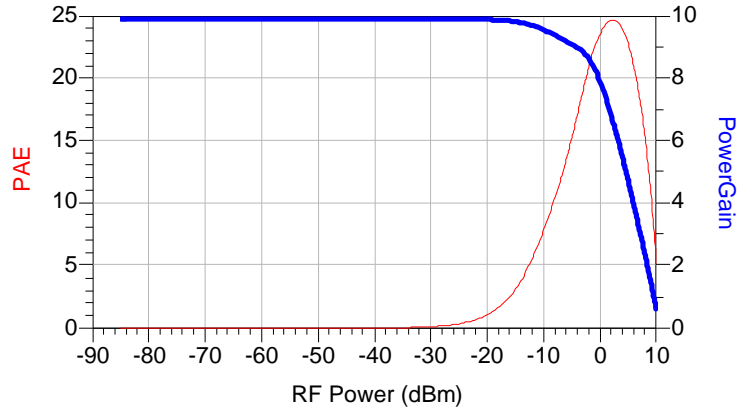


Figure 5.48 Power added efficiency and power gain vs. input power

Figure 5.49 shows how the DC current drawn from the supply, and correspondingly the DC power consumption, change as the input power changes. As expected, DC current and DC power consumption increase as the input RF power increases. The increase starts at -25dBm which means it is at this value that the input RF power starts to change the DC characteristics of the amplifier. At small input powers, only 4.4mA is withdrawn from the 1.8V supply yielding a DC power consumption of about 8mW which is low when compared with some of the results from the literature shown in Table 5.1. This power increases to 22mW with only 12mA drawn from the supply which still represents a superior result if compared with other chips available in the market as shown in Table 3.1, bearing in mind that the power amplifier consumes most of the power in a transmitter. Note how the DC power consumption of the final design agrees with that predicted by the initial optimisation study of Figure 5.27 (despite the slight increase due to the current in the auxiliary transistors added for linearisation).

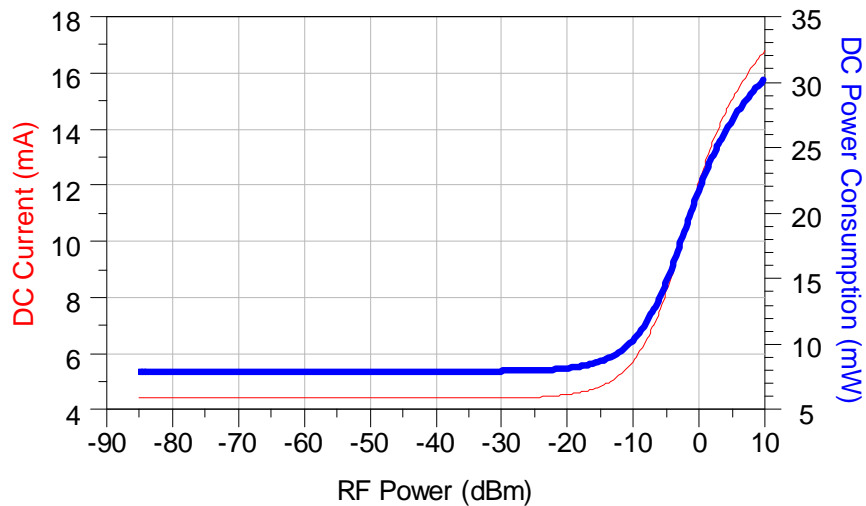


Figure 5.49 DC current and DC power consumption vs. input RF power

In Figure 5.50, a one tone simulation to determine the 1dB compression point shows that the amplifier compresses at about -3.5dBm input power. Figure 5.51 presents the result of a two tone simulation showing that the amplifier achieves an IIP_3 of 6dBm.

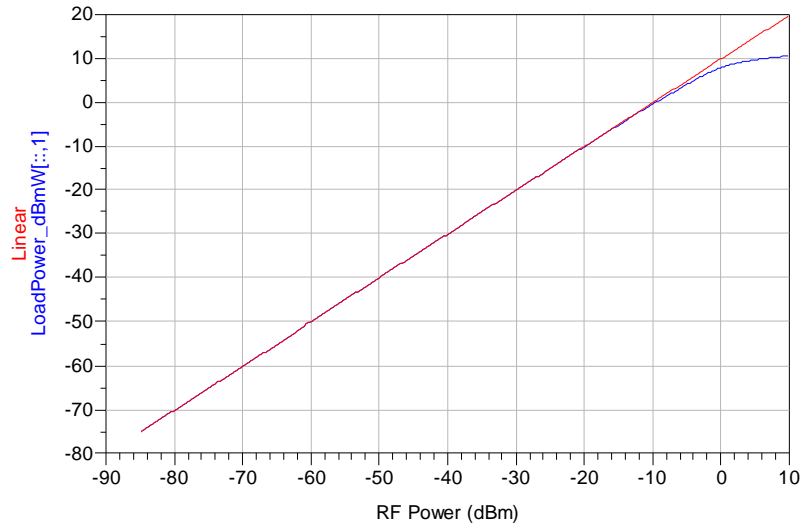


Figure 5.50 One-tone harmonic balance simulation results of matched amplifier

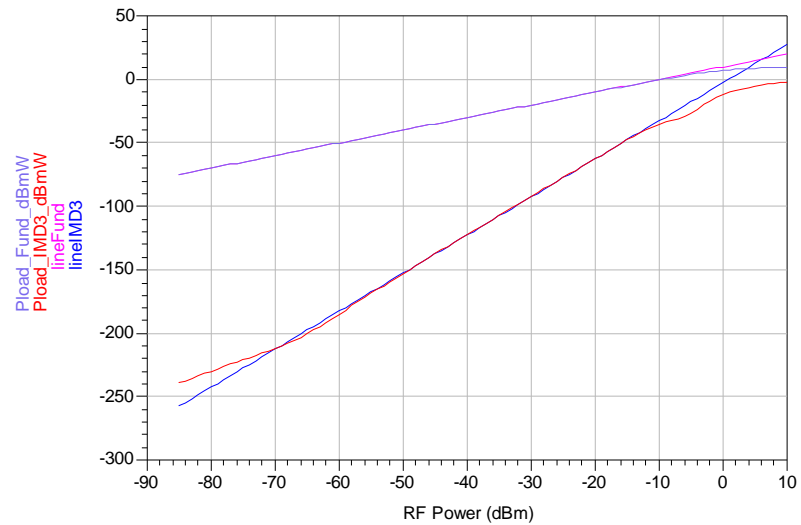


Figure 5.51 Two-tone simulation results of matched amplifier

Obtaining the results above involved no ambiguous automatic optimisation in the simulator and did not require any iteration at the end of the design flow. Optimising the core transistor circuit and the matching was based entirely on visual analysis of the results presented in Sections 5.5.2 and 5.5.3.2. This demonstrates the significance of the proposed design methodology which enabled the exploration of the available space of possibilities in the design and allowed the designer to make efficient trade-off choices, making this design possible.

It is not possible to directly relate the above results of the low noise amplifier to the requirements of the ZigBee standard in the receiver. This is because the standard does not provide specifications for particular circuits in the receiver but for the receiver as a whole. For example, the sensitivity was defined in Section 2.3.1 as the lowest input power the receiver can detect while achieving a Packet Error Rate (PER) below a pre-specified maximum. PER was defined as the ratio of error in received packets when compared with the transmitted packets. Therefore, in order to quantify PER, the received signal has to be demodulated and processed in baseband before the received packets can be compared with the transmitted packets. In another example, the 30dB requirement for alternate channel rejection means that the adjacent channel should be attenuated by a factor of 1000. This is usually achieved through a combination of several filtering and mixing stages in the receiver beyond the LNA.

Therefore, in order to assess the above low noise amplifier for ZigBee requirements, its performance is compared with the low noise amplifier sections of recently published 2.4GHz full ZigBee transceiver designs. Performance parameters of concern are noise figure, gain, reverse isolation, input and output return loss, IIP_3 and 1dB compression point, where LNA-only data is available.

This difficulty in assessing the compatibility of the amplifier with the requirements of the standard does not apply to the power amplifier. This is because the power amplifier is the last element in the transmitter path, and hence its compatibility to the standard can be assessed with reference to the transmitter's output specifications.

Table 5.4 presents a summary of results of the designed amplifier and comparison with other ZigBee front-end designs. Both low noise and power amplifier comparison is undertaken on worst case basis. The idea is, if the designed amplifier in this work can perform better than the worst published ZigBee-compatible design, then it must also be ZigBee-compatible. Reviewed designs are [Nguyen 2005], [Ho 2006], [Nguyen 2006], [Beyer 2006], [Cimino 2007], [Choi 2003], [Hwang 2007], [Zito 2006], [Lapuyade 2006], [Kluge 2006] and [Sarhangian 2007].

Table 5.4 Amplifier results comparison with ZigBee-compatible LNA and PA

Comparison Parameter	This work	Worst case value	Reported in
IIP ₃ (dBm)	6	-9	[Ho 2006]
1dBc (dBm)	-3	-21	[Zito 2006]
LNA			
Gain (dB)	10	10	[Lapuyade 2006]
Reverse Isolation (dB)	-42	-28	[Ho 2006]
Input return loss (dB)	-1	-17	[Cimino 2007]
Output return loss (dB)	-1	-12	[Nguyen 2006]
Power Consumption (mW)	7.2	5.4	[Choi 2003]
NF (dB)	1.6	4.9	[Ho 2006]
PA			
Maximum Output power (dBm)	10	0	[Nguyen 2006]
Power Consumption (mW)	22@6dBm output	26@6dBm output	[Zito 2006]
PAE (mW)	25%@6dBm output	15% @6dBm output	[Zito 2006]

Due to its dual functionality target, there are no grounds of comparison between the amplifier designed in this work and LNA-only or PA-only designs. However, what is important is that the amplifier designed achieves acceptable performance satisfying each functionality subject to the ZigBee requirements. Table 5.4 shows that obtained results for this amplifier design in both low noise and power amplifier operations are inline with the requirements of ZigBee. This is with exception to the input and output return losses which are clearly significantly much higher than those of other designs. Although reverse isolation is very high and ZigBee does not have any specifications for input and output return loss, this may still cause problems due to the reflected power back to the antenna. As explained earlier, this result was due to the required trade-off in matching the input and output which appeared to have the worst case possible. In Chapter 6 a proposal is made to develop a solution to this problem as part of the future developments based on this thesis.

5.6 Summary

In this chapter, the design of a dual functionality LNA/PA and the development of its design methodology were described. The chapter started by first looking at several LNA architectures, beginning with single-stage amplifiers considering their noise, matching and gain characteristics and then moving to commonly used LNA architectures. The purpose was to identify an architecture where required trade-offs between low noise and power amplifier performance requirements can be attained. The architecture found to be most suitable was the common-source common-gate cascode architecture. Following this discussion, an extensive review was given to LNA design approaches developed by designers based on the cascode architecture. The

analysis first looked at the diversity and requirements of design approaches and then discussed the limitations of the current available methods. This was followed by a review of some highly performing low noise amplifiers based on the cascode architecture from published work. Focus was then placed on the proposed circuit techniques that helped achieving these high performance levels.

Non-linearity of amplifiers was discussed by first explaining circuit factors which affect the nonlinear performance, including bias circuitry and matching. This was followed by a review of low power linearisation techniques with focus on the derivative superposition and harmonic termination methods. The performance of some recently published designs where these techniques were employed was then discussed.

To broaden the discussion into the effect of some frequency components affecting the nonlinear performance as a result of being fed-back to the input, an extended analysis was conducted using polynomial substitutions of an assumed amplifier with up to 7th-order significant nonlinear terms. The analysis was based on first producing all frequency components which result from any single-order of nonlinearity when excited with a two-tone input, and then individually substituting these components in all other orders of nonlinearity by adding them to the original two-tone input.

The design methodology of the dual functionality LNA/PA amplifier was then introduced by first laying out the major steps of the design flow and then implementing these steps individually. The design methodology was based on simultaneous visual analysis of the effect on low and high power performance from variations to various core amplifier components, with focus on the sizes of the common-source and common-gate transistor stages. A design example then demonstrated the effectiveness and usefulness of the proposed design methodology in designing an amplifier, the performance of which is sufficient for the requirements of both an LNA and PA of the ZigBee standard.

Chapter 6 CONCLUSIONS AND FUTURE DEVELOPMENTS

6.1 Introduction

The focus of the work presented in this thesis was to provide solutions to contribute to the improvement of low power transceivers used in sensor and control networks. This was realised by carrying out work on three fronts; application requirement analysis, circuit design, and device modelling. The outcome of these studies feeds back into the design of front-end transceiver circuits with the aim of enhancing performance levels while maintaining or reducing power requirements.

This final chapter presents a summary of the thesis, its findings and offerings. Issues of future development and improvements building on this work are highlighted and some further research ideas are proposed. The chapter ends with concluding remarks.

6.2 Thesis Summary

Chapter 1 of this thesis introduced the context of the work, setting out the issues that are taken into consideration in the following chapters of the thesis. It discussed the expansion of wireless communications, the challenges associated with that, and the general form of required solutions. In Chapter 2, the technological background of the research was presented. The aim was to provide a better understanding of the technical issues involved in the development work. The chapter discussed the selection of ZigBee as the wireless standard to investigate, and presented a review of general transceiver front-end requirements, and the means by which they can be achieved. Particular focus was given to microwave transistors, low noise and power amplifiers and the nonlinearity phenomena. It also introduced the test application, the Tyre Pressure Monitoring System (TPMS) focusing on its transceiver requirements and reviewed some of the circuit design and analysis techniques used in the thesis.

In Chapter 3, following discussion of the ZigBee standard and current TPMS designs, a compatibility analysis was conducted between ZigBee and TPMS, identifying the benefits a

ZigBee implementation can bring to TPMS and highlighting areas where improvements are needed. The chapter also looked at ZigBee's modulation scheme, the Quadrature Phase Shift Keying (QPSK), and introduced a symbolic method to describe the options of phase selection in QPSK. A demonstration, with the aid of an implementation in Matlab, showed how a full QPSK modulator and demodulator can be built using non mathematically-complicated components.

In order to better understand the nonlinear behaviour of the transistor used in the amplifier design work in this thesis, Chapter 4 presented a new concept of transistor nonlinear modelling in which each of the transistor's nonlinear elements was represented individually, with the aim to quantify the individual contribution to the total distortion from each nonlinear element. Three different models based on different representation methods were built, and their accuracies and performances were assessed. The linearisation of nonlinear elements, quantification procedures and quantification results were then discussed. Chapter 5 was dedicated to the design of a dual functionality LNA/PA amplifier for ZigBee. It first looked at different LNA architectures and assessed their suitability for the required trade-offs of that design. Following a review of the state of the art LNA designs, their design methodologies and linearisation techniques, a new design methodology targeted at the dual functionality amplifier was introduced. The design methodology aimed at visually exploring the full potential of available components for the design and the extent of trade-offs required. A design example was provided to demonstrate the effectiveness and usefulness of the design methodology.

Putting the work of this thesis altogether, everything falls back into one context; improving the capabilities of wireless transmission radios, with demonstrative effects on the test application, the TPMS. First, the grounds on which the compatibility analysis in Chapter 3 was conducted can be reused to assess the capability of any network standard on the requirements of an application. Beyond that, the arguments on which a network-standard based TPMS was deemed better than a point-to-point based one can be used to examine the performance of a wide range of non-standardised wireless communication systems leading to the potential standardisation of every form of wireless communication in use today. The description of the QPSK phase selection options and implementation in Chapter 3 presents an interesting feature that can be useful to TPMS and other communication systems. This work can be utilised in enhancing the security of communication systems and for identification. In the TPMS, it means that the tyre modules in a vehicle and its central receiver can be programmed to use the same phase shifting scheme to ensure that signals between them cannot be correctly interpreted by any other receiver in another vehicle.

The transistor modelling in Chapter 4 adds a significant advantage to wireless communications for TPMS and beyond. Accurate device models are a key for first-time-right designs and subsequent satisfaction of fast time to market requirements.

The importance of power efficiency in the TPMS application was discussed with particular focus in Chapter 2. Chapter 3 presented some data sheet information for ZigBee compliant chips made by leading manufacturers and showed that current power consumption rates of these chips are not suitable for the TPMS application. As it is well known that amplifiers are the most power-hungry components of a transceiver, combining the low noise amplification and power amplification functionalities in Chapter 5 led to a significant reduction in the power consumed in the transceiver since there is a single supply voltage and a single current consumed for both these functionalities. This is particularly significant for TPMS as it means fully portable systems with acceptable power consumption rates can now be built. The grounds on which the LNA/PA design methodology in Chapter 5 was based present a new concept in circuit design methodology. The concept, based on 3D visualisation of the design space and involved trade-offs, can benefit circuit design activities for many more circuits and applications beyond RF amplifiers and TPMS, if properly adapted.

6.3 Contributions

The contributions of this work to the field can be summarised as follows:

- Following the identification of ZigBee as the standard most suitable for low power applications and the study of the current approaches for building a TPMS, the first contribution was a proposal for a new form of a TPMS based on a ratified wireless networking standard, implemented using ZigBee. This was backed by a comprehensive compatibility analysis of the offerings of ZigBee to the requirements of TPMS [Abuelmaatti 2004].
- A symbolic method based on visual observation and manipulation of phasors was developed to provide an alternative approach to describe the selection of phases in the QPSK modulation. A system level implementation based on that approach was proposed and implemented in Matlab to demonstrate how sophisticated modulation schemes like QPSK can be reliably built using simple basic low level components, with no need for mathematically-complicated high level blocks [Abuelma'atti 2007].

- A new form of table-based transistor nonlinear model was proposed. The model was the first to represent each nonlinear element in the model individually and independently from other nonlinear elements. Three different variations of the model were built using Symbolically-Defined Devices in ADS. The differences between the models were in the representation of the nonlinear elements [Abuelmaatti 2006a, Abuelmaatti 2006b]. All models produced a good fit in S-parameter simulation, but their differences were used to explain their varying performances in harmonic balance simulation.
- A technique based on sweep indices implemented in coded functions using ADS data display window equation components was developed to automate the extraction of the models nonlinear elements over a wide range of terminal voltages with smaller steps for enhanced accuracy. This enabled the generation of 3D plots describing the nonlinear behaviour of the models intrinsic elements as a function of its terminal voltages
- A new superposition method was proposed for quantifying the individual contribution to distortion from the transistors various nonlinear elements using nonlinear modelling. The method was based on linearising the nonlinear characteristics of intrinsic model elements and then switching between the linear and nonlinear representations of each element individually or in combinations while assessing the transistor's total nonlinear distortion every time.
- A novel idea was proposed to combine the functionality of the low noise amplifier and the power amplifier in one amplifier for the ZigBee standard. The idea originated from the fact that the required output power in ZigBee is relatively low. This introduced a new investigation in which a common low noise amplifier architecture was examined for operation with higher than usual input powers. This initiated exploration of a new level of trade-offs between low noise and power amplification performance parameters, and lead to a new set of results describing the behaviour of this architectures components at high input powers. New measurements were introduced to the LNA architecture such as power added efficiency (PAE) which is normally a measurement of power amplifiers. Combined analysis of low and high power matching conditions at the input and output of the amplifier for both modes of operation were also introduced.

- A unique design methodology was introduced for the cascode architecture of low noise amplifiers. The methodology facilitated simultaneous analysis of several performance parameters with respect to the two most influential components in the architecture, the common-source and common-gate transistors. Simultaneously analysing both transistors and considering their common current, the representation was based on visual observation, in 3D graphs, of how performance parameters vary with related design parameters. Taking into account the multiple dependencies of performance parameters on design parameters, enables the designer to make satisfactory trade-off decisions. Components were simulated between their minimum and maximum values to explore their full potential and uncover areas of their operation where better results may be obtained, overcoming the weakness of mathematical equation-based design approaches where this is not possible. The design methodology also gives the designer the ability to reuse the results of this analysis and change trade-off choices if another design with different constraints was needed.
- Based on the proposed methodology, an amplifier design satisfying both requirements of low noise and power amplification for ZigBee was designed. The amplifier employs some low power linearisation techniques but uses only the basic cascode architecture components, yet achieves excellent performance compared with recently published designs.

6.4 Future Work

So far, this chapter reiterated this thesis aims, findings and main results and considered the novel contributions of the work conducted. While the contributions of this research are valuable, it raises, as research should, some interesting questions and ideas for further research. This section deals with the most significant of these questions and ideas.

6.4.1 ZigBee vs. TPMS analysis

The ZigBee offerings vs. TPMS requirements analysis conducted in Section 3.3.1 did not take into account the effects of the practical conditions of vehicle operation on the robustness of the existing wireless communications channel. These electromagnetic effects are may be due to the fast rotation speed of the tyre, generated heat and sudden breaking. In order to make this analysis more conclusive and enhance its validity with respect to real operating conditions, these effects maybe addressed.

6.4.2 Transistor non-Linear Modelling

Table-based implementation is a very convenient and accurate way of building transistor models. The only drawback of this method is the length of time required for generation and manipulation of data in the tables. The work presented in this thesis proposed a set of techniques to automate most of these processes. The results achieved were notable in reaching the goal of nonlinear quantification. However, based on observations made when doing this work and on its results, there are still issues in table-based models to be investigated, in effort to enhance their production automation, accuracy and performance. These can be summarised as follows:

- 1- More study needs to be conducted on the reasons for convergence or non-convergence in harmonic balance simulation. This need is based on some unexplained observations made when performing harmonic balance simulations over a sweep of input powers at different bias levels. These observations include:
 - The simulation switches between converging or producing an error as the input power sweep step is changed. For example, it may converge if the step was set to 5dBm but does not converge if it is reduced, but also the opposite happens.
 - Changing the interpolation in the DAC from cubic to cubic spline or vice versa had a significant but unstable effect on convergence. For example, the simulation may converge if the C_{gs} table was set to cubic interpolation at V_{GS} of 0.7V, but if V_{GS} was changed to 0.8V, the interpolation need to be changed to cubic spline for it to converge.
 - A slight change in the bias voltage of the model significantly affects convergence. For example, the harmonic balance simulation may converge at V_{GS} of 0.72V but not 0.7V, even though parameter values at the 0.7V value are in the table and the ones at 0.72V are generated by interpolation.
- 2- It is beneficial to investigate ways to increase the accuracy of the table-based model, even beyond levels presented in this thesis. This can lead to increased usability of the model in several areas. For example, the nonlinearity of the model may be valid up to higher-orders of nonlinearity, achieving a good fit of the 7th-order intermodulation distortion or the 5th harmonic.
- 3- A written code for manual interpolations may prove significant in reducing the interpolation error in the linearisation of current characteristics as demonstrated in Section 4.4.2.1.

- 4- Further investigation need to be conducted in order to find a way to maintain separation between the representations of the model's nonlinear elements while taking into account their correlation. This may be done by introducing functions in the representation of each nonlinear element to describe the nature and weight of correlation between its characteristics and that of any other nonlinear parameter in the model. This would enhance the accuracy and representation of the nonlinearity quantification results.

6.4.3 LNA/PA Design Approach

With respect to the amplifier design methodology, three sets of recommendations can be made to build on the work presented in this thesis, these are as follows:

- 1- An improvement will be to extend the usefulness of the graphical visualisation-based design methodology by including more valuable components in the design of the amplifier such as the inter-stage matching inductor used to match the output of the first stage to the input of the second stage and the external capacitance between the gate and the source of the common-source transistor used for noise and linearity improvement. It remains to be investigated how the simultaneous analysis of the effect of the settings of the two cascode transistors on various performance parameters can be maintained in a similar fashion to the way presented in this thesis while accounting for, but not initially including, the effects of these two components.
- 2- When no isolation between the input and output is implemented in the circuit, the choice of the input and output matching networks will be cross-dependant, depending on which was designed first. For instance, if nonlinearity and output power were the most important performance parameters in a particular design, the designer may perform a source-pull simulation for IMD_3 and delivered power contours, choose a suitable trade-off and design the input matching network. This would normally be followed by a load-pull simulation for IMD_3 and delivered power contours. The load-pull simulation will take into account the matching network designed at the input. If a different trade-off choice was made at the input, different IMD_3 and delivered power contours will result in the output, hence another trade-off choice could be made at the output and the amplifier may have a better final resulting performance. Therefore, it would be very beneficial to find a way to program the simulator to run source-pull and load-pull simulations simultaneously taking into account only the core amplifier circuit and its terminations, and then tune the output contours as the designer moves a marker on the contours of impedances at the input. This way the designer may optimise the input and output matching circuits simultaneously and nothing will be left to chance.

- 3- Finally, the results of the matching analysis in Section 5.5.3.2 introduce a very important proposal for the design approach improvements. As presented, the optimum matching conditions for power and linearity were on opposite sides of the Smith chart for both the input and output. This represented a great challenge and meant that big sacrifices in either performance parameter had to be made. Since the source- and load-pull simulations are performed on the core amplifier circuit, a very useful investigation would be to try to find a way to predict the location of the contours on the Smith chart relative to the sizes and biases of the transistors (and maybe other components) in the core amplifier circuit, and involve this in the trade-off decisions taken in designing these components. This way, the location of the contours will not be left to chance, and a situation similar to that raised in Section 5.5.3.2 can be avoided. Moreover, if the location of the contours can be controlled by the settings of the core circuit components, then the designer can try to make sure that optimum impedances for several performance parameters will not be far apart, hence a less sacrificing trade-off decisions for the matching impedances can be made.

6.5 Final Remark

Wireless technology adds a significant value of convenience to humans' lives. Consumers will continue demanding highly performing wireless devices, at lower costs. This demand must be met by organisations involved in the field to capture the interest of consumers and hence the market, which is worth billions of pounds every year. Research aimed at this target is wide-ranging and very common in many research organisations and businesses throughout the world. The work presented in this thesis forms a tiny, yet significant, contribution to this global effort.

REFERENCES

- [Abidi 1986] Abidi, A. A., "High-frequency noise measurements on FET's with small dimensions," *IEEE Transactions on Electron Devices*, vol.33 (11), pp.1801 - 1805, (1986).
- [Abuelmaatti 2004] Abuelmaatti, A. M. T. and Beaumont, S., "A second generation Network-Based Tire Pressure Monitoring System using 802.15.4," In Proceedings of *The 5th EPSRC Annual Postgraduate Symposium on the Convergence of Telecommunications, Networking and Broadcasting (PGNet2004)*, pp.447-450, (2004).
- [Abuelmaatti 2006a] Abuelmaatti, A. M. T. and Thayne, I., "A New RF SiCMOS SDD Model for Quantifying Individual Contribution to Distortion from Transistor's Nonlinear Parameters," In Proceedings of *The 13th IEEE International Conference on Electronics, Circuits and Systems (ICECS 2006)*, Nice, France, (2006).
- [Abuelmaatti 2006b] Abuelmaatti, A. M. T., Thayne, I., McGregor, I. and Wasige, E., "A New Implementation for RF SiCMOS Transistor Model Using SDD for Quantifying Individual Contribution to Distortion from Transistor's Nonlinear Parameters," In Proceedings of *The 2006 Asia Pacific Microwave Conference (APMC 2006)*, Yokohama, Japan, (2006).
- [Abuelma'atti 2007] Abuelma'atti, A. M. T., Thayne, I. and Beaumont, S., "A new approach to QPSK: Mechanism and Implementation," In Proceedings of *The IEEE Wireless Communications and Networking Conference (WCNC2007)*, Hong Kong, (2007).
- [Alam 2006] Alam, S. K. and Degroat, J., "A 1.5-V 2.4 GHz differential CMOS low noise amplifier for bluetooth and wireless LAN applications," In Proceedings of *4th International IEEE North-East Workshop on Circuits and Systems, NEWCAS 2006*, pp.13-16, (2006).
- [Allstot 2004] Allstot, D. J., Li, X. and Shekhar, S., "Design Considerations for CMOS Low-Noise Amplifiers," In Proceedings of *2004 IEEE Radio Frequency Integrated Circuits (RFIC) Systems. Digest of Papers*, pp.97-100, (2004).
- [Amor 2006] Amor, M. B., Fakhfakh, A., Mnif, H. and Loulou, M., "Dual Band CMOS LNA Design With Current Reuse Topology," In Proceedings of *2006 International Conference on Design and Test of Integrated Systems in Nanoscale Technology, DTIS 2006*, pp.57 - 61, (2006).
- [Andreani 2001] Andreani, P. and Sjolund, H., "Noise optimization of an inductively degenerated CMOS low noise amplifier," *IEEE Transactions on Circuits and Systems-II: Analog and Digital Signal Processing*, vol.48 (September), pp.835-841, (2001).

- [Anritsu 2003] Anritsu, "Fundamentals of Interference in Wireless Networks," Anritsu Application Note, <http://www.us.anritsu.com/downloads/files/11410-00302.pdf>, (2003).
- [Aparin 1999] Aparin, V. and Persico, C., "Effect of out-of-band Terminations on Intermodulation Distortion in Common-Emitter Circuits," In Proceedings of *IEEE MTT-S International Microwave Symposium Digest*, vol.3, pp.977-980, (1999).
- [Aparin 2004] Aparin, V., Brown, G. and Larson, L. E., "Linearization of CMOS LNA's via optimum gate biasing," In Proceedings of *the 2004 International Symposium Circuits and Systems, ISCAS '04*, vol.4, pp.IV-748-51, (2004).
- [Aparin 2005] Aparin, V. and Larson, L. E., "Modified Derivative superposition method for Linearizing FET Low-Noise Amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol.53 (2), pp.571-580, (2005).
- [Asgaran 2006] Asgaran, S., Deen, M. J. and Chen, C.-H., "A 4-mW monolithic CMOS LNA at 5.7 GHz with the gate resistance used for input matching," *IEEE Microwave and Wireless Components Letters*, vol.16 (April), pp.188-190, (2006).
- [Asgaran 2007] Asgaran, S., Deen, M. J. and Chen, C.-H., "Design of the Input Matching Network of RF CMOS LNAs for Low-Power Operation," *IEEE Transactions On Circuits And Systems-I*, vol.54 (3), pp.544 - 554, (2007).
- [Baki 2005] Baki, R. A., Tsang, T. K. K. and El-Gamal, M. N., "Distortion in RF CMOS Short-Channel Low-Noise Amplifiers," In Proceedings of *3rd International IEEE Northeast Workshop on Circuits and Systems Conference, NEWCAS 2005*, pp.369-372, (2005).
- [Baki 2006] Baki, R. A., Tsang, T. K. K. and El-Gamal, M. N., "Distortion in RF CMOS short-channel low-noise amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol.54 (1), pp.46 - 56, (2006).
- [Benton 1992] Benton, R., Nijjar, M., Woo, C., Podell, A., Horvath, G., Wilson, E. and Mitchell, S., "GaAs MMICs for an integrated GPS front-end," In Proceedings of *14th Annual IEEE Gallium Arsenide Integrated Circuit (GaAs IC) Symposium*, pp.123 - 126, (1992).
- [Beru 2000] Beru Aktiengesellschaft, "Electronic tire pressure monitoring system," White Paper, www.beru.com, (2000).
- [Beyer 2006] Beyer, S., Jaehne, R., Kluge, W. and Eggert, D., "A 2.4GHz direct modulated 0.18 μ m CMOS IEEE 802.15.4 compliant Transmitter for ZigBee™," In Proceedings of *IEEE Custom Integrated Circuits Conference 2006*, pp.121 - 124, (2006).
- [Bluetooth-SIG 2007] Bluetooth SIG, "Bluetooth Special Interest Group," <http://www.bluetooth.org>, (Accessed: 2007).
- [Van den Bosch 1999] Van den Bosch, S. and Martens, L., "Approximation of State Functions in Measurement-Based Transistor Model," *IEEE Transactions on Microwave Theory and Techniques*, vol.47 (1), pp.14-17, (1999).
- [Calvo 1993] Calvo, M., Winson, P. and Snider, A. D., "Mechanistic interpretation of the transcapacitance element," In Proceedings of *IEEE Southeastcon '93*, pp.1-4, (1993).

- [De Carvalho 1999] De Carvalho, N. B. and Pedro, J. C., "Large- and Small-Signal IMD Behaviour of Microwave Power Amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol.47 (12), pp.2364 - 2374, (1999).
- [Cha 2000] Cha, C.-Y. and Lee, S.-G., "A low power, high gain LNA topology," In Proceedings of 2000 2nd International Conference on Microwave and Millimeter Wave Technology, *ICMMT 2000*, pp.420-3, (2000).
- [Cha 2003] Cha, C.-Y. and Lee, S.-G., "A 5.2-GHz LNA in 0.35- CMOS utilizing inter-stage series resonance and optimizing the substrate resistance," *IEEE Journal of Solid-State Circuits*, vol.38, (April), pp.669-672, (2003).
- [Chen 2002] Chen, J. and Shi, B., "Impact of intrinsic channel resistance on noise performance of CMOS LNA," *IEEE Electron Device Letters*, vol.23 (1), pp.34-36, (2002).
- [Chen 2006a] Chen, W.-N., Cheng, K.-H. and Chang, T.-D., "A 10 GHz with low power and low noise figure amplifier using 0.18 CMOS technology," In Proceedings of *ICSICT-2006: 2006 8th International Conference on Solid-State and Integrated Circuit Technology*, pp.1556-1558, (2006).
- [Chen 2006b] Chen, Y.-y., Zhang, R.-x., Shi, C.-q. and Lai, Z.-s., "A 0.18 μ m CMOS LNA for UHF-RFID reader application," In Proceedings of 8th IEEE International Conference on Solid-State and Integrated Circuit Technology, (2006).
- [Chiu 2005] Chiu, H.-W., Lu, S.-S. and Lin, Y.-S., "A 2.17-dB NF 5-GHz-band monolithic CMOS LNA with 10-mW dc power consumption," *IEEE Transaction on Microwave Theory Technology*, vol.53 (3), pp.813-824, (2005).
- [Choi 2003] Choi, P., Park, H. C., Kim, S., Park, S., Nam, I., Kim, T. W., Park, S., Shin, S., Kim, M. S., Kang, K., Ku, Y., Choi, H., Park, S. M. and Lee, K., "An Experimental Coin-Sized Radio for Extremely Low-Power WPAN (IEEE 802.15.4) Application at 2.4 GHz," *IEEE Journal of Solid-State Circuits*, vol.38 (12), pp.2258 - 2268, (2003).
- [Cimino 2007] Cimino, M., Lapuyade, H., De Matos, M., Taris, T., Deval, Y. and Begueret, J. B., "A RF circuit design methodology dedicated to critical applications," In Proceedings of 2007. *RME. Ph.D. Research in Microelectronics and Electronics Conference*, pp.53 - 56, (2007).
- [Colantonio 2004] Colantonio, P., Giannini, F., Limiti, E. and Teppati, V., "An Approach to Harmonic Load- and Source-Pull Measurements for High-Efficiency PA Design," *IEEE Transactions on Microwave Theory and Techniques*, vol.52 (1, Part 1), pp.191 - 198, (2004).
- [Conde 2002] Ortiz-Conde, A., Garcia Sanchez, F. J., Liou, J. J., Cerdeira, A., Estrada, M. and Yue, Y., "A review of recent MOSFET threshold voltage extraction methods," *Microelectronics Reliability*, vol.42 (4-5), pp.583-596, (2002).
- [Crols 1998] Crols, J. and Steyeart, M., "Low-IF Topologies for High-Performance Analog Front Ends of Fully Integrated Receivers," *IEEE Transactions on Circuits and Systems-II: Analog and Digital Signal Processing*, vol.45, pp.269-282, (1998).
- [CyberInI 2007] CyberInI LLC., "Stepping up to Wireless ?," <http://digital.cyberini.com/2007/03/19/stepping-up-to-wireless-.aspx>, (Accessed: 2007).

-
- [Dambrine 1988] Dambrine, G., Cappy, A., Heliodore, F. and Playez, E., "A new method for determining the FET small-signal equivalent circuit," *IEEE Transactions on Microwave Theory and Techniques*, vol.36 (7), pp.1151 - 1159, (1988).
- [Egels 2004] Egels, M., Gaubert, J., Pannier, P. and Bourdel, S., "Design method for fully integrated CMOS RF LNA," *Electronics Letters*, vol.40 (24), pp.1513-14, (2004).
- [Enz 2000] Enz, C. and Cheng, Y., "MOS transistor modeling for RF IC design," *IEEE Journal of Solid-State Circuits*, vol.35 (2), pp.186 - 201, (2000).
- [Everard 2001] Everard, J., "Fundamentals of RF Circuit Design," John Wiley & Sons, Ltd, ISBN: 9780471497936, (2001).
- [Fouad 2002] Fouad, H., Sharaf, K., El-Diwanly, E. and El-Hennawy, H., "An RF CMOS cascode LNA with current reuse and inductance source degeneration," In *Proceedings of Nineteenth National Radio Science Conference, 19th NRSC*, pp.450-457, (2002).
- [Gao 2005] Gao, S., Butterworth, P., Ooi, S. and Sambell, A., "High-Efficiency Power Amplifier Design Including Input Harmonic Termination," *IEEE Microwave and Wireless Components Letters*, vol.16 (2), pp.81-83, (2005).
- [Gatta 2001a] Gatta, F., Sacchi, E., Svelto, F., Vilmercati, P. and Castello, R., "A 2-dB noise figure 900-MHz differential CMOS LNA," *IEEE Journal on Solid-State Circuits*, vol.36 (10), (2001).
- [Gatta 2001b] Gatta, F., Sacchi, E., Svelto, F., Vilmercati, P. and Castello, R., "A 2-dB Noise Figure 900-MHz Differential CMOS LNA," *IEEE Journal Of Solid-State Circuits*, vol.36 (10), pp.1444 - 1452, (2001).
- [Ge 1998] Ge, Y. and Mayaram, K., "A Comparative Analysis of CMOS Low Noise Amplifiers For RF Applications," In *Proceedings of the 1998 IEEE International Symposium on Circuits and Systems, ISCAS '98*, vol.4, pp.349 - 352, (1998).
- [Gianesello 2007] Gianesello, F., Gloria, D., Raynaud, C. and Boret, S., "5GHz 1.4 dB NF CMOS LNA integrated in 130 nm High Resistivity SOI Technology," In *Proceedings of 2007 IEEE International Symposium on Integrated Circuits, ISIC-2007*, pp.120-123.
- [Glover 2005] Glover, I. A., Pennock, S. and Shepherd, P., "Microwave Devices, Circuits and Subsystems for Communications Engineering: Devices, Circuits and Subsystems: Vol. 1," John Wiley & Sons, ISBN: 978-0471899648, (2005).
- [Gonzalez 1996] Gonzalez, G., "Microwave Transistor Amplifiers: Analysis and Design," Prentice Hall, 2nd edition, ISBN: 0132543354, (1996).
- [Goo 2002] Goo, J.-S., Ahn, H.-T., Ladwig, D. J., Yu, Z., Lee, T. H. and Dutton, R. W., "A Noise Optimization Technique for Integrated Low-Noise Amplifiers," *IEEE Journal Of Solid-State Circuits*, vol.37 (8), pp.994 - 1002, (2002).
- [Govind 2004] Govind, V., Dalmia, S. and Swaminathan, M., "Design of integrated low noise amplifiers (LNA) using embedded passives in organic substrates," *IEEE Transactions on Advanced Packaging*, vol.27 (1), pp.79-89, (2004).
- [Guo 2002] Guo, W. and Huang, D., "The noise and linearity optimization for a 1.9-GHz CMOS low noise amplifier," In *Proceedings of 2002 IEEE Asia-Pacific Conference on ASIC*, pp.253 - 257, (2002).

-
- [Gutierrez 2001] Gutierrez, J. A., Naeve, M., Callaway, E., Bourgeois, M., Mitter, V. and Heile, B., "IEEE 802.15.4: A developing standard for low-power, low-cost wireless personal area networks," *IEEE Network*, vol.15 (5), pp.12-19, (2001).
- [Gutierrez 2003] Gutierrez, J. A., Callaway, E. H. and Barrett, R., "IEEE 802.15.4 Low-Rate Wireless Personal Area Networks: Enabling Wireless Sensor Networks," IEEE Press, ISBN: 0738135577, (2003).
- [Ha 1981] Ha, T. T., "Solid-State Microwave Amplifier Design," John Wiley & Sons, New York, ISBN: 0471089710, (1981).
- [Ho 2006] Ho, D. and Mirabbasi, S., "Low-Voltage Low-Power Low-Noise Amplifier for Wireless Sensor Networks," In Proceedings of *2006 Canadian Conference on Electrical and Computer Engineering, CCECE '06*, pp.1494-1497, (2006).
- [Hooge 1969] Hooge, F.N., "1/f noise is no surface effect", *Physics Letters*, vol. 29(A), pp. 139, (1969).
- [Hooge 1981] Hooge, F.N., Kleinpenning, T.G.M., and Vandamme, L.K.J., "Experimental studies on 1/f noise", *Reports on Progress in Physics*, vol. 44 (5), pp. 497-532, (1981).
- [Huang 1997] Huang, Q., Piazza, F., Orsatti, P. and Ohguro, T., "The impact of scaling down to deep-submicron on CMOS RF circuits," In Proceedings of *the 23rd European Solid-State Circuits Conference, ESSCIRC'97*, pp.132 - 135, (1997).
- [Huang 2006] Huang, G., Kim, T.-S., Kim, B.-S., Yu, M. and Ye, Y., "Post linearization of CMOS LNA using double cascode FETS," In Proceedings of *2006 IEEE International Symposium on Circuits and Systems*, pp.4499-4502, (2006).
- [Hwang 2007] Hwang, Y.-S. and Yoo, H.-J., "A Low Power Folded RF Front-End with Merged LNA and Mixer for ZigBee/Bluetooth," In Proceedings of *2007 IEEE Radio and Wireless Symposium*, pp.85-86, (2007).
- [IEEE-SA 1999] IEEE-SA, "Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," ANSI/IEEE Std 802.11-1999, IEEE Standards Association, (1999).
- [IEEE-SA 2002] IEEE-SA, "Part 15.1: Wireless Personal Area Networks Medium Access Control (MAC) and Physical Layer (PHY) Specifications," IEEE Std 802.15.1-2002, IEEE Standards Association, (2002).
- [IEEE-SA 2003a] IEEE-SA, "Part 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low Rate Wireless Personal Area Networks (LR-WPANs)," IEEE 802.15.4-2003, IEEE Standards Association, (2003).
- [IEEE-SA 2003b] IEEE-SA, "Part 15.2: Coexistence of Wireless Personal Area Networks with Other Wireless Devices Operating in Unlicensed frequency Bands," IEEE Std 802.15.2™-2003, IEEE Standards Association, (2003).
- [IEEE-SA 2003c] IEEE-SA, "IEEE Standard for Information technology--Telecommunications and information exchange between systems--Local and metropolitan area networks--Specific requirements Part 15.3: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for High Rate Wireless Personal Area Networks (WPAN)," IEEE 802.15.3-2003, IEEE Standards Association, (2003).

- [IEEE-SA 2004] IEEE-SA, "IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems," IEEE 802.16-2004, IEEE Standards Association, (2004).
- [IEEE-SA 2007a] IEEE-SA, "IEEE 802.11 WIRELESS LOCAL AREA NETWORKS - The Working Group for WLAN Standards," IEEE Standards Association, <http://www.ieee802.org/11/>, (Accessed: 2007).
- [IEEE-SA 2007b] IEEE-SA, "IEEE 802.15 Working Group for WPAN," IEEE Standards Association, <http://www.ieee802.org/15/>, (Accessed: 2007).
- [IEEE-SA 2007c] IEEE-SA, "The IEEE 802.16 Working Group on Broadband Wireless Access Standards," IEEE Standards Association, <http://www.ieee802.org/16/>, (Accessed: 2007).
- [IEEE-SA 2007d] IEEE-SA, "IEEE P1451.5 Project - Draft Standard for A Smart Transducer Interface for Sensors and Actuators - Wireless Communication Protocols and Transducer Electronic Data Sheets (TEDS) Formats," IEEE Standards Association, <http://grouper.ieee.org/groups/1451/5/>, (Accessed: 2007).
- [IEEE-WSZ 2007] IEEE Wireless Standards Zone," <http://standards.ieee.org/wireless>, (Accessed: 2007).
- [Jargon 2004] Jargon, J. A., DeGroot, D. C. and Gupta, K. C., "Frequency-Domain Models for Nonlinear Microwave Devices Based on Large-Signal Measurements," *Journal of Research of the National Institute of Standards and Technology*, vol.109 (4), pp.407-427, (2004).
- [Jen 1999] Jen, S. H.-M.,ENZ, C. C., Pehlke, D. R., Schroter, M. and Sheu, B. J., "Accurate modeling and parameter extraction for MOS transistors valid up to 10 GHz," *IEEE Transactions on Electron Devices*, vol.46 (11), pp.2217 - 2227, (1999).
- [Jin 1998] Jin, X., Ou, J.-J., Chen, C.-H., Liu, W., Deen, M. J., Gray, P. R. and Hu, C., "An effective gate resistance model for CMOS RF and noise modeling," In *Proceedings of 1998 International Electron Devices Meeting, IEDM '98 Technical Digest*, pp.961 - 964, (1998).
- [Jindal 1984] Jindal, R. P., "Noise associated with distributed resistance of MOSFET gate structures in integrated circuits," *IEEE Transactions on Electron Devices*, vol. 31(10), pp. 1505–1509, (1984).
- [Kim 1999] Kim, H.-S., Li, X. and Ismail, M., "A 2.4GHz CMOS Low Noise Amplifier using an Inter-stage Matching Inductor," In *Proceedings of 42nd Midwest Symposium on Circuits and Systems*, vol.2, pp.1040 - 1043, (1999).
- [Kim 2000] Kim, B., Ko, J.-S. and Lee, K., "A new linearization technique for MOSFET RF amplifier using multiple gated transistors," *IEEE Microwave and Guided Wave Letters*, vol.10 (9), pp.371 - 373, (2000).
- [Kim 2001] Kim, B., Ko, J.-S. and Lee, K., "Highly linear CMOS RF MMIC Amplifier Using Multiple Gated Transistors and its Volterra series Analysis," In *Proceedings of IEEE MTT-S International Microwave Symposium Digest*, vol.3, pp.515-518, (2001).

- [Kim 2004] Kim, T. W., Kim, B. and Lee, K., "Highly Linear receiver front-end adopting MOSFET transconductance linearisation by multiple gated transistors," *IEEE Journal of Solid State Circuits*, vol.39 (1), pp.223- 229, (2004).
- [Kim 2006a] Kim, T.-S. and Kim, B.-S., "Post-linearization of cascode CMOS low noise amplifier using folded PMOS IMD sinker," *IEEE Microwave and Wireless Components Letters*, vol.16 (April), pp.182-184, (2006).
- [Kim 2006b] Kim, N., Aparin, V., Barnett, K. and Persico, C., "A Cellular-Band CDMA 0.25-m CMOS LNA Linearized Using Active Post-Distortion," *IEEE Journal of Solid-State Circuits*, vol.41 (7), pp.1530 - 1534, (2006).
- [Kluge 2006] Kluge, W., Poegel, F., Roller, H., Lange, M., Ferchland, T., Dathe, L. and Eggert, D., "A Fully Integrated 2.4-GHz IEEE 802.15.4-Compliant Transceiver for ZigBee™ Applications," *IEEE Journal of Solid-State Circuits*, vol.41 (12), pp.2767 - 2775, (2006).
- [Kobayashi 1994] Kobayashi, K. W. and Oki, A. K., "A low-noise baseband 5-GHz direct-coupled HBT amplifier with common-base active input match," *IEEE Microwave and Guided Wave Letters*, vol.4 (11), pp.373 - 375, (1994).
- [Koolivand 2006] Koolivand, Y., Shoaie, O., Fotowat-Ahmadi, A., Zahabi, A. and Jabedar-Maralani, P., "Nonlinearity Analysis in ISD CMOS LNA's Using Volterra Series," In Proceedings of *ACM Great Lakes Symposium on VLSI 2006, GLSVLSI'06*, pp.135-139, (2006).
- [Kwon 2002] Kwon, I., Je, M., Lee, K. and Shin, H., "A simple and analytical parameter-extraction method of a microwave MOSFET," *IEEE Transactions on Microwave Theory and Techniques*, vol.50 (6), pp.1503 - 1509, (2002).
- [Kwon 2005] Kwon, I. and Lee, K., "An integrated low power highly linear 2.4-GHz CMOS receiver front-end based on current amplification and mixing," *IEEE Microwave and Wireless Components Letters*, vol.15 (1), pp.36 - 38, (2005).
- [Lambrianou 1985] Lambrianou, G. M. and Aitchison, C. S., "Optimization of third-order intermodulation product and output power from an X-band MESFET amplifier using Volterra series analysis," *IEEE Transactions on Microwave Theory and Techniques*, vol.MTT-33 (December), pp.1395-1403, (1985).
- [Lapuyade 2006] Cimino, M., De Matos, M., Lapuyade, H., Taris, T., Deval, Y. and Begueret, J.-B., "A Low-Power and Low Silicon Area Testable CMOS LNA Dedicated to 802.15.4 Sensor Network Applications," In Proceedings of *13th IEEE International Conference on Electronics, Circuits and Systems, ICECS '06*, pp.383-386, (2006).
- [Lavasani 2003] Lavasani, S. H. M. and Kiaei, S., "A new method to stabilize high frequency high gain CMOS LNA," In Proceedings of *2003 10th IEEE International Conference on Electronics, Circuits and Systems, ICECS 2003*, vol.3, pp.982 - 985, (2003).
- [Le 2004] Le, K. T., "Designing a ZigBee-ready IEEE 802.15.4 compliant radio transceiver," *R.F. Design*, vol.27 (11), pp.42, 44, 46, 48, 50, (2004).
- [Le 2005] Le, K. T., "ZigBee SoCs provide cost-effective solutions," *EE Times* online, <http://www.eetimes.com/showArticle.jhtml?articleID=173600329>, (Accessed: 2007).

- [Le 2006] Le, K. T., "Zigbee System-on-Chip (SoC) Design," *High Frequency Electronics*, pp.16-25, (2006).
- [Lee 1997] Lee, S., Yu, H. K., Kim, C. S., Koo, J. G. and Nam, K. S., "A Novel Approach to Extracting Small-Signal Model Parameters of Silicon MOSFET's," *IEEE Microwave and Guided Wave Letters*, vol.7 (3), pp.75 - 77, (1997).
- [Lee 2000] Lee, S. and Yu, H. K., "A Semianalytical Parameter Extraction of a SPICE BSIM3v3 for RF MOSFET's Using S-parameters," *IEEE Transactions on Microwave Theory and Techniques*, vol.48 (3), pp.412-416, (2000).
- [Lee 2002] Lee, S., "A parameter extraction method for a small-signal MOSFET model including substrate parameters," In Proceedings of *2002 IEEE International Conference on Semiconductor Electronics, ICSE 2002*, pp.255 - 260, (2002).
- [Lee 2004] Lee, S.-H., Juang, Y.-Z., Chiu, C.-F. and Chiou, H.-K., "A novel low noise design method for CMOS L-degeneration cascoded LNA," In Proceedings of *The 2004 IEEE Asia-Pacific Conference on Circuits and Systems*, pp.273-276, (2004).
- [Leman 2004] Leman, D., "Technology approaches for battery-based and battery-less Tire-Pressure Monitoring Systems," In Proceedings of *Automotive Sensors and Applications*, Detroit, (2004).
- [Leong 2007] Leong, P. B., Cheung, D., Kumar, N., Lin, L., Tan, S., Wang, H., Wang, S., Cahn, S. K., Chien, G., Cho, T., Song, Y. H. and Tse, L., "World's First 90 nm CMOS Single-Chip Bluetooth and WLAN with Integrated RF," *2007 IEEE International Symposium on Integrated Circuits, ISIC-2007*, pp.377-380, (2007).
- [Li 1999] Li, X., Kim, H.-S., Ismail, M. and Olsson, H., "A Novel design approach for GHz CMOS Low Noise Amplifiers," In Proceedings of *1999 IEEE Radio and Wireless Conference, RAWCON 99*, pp.285-8, (1999).
- [Li 2002] Li, Q. and Yuan, J. S., "Linearity analysis and design optimisation for 0.18 μ m CMOS RF mixer," *IEE Proceedings-Circuits Devices System*, vol.149 (2), pp.112-118, (2002).
- [Liang 1999] Liang, C.-P., Jong, J.-H., Stark, W. E. and East, J. R., "Nonlinear Amplifier Effects in Communications Systems," *IEEE Transactions on Microwave Theory and Techniques*, vol.47 (8), pp.1461 - 1466, (1999).
- [Liao 2003] Liao, C.-H. and Chuang, H.-R., "A 5.7-GHz 0.18- μ m CMOS Gain-Controlled Differential LNA with Current reuse for WLAN receiver," *IEEE Microwave and Wireless Components Letters*, vol.13 (12), pp.526-528, (2003).
- [Linten 2005] Linten, D., Thijis, S., Natarajan, M. I., Wambacq, P., Jearnsaksiri, W., Ramos, J., Mercha, A., Jenei, S., Donnay, S. and Decoutere, S., "A 5-GHz fully integrated ESD-protected low-noise amplifier in 90-nm RF CMOS," *IEEE Journal of Solid-State Circuits*, vol.40 (7), pp.1434-1442, (2005).
- [Liou 2003] Liou, J. J. and Schwierz, F., "RF MOSFET: Recent advances, current status and future trends," *Solid-State Electronics*, vol.47 (11), pp.1881-1895, (2003).
- [Lovelace 1994] Lovelace, D., Costa, J. and Camilleri, N., "Extracting small-signal model parameters of silicon MOSFET transistors," In Proceedings of *IEEE MTT-S International Microwave Symposium Digest*, vol.2, pp.865 - 868, (1994).

- [Maas 1998] Maas, S., "Nonlinear Microwave Circuits," IEEE Press, ISBN: 0780334035, (1998).
- [Maas 1999a] Maas, S., "What You Need To Know About Volterra-Series Analysis," http://www.eltm.ru/store/AWR/MWO_AN_Awr_VS_General.PDF, (1999).
- [Maas 1999b] Maas, S. A., "FET Models for Volterra Series Analysis," Press release, Applied Wave Research Inc., El Segundo, CA, (1999).
- [Maeda 1995] Maeda, M., Masato, H., Takehara, H., Nakamura, M., Morimoto, S., Fujimoto, H., Ota, Y. and Ishikawa, O., "Source Second-Harmonic Control for High Efficiency Power Amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol.43 (12 pt 2), pp.2952-2958, (1995).
- [Maxim 1996] Maxim, "MAX 2450 Data Sheet," (1996).
- [Maxim 2000] Maxim, "QPSK Modulation Demystified," Maxim application note 686, (2000).
- [McWhorter 1957] McWhorter, A. L., "Semiconductor Interface Physics", *University of Philadelphia Press*, Philadelphia, (1957).
- [Miers 1992] Miers, T. H. and Hirsch, V. A., "A Thorough Investigation Of Dynamic Bias On Linear Gam Fet Power Amplifier Performance," In *Proceedings of IEEE MTT-S International Microwave Symposium Digest*, vol.2, pp.537-540, (1992).
- [Minowa 1990] Minowa, M., Onoda, M., Fukuda, E. and Daido, Y., "Backoff improvement of an 800-MHz GaAs FET amplifier for a QPSK transmitter using an adaptive nonlinear distortion canceller," In *Proceedings of IEEE 40th Vehicular Technology Conference*, pp.542 - 546, (1990).
- [Mnif 2001] Mnif, K., "A Smart Tire Pressure Monitoring System," *Sensors Magazine*, November, (2001).
- [Montenegro 2007] Montenegro, G., Kushalnagar, N., Hui, J. and Culler, D., "RFC4944: Transmission of IPv6 Packets over IEEE 802.15.4 Networks." IETF, Request for Comments, <http://www.ietf.org/rfc/rfc4944.txt>, (2007).
- [Nadezhin 2003] Nadezhin, D., Gavrilov, S., Glebov, A., Egorov, Y., Zolotov, V., Blaauw, D., Becer, M., Ardelea, A. and Patel, A., "SOI transistor model for fast transient simulation," In *Proceedings of 2003 IEEE/ACM International Conference on Computer-Aided Design, Digest of Technical Papers*, pp.120-127, (2003).
- [Nguyen 2004] Nguyen, T.-K., Kim, C.-H., Ihm, G.-J., Yang, M.-S. and Lee, S.-G., "CMOS Low-Noise Amplifier Design Optimization Techniques," *IEEE Transactions On Microwave Theory And Techniques*, vol.52 (5), pp.1433 - 1442, (2004).
- [Nguyen 2005] Nguyen, T.-K., Han, S.-K. and Lee, S.-G., "Ultra-low-power 2.4 GHz image-rejection low-noise amplifier," *Electronics Letters*, vol.41 (15), pp.842 - 843, (2005).
- [Nguyen 2006] Nguyen, T.-K., Oh, N.-J., Han, S.-K. and Lee, S.-G., "A low power CMOS RF transmitter front-end for 2.4 GHz ZigBee applications," In *Proceedings of 2006 IEEE Radio and Wireless Symposium*, pp.43- 46, (2006).

-
- [Nojima 1985] Nojima, T. and Konno, T., "Cuber predistortion linearizer for relay equipment in 800 MHz band land mobile telephone system," *IEEE Transactions on Vehicular Technology*, vol.VT-34 (4), pp.169 - 177, (1985).
- [Notor 2003] Notor, J., Caviglia, A. and Levy, G., "CMOS RFIC Architectures for IEEE 802.15.4 Networks," White Paper, Cadence Design Systems, Inc., <http://www.cadence.com/whitepapers/CMOSRFICArchforIEEE80215.pdf>, (2003).
- [Ock 2002] Ock, S., Han, K., Lee, J.-R. and Kim, B., "A modified cascode type Low Noise Amplifier using dual common source transistors," In Proceedings of *IEEE MTT-S International Microwave Symposium Digest*, vol.3, pp.1423-1426, (2002).
- [Park 2001] Park, P., Kim, C. S. and Yu, H. K., "Linearity, Noise Optimisation for Two Stage RF CMOS LNA," In Proceedings of *IEEE Region 10 International Conference on Electrical and Electronic Technology. TENCON 2001*, vol.2, pp.756-8, (2001).
- [Pedro 1994] Pedro, J. C. and Perez, J., "Accurate simulation of GaAs MESFET's intermodulation distortion using a new drain-source current model," *IEEE Transactions on Microwave Theory and Techniques*, vol.42 (1), pp.25-33, (1994.).
- [Pinel 2002] Pinel, S., Chakraborty, S., Venkataraman, S., Bhatia, R., Mandal, S., Nuttinck, S., Larson, B. and Laskar, J., "Development of SOI based MMIC's for Wireless LAN Applications," In Proceedings of *2002 IEEE MTT-S International Microwave Symposium Digest*, vol.2, pp.1053-1056, (2002).
- [Poazar 2005] Pozar, D. M., "Microwave Engineering," Wiley, Third Edition, ISBN: 0-471-44878-8, (2005).
- [Rafla 1999] Rafla, R. A. and El-Gamal, M. N., "Design of a 1.5V CMOS Integrated 3 GHz LNA," In Proceedings of *ISCAS '99. Proceedings of the 1999 IEEE International Symposium on Circuits and Systems*, vol.2, pp.440 - 443, (1999).
- [Ragheb 2007] Ragheb, T., Nieuwoudt, A. and Massoud, Y., "Parasitic-aware analytical modeling of integrated CMOS inductively degenerated narrow-band low noise amplifiers," *Analog Integrated Circuits and Signal Processing*, vol.51 (1), pp.11-17, (2007).
- [Razavi 1997a] Razavi, B., "RF Microelectronics," Prentice Hall, 1st edition, ISBN: 0138875715, (1997).
- [Razavi 1997b] Razavi, B., "Design Considerations for Direct-Conversion Receivers," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol.44 (6), pp.428 - 435, (1997).
- [Roy 2006] Roy, N., Najmabadi, M., Raut, R. and Devabhaktuni, V., "A systematic approach towards the implementation of a low-noise amplifier in sub-micron CMOS technology," In Proceedings of *2006 IEEE Canadian Conference on Electrical and Computer Engineering*, (2006).
- [Sarhangian 2007] Sarhangian, S. and Atarodi, S. M., "A Low-Power CMOS Low-IF Receiver Front-End for 2450-MHz Band IEEE 802.15.4 ZigBee Standard," In Proceedings of *2007. IEEE International Symposium on Circuits and Systems, 2007 ISCAS*, pp.433 - 436, (2007).

- [Schmale 1997] Schmale, I., "An Improved Physics-based Nonquasistatic FET model," In *Proceedings of 27th European Microwave Conference and Exhibition*, pp.328 - 3, (1997).
- [Schroeder 1998] Schroeder, D. K., "Semiconductor material and device characterization," Wiley-Interscience, New York, 2nd ed, ISBN: 0471241393, (1998).
- [Schwierz 2000] Schwierz, F., "Microwave transistors - the last 20 years," In *Proceedings of the 2000 Third IEEE International Caracas Conference on Devices, Circuits and Systems, ICCDCS*, pp.D28/1 - D28/7, Cancun, Mexico, (2000).
- [Seidel 1971] Seidel, H., "A microwave feedforward experiment," *Bell Systems Technical Journal*, vol.50, pp.2879-2916, (1971).
- [Shaeffer 1997] Shaeffer, D. K. and Lee, T. H., "A 1.5-V, 1.5-GHz CMOS Low Noise Amplifier," *IEEE Journal of Solid-State Circuits*, vol.32 (5), pp.745-759, (1997).
- [Sheng 1991] Sheng, N. H., Ho, W. J., Wang, N. L., Pierson, R. L., Asbeck, P. M. and Edwards, W. L., "A 30 GHz bandwidth AlGaAs-GaAs HBT direct-coupled feedback amplifier," *IEEE Microwave and Guided Wave Letters*, vol.1 (8), pp.208 - 210, (1991).
- [Siemens 2007] Siemens VDO, "Tire Information Systems: Tire pressure monitoring and intelligent tire," http://www.vdo.com/products_solutions/cars/safety/tire-info-system/index.htm, (Accessed: 2007).
- [Singh 2006] Singh, R. D. and Yu, K.-W., "A Linear Mode CMOS Power Amplifier with Self-Linearizing Bias," In *Proceedings of 2006 IEEE Asian Solid-State Circuits Conference, ASSCC 2006*, pp.251 - 254, (2006).
- [Sivonen 2005] Sivonen, P. and Pärssinen, A., "Analysis and Optimization of Packaged Inductively degenerated Common-Source Low-Noise Amplifiers With ESD Protection," *IEEE Transactions On Microwave Theory And Techniques*, vol.53 (4), pp.1304 - 1313, (2005).
- [Snider 1967] Snider, D. M., "A theoretical analysis and experimental confirmation of the optimally loaded and overdriven rf power amplifier," *IEEE Transactions on Electron Devices*, vol.ED-14 (12), pp.851-857, (1967).
- [Sun 2005] Sun, I.-S. M., Ng, W. T., Kanekiyo, K., Kobayashi, T., Mochizuki, H., Toita, M., Imai, H., Ishikawa, A., Tamura, S. and Takasuka, K., "Lateral high-speed bipolar transistors on SOI for RF SoC applications," *IEEE Transactions on Electron Devices*, vol.52 (7), pp.1376 - 1383, (2005).
- [Sun 2006] Sun, K.-J., Tsai, Z.-M., Lin, K.-Y. and Wang, H., "A noise optimization formulation for CMOS low-noise amplifiers with on-chip low-Q inductors," *IEEE Transactions on Microwave Theory and Techniques*, vol.54 (2006), pp.1554-1560, (2006).
- [Taniguchi 2001] Taniguchi, E., Maeda, K., Ikushima, T., Sadahiro, K., Itoh, K., Suematsu, N. and Takagi, T., "Dual Bias Feed SiGe HBT Low Noise Linear Amplifier," In *Proceedings of IEEE MTT-S International Microwave Symposium Digest*, vol.3, pp.285-288, (2001).
- [Terada 2001] Terada, K., Nishiyama, K. and Hatanaka, K.-I., "Comparison of MOSFET-threshold-voltage extraction methods," *Solid-State Electronics*, vol.45 (1), pp.35-40, (2001).

- [Tong 2004] Tong, A. F., Yeo, K. S., Jia, L., Geng, C. Q., Ma, J.-G. and Do, M. A., "Simple and accurate extraction methodology for RF MOSFET valid up to 20 GHz," *IEEE Proceedings: Circuits, Devices and Systems*, vol.151 (6), pp.587-592, (2004).
- [Toole 2004] Toole, B., Plett, C. and Cloutier, M., "RF Circuit Implications of Moderate Inversion Enhanced Linear Region in MOSFETs," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol.51 (2), pp.319-328, (2004).
- [Tulunay 2004] Tulunay, G. and Balkir, S., "A compact optimization methodology for single-ended LNA," In *Proceedings of IEEE International Symposium on Circuits and Systems*, pp.V-273-V-276, (2004).
- [Valenza 2004] Valenza, M., Hoffmann, A., Sodini, D., Laigle, A., Martinez, F. and Rigaud, D., "Overview of the impact of downscaling technology on 1/f noise in p-MOSFETs to 90 nm" *IEE Proceedings - Circuits, Devices and Systems*, vol.151(2), pp.102 – 110, (2004).
- [Vidojkovic 2004] Vidojkovic, V., Tang, J. v. d., Hanssen, E., Leeuwenburgh, A. and Roemund, A. v., "Fully-integrated DECT/Bluetooth multi-band LNA in 0.18 CMOS," In *Proceedings of 2004 IEEE International Symposium on Circuits and Systems*, pp.I-565-I-568, (2004).
- [Viztmuller 1995] Viztmuller, P., "RF Design Guide: Systems, Circuits, and Equations," Artech House, 1st Edition, ISBN: 978-0890067543, (1995).
- [Voinigescu 2004] Voinigescu, S. P., Tazlauanu, M., Ho, P. C. and Yang, M. T., "Direct extraction methodology for geometry-scalable RF-CMOS models," In *Proceedings of The 2004 International Conference on Microelectronic Test Structures, ICMTS '04*, pp.235 - 240, (2004).
- [Wambacq 1998] Wambacq, P. and Sansen, W. M. C., "Distortion Analysis of Analog Integrated Circuits," Springer, 1st Edition, ISBN: 978-0792381860, (1998).
- [Webster 1996] Webster, D. R., Haigh, D. G., Scott, J. B. and Parker, A. E., "Derivative superposition-a linearisation technique for ultra broadband systems," *IEE Colloquium Wideband Circuits, Modelling and Techniques*, (May), pp.3/1 - 3/14, (1996).
- [Weiner 1980] Weiner, D. D. and Spina, J. F., "Sinusoidal analysis and modeling of weakly nonlinear circuits," van Nostrand Reinhold, New York, (1980).
- [Wilkinson 1992] Wilkinson, B., "Digital Systems Design," Prentice Hall International, 2 Sub edition, ISBN: 0132202867, (1992).
- [Wolfram 2007] Wolfram Research Inc., "Mathematica," <http://www.wolfram.com/>, (Accessed: 2007).
- [Wu 2006] Wu, D.-K., Huang, R. and Wang, Y.-Y., "A low-voltage and low-power CMOS LNA using forward-body-bias NMOS at 5GHz," In *Proceedings of IEEE 8th International Conference on Solid-State and Integrated Circuit Technology*, (2006).
- [Yang 2001] Yang, X., Wu, T. and McMacken, J., "Design of LNA at 2.4 GHz Using 0.25um CMOS Technology," In *Proceedings of 2001 Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems*, (2001).

- [Yang 2003] Yang, M.-S. and Lee, S.-G., "Linearity Improvement Technique for the Common-Source Transconductance Stage," In *Proceedings of IEEE, Midwest International Conference on Circuit And System (MWSCAS)*, Cairo, Egypt, (2003).
- [Yao 2006] Yao, T., Gordon, M., Yau, K., Yang, M. T. and Voinigescu, S. P., "60-GHz PA and LNA in 90-nm RF-CMOS," In *Proceedings of 2006 IEEE Radio Frequency Integrated Circuits(RFIC) Symposium*, pp.125-128, (2006).
- [Young 1999] Young, P. H., "Electronic Communication Techniques," Prentice Hall, Fourth Edition, ISBN: 0137799845, (1999).
- [Zencir 2003] Zencir, E., Huang, T.-H., Dogan, N. S. and Arvas, E., "A low-power UHF differential LNA in 0.35- μ m SOI CMOS," In *Proceedings of 2003 IEEE International Radio Frequency Integrated Circuits Symposium*, pp.111-114, (2003).
- [Zencir 2005] Zencir, E., Huang, T.-H., Tekin, A., Dogan, N. S. and Arvas, E., "UHF RF Front-End Circuits in 0.35- μ m Silicon on Insulator (SOI) CMOS," *Journal of Analog Integrated Circuits and Signal Processing*, vol.45 (3), pp.231-245, (2005).
- [Zhang 2003] Zhang, C., Huang, D. and Lou, D., "Optimization of cascode CMOS low noise amplifier using inter-stage matching network," In *Proceedings of 2003 IEEE Conference on Electron Devices and Solid-State Circuits*, pp.465 - 468, (2003).
- [Zhenying 2003] Zhenying, L., Rustagi, S. C., Li, M. F. and Lian, Y., "A 1V, 2.4GHz fully integrated LNA using 0.18 μ m CMOS technology," In *Proceedings of IEEE 5th International Conference on ASIC*, vol.2, pp.1062-5, (2003).
- [van der Ziel 1969] van der Ziel, A., "Representation of noise in linear two-ports," *Proceedings of the IEEE*, vol.57 (6), pp.1211- 1211, (1969).
- [van der Ziel 1986] van der Ziel, A., "Noise in Solid State Devices and Circuits," Wiley-Interscience, New York, ISBN: 0471832340, (1986).
- [Zhang 2001] Zhang, Z. and Lau, J., "Experimental study on MOSFET's flicker noise under switching conditions and modelling in RF applications" In *Proceedings of IEEE Conference on Custom Integrated Circuits, 2001*, pp. 393-396, (2001).
- [Zigbee 2007] Zigbee, "ZigBee Alliance," <http://www.zigbee.org>, (Accessed: 2007).
- [Zito 2006] Zito, D., Pepe, D. and Neri, B., "Low-Power RF Transceiver for IEEE 802.15.4 (ZigBee) Standard Applications," In *Proceedings of 13th IEEE International Conference on Electronics, Circuits and Systems, ICECS 2006*, pp.1312 - 1315, (2006).

Appendix A EXPLANATION OF CODE FUNCTIONS USED IN TRANSISTOR NONLINEAR MODELLING

A.1 Automated Parameter Extraction

This section presents the ADS data display window coding functions used for the automated parameter extraction of the extrinsic and intrinsic model parameters, and for writing the extracted values to data tables.

A.2.1 Extraction of Extrinsic Parameters

According to the extraction method in [Tong 2004], the extrinsic resistances R_s , R_d and R_g are extracted using Equations (A.1), (A.2) and (A.3), respectively. The ADS data display window functions used to implement this extraction are detailed in Figure A.1.

$$R_s = \text{real}(Z_{12}) = \text{real}(Z_{21}) \quad (\text{A.1})$$

$$R_d = \text{real}(Z_{22}) - R_s \quad (\text{A.2})$$

$$R_g = \text{real}(Z_{11}) - R_s \quad (\text{A.3})$$

Eqn Rd_plot=real(Z(2,2))-Rs

Eqn Rg_plot=real(Z(1,1))-Rs

Eqn Rs1_plot=real(Z(2,1))

Eqn Rs2_plot=real(Z(1,2))

Eqn Rs_plot=(Rs1_plot+Rs2_plot)/2

(a)

Eqn mid_index=find_index(freq, 5GHz)

Eqn Rd=Rd_plot[mid_index]

Eqn Rg=Rg_plot[mid_index]

Eqn Rs1=Rs1_plot[mid_index]

Eqn Rs2=Rs2_plot[mid_index]

Eqn Rs=(Rs1+Rs2)/2

(b)

Figure A.1 Automated extraction of extrinsic resistances (a) Initial extraction equations and (b) Equations to calculate the final values

In Figure A.1.a, `Rx_plot` is a variable used to write the Z-parameter equations to evaluate the extrinsic resistances (where `x` is either `d`, `g`, or `s` for the drain, gate or source resistance, respectively), reserving the variable name `Rx` for the final value of each resistor. `Rs1` and `Rs2` represent the two different approaches to evaluating R_s which are suppose to yield equal values, but as can be seen in Figure 4.2.c `Rs1` and `Rs2` slightly differ and hence `Rs_plot` takes the average. In Figure A.1.b, the index number for the mid point (at 5GHz) is found using the `find_index` function. This constitutes an average value for the resistor and can be taken as a final value. The final values of R_d and R_g are therefore taken as their values at this point. The final value of R_s is the average of the two averages of `Rs1` and `Rs2`.

A.2.2 Extraction of Intrinsic Parameters

This section details the extraction of the intrinsic parameters of the transistor small signal model. These include: g_m , g_{ds} , C_{gs} , C_{gd} , R_{sub} , C_{ds} and C_{jd} .

A.1.2.1 Extraction of g_m and g_{ds}

The extraction of the transconductance, g_m and the output conductance, g_{ds} is detailed in Figure A.2 based on Equations (A.4) and (A.5)

$$g_m = \text{real}(Y_{21})_{\omega=0} \quad (\text{A.4})$$

$$g_{ds} = \text{real}(Y_{22})_{\omega^2=0} \quad (\text{A.5})$$

Eqn <code>gm_plot=real(Y(2,1))</code>	Eqn <code>gds_plot=real(Y(2,2))</code>
Eqn <code>w_gm_plot=what(gm_plot)</code>	Eqn <code>w_gds_plot=what(gds_plot)</code>
Eqn <code>gm=gm_plot[:,::,0]</code>	Eqn <code>gds=gds_plot[:,::,0]</code>
Eqn <code>w_gm=what(gm)</code>	Eqn <code>w_gds=what(gds)</code>
(a)	(b)

<code>w_gm</code>	<code>w_gds</code>	<code>w_gm_plot</code>	<code>w_gds_plot</code>
Dependency : [VDS,VGS] Num. Points : [41, 37] Matrix Size : scalar Type : Real	Dependency : [VDS,VGS] Num. Points : [41, 37] Matrix Size : scalar Type : Real	Dependency : [VDS,VGS,freq] Num. Points : [41, 37, 11] Matrix Size : scalar Type : Real	Dependency : [VDS,VGS,freq] Num. Points : [41, 37, 11] Matrix Size : scalar Type : Real

(c)

Figure A.2 Extraction of g_m and g_{ds} (a) g_m , (b) g_{ds} and (c) dependency table on swept parameters

In the first line of Figure A.2.a and Figure A.2.b, `gm_plot` and `gds_plot` write the extraction equations for g_m and g_{ds} , respectively. Their dependency as shown in the table in Figure A.2.c is on VDS, VGS and the frequency, `freq`. The graph of `gm_plot` and `gds_plot` is a 4-way relationship displaying the variable versus frequency at various gate and drain voltage values. According to Equations (A.4) and (A.5) the value chosen for g_m and g_{ds} at any bias point should be its value at zero frequency. In Figure A.2.a, the equation

`gm=gm_plot[:, :, 0]` accounts for this with the aid of the index number feature. It ensures that the `gm` parameter will take from `gm_plot` all values of the variable resulting from variance against all points of the first dependant, `VDS`, (identified by “:” which means the whole range) and the second dependant, `VGS`, but should only take values at the first point of the third dependant, `freq`, (identified by the index number of this point that is 0). `gds` is treated similarly as shown in Figure A.2.b.

A.1.2.2 Extraction of C_{gs} and C_{gd}

According to [Tong 2004], the values of C_{gs} and C_{gd} are extracted using Equations (A.6) and (A.7). The implementation of these equations in ADS data display window is shown in Figure A.3

$$C_{gs} = -\frac{\text{imag}(Y_{11}) + \text{imag}(Y_{12})}{\omega} \quad (\text{A.6})$$

$$C_{gd} = -\frac{\text{imag}(Y_{12})}{\omega} \quad (\text{A.7})$$

Eqn `Cgs_plot=(im(Y(1,1))+im(Y(1,2)))/(2*pi*freq)`

Eqn `Cgd_plot=-im(Y(1,2))/(2*pi*freq)`

Figure A.3 Extraction of C_{gs} and C_{gd}

Figure A.3 is a direct and straightforward implementation of Equations (A.6) and (A.7). This is since no dependencies or selections are involved in the extraction of these two parameters. The situation is very different in the extraction of R_{sub} as will be seen in the next section.

A.1.2.3 Extraction of R_{sub}

According to [Tong 2004], $Y_{sub} = Y_{22} - g_{ds} - j\omega C_{ds} - j\omega C_{gd}$ and R_{sub} is the slope of the $\omega^2 / \text{real}(Y_{sub})$ versus ω^2 plot. The extraction of R_{sub} is described in Figure A.4.

In Figure A.4.a, $\text{real}(Y_{sub})$ is calculated by subtracting g_{ds} from the real part of Y_{22} . Before this subtraction is performed, the dependency of `gds` has to be expanded. As can be seen from the table in Figure A.4.b, Y_{22} is dependant on `VDS`, `VGS` and `freq` while the dependency on `freq` was removed from `gds` as the value for `gds` was taken at 0 frequency as previously explained. ADS data display window function rules dictate that in order for the subtraction to work, the two elements in the equation must have the same number of dependencies. The `[expand]` function does that by spreading the dependency of the parameter to the total number of swept variables in the simulation. As can be seen in Figure A.4.c, for a range of dependant voltages, the `expand` function does not change the relationship between the frequency and `gds`, instead, it copies the value of `gds` to all points of the frequency sweep, so `gds` is constant

for all the frequency range. This does not have any implication on the physical or practical meaning of the result, but it is only a work-around to satisfy the condition of the ADS equation solver that elements must have the same dependencies. In the rest of the equations of Figure A.4.a, the angular frequency is calculated and `Rsub_plot` is used to evaluate $\omega^2 / \text{real}(Y_{sub})$. `Rsub_plot` is plotted against ω^2 in Figure A.4.d for one bias point.

Eqn `ReYsub=real(Y(2,2))-exp_gds`

Eqn `exp_gds=expand(gds)`

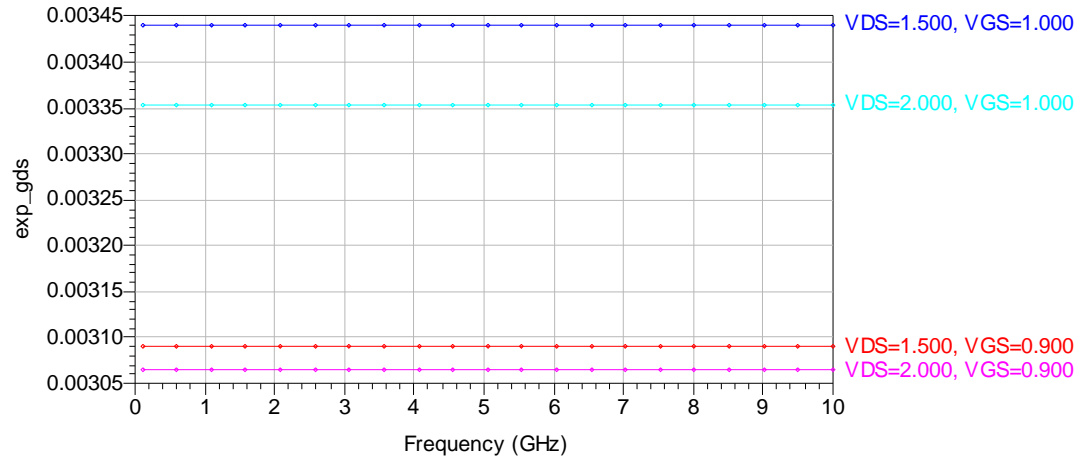
Eqn `AngFreqSquare=pow((2*pi*freq),2)`

Eqn `Rsub_plot=AngFreqSquare/ReYsub`

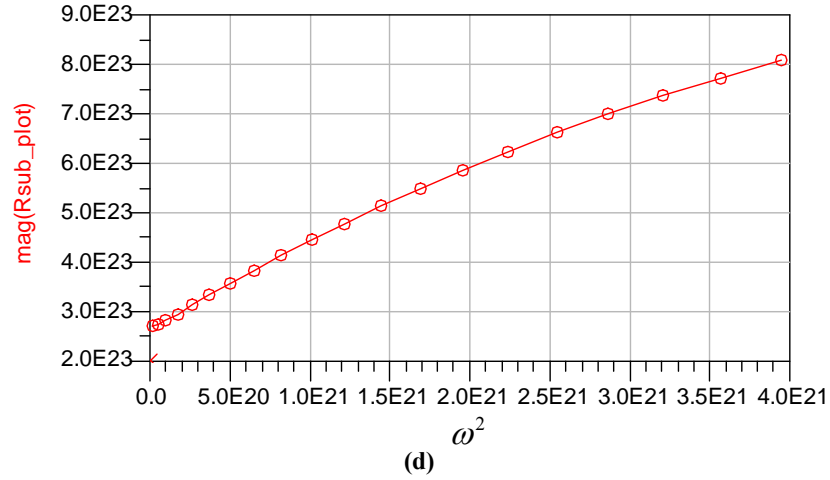
(a)

Info_freq	Info_gds	Info_exp_gds	Info_AngFreqSquare
Dependency : [VDS,VGS,freq] Num. Points : [41, 37, 21] Matrix Size : scalar Type : Real	Dependency : [VDS,VGS] Num. Points : [41, 37] Matrix Size : scalar Type : Real	Dependency : [VDS,VGS,] Num. Points : [41, 37, 1] Matrix Size : scalar Type : Real	Dependency : [VDS,VGS,freq] Num. Points : [41, 37, 21] Matrix Size : scalar Type : Real
Info_ReYsub	Info_Y22	Info_Rsub	Info_Rsub_plot
Dependency : [VDS,VGS,freq] Num. Points : [41, 37, 21] Matrix Size : scalar Type : Real	Dependency : [VDS,VGS,freq] Num. Points : [41, 37, 21] Matrix Size : scalar Type : Complex	Dependency : [VDS,VGS] Num. Points : [41, 37] Matrix Size : scalar Type : Real	Dependency : [VDS,VGS,freq] Num. Points : [41, 37, 21] Matrix Size : scalar Type : Real

(b)



(c)



```
Eqn Rsub_freq_1=5.5GHz
Eqn Rsub_freq_2=9.5GHz
Eqn idx_Rsub_1=find_index(freq[0,0,:], Rsub_freq_1)
Eqn Rsub_point_1=mag(Rsub_plot)[:::,idx_Rsub_1]
Eqn idx_Rsub_2=find_index(freq[0,0,:], Rsub_freq_2)
Eqn Rsub_point_2=mag(Rsub_plot)[:::,idx_Rsub_2]
```

(e)

```
Eqn Rsub=(Rsub_point_2-Rsub_point_1)
      /(AngFreqSquare[idx_Rsub_2]-AngFreqSquare[idx_Rsub_1])
```

(f)

Figure A.4 Equations and plots for the extraction of R_{sub}

The first two equations in Figure A.4.e specify the two points on the frequency curve between which the slope is calculated. `idx_Rsub_1=find_index(freq[0,0,:], Rsub_freq_1)`, finds the index point of the first frequency value selected using the `find_index` function. The `find_index` function returns the index number of the point in the sweep searched for, and because it is a single-dimensional function, it can only search one sweep list at a time. The first item in the function is the list in which the value searched is located and the second is the value searched for. To specify the correct list, and make it of a single dimension, `freq[0,0,:]` is used. This is because as can be seen from the dependency table in Figure A.4.b, `freq` is dependant on `VGS`, `VDS` and `freq`, so the expression `[0,0,:]` eliminates the first two dependencies and keeps the full list of the third (`freq`) which need to be searched. This also ensures that this step is independent of the bias condition, avoiding any conflicts when the process is repeated for other bias points, thus facilitating automation.

`Rsub_point_1=mag(Rsub_plot)[:::,:::,idx_Rsub_1]` finds the value of `mag(Rsub_plot)` at the specified frequency (pointed to by its index number at `idx_Rsub_1`) for all the swept points of VGS and VDS (specified by “: :” at their dependency location in the function), according to the dependency table in Figure A.4.b. This ensures that the extraction process is automatically repeated for the gate and drain voltage ranges, and is not sensitive to any change in the specified range or the size of step desired. Similarly, the second value of `mag(Rsub_plot)` for the calculation of the slope is found. Finally, in Figure A.4.f the slope of $\omega^2/\text{real}(Y_{sub})$ versus ω^2 plot (the value of `Rsub`) is found by calculating the difference between the two values of `mag(Rsub_plot)` and dividing by the difference of their corresponding ω^2 . `Rsub` now becomes dependant only on VGS and VDS, ensuring the ease of automatically repeating the procedures for other simulated bias points.

A.1.2.4 Extraction of C_{jd} and C_{ds}

According to [Tong 2004], the values of C_{jd} and C_{ds} are extracted using Equations (A.8) and (A.9). Based on these equations, the automated extraction of these parameters in ADS data display window is shown in Figure A.5

$$C_{jd} = \left(\frac{\omega^2 R_{sub}}{\text{real}(Y_{sub})} - \omega^2 R_{sub}^2 \right)^{-1/2} \quad (\text{A.8})$$

$$C_{ds} = \frac{\text{imag}(Y_{22})}{\omega} - C_{gd} - \frac{C_{jd}}{1 + \omega^2 R_{sub}^2 C_{jd}^2} \quad (\text{A.9})$$

Eqn `exp_Rsub=expand(Rsub)`

Eqn `Cjd_plot=pow((((AngFreqSquare*exp_Rsub)/ReYsub)
-AngFreqSquare*pow(exp_Rsub,2)), -0.5)`

(a)

Eqn `exp_Cjd=expand(Cjd)`

Eqn `exp_Cgd=expand(Cgd)`

Eqn `Cds_plot=(im(Y(2,2))/(2*pi*freq))-exp_Cgd-
-(exp_Cjd/(1+(AngFreqSquare*pow(exp_Rsub,2)
*pow(exp_Cjd,2))))`

(b)

Info_Rsub	Info_ReYsub	Info_Y22	Info_Cjd	Info_Cgd
Dependency : [VDS,VGS] Num. Points : [41, 37] Matrix Size : scalar Type : Real	Dependency : [VDS,VGS,freq] Num. Points : [41, 37, 21] Matrix Size : scalar Type : Real	Dependency : [VDS,VGS,freq] Num. Points : [41, 37, 21] Matrix Size : scalar Type : Complex	Dependency : [VDS,VGS] Num. Points : [41, 37] Matrix Size : scalar Type : Real	Dependency : [VDS,VGS] Num. Points : [41, 37] Matrix Size : scalar Type : Real

(c)

Figure A.5 The extraction of C_{jd} and C_{ds}

Again, the names `Cjd_plot` and `Cds_plot` are used to evaluate C_{jd} and C_{ds} versus frequency. For C_{jd} in Figure A.5.a, the expanded version of `Rsub` is used due to the dependency mismatch between it and `ReYsub` as shown in the table in Figure A.5.c. Similarly, for C_{ds} in Figure (b), the expanded versions of `Cjd`, `Cgd` and `Rsub` are used due to the dependency mismatch between them and `im(Y22)`.

A.2.3 Finalising Intrinsic Capacitance Values

Figure A.7 explains the automation of the process to calculate the average value of intrinsic capacitances against frequency.

```
Eqn Cap_freq_1=3GHz
Eqn Cap_freq_2=6.5GHz
Eqn Cap_freq_3=10GHz

Eqn idx_Cap_1=find_index(freq[0,0,:], Cap_freq_1)
Eqn idx_Cap_2=find_index(freq[0,0,:], Cap_freq_2)
Eqn idx_Cap_3=find_index(freq[0,0,:], Cap_freq_3)

Eqn Cgs=(Cgs_plot[:, :, :, idx_Cap_1]+Cgs_plot[:, :, :, idx_Cap_2]+Cgs_plot[:, :, :, idx_Cap_3])/3
Eqn Cgd=(Cgd_plot[:, :, :, idx_Cap_1]+Cgd_plot[:, :, :, idx_Cap_2]+Cgd_plot[:, :, :, idx_Cap_3])/3
Eqn Cjd=(mag(Cjd_plot[:, :, :, idx_Cap_1]+mag(Cjd_plot[:, :, :, idx_Cap_2]+mag(Cjd_plot[:, :, :, idx_Cap_3])/3
Eqn Cds=(mag(Cds_plot[:, :, :, idx_Cap_1]+mag(Cds_plot[:, :, :, idx_Cap_2]+mag(Cds_plot[:, :, :, idx_Cap_3])/3
```

Figure A.6 Taking the average of the capacitance values versus frequency

The first group of equations in Figure A.7 selects three frequency values spread over the frequency sweep range from which points for the calculation of the average are taken. These frequencies represent the beginning, middle and end of the frequency range. The second group of equations in Figure A.7 find the indices of these points by searching an independent `freq` variable as previously explained. The last group of equations in Figure A.7 evaluate `Cx_plot` at the set frequency values and takes the average by dividing the combined value by 3. All final capacitance values now become independent of frequency. The most important feature of this setup is that it ensures the repetition of this averaging process for all combinations of gate and drain voltages irrespective of the range and the step used. This was facilitated by the use of “: : , : :” in the third group of equations in Figure A.7, indicating that this process should be performed at every combination of the gate and drain voltage sweep lists.

A.2.4 Writing Extracted Parameters Values to Data Tables

Figure A.7, demonstrates how the `[write_var]` function is used to transfer the extracted parameter values to text files to form the data tables, using C_{gs} as an example. The main drawback of the `write_var` function is that it is a single-dimensional function which means it

does not accept any of the extracted parameters in their usual form since they are all multi-dimensional (dependant on both VGS and VDS).

Eqn	VDS_info=what(VDS)			
Eqn	VGS_info=what(VGS)			
Eqn	Cgs_info=what(Cgs)			
Eqn	VDS_point=40			
Eqn	indep_VGS=VGS[VDS_point, :]			
Eqn	Cgs_vs_VGS=Cgs[VDS_point, :]			
Eqn	write_var_Cgs=write_var("Cgs_VDS.txt", "A", Cgs_list_name, " ", "f", 21, indep_VGS, Cgs_vs_VGS)			
Eqn	Cgs_list_name=strcat("End\n", " ", "\n", "VAR VDS(1)=1\n", "Begin TestData\n", "%VGS(1) Cgs(1)")			

VDS_info	VGS_info	Cgs_info
Dependency : [VDS] Num. Points : [41] Matrix Size : scalar Type : Real	Dependency : [VDS,VGS] Num. Points : [41, 37] Matrix Size : scalar Type : Real	Dependency : [VDS,VGS] Num. Points : [41, 37] Matrix Size : scalar Type : Real

Cgs_vs_VGS_info	indep_VGS_info
Dependency : [VGS] Num. Points : [37] Matrix Size : scalar Type : Real	Dependency : [VGS] Num. Points : [37] Matrix Size : scalar Type : Real

Figure A.7 Writing the extracted data to a text file

In its current form, if Cgs was displayed in a list, it will be displayed with VDS leading VGS. This means the table will display the swept values of VGS (and their corresponding Cgs values) as a subsection to the swept values of VDS. This is in line with their dependencies as shown in the table of Figure A.7 where VGS and Cgs are dependant on both VDS and VGS but VDS is dependant only on itself. The way to deal with the single-dimensional feature of the write_var function is to remove the dependency of VGS and Cgs on VDS by choosing one VDS point at a time to work on, as shown in Figure A.7. When the value in VDS_point is changed, everything on the data display is re-evaluated and the new list of VGS and its corresponding Cgs is added to the table. Cgs_list_name writes the headers and the ends of every section of the data so they are added before and after every list respectively, eliminating the need for doing this manually.

A.2 Generating I_{gm} and I_{gds} for DC Current Model

This section describes the coded function used to automatically split the DC current characteristics into g_m - and g_{ds} -related components. This starts by first accurately determining the threshold voltage of the transistor, and then using this to determine the point at which the transistor enters saturation, defining the point at which the current should be split.

A.2.1 Accurate V_T identification

Figure A.8 details the data display window coding equations that were used to automatically obtain the linear extrapolation of the I_{DS} vs. V_{GS} characteristics at its maximum g_m point. In Figure A.8.a, the VDS point at which the characteristics is plotted is specified.

VDSindex=find_index (indepVDS,VDSpoint) finds the index of the selected VDS value, to be used in subsequent equations. indepVDS=VDS[0,: :] is used to remove the dependence of VDS on VGS. This is because the find_index function is only capable of operating on one sweep list (single dimensional), and this is a double sweep where VGS is leading VDS. Also, the VGS point at which maximum transconductance, MAXgm, occurs is saved to a variable and its index is found.

```
Eqn VDSpoint=0.01
Eqn indepVDS=VDS[0,: :]
Eqn VDSindex=find_index(indepVDS,VDSpoint)
Eqn MAXgmVGS=0.78
Eqn MAXgmVGSindex=find_index(VGS,MAXgmVGS)
```

(a)

```
Eqn IdsChange=Ids[(MAXgmVGSindex+1),VDSindex]-Ids[(MAXgmVGSindex-1),VDSindex]
Eqn VGSCheck=VGS[(MAXgmVGSindex+1)]-VGS[(MAXgmVGSindex-1)]
Eqn Slope=IdsChange/VGSCheck
Eqn intLinearLine=Slope*(VGS)
```

(b)

```
Eqn MAXgmVGSpoint=Ids[MAXgmVGSindex, VDSindex]
Eqn MAXgmPoint=Ids[MAXgmVGSindex, VDSindex]
Eqn MAXgmLinearLinePoint=intLinearLine[MAXgmVGSindex]
Eqn LineDifference=MAXgmLinearLinePoint-MAXgmPoint
```

(c)

```
Eqn LinearLine=intLinearLine-LineDifference
```

(d)

Figure A.8 Equations for plotting the linear extrapolation of I_{DS} vs. V_{GS} characteristics at maximum g_m

In Figure A.8.b, the slope of the I_{DS} - V_{GS} characteristics is calculated at the point at which g_m is maximum. For ultimate accuracy, both the IdsChange and VGSCheck are calculated from just one point above and below the MAXgm point. An initial linear line (intLinearLine) is plotted using the calculated slope. This line has the same slope as the I_{DS} - V_{GS} characteristic curve at MAXgm but it does not intersect it at this point. Instead, it crosses the VGS axis at the VGS=0 point, as shown in Figure A.9. In order to make that line cross the MAXgm point, in Figure A.8.c. the difference in I_{DS} between the maximum g_m point on the characteristic curve and the corresponding point on the initial linear line is calculated. The final linear line in Figure A.8.d is evaluated as the initial line minus that difference. The final linear line crosses the

characteristic curve at the maximum g_m point as required and extrapolates the characteristics in this area crossing the V_{GS} axis as shown in Figure A.9.

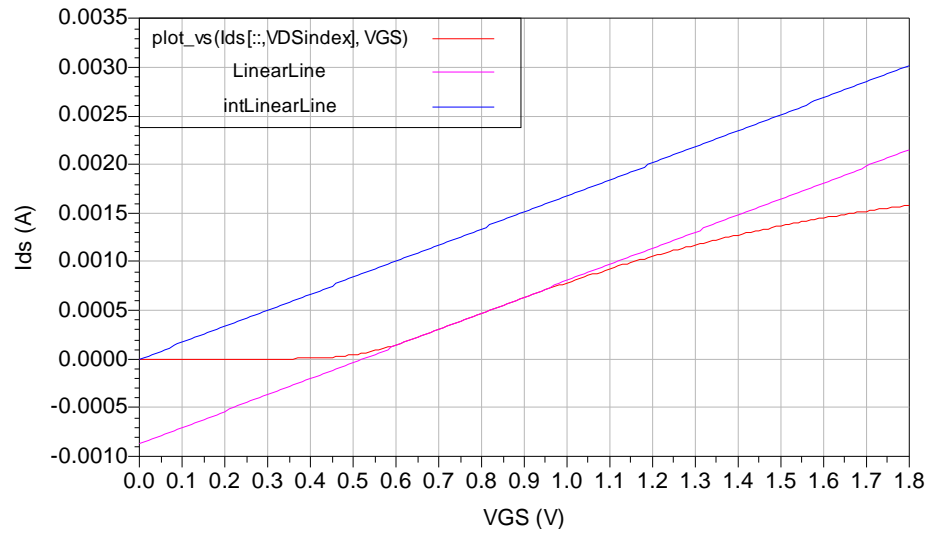


Figure A.9 V_T extraction using ELR method

The above code was designed to be very adaptable in that it can be used to accurately find the threshold voltage of any transistor, extremely quickly, and accurately, no matter what type it is.

A.2.2 Splitting DC current to I_{g_m} and $I_{g_{ds}}$

Few lines of the transistor's current characteristics are shown in Figure A.10 for demonstration.

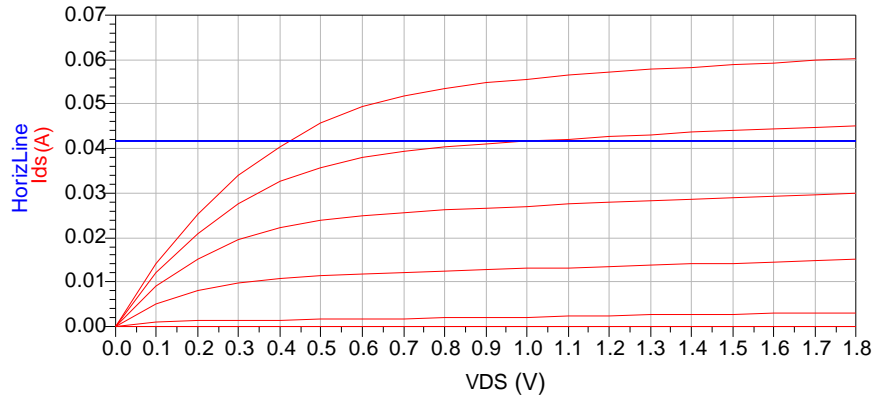


Figure A.10 Transistor's DC characteristics

The ADS data display window coding functions used to split this current into I_{g_m} and $I_{g_{ds}}$ are shown in Figure A.11.


```

Eqn VT=0.52
Eqn VGSindex=36
Eqn SatPoint=VDS-VGS[VGSindex]+VT
Eqn SatPointAdj=SatPoint[VGSindex,:]
Eqn SatPointIndex=find_index(SatPointAdj, 0)
Eqn Ids=IDS.i[:,::,0]
Eqn IDSKnee=Ids[VGSindex, SatPointIndex]
Eqn HorizLine=Ids[VGSindex, SatPointIndex]
Eqn Difference=Ids[VGSindex,:]-HorizLine
Eqn lgds= if (Difference<0) then 0 else Difference
Eqn lgm=Ids[VGSindex,:]-lgds
    
```

(a)

VGS_Info	VDS_Info	IDS_Info	Ids_Info
Dependency : [VGS] Num. Points : [37] Matrix Size : scalar Type : Real	Dependency : [VGS,VDS] Num. Points : [37, 19] Matrix Size : scalar Type : Real	Dependency : [VGS,VDS,freq] Num. Points : [37, 19, 1] Matrix Size : scalar Type : Real	Dependency : [VGS,VDS] Num. Points : [37, 19] Matrix Size : scalar Type : Real

(b)

SatPoint_Info	SatPointAdj_Info
Dependency : [VGS,VDS] Num. Points : [37, 19] Matrix Size : scalar	Dependency : [VDS] Num. Points : [19] Matrix Size : scalar

(c)

Figure A.11 Coded functions for the split of the transistor's current characteristics

In Figure A.11.a, the variable VT is used to store the value of V_T obtained from the implementation of the ELR method. $VGSindex=x$ is a variable used to specify the VGS curve selected to operate on using its index number in the VGS sweep list. $SatPoint=VDS-VGS[VGSindex]+VT$ is based on the condition for saturation in the MOSFET transistor which is specified as $V_{DS} = V_{GS} - V_T$. It subtracts the VGS value of the chosen VGS curve (specified using $VGS[VGSindex]$) and the threshold voltage from all the points on the VDS sweep list. $SatPointIndex=find_index(SatPointAdj, 0)$ then searches all resultant values for the point when $SatPoint$ is equal to 0 and saves the index of that point in $SatPointIndex$. Because $SatPoint$ is dependant on both VGS and VDS as in Figure A.11.c, and that the `find_index` function is only capable of operating on one sweep list, $SatPointAdj$ reduces this dependency to only VDS. $Ids=IDS[:,::,0]$ removes the dependency of IDS on $freq$ making Ids dependant on the full sweep lists of VGS and VDS only as in Figure A.11.b. The knee point of the selected VGS curve is then specified by the index of the chosen VGS curve and the index of the point on that curve when $V_{DS} = V_{GS} - V_T$ ($SatPointIndex$). Both indices are used to plot a horizontal line over all the points of the VDS sweep list and crosses the respective VGS curve at its obtained knee point. $Difference=Ids[VGSindex,:]-HorizLine$ calculates the difference between the

current in the chosen V_{GS} curve and `HorizLine` for all the V_{DS} values. As can be seen from Figure A.10, if channel modulation (the effect of V_{DS} on I_{DS}) did not exist, the current after the saturation point will be equivalent to the horizontal line and any additional current is due to g_{ds} . As previously mentioned, the current in the linear region is added to I_{gm} . `Igds=if(Difference<0) then 0 else Difference` then compares the calculated `Difference` to 0. If `Difference` was less than 0, that means the transistor is in the linear region and no current needs to be added from that area to I_{gds} , so I_{gds} is 0. If `Difference` was larger than 0, then the transistor is in the saturation region and that difference current is due to g_{ds} so the value in `Difference` is given to I_{gds} . `Igm=Ids[VGSindex,:]-Igds` calculates the g_m current by deducting the I_{gds} current from the total current of the chosen V_{GS} curve at every point of the V_{DS} sweep list. Because I_{gds} is 0 in the linear region, all the linear region current goes to I_{gm} and effectively, I_{gm} is equal to the horizontal line in the saturation region. To choose another V_{GS} curve, the value in `VGSindex` is manually changed. When that value is changed, all parts of the code are automatically re-executed and values of I_{gm} and I_{gds} for the new V_{GS} curve are calculated.

A.3 Linearising g_m by Modifying its Data Table

In Figure A.12.a, the variable `VDSpoint` is used to allow the user to select the value of simulated drain voltages where the I_{gm} - V_{GS} characteristics used for linearisation is to be plot. Generally, any point in the saturation area of the highest V_{GS} characteristic curve is valid. Here, the highest V_{DS} is selected for convenience. `VDSindex=find_index(indepVDS,VDSpoint)` finds the index of the selected V_{DS} value where `indepVDS` is used to remove the dependence of V_{DS} on V_{GS} as previously explained in Section A.2.1. Note how `VDSindex` is used to plot I_{gm} vs. V_{GS} at the selected value of V_{DS} in Figure 4.29. `midVGS` is a variable used to store the value of the chosen V_{GS} bias point at which the tangent line will be drawn. Here, it is done in the middle of simulated V_{GS} values to try to get as maximally valid linearisation as possible at any bias point, but any other V_{GS} value can also be selected. `midVGSindex=find_index(VGS,midVGS)` finds the index number of this point to be used in subsequent equations.

The equations in Figure A.12.b find the slope of the I_{gm} - V_{GS} characteristic curve at the chosen point (V_{GS} bias point) to be used for drawing the linear tangent. `IgmChange=TableIgm[(midVGSindex+1),VDSindex]-TableIgm[midVGSindex,VDSindex]` calculates

the change in I_{gm} . The change is calculated between the selected bias point (specified by `midVGSindex` and `VDSindex`) and the next swept point on the VGS scale (specified by `(midVGSindex+1), VDSindex`). This setup ensures that this process is adaptable to any change in selected bias points. Similarly, $VGS_{Change} = VGS[(midVGSindex+1)] - VGS[midVGSindex]$ calculates the change in VGS. Calculating the change in this way, over one point only, ensures that the slope of the characteristics will be calculated from only this tiny portion of the line, increasing accuracy. The slope of the characteristic line is calculated in $Slope = I_{gm}Change / VGS_{Change}$ by dividing the change in I_{gm} by the change in VGS.

The equations in Figure A.12.c plot the linear $I_{gm} - V_{GS}$ characteristic line. The first 4 equations plot an initial linear line using the calculated slope, and make it overlap the characteristic line centralised at the chosen bias point in a similar fashion to the implementation of the ELR method explained in Section A.2.1. $LinearI_{gm} = \text{if } (LinearLine < 0) \text{ then } 0 \text{ else } LinearLine$ removes the negative part of the $LinearLine$ and establishes 0 characteristics for I_{gm} from the VGS point at which the $LinearLine$ crosses 0 I_{gm} . This is done by comparing the value of $LinearLine$ to 0; if it was less than 0, $LinearI_{gm}$ is given 0, if it was equal or more than 0, $LinearI_{gm}$ is given the value of $LinearLine$, producing the linearised characteristic in Figure 4.30

```
Eqn VDSpoint=4
Eqn VDSindex=find_index(indepVDS,VDSpoint)
Eqn midVGS=1.5
Eqn midVGSindex=find_index(VGS,midVGS)
```

(a)

```
Eqn IgmChange=TableIgm[(midVGSindex+1),VDSindex]-TableIgm[midVGSindex,VDSindex]
Eqn VGSChange=VGS[(midVGSindex+1)]-VGS[midVGSindex]
Eqn Slope=IgmChange/VGSChange
```

(b)

```
Eqn intLinearLine=Slope*(VGS)
Eqn midIgmPoint=TableIgm[midVGSindex, VDSindex]
Eqn LineDifference=midLinearLinePoint-midIgmPoint
Eqn LinearLine=intLinearLine-LineDifference
Eqn LinearIgm= if (LinearLine<0) then 0 else LinearLine
```

(c)

Eqn $VDSSatPoint = VDS[SatPointIndex]$

Eqn $VDSSat = VDSSatPoint[VGSindex]$

Eqn $LinearIgmVGS = \text{if } (VDS[VGSindex, :] < VDSSat) \text{ then } Ids[VGSindex, :] \text{ else } LinearIgm[VGSindex]$

VDSSatPoint_Info	VDSSat_Info
Dependency : [VGS] Num. Points : [61]	Dependency : [] Num. Points : [1]

(d)

Eqn $TableIgmSelect = TableIgm[VGSindex, :]$

(e)

Figure A.12 Linearising g_m by modifying the I_{gm} data table

I_{gm} levels will be changed in every V_{GS} curve only after saturation; this is performed in the equations of Figure A.12.d. $VDSSatPoint = VDS[SatPointIndex]$ finds the V_{DS} value at which saturation occurs. $SatPointIndex$ was obtained and defined in Figure A.11. Note that since VDS is dependant on VGS , this will be repeated for every VGS characteristic curve (61 points), as indicated in the dependency table shown in the table in Figure A.12.d. However, as discussed earlier, it is required that every VGS curve is handled independently. $VDSSat = VDSSatPoint[VGSindex]$ corrects this by finding the point on the VDS scale at which the transistor enters saturation on the specific VGS curve being manipulated (specified using $VGSindex$). In $LinearIgmVGS = \text{if } (VDS[VGSindex, :] < VDSSat) \text{ then } Ids[VGSindex, :] \text{ else } LinearIgm[VGSindex]$, the Igm current level in each VGS curve is modified to the new value taken from the linear tangent line after the saturation point (on the VDS scale). This is done by searching all VDS values if they were smaller or larger than the saturation point of that VGS curve. If it is below, the value of $LinearIgmVGS$ is taken from the original nonlinear current table ($Ids[VGSindex, :]$), if it is on or after saturation, the value of Igm is given its value from the linearised $I_{gm}-V_{GS}$ characteristic line in Figure 4.30. $[VGSindex, :]$ indicates that the whole (on the VDS scale, specified by ‘:’) of a single VGS curve (specified by $VGSindex$) is being handled at a time. In Figure A.12.e, a single VGS curve is selected using $TableIgmSelect = TableIgm[VGSindex, :]$ to demonstrate how the current changes, the resultant curve is shown in Figure 4.31.

Appendix B IIP₃ DERIVATION FOR A COMMON-SOURCE MOSFET USING VOLTERRA AND TAYLOR SERIES

In this appendix, a full derivation is provided for the IIP₃ of a common-source transistor with an input matching impedance and a source degeneration inductance. The purpose of this is to demonstrate how Volterra and Taylor series can be combined to describe the nonlinearities of the transistor considering both its memory and memory-less nonlinear parameters.

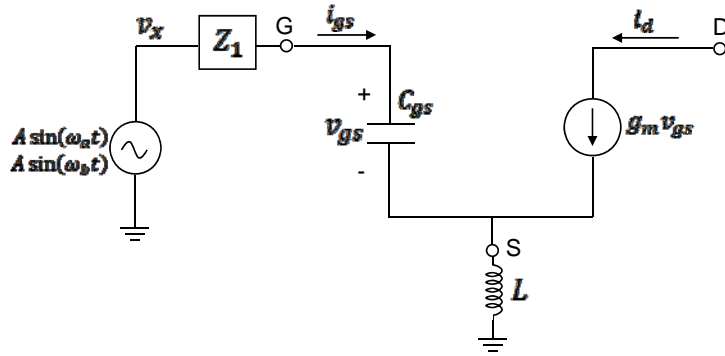


Figure B.1 A simplified MOSFET model with input matching impedance

Nonlinearity of frequency-independent part (Taylor series)

$$i_d = g_1 v_{gs} + g_2 v_{gs}^2 + g_3 v_{gs}^3 + \dots \quad (\text{B.1})$$

Nonlinearity of frequency-dependant part (Volterra series)

$$v_{gs} = A_1(s)v_x + A_2(s_1, s_2)v_x^2 + A_3(s_1, s_2, s_3)v_x^3 + \dots \quad (\text{B.2})$$

Nonlinearity of both parts (Volterra series)

$$i_d = B_1(s)v_x + B_2(s_1, s_2)v_x^2 + B_3(s_1, s_2, s_3)v_x^3 + \dots \quad (\text{B.3})$$

The Volterra series in (B.3) effectively combines the Taylor series in (B.1) and the Volterra series in (B.2) in such a manner where $B_1(s)$ is determined in terms of the 1st-order terms in (B.1) and (A.2), $B_2(s_1, s_2)$ is determined in terms of the 2nd- and lower-order terms of (B.1) and (B.2), and $B_3(s_1, s_2, s_3)$ is determined in terms of and the 3rd- and lower-order terms of (B.1) and (B.2).

First Step: To determine $B_1(s)$, $B_2(s_1, s_2)$ and $B_3(s_1, s_2, s_3)$ in terms of g_1 , g_2 , g_3 , $A_1(s)$, $A_2(s_1, s_2)$ and $A_3(s_1, s_2, s_3)$. This is done by substituting (B.2) in (B.1) and comparing the resulting terms with (B.3).

(B.2) in (B.1) gives

$$\begin{aligned} i_d = & g_1[A_1(s)v_x + A_2(s_1, s_2)v_x^2 + A_3(s_1, s_2, s_3)v_x^3] \\ & + g_2[A_1(s)v_x + A_2(s_1, s_2)v_x^2 + A_3(s_1, s_2, s_3)v_x^3]^2 \\ & + g_3[A_1(s)v_x + A_2(s_1, s_2)v_x^2 + A_3(s_1, s_2, s_3)v_x^3]^3 \end{aligned} \quad (B.4)$$

Collecting v_x terms from (B.4)

$$i_d = g_1 A_1(s) v_x \quad (B.5)$$

Comparing with (B.3)

$$\boxed{B_1(s) = g_1 A_1(s)} \quad (B.6)$$

Collecting v_x^2 terms from (B.4)

$$i_d = g_1 A_2(s_1, s_2) v_x^2 + g_2 [A_1(s)]^2 v_x^2 = g_1 A_2(s_1, s_2) v_x^2 + g_2 A_1(s_1) A_1(s_2) v_x^2 \quad (B.7)$$

Comparing with (B.3)

$$\boxed{B_2(s_1, s_2) = g_1 A_2(s_1, s_2) + g_2 A_1(s_1) A_1(s_2)} \quad (B.8)$$

Collecting v_x^3 terms from (B.4)

$$\begin{aligned} i_d = & g_1 A_3(s_1, s_2, s_3) v_x^3 + 2g_2 A_1(s) A_2(s_1, s_2) v_x^3 + g_3 [A_1(s)]^3 v_x^3 \\ i_d = & g_1 A_3(s_1, s_2, s_3) v_x^3 + 2g_2 \overline{A_1(s_1) A_2(s_2, s_3)} v_x^3 + g_3 A_1(s_1) A_1(s_2) A_1(s_3) v_x^3 \end{aligned} \quad (B.9)$$

Comparing with (B.3)

$$\boxed{B_3(s_1, s_2, s_3) = g_1 A_3(s_1, s_2, s_3) + 2g_2 \overline{A_1(s_1) A_2(s_2, s_3)} + g_3 A_1(s_1) A_1(s_2) A_1(s_3)} \quad (B.10)$$

Second Step: To determine $A_1(s)$, $A_2(s_1, s_2)$ and $A_3(s_1, s_2, s_3)$ using the harmonic input method

This needs working out KVL and KCL for the small signal circuit,

$$v_x = i_{gs} Z_1 + v_{gs} + (i_d + i_{gs}) sL$$

$$i_{gs} = v_{gs} s C_{gs}$$

$$v_x = v_{gs} s C_{gs} Z_1 + v_{gs} + (i_d + v_{gs} s C_{gs}) sL$$

$$v_x = v_{gs} s C_{gs} Z_1 + v_{gs} + i_d sL + v_{gs} s^2 C_{gs} L$$

$$v_x - i_d sL = v_{gs} (s C_{gs} Z_1 + 1 + s^2 C_{gs} L)$$

$$v_{gs} = \frac{v_x - i_d sL}{s C_{gs} Z_1 + 1 + s^2 C_{gs} L}$$

$$\text{Letting } a(s) = s C_{gs} Z_1 + 1 + s^2 C_{gs} L \text{ and } b(s) = sL \quad (B.11)$$

Then

$$\boxed{v_{gs} = \frac{v_x - i_d b(s)}{a(s)}} \quad (B.12)$$

To apply the harmonic input Method:

- 1- Let $v_x = e^{st}$ for $A_1(s)$, $v_x = e^{s_1 t} + e^{s_2 t}$ for $A_2(s_1, s_2)$ and $v_x = e^{s_1 t} + e^{s_2 t} + e^{s_3 t}$ for $A_3(s_1, s_2, s_3)$ in (B.2) and right hand side of (B.12)

- 2- Substitute new (B.2) in (B.1)
- 3- Substitute new (B.2) and new (B.1) in (B.12)
- 4- Compare e^{st} terms on both sides of new (B.12)

$A_1(s)$

In every step, only substitutions that would eventually result in e^{st} terms on either side of (B.12)¹ (B.12 reevaluated for determination of $A_1(s)$) are processed.

$$(B.2)^1: \quad v_{gs} = A_1(s)e^{st}$$

Substituting (B.2)¹ in (B.1) (Considering (B.5) to locate 1st-order terms)

$$(B.1)^1: \quad i_d = g_1 A_1(s)e^{st}$$

Substituting (B.2)¹ and (B.1)¹ in (B.12)

(B.12)¹:

$$A_1(s)e^{st} = \frac{e^{st} - g_1 A_1(s)e^{st}b(s)}{a(s)}$$

$$A_1(s)e^{st} + \frac{g_1 A_1(s)e^{st}b(s)}{a(s)} = \frac{e^{st}}{a(s)}$$

$$A_1(s)e^{st}a(s) + g_1 A_1(s)e^{st}b(s) = e^{st}$$

$$(B.12)^1: \quad (A_1(s)e^{st}(a(s) + g_1 b(s))) = e^{st}$$

Comparing e^{st} terms on both sides of (B.12)¹

$$A_1(s)(a(s) + g_1 b(s)) = 1$$

$$\boxed{A_1(s) = \frac{1}{a(s) + g_1 b(s)}} \quad (B.13)$$

$A_2(s_1, s_2)$

In every step, only substitutions that would eventually result in $e^{(s_1+s_2)t}$ terms on either side of (B.12)² (B.12 reevaluated for determination of $A_2(s_1, s_2)$) are processed. $e^{(s_1+s_2)t}$ terms will come wherever v_x^2 is present since $(e^{s_1 t} + e^{s_2 t})^2 = e^{2s_1 t} + e^{2s_2 t} + 2e^{(s_1+s_2)t}$.

$$(B.2)^2: \quad v_{gs} = A_2(s_1, s_2)v_x^2 = A_2(s_1, s_2) * 2e^{(s_1+s_2)t} = 2A_2(s_1, s_2)e^{(s_1+s_2)t}$$

Substituting (B.2)² in (B.1) (Considering (B.7) to locate 2nd-order terms)

$$(B.1)^2: \quad i_d = 2g_1 A_2(s_1, s_2)e^{(s_1+s_2)t} + 2g_2 A_1(s_1)A_1(s_2)e^{(s_1+s_2)t}$$

(B.12)²:

$$2A_2(s_1, s_2)e^{(s_1+s_2)t} = \frac{-[2g_1 A_2(s_1, s_2)e^{(s_1+s_2)t} + 2g_2 A_1(s_1)A_1(s_2)e^{(s_1+s_2)t}]b(s)}{a(s)}$$

Dividing (B.12)² by 2 and moving $\frac{-g_1 A_2(s_1, s_2) e^{(s_1+s_2)t} b(s)}{a(s)}$ to its left hand side

$$A_2(s_1, s_2) e^{(s_1+s_2)t} + \frac{g_1 A_2(s_1, s_2) e^{(s_1+s_2)t} b(s)}{a(s)} = \frac{2g_2 A_1(s_1) A_1(s_2) e^{(s_1+s_2)t} b(s)}{a(s)}$$

Simplifying

$$\frac{A_2(s_1, s_2) e^{(s_1+s_2)t} a(s) + g_1 A_2(s_1, s_2) e^{(s_1+s_2)t} b(s)}{a(s)} = \frac{2g_2 A_1(s_1) A_1(s_2) e^{(s_1+s_2)t} b(s)}{a(s)}$$

$$A_2(s_1, s_2) e^{(s_1+s_2)t} [a(s) + g_1 b(s)] = 2g_2 A_1(s_1) A_1(s_2) e^{(s_1+s_2)t} b(s)$$

Using (B.13),

$$A_2(s_1, s_2) e^{(s_1+s_2)t} \left[\frac{1}{A_1(s)} \right] = 2g_2 A_1(s_1) A_1(s_2) e^{(s_1+s_2)t} b(s)$$

With 2nd-order harmonic input $A_1(s) \Rightarrow A_1(s_1 + s_2)$ and $b(s) \Rightarrow b(s_1 + s_2)$

$$(B.12)^2: \quad A_2(s_1, s_2) e^{(s_1+s_2)t} = -A_1(s_1 + s_2) [g_2 A_1(s_1) A_1(s_2) e^{(s_1+s_2)t}] b(s_1 + s_2)$$

Comparing $e^{(s_1+s_2)t}$ terms on both sides of (B.12)²

$$\boxed{A_2(s_1, s_2) = -A_1(s_1 + s_2) g_2 A_1(s_1) A_1(s_2) b(s_1 + s_2)} \quad (B.14)$$

$$A_3(s_1, s_2, s_3)$$

In every step, only substitutions that would eventually result in $e^{(s_1, s_2, s_3)t}$ terms on either side of (B.12)³ (B.13 reevaluated for determination of $A_3(s_1, s_2, s_3)$) are processed. $e^{(s_1, s_2, s_3)t}$ terms come wherever v_x^3 is present having that $(e^{s_1 t} + e^{s_2 t} + e^{s_3 t})^3$ will produce $6e^{(s_1, s_2, s_3)t}$

$$(B.2)^3: \quad v_{gs} = A_3(s_1, s_2, s_3) e^{(s_1, s_2, s_3)t}$$

Substituting (B.2)³ in (B.1) (Considering (B.9) to locate 3rd-order terms)

$$(B.1)^3: \quad i_d = 6g_1 A_3(s_1, s_2, s_3) e^{(s_1, s_2, s_3)t} + 6g_2 \overline{A_1(s_1) A_2(s_2, s_3)} e^{(s_1, s_2, s_3)t} + 6g_3 A_1(s_1) A_1(s_2) A_1(s_3) e^{(s_1, s_2, s_3)t}$$

Substituting (B.2)³ and (B.1)³ in (B.12)

(B.12)³:

$$6A_3(s_1, s_2, s_3) e^{(s_1, s_2, s_3)t} = \frac{-6e^{(s_1, s_2, s_3)t} [g_1 A_3(s_1, s_2, s_3) + 2g_2 \overline{A_1(s_1) A_2(s_2, s_3)} + g_3 A_1(s_1) A_1(s_2) A_1(s_3)] b(s)}{a(s)}$$

Dividing (B.12)³ by 6 and moving $\frac{-g_1 A_3(s_1, s_2, s_3) b(s) e^{(s_1, s_2, s_3)t}}{a(s)}$ to the left hand side

$$A_3(s_1, s_2, s_3) e^{(s_1, s_2, s_3)t} + \frac{g_1 A_3(s_1, s_2, s_3) b(s) e^{(s_1, s_2, s_3)t}}{a(s)} =$$

$$\frac{-6e^{(s_1, s_2, s_3)t} [2g_2 \overline{A_1(s_1)A_2(s_2, s_3)} + g_3 A_1(s_1)A_1(s_2)A_1(s_3)] b(s)}{a(s)}$$

Simplifying

$$\begin{aligned} A_3(s_1, s_2, s_3) e^{(s_1, s_2, s_3)t} a(s) + g_1 A_3(s_1, s_2, s_3) b(s) e^{(s_1, s_2, s_3)t} &= \\ e^{(s_1, s_2, s_3)t} [2g_2 \overline{A_1(s_1)A_2(s_2, s_3)} + g_3 A_1(s_1)A_1(s_2)A_1(s_3)] b(s) &= \\ A_3(s_1, s_2, s_3) e^{(s_1, s_2, s_3)t} [a(s) + g_1 b(s)] &= \\ e^{(s_1, s_2, s_3)t} [2g_2 \overline{A_1(s_1)A_2(s_2, s_3)} + g_3 A_1(s_1)A_1(s_2)A_1(s_3)] b(s) &= \end{aligned}$$

Using (B.13),

$$A_3(s_1, s_2, s_3) e^{(s_1, s_2, s_3)t} \left[\frac{1}{A_1(s)} \right] = e^{(s_1, s_2, s_3)t} [2g_2 \overline{A_1(s_1)A_2(s_2, s_3)} + g_3 A_1(s_1)A_1(s_2)A_1(s_3)] b(s)$$

With 3rd-order harmonic input $A_1(s) \Rightarrow A_1(s_1 + s_2 + s_3)$ and $b(s) \Rightarrow b(s_1 + s_2 + s_3)$

$$\begin{aligned} \text{(B.12)}^3: \quad A_3(s_1, s_2, s_3) e^{(s_1, s_2, s_3)t} &= \\ -e^{(s_1 + s_2 + s_3)t} A_1(s_1 + s_2 + s_3) [2g_2 \overline{A_1(s_1)A_2(s_2, s_3)} + g_3 A_1(s_1)A_1(s_2)A_1(s_3)] b(s_1 + s_2 + s_3) &= \end{aligned}$$

Comparing $e^{(s_1 + s_2 + s_3)t}$ terms on both sides of (B.12)³

$$\boxed{A_3(s_1, s_2, s_3) = -A_1(s_1 + s_2 + s_3) [2g_2 \overline{A_1(s_1)A_2(s_2, s_3)} + g_3 A_1(s_1)A_1(s_2)A_1(s_3)] b(s_1 + s_2 + s_3)} \quad (\text{B.15})$$

Third Step: To determine IIP³ from (B.3) using

$$A_{IP3} = \sqrt{\frac{4}{3} \left| \frac{B_1(s)}{B_3(s_1, s_2, s_3)} \right|} \quad (\text{B.16})$$

To do this for a two-tone input $\omega_a + \omega_b$, set $s_1 = s_2 = s_b$ and $s_3 = -s_a$ to represent the IMD3 component $2\omega_b - \omega_a$. Assume closely spaced frequencies, hence $s_a \approx s_b \approx s$. (B.16) then changes to

$$A_{IP3}(2\omega_b - \omega_a) = \sqrt{\frac{4}{3} \left| \frac{B_1(s_a)}{B_3(s_b, s_b, -s_a)} \right|}$$

$$\text{And having } IIP_3(2\omega_b - \omega_a) = \frac{A_{IP3}(2\omega_b - \omega_a)^2}{8 * \Re[Z_1(s_a)]}$$

Then

$$\boxed{IIP_3(2\omega_b - \omega_a) = \frac{1}{6\Re(Z_1(s_a))} \left| \frac{B_1(s_a)}{B_3(s_b, s_b, -s_a)} \right|} \quad (\text{B.17})$$

Applying the two-tone input to (B.10)

$$B_3(s_b, s_b, -s_a) = g_1 A_3(s_b, s_b, -s_a) + 2g_2 \overline{A_1(s_b)A_2(s_b, -s_a)} + g_3 A_1(s_b)A_1(s_b)A_1(-s_a)$$

$$B_3(s_b, s_b, -s_a) = g_1 A_3(s_b, s_b, -s_a) + 2g_2 \overline{A_1(s_b)A_2(s_b, -s_a)} + g_3 A_1(s) |A_1(s)|^2 \quad (\text{B.18})$$

$B_3(s_b, s_b, -s_a)$ is determined in terms of A_1 , A_2 and A_3 , which have to be resolved first

Applying the two-tone input to (B.15)

$$A_3(s_b, s_b, -s_a) = -A_1(s)b(s) \left[2g_2 \overline{A_1(s_b)A_2(s_b, -s_a)} + g_3 A_1(s)|A_1(s)|^2 \right] \quad (\text{B.19})$$

$$\text{Because } \overline{A_1(s_1)A_2(s_2, s_3)} = \frac{1}{3} [A_1(s_1)A_2(s_2, s_3) + A_1(s_2)A_2(s_1, s_3) + A_1(s_3)A_2(s_1, s_2)]$$

$$\text{Then } \overline{A_1(s_b)A_2(s_b, -s_a)} = \frac{1}{3} [A_1(s_b)A_2(s_b, -s_a) + A_1(s_b)A_2(s_b - s_a) + A_1(-s_a)A_2(s_b, s_b)]$$

$$A_1(s_b)A_2(s_b, -s_a) = \frac{1}{3} [2A_1(s_b)A_2(s_b, -s_a) + A_1(-s_a)A_2(s_b, s_b)] \quad (\text{B.20})$$

Substituting (B.14) into (B.20)

$$\begin{aligned} \overline{A_1(s_b)A_2(s_b, -s_a)} &= \\ -\frac{1}{3} [2A_1(s_b)A_1(s_b - s_a)g_2 A_1(s_b)A_2(-s_a)b(s_b - s_a) &+ A_1(-s_a)A_1(s_b + s_b)g_2 A_1(s_b)A_2(s_b)b(s_b + s_b)] \end{aligned}$$

$$\begin{aligned} \overline{A_1(s_b)A_2(s_b, -s_a)} &= \\ -\frac{1}{3} [2A_1(s_b)A_1(\Delta s)g_2 A_1(s_b)A_1(-s_a)b(\Delta s) &+ A_1(-s_a)A_1(2s_b)g_2 A_1(s_b)A_1(s_b)b(2s_b)] \end{aligned}$$

$$\overline{A_1(s_b)A_2(s_b, -s_a)} = -\frac{1}{3} g_2 A_1(s_b)^2 A_1(-s_a) [2A_1(\Delta s)b(\Delta s) + A_1(2s_b)b(2s_b)]$$

$$\overline{A_1(s_b)A_2(s_b, -s_a)} \approx -\frac{1}{3} g_2 A_1(s_b)^2 A_1(-s_a) A_1(2s_b)b(2s_b)$$

That is since $\Delta s = s_b - s_a \approx 0$ due to the closely spaced two-tones, which also implies $s_a \approx s_b$.

It follows

$$\overline{A_1(s_b)A_2(s_b, -s_a)} = -\frac{g_2 A_1(s)|A_1(s)|^2}{3} A_1(2s)b(2s) \quad (\text{B.21})$$

substituting (B.21) into (B.19)

$$A_3(s_b, s_b, -s_a) = -A_1(s)b(s) \left\{ 2g_2 \left[\frac{-g_2 A_1(s)|A_1(s)|^2}{3} A_1(2s)b(2s) \right] + g_3 A_1(s)|A_1(s)|^2 \right\}$$

$$A_3(s_b, s_b, -s_a) = -A_1(s)b(s) \left[\frac{-2g_2^2 A_1(s)|A_1(s)|^2}{3} A_1(2s)b(2s) + g_3 A_1(s)|A_1(s)|^2 \right]$$

$$A_3(s_b, s_b, -s_a) = -A_1(s)^2 |A_1(s)|^2 b(s) \left[\frac{-2g_2^2 A_1(2s)b(2s)}{3} + g_3 \right]$$

$$A_3(s_b, s_b, -s_a) = -A_1(s)^2 |A_1(s)|^2 b(s) \left[g_3 - \frac{2g_2^2 A_1(2s)b(2s)}{3} \right] \quad (\text{B.22})$$

Substituting (B.21) and (B.22) into (B.18)

$$B_3(s_b, s_b, -s_a) =$$

$$-g_1 A_1(s)^2 |A_1(s)|^2 b(s) \left[g_3 - \frac{2g_2^2 A_1(2s)b(2s)}{3} \right] + 2g_2 \left[-\frac{g_2 A_1(s)|A_1(s)|^2}{3} A_1(2s)b(2s) \right] + g_3 A_1(s) |A_1(s)|^2$$

$$B_3(s_b, s_b, -s_a) = A_1(s) |A_1(s)|^2 \left\{ -g_1 A_1(s) b(s) \left[g_3 - \frac{2g_2^2 A_1(2s)b(2s)}{3} \right] - \frac{2g_2^2 A_1(2s)b(2s)}{3} + g_3 \right\}$$

$$B_3(s_b, s_b, -s_a) = A_1(s) |A_1(s)|^2 \left\{ \left[g_3 - \frac{2g_2^2 A_1(2s)b(2s)}{3} \right] [1 - g_1 A_1(s) b(s)] \right\}$$

Substituting for A_1 (from (B.13)) in $\left[g_3 - \frac{2g_2^2 A_1(2s)b(2s)}{3} \right]$ of $B_3(s_b, s_b, -s_a)$

$$B_3(s_b, s_b, -s_a) = A_1(s) |A_1(s)|^2 \left\{ \left[g_3 - \frac{2g_2^2 \frac{b(2s)}{a(2s) + g_1 b(2s)}}{3} \right] [1 - g_1 A_1(s) b(s)] \right\}$$

$$B_3(s_b, s_b, -s_a) = A_1(s) |A_1(s)|^2 \left\{ \left[g_3 - \frac{2g_2^2 \frac{2sL}{2sC_{gs}Z_1(2s) + 1 + (2s)^2 C_{gs}L + g_1 2sL}}{3} \right] [1 - g_1 A_1(s) b(s)] \right\}$$

$$B_3(s_b, s_b, -s_a) = A_1(s) |A_1(s)|^2 \left\{ \left[g_3 - \frac{\left(\frac{2g_2^2}{3} \right) 2sL}{2sC_{gs}Z_1(2s) + 1 + 4s^2 C_{gs}L + g_1 2sL} \right] [1 - g_1 A_1(s) b(s)] \right\}$$

$$B_3(s_b, s_b, -s_a) = A_1(s) |A_1(s)|^2 \left\{ \left[g_3 - \frac{\left(\frac{2g_2^2}{3} \right)}{\frac{C_{gs}}{L} Z_1(2s) + \frac{1}{2sL} + 2sC_{gs}L + g_1} \right] [1 - g_1 A_1(s) b(s)] \right\}$$

$$\text{Letting } \varepsilon = g_3 - \frac{\left(\frac{2g_2^2}{3} \right)}{\frac{C_{gs}}{L} Z_1(2s) + \frac{1}{2sL} + 2sC_{gs}L + g_1} \quad (\text{B.23})$$

Then

$$B(s_b, s_b, -s_a) = A_1(s) |A_1(s)|^2 [1 - g_1 A_1(s) b(s)] \varepsilon \quad (\text{B.24})$$

Substituting (B.6) and (B.24) in (B.17)

$$IIP_3(2\omega_b - \omega_a) = \frac{1}{6\Re(Z_1(s_a))} \left| \frac{g_1 A(s)}{A(s) |A(s)|^2 [1 - g_1 A_1(s) b(s)] \varepsilon} \right|$$

$$IIP_3(2\omega_b - \omega_a) = \frac{g_1}{6\Re(Z_1(s_a)) |A(s)|^2 [1 - g_1 A_1(s) b(s)] \varepsilon} \quad (\text{B.25})$$

Finding $\Re(Z_1(s_a))$

$Z_1(s) = Z_{in}(-s)$ when the input is conjugately matched

In order to determine $Z_{in}(s)$, the input matching impedance $Z_1(s)$ is ignored and hence

$$v_x = v_{gs} + (i_d + i_{gs})sL \quad (\text{B.26})$$

$$\text{But also } v_{gs} = \frac{i_{gs}}{sC_{gs}}, i_d = g_1 v_{gs} \quad (\text{B.27})$$

Substituting (B.27) into (B.26)

$$\begin{aligned} v_x &= \frac{i_{gs}}{sC_{gs}} + g_1 v_{gs}sL + i_{gs}sL = \frac{i_{gs}}{sC_{gs}} + \frac{g_1 i_{gs}sL}{sC_{gs}} + i_{gs}sL = i_{gs} \left(\frac{1}{sC_{gs}} + \frac{g_1 sL}{sC_{gs}} + sL \right) \\ v_x &= i_{gs} \left(\frac{1 + g_1 sL + s^2 C_{gs} L}{sC_{gs}} \right) \end{aligned} \quad (\text{B.28})$$

From the small signal circuit

$$Z_{in}(s) = \frac{v_x}{i_{gs}} \quad (\text{B.29})$$

Substituting (B.28) into (B.29)

$$Z_{in}(s) = \frac{1 + g_1 sL + s^2 C_{gs} L}{sC_{gs}}$$

It follows

$$Z_1(s) = Z_{in}(-s) = \frac{1 + g_1 sL + s^2 C_{gs} L}{-sC_{gs}} \quad (\text{B.30})$$

Taking the real part of (B.30)

$$\Re(Z_1(s)) = \frac{-g_1 sL}{-sC_{gs}} = \frac{g_1 L}{C_{gs}} \quad (\text{B.31})$$

Using (B.11), (B.13) and (B.30) it can be algebraically shown that

$$g_1 A_1(s) b(s) = \frac{1}{2} \quad (\text{B.32})$$

Finding $|A_1(s)|^2$:

From (B.32),

$$A_1(s) = \frac{1}{2g_1 b(s)}$$

And since $b(s) = sL$ (from (B.11)), then

$$|A_1(s)|^2 = \left| \frac{1}{4g_1^2 s^2 L^2} \right| = \left| -\frac{1}{4g_1^2 \omega^2 L^2} \right| = \frac{1}{4g_1^2 \omega^2 L^2} \quad (\text{B.33})$$

Using (B.31) and (B.33)

$$\Re(Z_1(s))|A_1(s)|^2 = \frac{1}{4\omega^2 LC_{gs}g_1} \quad (\text{B.34})$$

Substituting (B.32) and (B.34) in (B.25)

$$IIP_3(2\omega_b - \omega_a) = \frac{g_1}{\frac{6}{4\omega^2 LC_{gs}g_1} \left[1 - \frac{1}{2}\right] \varepsilon}$$

Rearranging

$$IIP_3(2\omega_b - \omega_a) = \frac{4g_1^2\omega^2 LC_{gs}}{3\varepsilon}$$

Appendix C FULL RESULTS OF HARMONIC AND IMD FEEDBACK ANALYSIS

This Appendix lists the full results of the analysis in Section 5.4. The tables below show which frequency components were taken from each single-order of nonlinearity (with their magnitude coefficients) and fed back to the input, and the resulting 3rd-, 5th- or 7th-order intermodulation components (where available). These results were obtained using the math software Mathematica as explained in Section 5.4.1

C.1 Fed-back Components From 2nd-order Nonlinear Term

2ω from 2 nd $+\frac{1}{2}bA^2\cos[2\omega_1]$ $+\frac{1}{2}bA^2\cos[2\omega_2]$	IMD3	IMD5	IMD7
2 nd	$-4A^3b^2\sin[2\omega_1 - \omega_2]$		
3 rd	$+\frac{3}{4}A^3c\sin[2\omega_1 - \omega_2]$ $+\frac{3}{8}A^5b^2c\sin[2\omega_1 - \omega_2]$	$+\frac{3}{8}A^5b^2c\sin[3\omega_1 - 2\omega_2]$	
4 th	$-3A^5bd\sin[2\omega_1 - \omega_2]$ $-\frac{9}{16}A^7b^3d\sin[2\omega_1 - \omega_2]$	$-A^5bd\sin[3\omega_1 - 2\omega_2]$ $-\frac{3}{16}A^7b^3d\sin[3\omega_1 - 2\omega_2]$	$-\frac{3}{16}A^7b^3d\sin[4\omega_1 - 3\omega_2]$
5 th	$+\frac{8}{25}A^5e\sin[2\omega_1 - \omega_2]$ $+\frac{175}{32}A^7b^2e\sin[2\omega_1 - \omega_2]$ $+\frac{15}{32}A^9b^4e\sin[2\omega_1 - \omega_2]$	$+\frac{5}{8}A^5e\sin[3\omega_1 - 2\omega_2]$ $+\frac{105}{32}A^7b^2e\sin[3\omega_1 - 2\omega_2]$ $+\frac{15}{32}A^9b^4e\sin[3\omega_1 - 2\omega_2]$	$+\frac{35}{32}A^7b^2e\sin[4\omega_1 - 3\omega_2]$ $+\frac{15}{128}A^9b^4e\sin[4\omega_1 - 3\omega_2]$
6 th	$-\frac{495}{32}A^7bf\sin[2\omega_1 - \omega_2]$ $-\frac{75}{8}A^9b^3f\sin[2\omega_1 - \omega_2]$ $-\frac{8}{75}A^{11}b^5f\sin[2\omega_1 - \omega_2]$	$-\frac{225}{32}A^7bf\sin[3\omega_1 - 2\omega_2]$ $-\frac{45}{8}A^9b^3f\sin[3\omega_1 - 2\omega_2]$ $-\frac{75}{256}A^{11}b^5f\sin[3\omega_1 - 2\omega_2]$	$-\frac{45}{32}A^7bf\sin[4\omega_1 - 3\omega_2]$ $-\frac{45}{16}A^9b^3f\sin[4\omega_1 - 3\omega_2]$ $-\frac{75}{256}A^{11}b^5f\sin[4\omega_1 - 3\omega_2]$
7 th	$+\frac{64}{1365}A^7\sin[2\omega_1 - \omega_2]$ $+\frac{32}{3465}A^9b^2\sin[2\omega_1 - \omega_2]$ $+\frac{32}{256}A^{11}b^4\sin[2\omega_1 - \omega_2]$ $+\frac{525}{1024}A^{13}b^6\sin[2\omega_1 - \omega_2]$	$+\frac{245}{64}A^7\sin[3\omega_1 - 2\omega_2]$ $+\frac{819}{32}A^9b^2\sin[3\omega_1 - 2\omega_2]$ $+\frac{5005}{512}A^{11}b^4\sin[3\omega_1 - 2\omega_2]$ $+\frac{525}{1024}A^{13}b^6\sin[3\omega_1 - 2\omega_2]$	$+\frac{35}{64}A^7\sin[4\omega_1 - 3\omega_2]$ $+\frac{1365}{128}A^9b^2\sin[4\omega_1 - 3\omega_2]$ $+\frac{2695}{512}A^{11}b^4\sin[4\omega_1 - 3\omega_2]$ $+\frac{105}{512}A^{13}b^6\sin[4\omega_1 - 3\omega_2]$

$\omega_1 + \omega_2$ from 2 nd $+A^2 b \cos[\omega_1 + \omega_2]$	IMD3	IMD5	IMD7
2 nd			
3 rd	$+\frac{3}{4}A^3 c \sin[2\omega_1 - \omega_2]$		
4 th	$-\frac{2}{8}A^5 b d \sin[2\omega_1 - \omega_2]$		
5 th	$+\frac{25}{8}A^5 e \sin[2\omega_1 - \omega_2]$ $+\frac{35}{8}A^7 b^2 e \sin[2\omega_1 - \omega_2]$	$+\frac{5}{8}A^5 e \sin[3\omega_1 - 2\omega_2]$	
6 th	$-\frac{225}{16}A^7 b f \sin[2\omega_1 - \omega_2]$ $-\frac{15}{2}A^9 b^3 f \sin[2\omega_1 - \omega_2]$	$-\frac{45}{16}A^7 b f \sin[3\omega_1 - 2\omega_2]$	
7 th	$+\frac{735}{64}A^7 g \sin[2\omega_1 - \omega_2]$ $+\frac{1365}{32}A^9 b^2 g \sin[2\omega_1 - \omega_2]$ $+\frac{385}{32}A^{11} b^4 g \sin[2\omega_1 - \omega_2]$	$+\frac{245}{64}A^7 g \sin[3\omega_1 - 2\omega_2]$ $+\frac{273}{32}A^9 b^2 g \sin[3\omega_1 - 2\omega_2]$	$+\frac{35}{64}A^7 g \sin[4\omega_1 - 3\omega_2]$

$\omega_1 - \omega_2$ from 2 nd $-A^2 b \cos[\omega_1 - \omega_2]$	IMD3	IMD5	IMD7
2 nd	$-2A^3 b^2 \sin[2\omega_1 - \omega_2]$		
3 rd	$+\frac{3}{4}A^3 c \sin[2\omega_1 - \omega_2]$ $+\frac{3}{4}A^5 b^2 c \sin[2\omega_1 - \omega_2]$	$+\frac{3}{4}A^5 b^2 c \sin[3\omega_1 - 2\omega_2]$	
4 th	$-\frac{9}{2}A^5 b d \sin[2\omega_1 - \omega_2]$ $-\frac{3}{2}A^7 b^3 d \sin[2\omega_1 - \omega_2]$	$-\frac{3}{2}A^5 b d \sin[3\omega_1 - 2\omega_2]$ $-\frac{1}{2}A^7 b^3 d \sin[3\omega_1 - 2\omega_2]$	$-\frac{1}{2}A^7 b^3 d \sin[4\omega_1 - 3\omega_2]$
5 th	$+\frac{25}{8}A^5 e \sin[2\omega_1 - \omega_2]$ $+\frac{75}{8}A^7 b^2 e \sin[2\omega_1 - \omega_2]$ $+\frac{5}{4}A^9 b^4 e \sin[2\omega_1 - \omega_2]$	$+\frac{5}{8}A^5 e \sin[3\omega_1 - 2\omega_2]$ $+\frac{45}{8}A^7 b^2 e \sin[3\omega_1 - 2\omega_2]$ $+\frac{5}{4}A^9 b^4 e \sin[3\omega_1 - 2\omega_2]$	$+\frac{15}{8}A^7 b^2 e \sin[4\omega_1 - 3\omega_2]$ $+\frac{5}{16}A^9 b^4 e \sin[4\omega_1 - 3\omega_2]$
6 th	$+\frac{165}{8}A^7 b f \sin[2\omega_1 - \omega_2]$ $+\frac{75}{4}A^9 b^3 f \sin[2\omega_1 - \omega_2]$ $+\frac{15}{8}A^{11} b^5 f \sin[2\omega_1 - \omega_2]$	$+\frac{75}{8}A^7 b f \sin[3\omega_1 - 2\omega_2]$ $+\frac{45}{4}A^9 b^3 f \sin[3\omega_1 - 2\omega_2]$ $+\frac{15}{16}A^{11} b^5 f \sin[3\omega_1 - 2\omega_2]$	$+\frac{15}{8}A^7 b f \sin[4\omega_1 - 3\omega_2]$ $+\frac{45}{8}A^9 b^3 f \sin[4\omega_1 - 3\omega_2]$ $+\frac{15}{16}A^{11} b^5 f \sin[4\omega_1 - 3\omega_2]$
7 th	$+\frac{735}{64}A^7 g \sin[2\omega_1 - \omega_2]$ $+\frac{525}{8}A^9 b^2 g \sin[2\omega_1 - \omega_2]$ $+\frac{945}{32}A^{11} b^4 g \sin[2\omega_1 - \omega_2]$ $+\frac{105}{64}A^{13} b^6 g \sin[2\omega_1 - \omega_2]$	$+\frac{245}{64}A^7 g \sin[3\omega_1 - 2\omega_2]$ $+\frac{315}{8}A^9 b^2 g \sin[3\omega_1 - 2\omega_2]$ $+\frac{1365}{64}A^{11} b^4 g \sin[3\omega_1 - 2\omega_2]$ $+\frac{105}{64}A^{13} b^6 g \sin[3\omega_1 - 2\omega_2]$	$+\frac{35}{64}A^7 g \sin[4\omega_1 - 3\omega_2]$ $+\frac{525}{64}A^9 b^2 g \sin[4\omega_1 - 3\omega_2]$ $+\frac{735}{64}A^{11} b^4 g \sin[4\omega_1 - 3\omega_2]$ $+\frac{21}{32}A^{13} b^6 g \sin[4\omega_1 - 3\omega_2]$

C.2 Fed-back Components From 3rd-order Nonlinear Term

$2\omega_1 + \omega_2$ from 3 rd $+\frac{3}{4}A^3 c \sin[2\omega_1 + \omega_2]$	IMD3	IMD5	IMD7
2 nd			
3 rd	$+\frac{3}{4}A^3 c \sin[2\omega_1 - \omega_2]$ $-\frac{9}{16}A^5 c^2 \sin[2\omega_1 - \omega_2]$		
4 th			
5 th	$+\frac{25}{8}A^5 e \sin[2\omega_1 - \omega_2]$ $-\frac{255}{64}A^7 c e \sin[2\omega_1 - \omega_2]$ $+\frac{135}{64}A^9 c^2 e \sin[2\omega_1 - \omega_2]$ $-\frac{405}{512}A^{11} c^3 e \sin[2\omega_1 - \omega_2]$	$+\frac{5}{8}A^5 e \sin[3\omega_1 - 2\omega_2]$ $-\frac{15}{16}A^7 c e \sin[3\omega_1 - 2\omega_2]$	
6 th			
7 th	$+\frac{735}{64}A^7 g \sin[2\omega_1 - \omega_2]$	$+\frac{245}{64}A^7 g \sin[3\omega_1 - 2\omega_2]$	$+\frac{35}{64}A^7 g \sin[4\omega_1 - 3\omega_2]$

	$ \begin{aligned} & -\frac{5481}{256} A^9 c g \sin[2 \omega_1 - \omega_2] \\ & + \frac{10395}{512} A^{11} c^2 g \sin[2 \omega_1 - \omega_2] \\ & - \frac{48195}{4096} A^{13} c^3 g \sin[2 \omega_1 - \omega_2] \\ & + \frac{25515}{8192} A^{15} c^4 g \sin[2 \omega_1 - \omega_2] \\ & - \frac{25515}{32768} A^{17} c^5 g \sin[2 \omega_1 - \omega_2] \end{aligned} $	$ \begin{aligned} & -\frac{63}{8} A^9 c g \sin[3 \omega_1 - 2 \omega_2] \\ & + \frac{4725}{1024} A^{11} c^2 g \sin[3 \omega_1 - 2 \omega_2] \\ & - \frac{2835}{1024} A^{13} c^3 g \sin[3 \omega_1 - 2 \omega_2] \end{aligned} $	$ \begin{aligned} & -\frac{315}{256} A^9 c g \sin[4 \omega_1 - 3 \omega_2] \\ & + \frac{189}{1024} A^{11} c^2 g \sin[4 \omega_1 - 3 \omega_2] \end{aligned} $
--	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------

$2 \omega_1 - \omega_2$ from 3 rd $-\frac{3}{4} A^3 c \sin[2 \omega_1 - \omega_2]$	IMD3	IMD5	IMD7
2 nd			
3 rd	$ \begin{aligned} & +\frac{3}{4} A^3 c \sin[2 \omega_1 - \omega_2] \\ & -\frac{9}{4} A^5 c^2 \sin[2 \omega_1 - \omega_2] \\ & -\frac{81}{256} A^9 c^4 \sin[2 \omega_1 - \omega_2] \end{aligned} $	$ \begin{aligned} & -\frac{9}{8} A^5 c^2 \sin[3 \omega_1 - 2 \omega_2] \\ & +\frac{27}{64} A^7 c^3 \sin[3 \omega_1 - 2 \omega_2] \end{aligned} $	$ +\frac{27}{64} A^7 c^3 \sin[4 \omega_1 - 3 \omega_2] $
4 th			
5 th	$ \begin{aligned} & +\frac{25}{8} A^5 e \sin[2 \omega_1 - \omega_2] \\ & -\frac{135}{16} A^7 c e \sin[2 \omega_1 - \omega_2] \\ & +\frac{405}{128} A^9 c^2 e \sin[2 \omega_1 - \omega_2] \\ & -\frac{128}{405} A^{11} c^3 e \sin[2 \omega_1 - \omega_2] \\ & -\frac{1215}{1215} A^{15} c^5 e \sin[2 \omega_1 - \omega_2] \end{aligned} $	$ \begin{aligned} & +\frac{5}{8} A^5 e \sin[3 \omega_1 - 2 \omega_2] \\ & -\frac{45}{8} A^7 c e \sin[3 \omega_1 - 2 \omega_2] \\ & +\frac{405}{128} A^9 c^2 e \sin[3 \omega_1 - 2 \omega_2] \\ & -\frac{405}{256} A^{11} c^3 e \sin[3 \omega_1 - 2 \omega_2] \\ & +\frac{405}{1024} A^{13} c^4 e \sin[3 \omega_1 - 2 \omega_2] \end{aligned} $	$ \begin{aligned} & -\frac{45}{32} A^7 c e \sin[4 \omega_1 - 3 \omega_2] \\ & +\frac{405}{128} A^9 c^2 e \sin[4 \omega_1 - 3 \omega_2] \\ & -\frac{135}{512} A^{11} c^3 e \sin[4 \omega_1 - 3 \omega_2] \\ & +\frac{405}{1024} A^{13} c^4 e \sin[4 \omega_1 - 3 \omega_2] \end{aligned} $
6 th			
7 th	$ \begin{aligned} & +\frac{735}{64} A^7 g \sin[2 \omega_1 - \omega_2] \\ & -\frac{8715}{256} A^9 c g \sin[2 \omega_1 - \omega_2] \\ & +\frac{14175}{512} A^{11} c^2 g \sin[2 \omega_1 - \omega_2] \\ & -\frac{25515}{1024} A^{13} c^3 g \sin[2 \omega_1 - \omega_2] \\ & +\frac{42525}{8192} A^{15} c^4 g \sin[2 \omega_1 - \omega_2] \\ & -\frac{25515}{8192} A^{17} c^5 g \sin[2 \omega_1 - \omega_2] \\ & -\frac{76545}{1048576} A^{21} c^7 g \sin[2 \omega_1 - \omega_2] \end{aligned} $	$ \begin{aligned} & +\frac{245}{64} A^7 g \sin[3 \omega_1 - 2 \omega_2] \\ & -\frac{1575}{64} A^9 c g \sin[3 \omega_1 - 2 \omega_2] \\ & +\frac{2835}{128} A^{11} c^2 g \sin[3 \omega_1 - 2 \omega_2] \\ & -\frac{17955}{1024} A^{13} c^3 g \sin[3 \omega_1 - 2 \omega_2] \\ & +\frac{25515}{4096} A^{15} c^4 g \sin[3 \omega_1 - 2 \omega_2] \\ & -\frac{25515}{16384} A^{17} c^5 g \sin[3 \omega_1 - 2 \omega_2] \\ & +\frac{76545}{262144} A^{19} c^6 g \sin[3 \omega_1 - 2 \omega_2] \end{aligned} $	$ \begin{aligned} & +\frac{35}{64} A^7 g \sin[4 \omega_1 - 3 \omega_2] \\ & -\frac{315}{32} A^9 c g \sin[4 \omega_1 - 3 \omega_2] \\ & +\frac{4725}{256} A^{11} c^2 g \sin[4 \omega_1 - 3 \omega_2] \\ & -\frac{16065}{2048} A^{13} c^3 g \sin[4 \omega_1 - 3 \omega_2] \\ & +\frac{25515}{4096} A^{15} c^4 g \sin[4 \omega_1 - 3 \omega_2] \\ & -\frac{25515}{65536} A^{17} c^5 g \sin[4 \omega_1 - 3 \omega_2] \\ & +\frac{76545}{262144} A^{19} c^6 g \sin[4 \omega_1 - 3 \omega_2] \end{aligned} $

$3 \omega_1$ from 3 rd $+\frac{1}{4} A^3 c \sin[3 \omega_1]$ $+\frac{1}{4} A^3 c \sin[3 \omega_2]$	IMD3	IMD5	IMD7
2 nd			
3 rd	$ \begin{aligned} & +\frac{3}{4} A^3 c \sin[2 \omega_1 - \omega_2] \\ & -\frac{3}{8} A^5 c^2 \sin[2 \omega_1 - \omega_2] \end{aligned} $	$ \begin{aligned} & -\frac{3}{16} A^5 c^2 \sin[3 \omega_1 - 2 \omega_2] \\ & +\frac{3}{32} A^7 c^3 \sin[3 \omega_1 - 2 \omega_2] \end{aligned} $	$ +\frac{3}{32} A^7 c^3 \sin[4 \omega_1 - 3 \omega_2] $
4 th			
5 th	$ \begin{aligned} & +\frac{25}{8} A^5 e \sin[2 \omega_1 - \omega_2] \\ & -\frac{75}{32} A^7 c e \sin[2 \omega_1 - \omega_2] \\ & +\frac{45}{64} A^9 c^2 e \sin[2 \omega_1 - \omega_2] \\ & -\frac{45}{256} A^{11} c^3 e \sin[2 \omega_1 - \omega_2] \end{aligned} $	$ \begin{aligned} & +\frac{5}{8} A^5 e \sin[3 \omega_1 - 2 \omega_2] \\ & -\frac{25}{16} A^7 c e \sin[3 \omega_1 - 2 \omega_2] \\ & +\frac{45}{64} A^9 c^2 e \sin[3 \omega_1 - 2 \omega_2] \\ & -\frac{45}{512} A^{11} c^3 e \sin[3 \omega_1 - 2 \omega_2] \\ & +\frac{15}{512} A^{13} c^4 e \sin[3 \omega_1 - 2 \omega_2] \end{aligned} $	$ \begin{aligned} & -\frac{25}{64} A^7 c e \sin[4 \omega_1 - 3 \omega_2] \\ & +\frac{45}{64} A^9 c^2 e \sin[4 \omega_1 - 3 \omega_2] \\ & -\frac{15}{512} A^{11} c^3 e \sin[4 \omega_1 - 3 \omega_2] \\ & +\frac{15}{512} A^{13} c^4 e \sin[4 \omega_1 - 3 \omega_2] \end{aligned} $
6 th			
7 th	$ \begin{aligned} & +\frac{735}{64} A^7 g \sin[2 \omega_1 - \omega_2] \\ & -\frac{3087}{256} A^9 c g \sin[2 \omega_1 - \omega_2] \end{aligned} $	$ \begin{aligned} & +\frac{245}{64} A^7 g \sin[3 \omega_1 - 2 \omega_2] \\ & -\frac{2205}{256} A^9 c g \sin[3 \omega_1 - 2 \omega_2] \end{aligned} $	$ \begin{aligned} & +\frac{35}{64} A^7 g \sin[4 \omega_1 - 3 \omega_2] \\ & -\frac{441}{128} A^9 c g \sin[4 \omega_1 - 3 \omega_2] \end{aligned} $

Appendix C: FULL RESULTS OF HARMONIC AND IMD FEEDBACK ANALYSIS

	$+ \frac{6615 A^{11} c^2 g \sin[2 \omega_1 - \omega_2]}{1024}$ $+ \frac{1575 A^{15} c^4 g \sin[2 \omega_1 - \omega_2]}{4096}$ $- \frac{525 A^{17} c^5 g \sin[2 \omega_1 - \omega_2]}{8192}$ $- \frac{4725 A^{13} c^3 g \sin[2 \omega_1 - \omega_2]}{2048}$	$+ \frac{1323}{256} A^{11} c^2 g \sin[3 \omega_1 - 2 \omega_2]$ $- \frac{6825 A^{13} c^3 g \sin[3 \omega_1 - 2 \omega_2]}{4096}$ $+ \frac{945 A^{15} c^4 g \sin[3 \omega_1 - 2 \omega_2]}{2048}$ $- \frac{525 A^{17} c^5 g \sin[3 \omega_1 - 2 \omega_2]}{16384}$ $+ \frac{525 A^{19} c^6 g \sin[3 \omega_1 - 2 \omega_2]}{65536}$	$+ \frac{2205}{512} A^{11} c^2 g \sin[4 \omega_1 - 3 \omega_2]$ $- \frac{3675 A^{13} c^3 g \sin[4 \omega_1 - 3 \omega_2]}{4096}$ $+ \frac{945 A^{15} c^4 g \sin[4 \omega_1 - 3 \omega_2]}{2048}$ $- \frac{525 A^{17} c^5 g \sin[4 \omega_1 - 3 \omega_2]}{32768}$ $+ \frac{525 A^{19} c^6 g \sin[4 \omega_1 - 3 \omega_2]}{65536}$
--	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

$\omega_1 - 2 \omega_2$ from 3 rd + $\frac{3}{4} A^3 c \sin[\omega_1 - 2 \omega_2]$	IMD3	IMD5	IMD7
2 nd			
3 rd	$+ \frac{3}{4} A^3 c \sin[2 \omega_1 - \omega_2]$ $- \frac{9}{8} A^5 c^2 \sin[2 \omega_1 - \omega_2]$	$- \frac{9}{16} A^5 c^2 \sin[3 \omega_1 - 2 \omega_2]$	
4 th			
5 th	$+ \frac{25}{8} A^5 e \sin[2 \omega_1 - \omega_2]$ $- \frac{45}{8} A^7 c e \sin[2 \omega_1 - \omega_2]$ $+ \frac{315}{128} A^9 c^2 e \sin[2 \omega_1 - \omega_2]$ $- \frac{405}{256} A^{11} c^3 e \sin[2 \omega_1 - \omega_2]$	$+ \frac{5}{8} A^5 e \sin[3 \omega_1 - 2 \omega_2]$ $- \frac{15}{4} A^7 c e \sin[3 \omega_1 - 2 \omega_2]$ $+ \frac{135}{128} A^9 c^2 e \sin[3 \omega_1 - 2 \omega_2]$ $- \frac{405}{512} A^{11} c^3 e \sin[3 \omega_1 - 2 \omega_2]$	$- \frac{15}{16} A^7 c e \sin[4 \omega_1 - 3 \omega_2]$ $+ \frac{135}{128} A^9 c^2 e \sin[4 \omega_1 - 3 \omega_2]$
6 th			
7 th	$+ \frac{735}{64} A^7 g \sin[2 \omega_1 - \omega_2]$ $- \frac{105}{4} A^9 c g \sin[2 \omega_1 - \omega_2]$ $+ \frac{23625 A^{11} c^2 g \sin[2 \omega_1 - \omega_2]}{1024}$ $- \frac{8505}{512} A^{13} c^3 g \sin[2 \omega_1 - \omega_2] +$ $+ \frac{31185 A^{15} c^4 g \sin[2 \omega_1 - \omega_2]}{8192}$ $- \frac{25515 A^{17} c^5 g \sin[2 \omega_1 - \omega_2]}{16384}$	$+ \frac{245}{64} A^7 g \sin[3 \omega_1 - 2 \omega_2]$ $- \frac{4725}{256} A^9 c g \sin[3 \omega_1 - 2 \omega_2]$ $+ \frac{6615}{512} A^{11} c^2 g \sin[3 \omega_1 - 2 \omega_2]$ $- \frac{46305 A^{13} c^3 g \sin[3 \omega_1 - 2 \omega_2]}{4096}$ $+ \frac{8505 A^{15} c^4 g \sin[3 \omega_1 - 2 \omega_2]}{4096}$ $- \frac{25515 A^{17} c^5 g \sin[3 \omega_1 - 2 \omega_2]}{32768}$	$+ \frac{35}{64} A^7 g \sin[4 \omega_1 - 3 \omega_2]$ $- \frac{945}{128} A^9 c g \sin[4 \omega_1 - 3 \omega_2]$ $+ \frac{4725}{512} A^{11} c^2 g \sin[4 \omega_1 - 3 \omega_2]$ $- \frac{945}{256} A^{13} c^3 g \sin[4 \omega_1 - 3 \omega_2]$ $+ \frac{8505 A^{15} c^4 g \sin[4 \omega_1 - 3 \omega_2]}{4096}$

$\omega_1 + 2 \omega_2$ from 3 rd + $\frac{3}{4} A^3 c \sin[\omega_1 + 2 \omega_2]$	IMD3	IMD5	IMD7
2 nd			
3 rd	$+ \frac{3}{4} A^3 c \sin[2 \omega_1 - \omega_2]$		
4 th			
5 th	$+ \frac{25}{8} A^5 e \sin[2 \omega_1 - \omega_2]$ $- \frac{15}{8} A^7 c e \sin[2 \omega_1 - \omega_2]$ $+ \frac{135}{64} A^9 c^2 e \sin[2 \omega_1 - \omega_2]$	$+ \frac{5}{8} A^5 e \sin[3 \omega_1 - 2 \omega_2]$ $- \frac{15}{64} A^7 c e \sin[3 \omega_1 - 2 \omega_2]$	
6 th			
7 th	$+ \frac{735}{64} A^7 g \sin[2 \omega_1 - \omega_2]$ $- \frac{945}{64} A^9 c g \sin[2 \omega_1 - \omega_2]$ $+ \frac{19089 A^{11} c^2 g \sin[2 \omega_1 - \omega_2]}{1024}$ $- \frac{2835}{512} A^{13} c^3 g \sin[2 \omega_1 - \omega_2]$ $+ \frac{25515 A^{15} c^4 g \sin[2 \omega_1 - \omega_2]}{8192}$	$+ \frac{245}{64} A^7 g \sin[3 \omega_1 - 2 \omega_2]$ $- \frac{1071}{256} A^9 c g \sin[3 \omega_1 - 2 \omega_2]$ $+ \frac{945}{256} A^{11} c^2 g \sin[3 \omega_1 - 2 \omega_2]$ $- \frac{2835 A^{13} c^3 g \sin[3 \omega_1 - 2 \omega_2]}{4096}$	$+ \frac{35}{64} A^7 g \sin[4 \omega_1 - 3 \omega_2]$ $- \frac{63}{128} A^9 c g \sin[4 \omega_1 - 3 \omega_2]$

C.3 Fed-Back Components From 4th-order Nonlinear Term

$2\omega_1 - 2\omega_2$ from 4 th $-\frac{3}{4}A^4d\cos[2\omega_1 - 2\omega_2]$	IMD3	IMD5	IMD7
2 nd	$-24A^5bd\sin[2\omega_1 - \omega_2]$	$-24A^5bd\sin[3\omega_1 - 2\omega_2]$	
3 rd	$+\frac{3}{4}A^3c\sin[2\omega_1 - \omega_2]$		$+\frac{27}{64}A^9cd^2\sin[4\omega_1 - 3\omega_2]$
4 th	$-\frac{27}{8}A^7d^2\sin[2\omega_1 - \omega_2]$ $-\frac{81}{128}A^{13}d^4\sin[2\omega_1 - \omega_2]$	$-\frac{27}{8}A^7d^2\sin[3\omega_1 - 2\omega_2]$ $-\frac{81}{128}A^{13}d^4\sin[3\omega_1 - 2\omega_2]$	$-\frac{9}{8}A^7d^2\sin[4\omega_1 - 3\omega_2]$
5 th	$+\frac{25}{8}A^5e\sin[2\omega_1 - \omega_2]$ $+\frac{135}{64}A^{11}d^2e\sin[2\omega_1 - \omega_2]$	$+\frac{5}{8}A^5e\sin[3\omega_1 - 2\omega_2]$ $+\frac{135}{128}A^{11}d^2e\sin[3\omega_1 - 2\omega_2]$	$+\frac{405}{128}A^{11}d^2e\sin[4\omega_1 - 3\omega_2]$ $+\frac{405}{1024}A^{17}d^4e\sin[4\omega_1 - 3\omega_2]$
6 th	$-\frac{225}{16}A^9df\sin[2\omega_1 - \omega_2]$ $-\frac{3645}{512}A^{15}d^3f\sin[2\omega_1 - \omega_2]$ $-\frac{3645}{8192}A^{21}d^5f\sin[2\omega_1 - \omega_2]$	$-\frac{225}{16}A^9df\sin[3\omega_1 - 2\omega_2]$ $-\frac{3645}{512}A^{15}d^3f\sin[3\omega_1 - 2\omega_2]$ $-\frac{3645}{8192}A^{21}d^5f\sin[3\omega_1 - 2\omega_2]$	$-\frac{225}{32}A^9df\sin[4\omega_1 - 3\omega_2]$ $-\frac{1215}{512}A^{15}d^3f\sin[4\omega_1 - 3\omega_2]$
7 th	$+\frac{735}{64}A^7g\sin[2\omega_1 - \omega_2]$ $+\frac{10395}{512}A^{13}d^2g\sin[2\omega_1 - \omega_2]$ $+\frac{25515}{8192}A^{19}d^4g\sin[2\omega_1 - \omega_2]$	$+\frac{245}{64}A^7g\sin[3\omega_1 - 2\omega_2]$ $+\frac{6615}{512}A^{13}d^2g\sin[3\omega_1 - 2\omega_2]$ $+\frac{8505}{4096}A^{19}d^4g\sin[3\omega_1 - 2\omega_2]$	$+\frac{35}{64}A^7g\sin[4\omega_1 - 3\omega_2]$ $+\frac{4725}{256}A^{13}d^2g\sin[4\omega_1 - 3\omega_2]$ $+\frac{25515}{4096}A^{19}d^4g\sin[4\omega_1 - 3\omega_2]$ $+\frac{76545}{262144}A^{25}d^6g\sin[4\omega_1 - 3\omega_2]$

$2\omega_1 + 2\omega_2$ from 4 th $-\frac{3}{4}A^4d\cos[2\omega_1 + 2\omega_2]$	IMD3	IMD5	IMD7
2 nd			
3 rd	$+\frac{3}{4}A^3c\sin[2\omega_1 - \omega_2]$		
4 th	$-\frac{3}{8}A^7d^2\sin[2\omega_1 - \omega_2]$		
5 th	$+\frac{25}{8}A^5e\sin[2\omega_1 - \omega_2]$ $+\frac{135}{64}A^{11}d^2e\sin[2\omega_1 - \omega_2]$	$+\frac{5}{8}A^5e\sin[3\omega_1 - 2\omega_2]$	
6 th	$-\frac{135}{32}A^9df\sin[2\omega_1 - \omega_2]$ $-\frac{405}{512}A^{15}d^3f\sin[2\omega_1 - \omega_2]$	$-\frac{27}{32}A^9df\sin[3\omega_1 - 2\omega_2]$	
7 th	$+\frac{735}{64}A^7g\sin[2\omega_1 - \omega_2]$ $+\frac{4725}{256}A^{13}d^2g\sin[2\omega_1 - \omega_2]$ $+\frac{25515}{8192}A^{19}d^4g\sin[2\omega_1 - \omega_2]$	$+\frac{245}{64}A^7g\sin[3\omega_1 - 2\omega_2]$ $+\frac{945}{256}A^{13}d^2g\sin[3\omega_1 - 2\omega_2]$	$+\frac{35}{64}A^7g\sin[4\omega_1 - 3\omega_2]$

4ω from 4 th $+2dA^4\cos[2\omega_1]$ $+2dA^4\cos[2\omega_2]$	IMD3	IMD5	IMD7
2 nd	$-4A^5bd\sin[2\omega_1 - \omega_2]$		
3 rd	$+\frac{3}{4}A^3c\sin[2\omega_1 - \omega_2]$ $+6A^9cd^2\sin[2\omega_1 - \omega_2]$	$+6A^9cd^2\sin[3\omega_1 - 2\omega_2]$	
4 th	$-12A^7d^2\sin[2\omega_1 - \omega_2]$ $-36A^{13}d^4\sin[2\omega_1 - \omega_2]$	$-4A^7d^2\sin[3\omega_1 - 2\omega_2]$ $-12A^{13}d^4\sin[3\omega_1 - 2\omega_2]$	$-12A^{13}d^4\sin[4\omega_1 - 3\omega_2]$
5 th	$+\frac{25}{8}A^5e\sin[2\omega_1 - \omega_2]$ $+\frac{175}{2}A^{11}d^2e\sin[2\omega_1 - \omega_2]$ $+120A^{17}d^4e\sin[2\omega_1 - \omega_2]$	$+\frac{5}{8}A^5e\sin[3\omega_1 - 2\omega_2]$ $+\frac{105}{2}A^{11}d^2e\sin[3\omega_1 - 2\omega_2]$ $+120A^{17}d^4e\sin[3\omega_1 - 2\omega_2]$	$+\frac{35}{2}A^{11}d^2e\sin[4\omega_1 - 3\omega_2]$ $+30A^{17}d^4e\sin[4\omega_1 - 3\omega_2]$

Appendix C: FULL RESULTS OF HARMONIC AND IMD FEEDBACK ANALYSIS

6 th	$-\frac{495}{8} A^9 d f \sin[2 \omega_1 - \omega_2]$ $-600 A^{15} d^3 f \sin[2 \omega_1 - \omega_2]$ $-600 A^{21} d^5 f \sin[2 \omega_1 - \omega_2]$	$-\frac{225}{8} A^9 d f \sin[3 \omega_1 - 2 \omega_2]$ $-360 A^{15} d^3 f \sin[3 \omega_1 - 2 \omega_2]$ $-300 A^{21} d^5 f \sin[3 \omega_1 - 2 \omega_2]$	$-\frac{45}{8} A^9 d f \sin[4 \omega_1 - 3 \omega_2]$ $-180 A^{15} d^3 f \sin[4 \omega_1 - 3 \omega_2]$ $-300 A^{21} d^5 f \sin[4 \omega_1 - 3 \omega_2]$
7 th	$+\frac{735}{64} A^7 g \sin[2 \omega_1 - \omega_2]$ $+\frac{1365}{2} A^{13} d^2 g \sin[2 \omega_1 - \omega_2]$ $+3465 A^{19} d^4 g \sin[2 \omega_1 - \omega_2]$ $+2100 A^{25} d^6 g \sin[2 \omega_1 - \omega_2]$	$+\frac{245}{64} A^7 g \sin[3 \omega_1 - 2 \omega_2]$ $+\frac{819}{2} A^{13} d^2 g \sin[3 \omega_1 - 2 \omega_2]$ $+\frac{5005}{2} A^{19} d^4 g \sin[3 \omega_1 - 2 \omega_2]$ $+2100 A^{25} d^6 g \sin[3 \omega_1 - 2 \omega_2]$	$+\frac{35}{64} A^7 g \sin[4 \omega_1 - 3 \omega_2]$ $+\frac{1365}{2} A^{13} d^2 g \sin[4 \omega_1 - 3 \omega_2]$ $+\frac{2695}{2} A^{19} d^4 g \sin[4 \omega_1 - 3 \omega_2]$ $+840 A^{25} d^6 g \sin[4 \omega_1 - 3 \omega_2]$

$3 \omega_1 + \omega_2$ from 4 th $-\frac{1}{2} A^4 d \cos[3 \omega_1 + \omega_2]$	IMD3	IMD5	IMD7
2 nd			
3 rd	$+\frac{3}{4} A^3 c \sin[2 \omega_1 - \omega_2]$		
4 th	$-\frac{3}{4} A^7 d^2 \sin[2 \omega_1 - \omega_2]$	$-\frac{1}{4} A^7 d^2 \sin[3 \omega_1 - 2 \omega_2]$	
5 th	$+\frac{25}{8} A^5 e \sin[2 \omega_1 - \omega_2]$ $+\frac{15}{16} A^{11} d^2 e \sin[2 \omega_1 - \omega_2]$	$+\frac{5}{8} A^5 e \sin[3 \omega_1 - 2 \omega_2]$	
6 th	$-\frac{153}{32} A^9 d f \sin[2 \omega_1 - \omega_2]$ $-\frac{45}{64} A^{15} d^3 f \sin[2 \omega_1 - \omega_2]$	$-\frac{75}{32} A^9 d f \sin[3 \omega_1 - 2 \omega_2]$ $-\frac{15}{64} A^{15} d^3 f \sin[3 \omega_1 - 2 \omega_2]$	$-\frac{15}{32} A^9 d f \sin[4 \omega_1 - 3 \omega_2]$
7 th	$+\frac{735}{64} A^7 g \sin[2 \omega_1 - \omega_2]$ $+\frac{525}{64} A^{13} d^2 g \sin[2 \omega_1 - \omega_2]$ $+\frac{315}{512} A^{19} d^4 g \sin[2 \omega_1 - \omega_2]$	$+\frac{245}{64} A^7 g \sin[3 \omega_1 - 2 \omega_2]$ $+\frac{105}{64} A^{13} d^2 g \sin[3 \omega_1 - 2 \omega_2]$	$+\frac{35}{64} A^7 g \sin[4 \omega_1 - 3 \omega_2]$

$3 \omega_1 - \omega_2$ from 4 th $+\frac{1}{2} A^4 d \cos[3 \omega_1 - \omega_2]$	IMD3	IMD5	IMD7
2 nd	$-\frac{4}{3} A^5 b d \sin[2 \omega_1 - \omega_2]$	$-\frac{4}{3} A^5 b d \sin[3 \omega_1 - 2 \omega_2]$	
3 rd	$+\frac{3}{4} A^3 c \sin[2 \omega_1 - \omega_2]$		
4 th	$-\frac{9}{4} A^7 d^2 \sin[2 \omega_1 - \omega_2]$ $-\frac{3}{16} A^{13} d^4 \sin[2 \omega_1 - \omega_2]$	$-\frac{9}{4} A^7 d^2 \sin[3 \omega_1 - 2 \omega_2]$ $-\frac{3}{16} A^{13} d^4 \sin[3 \omega_1 - 2 \omega_2]$	$-\frac{3}{4} A^7 d^2 \sin[4 \omega_1 - 3 \omega_2]$
5 th	$+\frac{25}{8} A^5 e \sin[2 \omega_1 - \omega_2]$ $+\frac{15}{16} A^{11} d^2 e \sin[2 \omega_1 - \omega_2]$	$+\frac{5}{8} A^5 e \sin[3 \omega_1 - 2 \omega_2]$ $+\frac{5}{32} A^{11} d^2 e \sin[3 \omega_1 - 2 \omega_2]$	$+\frac{15}{32} A^{11} d^2 e \sin[4 \omega_1 - 3 \omega_2]$
6 th	$-\frac{75}{8} A^9 d f \sin[2 \omega_1 - \omega_2]$ $-\frac{135}{64} A^{15} d^3 f \sin[2 \omega_1 - \omega_2]$ $-\frac{15}{256} A^{21} d^5 f \sin[2 \omega_1 - \omega_2]$	$-\frac{75}{8} A^9 d f \sin[3 \omega_1 - 2 \omega_2]$ $-\frac{135}{64} A^{15} d^3 f \sin[3 \omega_1 - 2 \omega_2]$ $-\frac{15}{256} A^{21} d^5 f \sin[3 \omega_1 - 2 \omega_2]$	$-\frac{75}{16} A^9 d f \sin[4 \omega_1 - 3 \omega_2]$ $-\frac{45}{64} A^{15} d^3 f \sin[4 \omega_1 - 3 \omega_2]$
7 th	$+\frac{735}{64} A^7 g \sin[2 \omega_1 - \omega_2]$ $+\frac{2205}{256} A^{13} d^2 g \sin[2 \omega_1 - \omega_2]$ $+\frac{315}{512} A^{19} d^4 g \sin[2 \omega_1 - \omega_2]$	$+\frac{245}{64} A^7 g \sin[3 \omega_1 - 2 \omega_2]$ $+\frac{945}{256} A^{13} d^2 g \sin[3 \omega_1 - 2 \omega_2]$ $+\frac{35}{256} A^{19} d^4 g \sin[3 \omega_1 - 2 \omega_2]$	$+\frac{35}{64} A^7 g \sin[4 \omega_1 - 3 \omega_2]$ $+\frac{525}{128} A^{13} d^2 g \sin[4 \omega_1 - 3 \omega_2]$ $+\frac{105}{256} A^{19} d^4 g \sin[4 \omega_1 - 3 \omega_2]$

4ω from 4 th $-\frac{1}{8} A^4 d \cos[4 \omega_1]$ $-\frac{1}{8} A^4 d \cos[4 \omega_2]$	IMD3	IMD5	IMD7
2 nd			
3 rd	$+\frac{3}{4} A^3 c \sin[2 \omega_1 - \omega_2]$		$+\frac{3}{128} A^9 c d^2 \sin[4 \omega_1 - 3 \omega_2]$
4 th	$-\frac{3}{16} A^7 d^2 \sin[2 \omega_1 - \omega_2]$	$-\frac{3}{16} A^7 d^2 \sin[3 \omega_1 - 2 \omega_2]$	$-\frac{1}{16} A^7 d^2 \sin[4 \omega_1 - 3 \omega_2]$
5 th	$+\frac{25}{8} A^5 e \sin[2 \omega_1 - \omega_2]$ $+\frac{15}{128} A^{11} d^2 e \sin[2 \omega_1 - \omega_2]$	$+\frac{5}{8} A^5 e \sin[3 \omega_1 - 2 \omega_2]$ $+\frac{15}{256} A^{11} d^2 e \sin[3 \omega_1 - 2 \omega_2]$	$+\frac{45}{256} A^{11} d^2 e \sin[4 \omega_1 - 3 \omega_2]$

			$+\frac{15 A^{17} d^4 e \sin[4 \omega_1 - 3 \omega_2]}{8192}$
6 th	$-\frac{45}{32} A^9 d f \sin[2 \omega_1 - \omega_2]$ $-\frac{135 A^{15} d^3 f \sin[2 \omega_1 - \omega_2]}{4096}$	$-\frac{45}{32} A^9 d f \sin[3 \omega_1 - 2 \omega_2]$ $-\frac{135 A^{15} d^3 f \sin[3 \omega_1 - 2 \omega_2]}{4096}$	$-\frac{45}{64} A^9 d f \sin[4 \omega_1 - 3 \omega_2]$ $-\frac{45 A^{15} d^3 f \sin[4 \omega_1 - 3 \omega_2]}{4096}$
7 th	$+\frac{735}{64} A^7 g \sin[2 \omega_1 - \omega_2]$ $+\frac{1155 A^{13} d^2 g \sin[2 \omega_1 - \omega_2]}{1024}$ $+\frac{945 A^{19} d^4 g \sin[2 \omega_1 - \omega_2]}{65536}$	$+\frac{245}{64} A^7 g \sin[3 \omega_1 - 2 \omega_2]$ $+\frac{735 A^{13} d^2 g \sin[3 \omega_1 - 2 \omega_2]}{1024}$ $+\frac{315 A^{19} d^4 g \sin[3 \omega_1 - 2 \omega_2]}{32768}$	$+\frac{35}{64} A^7 g \sin[4 \omega_1 - 3 \omega_2]$ $+\frac{525}{64} A^{13} d^2 g \sin[4 \omega_1 - 3 \omega_2]$ $+\frac{512}{945 A^{19} d^4 g \sin[4 \omega_1 - 3 \omega_2]}{32768}$ $+\frac{525 A^{25} d^6 g \sin[4 \omega_1 - 3 \omega_2]}{4194304}$

$\omega_1 - 3 \omega_2$ from 4 th $+\frac{1}{2} A^4 d \cos[\omega_1 - 3 \omega_2]$	IMD3	IMD5	IMD7
2 nd			
3 rd	$+\frac{3}{4} A^3 c \sin[2 \omega_1 - \omega_2]$		
4 th	$-\frac{3}{4} A^7 d^2 \sin[2 \omega_1 - \omega_2]$	$-\frac{3}{4} A^7 d^2 \sin[3 \omega_1 - 2 \omega_2]$	$-\frac{1}{4} A^7 d^2 \sin[4 \omega_1 - 3 \omega_2]$
5 th	$+\frac{25}{8} A^5 e \sin[2 \omega_1 - \omega_2]$ $+\frac{15}{16} A^{11} d^2 e \sin[2 \omega_1 - \omega_2]$	$+\frac{5}{8} A^5 e \sin[3 \omega_1 - 2 \omega_2]$	
6 th	$-\frac{75}{16} A^9 d f \sin[2 \omega_1 - \omega_2]$ $-\frac{45}{64} A^{15} d^3 f \sin[2 \omega_1 - \omega_2]$	$-\frac{75}{16} A^9 d f \sin[3 \omega_1 - 2 \omega_2]$ $-\frac{45}{64} A^{15} d^3 f \sin[3 \omega_1 - 2 \omega_2]$	$-\frac{75}{32} A^9 d f \sin[4 \omega_1 - 3 \omega_2]$ $-\frac{15}{64} A^{15} d^3 f \sin[4 \omega_1 - 3 \omega_2]$
7 th	$+\frac{735}{64} A^7 g \sin[2 \omega_1 - \omega_2]$ $+\frac{2121}{256} A^{13} d^2 g \sin[2 \omega_1 - \omega_2]$ $+\frac{315}{512} A^{19} d^4 g \sin[2 \omega_1 - \omega_2]$	$+\frac{245}{64} A^7 g \sin[3 \omega_1 - 2 \omega_2]$ $+\frac{525}{256} A^{13} d^2 g \sin[3 \omega_1 - 2 \omega_2]$	$+\frac{35}{64} A^7 g \sin[4 \omega_1 - 3 \omega_2]$ $+\frac{105}{128} A^{13} d^2 g \sin[4 \omega_1 - 3 \omega_2]$

$\omega_1 + 3 \omega_2$ from 4 th $-\frac{1}{2} A^4 d \cos[\omega_1 + 3 \omega_2]$	IMD3	IMD5	IMD7
2 nd			
3 rd	$+\frac{3}{4} A^3 c \sin[2 \omega_1 - \omega_2]$		
4 th			
5 th	$+\frac{25}{8} A^5 e \sin[2 \omega_1 - \omega_2]$ $+\frac{15}{16} A^{11} d^2 e \sin[2 \omega_1 - \omega_2]$	$+\frac{5}{8} A^5 e \sin[3 \omega_1 - 2 \omega_2]$	
6 th	$-\frac{45}{32} A^9 d f \sin[2 \omega_1 - \omega_2]$	$-\frac{15}{32} A^9 d f \sin[3 \omega_1 - 2 \omega_2]$	$-\frac{3}{32} A^9 d f \sin[4 \omega_1 - 3 \omega_2]$
7 th	$+\frac{735}{64} A^7 g \sin[2 \omega_1 - \omega_2]$ $+\frac{525}{64} A^{13} d^2 g \sin[2 \omega_1 - \omega_2]$ $+\frac{315}{512} A^{19} d^4 g \sin[2 \omega_1 - \omega_2]$	$+\frac{245}{64} A^7 g \sin[3 \omega_1 - 2 \omega_2]$ $+\frac{105}{64} A^{13} d^2 g \sin[3 \omega_1 - 2 \omega_2]$	$+\frac{35}{64} A^7 g \sin[4 \omega_1 - 3 \omega_2]$

$\omega_1 + \omega_2$ from 4 th $+3 A^4 d \cos[\omega_1 + \omega_2]$	IMD3	IMD5	IMD7
2 nd			
3 rd	$+\frac{3}{4} A^3 c \sin[2 \omega_1 - \omega_2]$		
4 th	$-\frac{6}{4} A^7 d^2 \sin[2 \omega_1 - \omega_2]$		
5 th	$+\frac{25}{8} A^5 e \sin[2 \omega_1 - \omega_2]$ $+\frac{315}{8} A^{11} d^2 e \sin[2 \omega_1 - \omega_2]$	$+\frac{5}{8} A^5 e \sin[3 \omega_1 - 2 \omega_2]$	
6 th	$-\frac{675}{16} A^9 d f \sin[2 \omega_1 - \omega_2]$ $-\frac{405}{2} A^{15} d^3 f \sin[2 \omega_1 - \omega_2]$	$-\frac{135}{16} A^9 d f \sin[3 \omega_1 - 2 \omega_2]$	

7 th	$ \begin{aligned} & + \frac{735}{64} A^7 g \sin[2 \omega_1 - \omega_2] \\ & + \frac{12285}{32} A^{13} d^2 g \sin[2 \omega_1 - \omega_2] \\ & + \frac{31185}{32} A^{19} d^4 g \sin[2 \omega_1 - \omega_2] \end{aligned} $	$ \begin{aligned} & + \frac{245}{64} A^7 g \sin[3 \omega_1 - 2 \omega_2] \\ & + \frac{2457}{32} A^{13} d^2 g \sin[3 \omega_1 - 2 \omega_2] \end{aligned} $	$ + \frac{35}{64} A^7 g \sin[4 \omega_1 - 3 \omega_2] $
-----------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------

$\omega_1 - \omega_2$ from 4 th $-3 A^4 d \cos[\omega_1 - \omega_2]$	IMD3	IMD5	IMD7
2 nd	$-6 A^5 b d \sin[2 \omega_1 - \omega_2]$		
3 rd	$ \begin{aligned} & + \frac{3}{4} A^3 c \sin[2 \omega_1 - \omega_2] \\ & + \frac{27}{4} A^9 c d^2 \sin[2 \omega_1 - \omega_2] \end{aligned} $	$+ \frac{27}{4} A^9 c d^2 \sin[3 \omega_1 - 2 \omega_2]$	
4 th			
5 th	$ \begin{aligned} & + \frac{25}{8} A^5 e \sin[2 \omega_1 - \omega_2] \\ & + \frac{675}{8} A^{11} d^2 e \sin[2 \omega_1 - \omega_2] \\ & + \frac{405}{4} A^{17} d^4 e \sin[2 \omega_1 - \omega_2] \end{aligned} $	$ \begin{aligned} & + \frac{5}{8} A^5 e \sin[3 \omega_1 - 2 \omega_2] \\ & + \frac{405}{8} A^{11} d^2 e \sin[3 \omega_1 - 2 \omega_2] \\ & + \frac{405}{4} A^{17} d^4 e \sin[3 \omega_1 - 2 \omega_2] \end{aligned} $	$ \begin{aligned} & + \frac{135}{8} A^{11} d^2 e \sin[4 \omega_1 - 3 \omega_2] \\ & + \frac{405}{16} A^{17} d^4 e \sin[4 \omega_1 - 3 \omega_2] \end{aligned} $
6 th	$ \begin{aligned} & - \frac{495}{8} A^9 d f \sin[2 \omega_1 - \omega_2] \\ & - \frac{2025}{4} A^{15} d^3 f \sin[2 \omega_1 - \omega_2] \\ & - \frac{3645}{8} A^{21} d^5 f \sin[2 \omega_1 - \omega_2] \end{aligned} $	$ \begin{aligned} & - \frac{225}{8} A^9 d f \sin[3 \omega_1 - 2 \omega_2] \\ & - \frac{1215}{4} A^{15} d^3 f \sin[3 \omega_1 - 2 \omega_2] \\ & - \frac{3645}{16} A^{21} d^5 f \sin[3 \omega_1 - 2 \omega_2] \end{aligned} $	$ \begin{aligned} & - \frac{45}{8} A^9 d f \sin[4 \omega_1 - 3 \omega_2] \\ & - \frac{1215}{8} A^{15} d^3 f \sin[4 \omega_1 - 3 \omega_2] \\ & - \frac{3645}{16} A^{21} d^5 f \sin[4 \omega_1 - 3 \omega_2] \end{aligned} $
7 th	$ \begin{aligned} & + \frac{735}{64} A^7 g \sin[2 \omega_1 - \omega_2] \\ & + \frac{4725}{8} A^{13} d^2 g \sin[2 \omega_1 - \omega_2] \\ & + \frac{76545}{8} A^{19} d^4 g \sin[2 \omega_1 - \omega_2] \\ & + \frac{32}{76545} A^{25} d^6 g \sin[2 \omega_1 - \omega_2] \end{aligned} $	$ \begin{aligned} & + \frac{245}{64} A^7 g \sin[3 \omega_1 - 2 \omega_2] \\ & + \frac{2835}{8} A^{13} d^2 g \sin[3 \omega_1 - 2 \omega_2] \\ & + \frac{110565}{8} A^{19} d^4 g \sin[3 \omega_1 - 2 \omega_2] \\ & + \frac{64}{76545} A^{25} d^6 g \sin[3 \omega_1 - 2 \omega_2] \end{aligned} $	$ \begin{aligned} & + \frac{35}{64} A^7 g \sin[4 \omega_1 - 3 \omega_2] \\ & + \frac{4725}{32} A^{13} d^2 g \sin[4 \omega_1 - 3 \omega_2] \\ & + \frac{59535}{64} A^{19} d^4 g \sin[4 \omega_1 - 3 \omega_2] \\ & + \frac{15309}{32} A^{25} d^6 g \sin[4 \omega_1 - 3 \omega_2] \end{aligned} $

C.4 Fed-Back Components From 5th-order Nonlinear Term

$2 \omega_1 - 3 \omega_2$ from 5 th $+ \frac{5}{8} A^5 e \sin[2 \omega_1 - 3 \omega_2]$	IMD3	IMD5	IMD7
2 nd			
3 rd	$ \begin{aligned} & + \frac{3}{4} A^3 c \sin[2 \omega_1 - \omega_2] \\ & - \frac{15}{32} A^7 c e \sin[2 \omega_1 - \omega_2] \end{aligned} $	$- \frac{15}{16} A^7 c e \sin[3 \omega_1 - 2 \omega_2]$	$- \frac{15}{32} A^7 c e \sin[4 \omega_1 - 3 \omega_2]$
4 th			
5 th	$ \begin{aligned} & + \frac{25}{8} A^5 e \sin[2 \omega_1 - \omega_2] \\ & - \frac{25}{8} A^9 e^2 \sin[2 \omega_1 - \omega_2] \\ & + \frac{375}{8} A^{13} e^3 \sin[2 \omega_1 - \omega_2] \\ & - \frac{256}{1875} A^{17} e^4 \sin[2 \omega_1 - \omega_2] \\ & \quad \quad \quad 4096 \end{aligned} $	$ \begin{aligned} & + \frac{5}{8} A^5 e \sin[3 \omega_1 - 2 \omega_2] \\ & - \frac{75}{16} A^9 e^2 \sin[3 \omega_1 - 2 \omega_2] \\ & - \frac{1875}{2048} A^{17} e^4 \sin[3 \omega_1 - 2 \omega_2] \end{aligned} $	$ \begin{aligned} & - \frac{25}{8} A^9 e^2 \sin[4 \omega_1 - 3 \omega_2] \\ & + \frac{125}{512} A^{13} e^3 \sin[4 \omega_1 - 3 \omega_2] \\ & - \frac{1875}{4096} A^{17} e^4 \sin[4 \omega_1 - 3 \omega_2] \end{aligned} $
6 th			
7 th	$ \begin{aligned} & + \frac{735}{64} A^7 g \sin[2 \omega_1 - \omega_2] \\ & - \frac{7875}{512} A^{11} e g \sin[2 \omega_1 - \omega_2] \\ & + \frac{13125}{1024} A^{15} e^2 g \sin[2 \omega_1 - \omega_2] \\ & - \frac{13125}{2048} A^{19} e^3 g \sin[2 \omega_1 - \omega_2] \\ & + \frac{196875}{4096} A^{23} e^4 g \sin[2 \omega_1 - \omega_2] \\ & + \frac{131072}{328125} A^{27} e^5 g \sin[2 \omega_1 - \omega_2] \\ & \quad \quad \quad 1048576 \end{aligned} $	$ \begin{aligned} & + \frac{245}{64} A^7 g \sin[3 \omega_1 - 2 \omega_2] \\ & - \frac{2625}{128} A^{11} e g \sin[3 \omega_1 - 2 \omega_2] \\ & + \frac{13125}{4096} A^{15} e^2 g \sin[3 \omega_1 - 2 \omega_2] \\ & - \frac{39375}{4096} A^{19} e^3 g \sin[3 \omega_1 - 2 \omega_2] \\ & - \frac{328125}{524288} A^{27} e^5 g \sin[3 \omega_1 - 2 \omega_2] \end{aligned} $	$ \begin{aligned} & + \frac{35}{64} A^7 g \sin[4 \omega_1 - 3 \omega_2] \\ & - \frac{7875}{512} A^{11} e g \sin[4 \omega_1 - 3 \omega_2] \\ & + \frac{13125}{4096} A^{15} e^2 g \sin[4 \omega_1 - 3 \omega_2] \\ & - \frac{13125}{2048} A^{19} e^3 g \sin[4 \omega_1 - 3 \omega_2] \\ & + \frac{21875}{65536} A^{23} e^4 g \sin[4 \omega_1 - 3 \omega_2] \\ & - \frac{328125}{1048576} A^{27} e^5 g \sin[4 \omega_1 - 3 \omega_2] \end{aligned} $

$2\omega_1 + 3\omega_2$ from 5^{th} $-\frac{5}{8}A^5e\sin[2\omega_1 + 3\omega_2]$	IMD3	IMD5	IMD7
2^{nd}			
3^{rd}	$+\frac{3}{4}A^3c\sin[2\omega_1 - \omega_2]$		
4^{th}			
5^{th}	$+\frac{25}{8}A^5e\sin[2\omega_1 - \omega_2]$ $-\frac{25}{128}A^9e^2\sin[2\omega_1 - \omega_2]$ $+\frac{375}{256}A^{13}e^3\sin[2\omega_1 - \omega_2]$	$+\frac{5}{8}A^5e\sin[3\omega_1 - 2\omega_2]$	
6^{th}			
7^{th}	$+\frac{735}{64}A^7g\sin[2\omega_1 - \omega_2]$ $-\frac{1785}{512}A^{11}eg\sin[2\omega_1 - \omega_2]$ $+\frac{13125}{1024}A^{15}e^2g\sin[2\omega_1 - \omega_2]$ $-\frac{13125}{32768}A^{19}e^3g\sin[2\omega_1 - \omega_2]$ $+\frac{196875}{131072}A^{23}e^4g\sin[2\omega_1 - \omega_2]$	$+\frac{245}{64}A^7g\sin[3\omega_1 - 2\omega_2]$ $-\frac{105}{128}A^{11}eg\sin[3\omega_1 - 2\omega_2]$ $+\frac{2625}{1024}A^{15}e^2g\sin[3\omega_1 - 2\omega_2]$	$+\frac{35}{64}A^7g\sin[4\omega_1 - 3\omega_2]$ $-\frac{35}{512}A^{11}eg\sin[4\omega_1 - 3\omega_2]$

$3\omega_1 - 2\omega_2$ from 5^{th} $-\frac{5}{8}A^5e\sin[3\omega_1 - 2\omega_2]$	IMD3	IMD5	IMD7
2^{nd}			
3^{rd}	$+\frac{3}{4}A^3c\sin[2\omega_1 - \omega_2]$ $-\frac{15}{16}A^7ce\sin[2\omega_1 - \omega_2]$	$-\frac{15}{8}A^7ce\sin[3\omega_1 - 2\omega_2]$ $-\frac{375}{2048}A^{15}ce^3\sin[3\omega_1 - 2\omega_2]$	$-\frac{15}{16}A^7ce\sin[4\omega_1 - 3\omega_2]$
4^{th}			
5^{th}	$+\frac{25}{8}A^5e\sin[2\omega_1 - \omega_2]$ $-\frac{75}{16}A^9e^2\sin[2\omega_1 - \omega_2]$ $+\frac{375}{256}A^{13}e^3\sin[2\omega_1 - \omega_2]$ $-\frac{1875}{2048}A^{17}e^4\sin[2\omega_1 - \omega_2]$	$+\frac{5}{8}A^5e\sin[3\omega_1 - 2\omega_2]$ $-\frac{225}{32}A^9e^2\sin[3\omega_1 - 2\omega_2]$ $-\frac{1875}{1024}A^{17}e^4\sin[3\omega_1 - 2\omega_2]$ $-\frac{15625}{262144}A^{25}e^6\sin[3\omega_1 - 2\omega_2]$	$-\frac{75}{16}A^9e^2\sin[4\omega_1 - 3\omega_2]$ $+\frac{375}{512}A^{13}e^3\sin[4\omega_1 - 3\omega_2]$ $-\frac{1875}{2048}A^{17}e^4\sin[4\omega_1 - 3\omega_2]$
6^{th}			
7^{th}	$+\frac{735}{64}A^7g\sin[2\omega_1 - \omega_2]$ $-\frac{2625}{128}A^{11}eg\sin[2\omega_1 - \omega_2]$ $+\frac{13125}{1024}A^{15}e^2g\sin[2\omega_1 - \omega_2]$ $+\frac{39375}{4096}A^{19}e^3g\sin[2\omega_1 - \omega_2]$ $+\frac{196875}{131072}A^{23}e^4g\sin[2\omega_1 - \omega_2]$ $+\frac{328125}{524288}A^{27}e^5g\sin[2\omega_1 - \omega_2]$	$+\frac{245}{64}A^7g\sin[3\omega_1 - 2\omega_2]$ $-\frac{875}{32}A^{11}eg\sin[3\omega_1 - 2\omega_2]$ $+\frac{7875}{2048}A^{15}e^2g\sin[3\omega_1 - 2\omega_2]$ $+\frac{118125}{8192}A^{19}e^3g\sin[3\omega_1 - 2\omega_2]$ $-\frac{328125}{262144}A^{27}e^5g\sin[3\omega_1 - 2\omega_2]$ $-\frac{2734375}{134217728}A^{35}e^7g\sin[3\omega_1 - 2\omega_2]$	$+\frac{35}{64}A^7g\sin[4\omega_1 - 3\omega_2]$ $-\frac{2625}{128}A^{11}eg\sin[4\omega_1 - 3\omega_2]$ $+\frac{13125}{2048}A^{15}e^2g\sin[4\omega_1 - 3\omega_2]$ $-\frac{39375}{4096}A^{19}e^3g\sin[4\omega_1 - 3\omega_2]$ $+\frac{65625}{65536}A^{23}e^4g\sin[4\omega_1 - 3\omega_2]$ $-\frac{328125}{524288}A^{27}e^5g\sin[4\omega_1 - 3\omega_2]$

$3\omega_1 + 2\omega_2$ from 5^{th} $-\frac{5}{8}A^5e\sin[3\omega_1 + 2\omega_2]$	IMD3	IMD5	IMD7
2^{nd}			
3^{rd}	$+\frac{3}{4}A^3c\sin[2\omega_1 - \omega_2]$		
4^{th}			
5^{th}	$+\frac{25}{8}A^5e\sin[2\omega_1 - \omega_2]$ $-\frac{25}{32}A^9e^2\sin[2\omega_1 - \omega_2]$ $+\frac{375}{256}A^{13}e^3\sin[2\omega_1 - \omega_2]$	$+\frac{5}{8}A^5e\sin[3\omega_1 - 2\omega_2]$ $-\frac{25}{128}A^9e^2\sin[3\omega_1 - 2\omega_2]$	
6^{th}			

7 th	$ \begin{aligned} & + \frac{735}{64} A^7 g \sin[2 \omega_1 - \omega_2] \\ & - \frac{105}{16} A^{11} e g \sin[2 \omega_1 - \omega_2] \\ & + \frac{13125}{16} A^{15} e^2 g \sin[2 \omega_1 - \omega_2] \\ & - \frac{1024}{13125} A^{19} e^3 g \sin[2 \omega_1 - \omega_2] \\ & + \frac{8192}{196875} A^{23} e^4 g \sin[2 \omega_1 - \omega_2] \\ & + \frac{131072}{131072} \end{aligned} $	$ \begin{aligned} & + \frac{245}{64} A^7 g \sin[3 \omega_1 - 2 \omega_2] \\ & - \frac{1295}{512} A^{11} e g \sin[3 \omega_1 - 2 \omega_2] \\ & + \frac{2625}{512} A^{15} e^2 g \sin[3 \omega_1 - 2 \omega_2] \\ & - \frac{1024}{13125} A^{19} e^3 g \sin[3 \omega_1 - 2 \omega_2] \\ & - \frac{32768}{32768} \end{aligned} $	$ \begin{aligned} & + \frac{35}{64} A^7 g \sin[4 \omega_1 - 3 \omega_2] \\ & - \frac{105}{256} A^{11} e g \sin[4 \omega_1 - 3 \omega_2] \end{aligned} $
-----------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------

3 rd from 5 th + $\frac{25}{16} A^5 e \sin[3 \omega_1]$ + $\frac{25}{16} A^5 e \sin[3 \omega_2]$	IMD3	IMD5	IMD7
2 nd			
3 rd	$ \begin{aligned} & + \frac{3}{4} A^3 c \sin[2 \omega_1 - \omega_2] \\ & - \frac{75}{32} A^7 c e \sin[2 \omega_1 - \omega_2] \end{aligned} $	$ \begin{aligned} & - \frac{75}{64} A^7 c e \sin[3 \omega_1 - 2 \omega_2] \\ & + \frac{1875}{512} A^{11} c e^2 \sin[3 \omega_1 - 2 \omega_2] \end{aligned} $	$ + \frac{1875}{512} A^{11} c e^2 \sin[4 \omega_1 - 3 \omega_2] $
4 th			
5 th	$ \begin{aligned} & + \frac{25}{8} A^5 e \sin[2 \omega_1 - \omega_2] \\ & - \frac{1875}{128} A^9 e^2 \sin[2 \omega_1 - \omega_2] \\ & + \frac{28125}{128} A^{13} e^3 \sin[2 \omega_1 - \omega_2] \\ & - \frac{1024}{703125} A^{17} e^4 \sin[2 \omega_1 - \omega_2] \\ & - \frac{16384}{16384} \end{aligned} $	$ \begin{aligned} & + \frac{5}{8} A^5 e \sin[3 \omega_1 - 2 \omega_2] \\ & - \frac{625}{64} A^9 e^2 \sin[3 \omega_1 - 2 \omega_2] \\ & + \frac{28125}{64} A^{13} e^3 \sin[3 \omega_1 - 2 \omega_2] \\ & - \frac{1024}{703125} A^{17} e^4 \sin[3 \omega_1 - 2 \omega_2] \\ & - \frac{32768}{5859375} A^{21} e^5 \sin[3 \omega_1 - 2 \omega_2] \\ & + \frac{131072}{131072} \end{aligned} $	$ \begin{aligned} & - \frac{625}{256} A^9 e^2 \sin[4 \omega_1 - 3 \omega_2] \\ & + \frac{28125}{256} A^{13} e^3 \sin[4 \omega_1 - 3 \omega_2] \\ & - \frac{1024}{234375} A^{17} e^4 \sin[4 \omega_1 - 3 \omega_2] \\ & + \frac{32768}{5859375} A^{21} e^5 \sin[4 \omega_1 - 3 \omega_2] \\ & + \frac{131072}{131072} \end{aligned} $
6 th			
7 th	$ \begin{aligned} & + \frac{735}{64} A^7 g \sin[2 \omega_1 - \omega_2] \\ & - \frac{77175}{1024} A^{11} e g \sin[2 \omega_1 - \omega_2] \\ & + \frac{4134375}{16384} A^{15} e^2 g \sin[2 \omega_1 - \omega_2] \\ & - \frac{73828125}{131072} A^{19} e^3 g \sin[2 \omega_1 - \omega_2] \\ & + \frac{615234375}{1048576} A^{23} e^4 g \sin[2 \omega_1 - \omega_2] \\ & - \frac{5126953125}{8388608} A^{27} e^5 g \sin[2 \omega_1 - \omega_2] \\ & + \frac{8388608}{8388608} \end{aligned} $	$ \begin{aligned} & + \frac{245}{64} A^7 g \sin[3 \omega_1 - 2 \omega_2] \\ & - \frac{55125}{1024} A^{11} e g \sin[3 \omega_1 - 2 \omega_2] \\ & + \frac{826875}{4096} A^{15} e^2 g \sin[3 \omega_1 - 2 \omega_2] \\ & - \frac{106640625}{262144} A^{19} e^3 g \sin[3 \omega_1 - 2 \omega_2] \\ & + \frac{369140625}{524288} A^{23} e^4 g \sin[3 \omega_1 - 2 \omega_2] \\ & - \frac{5126953125}{16777216} A^{27} e^5 g \sin[3 \omega_1 - 2 \omega_2] \\ & + \frac{128173828125}{268435456} A^{31} e^6 g \sin[3 \omega_1 - 2 \omega_2] \\ & + \frac{268435456}{268435456} \end{aligned} $	$ \begin{aligned} & + \frac{35}{64} A^7 g \sin[4 \omega_1 - 3 \omega_2] \\ & - \frac{11025}{512} A^{11} e g \sin[4 \omega_1 - 3 \omega_2] \\ & + \frac{1378125}{8192} A^{15} e^2 g \sin[4 \omega_1 - 3 \omega_2] \\ & - \frac{57421875}{262144} A^{19} e^3 g \sin[4 \omega_1 - 3 \omega_2] \\ & + \frac{369140625}{524288} A^{23} e^4 g \sin[4 \omega_1 - 3 \omega_2] \\ & - \frac{5126953125}{33554432} A^{27} e^5 g \sin[4 \omega_1 - 3 \omega_2] \\ & + \frac{128173828125}{268435456} A^{31} e^6 g \sin[4 \omega_1 - 3 \omega_2] \\ & + \frac{268435456}{268435456} \end{aligned} $

4 th $\omega_1 + \omega_2$ from 5 th - $\frac{5}{16} A^5 e \sin[4 \omega_1 + \omega_2]$	IMD3	IMD5	IMD7
2 nd			
3 rd	$ - \frac{3}{4} A^3 c \sin[2 \omega_1 - \omega_2] $		
4 th			
5 th	$ \begin{aligned} & + \frac{25}{8} A^5 e \sin[2 \omega_1 - \omega_2] \\ & - \frac{75}{128} A^9 e^2 \sin[2 \omega_1 - \omega_2] \\ & + \frac{375}{1024} A^{13} e^3 \sin[2 \omega_1 - \omega_2] \end{aligned} $	$ \begin{aligned} & + \frac{5}{8} A^5 e \sin[3 \omega_1 - 2 \omega_2] \\ & - \frac{25}{64} A^9 e^2 \sin[3 \omega_1 - 2 \omega_2] \end{aligned} $	$ - \frac{25}{256} A^9 e^2 \sin[4 \omega_1 - 3 \omega_2] $
6 th			
7 th	$ \begin{aligned} & + \frac{735}{64} A^7 g \sin[2 \omega_1 - \omega_2] \\ & - \frac{4235}{1024} A^{11} e g \sin[2 \omega_1 - \omega_2] \\ & + \frac{13125}{4096} A^{15} e^2 g \sin[2 \omega_1 - \omega_2] \\ & - \frac{39375}{131072} A^{19} e^3 g \sin[2 \omega_1 - \omega_2] \\ & - \frac{196875}{2097152} A^{23} e^4 g \sin[2 \omega_1 - \omega_2] \\ & + \frac{2097152}{2097152} \end{aligned} $	$ \begin{aligned} & + \frac{245}{64} A^7 g \sin[3 \omega_1 - 2 \omega_2] \\ & - \frac{1575}{512} A^{11} e g \sin[3 \omega_1 - 2 \omega_2] \\ & + \frac{2625}{4096} A^{15} e^2 g \sin[3 \omega_1 - 2 \omega_2] \\ & - \frac{13125}{65536} A^{19} e^3 g \sin[3 \omega_1 - 2 \omega_2] \\ & - \frac{65536}{65536} \end{aligned} $	$ \begin{aligned} & + \frac{35}{64} A^7 g \sin[4 \omega_1 - 3 \omega_2] \\ & - \frac{315}{256} A^{11} e g \sin[4 \omega_1 - 3 \omega_2] \\ & - \frac{13125}{262144} A^{19} e^3 g \sin[4 \omega_1 - 3 \omega_2] \\ & - \frac{262144}{262144} \end{aligned} $

$4\omega_1 - \omega_2$ from 5^{th} $+\frac{5}{16}A^5e\sin[4\omega_1 - \omega_2]$	IMD3	IMD5	IMD7
2^{nd}			
3^{rd}	$+\frac{3}{4}A^3c\sin[2\omega_1 - \omega_2]$ $-\frac{15}{64}A^7ce\sin[2\omega_1 - \omega_2]$	$-\frac{15}{32}A^7ce\sin[3\omega_1 - 2\omega_2]$	$-\frac{15}{64}A^7ce\sin[4\omega_1 - 3\omega_2]$
4^{th}			
5^{th}	$+\frac{25}{8}A^5e\sin[2\omega_1 - \omega_2]$ $-\frac{25}{16}A^9e^2\sin[2\omega_1 - \omega_2]$ $+\frac{375}{16}A^{13}e^3\sin[2\omega_1 - \omega_2]$ $+\frac{1024}{1875}A^{17}e^4\sin[2\omega_1 - \omega_2]$ $-\frac{32768}{32768}$	$+\frac{5}{8}A^5e\sin[3\omega_1 - 2\omega_2]$ $-\frac{75}{32}A^9e^2\sin[3\omega_1 - 2\omega_2]$ $-\frac{1875}{16384}A^{17}e^4\sin[3\omega_1 - 2\omega_2]$	$-\frac{25}{16}A^9e^2\sin[4\omega_1 - 3\omega_2]$ $-\frac{1875}{32768}A^{17}e^4\sin[4\omega_1 - 3\omega_2]$
6^{th}			
7^{th}	$+\frac{735}{64}A^7g\sin[2\omega_1 - \omega_2]$ $-\frac{7875}{1024}A^{11}eg\sin[2\omega_1 - \omega_2]$ $+\frac{13125}{4096}A^{15}e^2g\sin[2\omega_1 - \omega_2]$ $+\frac{13125}{16384}A^{19}e^3g\sin[2\omega_1 - \omega_2]$ $-\frac{196875}{2097152}A^{23}e^4g\sin[2\omega_1 - \omega_2]$ $+\frac{328125}{33554432}A^{27}e^5g\sin[2\omega_1 - \omega_2]$ $-\frac{33554432}{33554432}$	$+\frac{245}{64}A^7g\sin[3\omega_1 - 2\omega_2]$ $-\frac{2625}{256}A^{11}eg\sin[3\omega_1 - 2\omega_2]$ $+\frac{11025}{16384}A^{15}e^2g\sin[3\omega_1 - 2\omega_2]$ $+\frac{39375}{32768}A^{19}e^3g\sin[3\omega_1 - 2\omega_2]$ $-\frac{328125}{16777216}A^{27}e^5g\sin[3\omega_1 - 2\omega_2]$	$+\frac{35}{64}A^7g\sin[4\omega_1 - 3\omega_2]$ $-\frac{7875}{1024}A^{11}eg\sin[4\omega_1 - 3\omega_2]$ $+\frac{2625}{16384}A^{15}e^2g\sin[4\omega_1 - 3\omega_2]$ $+\frac{13125}{16384}A^{19}e^3g\sin[4\omega_1 - 3\omega_2]$ $-\frac{328125}{33554432}A^{27}e^5g\sin[4\omega_1 - 3\omega_2]$

5ω from 5^{th} $-\frac{1}{16}A^5e\sin[5\omega_1]$ $-\frac{1}{16}A^5e\sin[5\omega_2]$	IMD3	IMD5	IMD7
2^{nd}			
3^{rd}	$+\frac{3}{4}A^3c\sin[2\omega_1 - \omega_2]$		
4^{th}			
5^{th}	$+\frac{25}{8}A^5e\sin[2\omega_1 - \omega_2]$ $-\frac{5}{64}A^9e^2\sin[2\omega_1 - \omega_2]$ $+\frac{15}{512}A^{13}e^3\sin[2\omega_1 - \omega_2]$	$+\frac{5}{8}A^5e\sin[3\omega_1 - 2\omega_2]$ $-\frac{15}{128}A^9e^2\sin[3\omega_1 - 2\omega_2]$	$-\frac{5}{64}A^9e^2\sin[4\omega_1 - 3\omega_2]$ $+\frac{15}{1024}A^{13}e^3\sin[4\omega_1 - 3\omega_2]$
6^{th}			
7^{th}	$+\frac{735}{64}A^7g\sin[2\omega_1 - \omega_2]$ $-\frac{735}{1024}A^{11}eg\sin[2\omega_1 - \omega_2]$ $+\frac{525}{2048}A^{15}e^2g\sin[2\omega_1 - \omega_2]$ $+\frac{315}{65536}A^{19}e^3g\sin[2\omega_1 - \omega_2]$ $-\frac{945}{1048576}A^{23}e^4g\sin[2\omega_1 - \omega_2]$ $+\frac{1048576}{1048576}$	$+\frac{245}{64}A^7g\sin[3\omega_1 - 2\omega_2]$ $-\frac{245}{256}A^{11}eg\sin[3\omega_1 - 2\omega_2]$ $+\frac{315}{4096}A^{15}e^2g\sin[3\omega_1 - 2\omega_2]$ $-\frac{945}{131072}A^{19}e^3g\sin[3\omega_1 - 2\omega_2]$	$+\frac{35}{64}A^7g\sin[4\omega_1 - 3\omega_2]$ $-\frac{735}{1024}A^{11}eg\sin[4\omega_1 - 3\omega_2]$ $+\frac{525}{4096}A^{15}e^2g\sin[4\omega_1 - 3\omega_2]$ $-\frac{315}{65536}A^{19}e^3g\sin[4\omega_1 - 3\omega_2]$ $+\frac{315}{524288}A^{23}e^4g\sin[4\omega_1 - 3\omega_2]$

$\omega_1 - 4\omega_2$ from 5^{th} $-\frac{5}{16}A^5e\sin[\omega_1 - 4\omega_2]$	IMD3	IMD5	IMD7
2^{nd}			
3^{rd}	$+\frac{3}{4}A^3c\sin[2\omega_1 - \omega_2]$		
4^{th}			
5^{th}	$+\frac{25}{8}A^5e\sin[2\omega_1 - \omega_2]$ $-\frac{25}{64}A^9e^2\sin[2\omega_1 - \omega_2]$ $+\frac{375}{1024}A^{13}e^3\sin[2\omega_1 - \omega_2]$	$+\frac{5}{8}A^5e\sin[3\omega_1 - 2\omega_2]$ $-\frac{75}{128}A^9e^2\sin[3\omega_1 - 2\omega_2]$	$-\frac{25}{64}A^9e^2\sin[4\omega_1 - 3\omega_2]$

6 th			
7 th	$ \begin{aligned} & + \frac{735}{64} A^7 g \sin[2 \omega_1 - \omega_2] \\ & - \frac{1575}{512} A^{11} e g \sin[2 \omega_1 - \omega_2] \\ & + \frac{13125 A^{15} e^2 g \sin[2 \omega_1 - \omega_2]}{4096} \\ & - \frac{13125 A^{19} e^3 g \sin[2 \omega_1 - \omega_2]}{65536} \\ & + \frac{196875 A^{23} e^4 g \sin[2 \omega_1 - \omega_2]}{2097152} \end{aligned} $	$ \begin{aligned} & + \frac{245}{64} A^7 g \sin[3 \omega_1 - 2 \omega_2] \\ & - \frac{525}{128} A^{11} e g \sin[3 \omega_1 - 2 \omega_2] \\ & + \frac{2625 A^{15} e^2 g \sin[3 \omega_1 - 2 \omega_2]}{4096} \\ & - \frac{39375 A^{19} e^3 g \sin[3 \omega_1 - 2 \omega_2]}{131072} \end{aligned} $	$ \begin{aligned} & + \frac{35}{64} A^7 g \sin[4 \omega_1 - 3 \omega_2] \\ & - \frac{1575}{512} A^{11} e g \sin[4 \omega_1 - 3 \omega_2] \\ & - \frac{13125 A^{19} e^3 g \sin[4 \omega_1 - 3 \omega_2]}{65536} \end{aligned} $

$\omega_1 + 4 \omega_2$ from 5 th $-\frac{5}{16} A^5 e \sin[\omega_1 + 4 \omega_2]$	IMD3	IMD5	IMD7
2 nd			
3 rd	$+\frac{3}{4} A^3 c \sin[2 \omega_1 - \omega_2]$		
4 th			
5 th	$ \begin{aligned} & + \frac{25}{8} A^5 e \sin[2 \omega_1 - \omega_2] \\ & + \frac{375 A^{13} e^3 \sin[2 \omega_1 - \omega_2]}{1024} \end{aligned} $	$+\frac{5}{8} A^5 e \sin[3 \omega_1 - 2 \omega_2]$	
6 th			
7 th	$ \begin{aligned} & + \frac{735}{64} A^7 g \sin[2 \omega_1 - \omega_2] \\ & - \frac{455}{512} A^{11} e g \sin[2 \omega_1 - \omega_2] \\ & + \frac{13125 A^{15} e^2 g \sin[2 \omega_1 - \omega_2]}{4096} \\ & + \frac{196875 A^{23} e^4 g \sin[2 \omega_1 - \omega_2]}{2097152} \end{aligned} $	$ \begin{aligned} & + \frac{245}{64} A^7 g \sin[3 \omega_1 - 2 \omega_2] \\ & - \frac{525 A^{11} e g \sin[3 \omega_1 - 2 \omega_2]}{1024} \\ & + \frac{2625 A^{15} e^2 g \sin[3 \omega_1 - 2 \omega_2]}{4096} \end{aligned} $	$ \begin{aligned} & + \frac{35}{64} A^7 g \sin[4 \omega_1 - 3 \omega_2] \\ & - \frac{105}{512} A^{11} e g \sin[4 \omega_1 - 3 \omega_2] \end{aligned} $

C.5 Fed-Back Components From 6th-order Nonlinear Term

$2 \omega_1 - 4 \omega_2$ from 6 th $+\frac{15}{32} A^6 f \cos[2 \omega_1 - 4 \omega_2]$	IMD3	IMD5	IMD7
2 nd			
3 rd	$+\frac{3}{4} A^3 c \sin[2 \omega_1 - \omega_2]$		
4 th	$-\frac{15}{64} A^9 d f \sin[2 \omega_1 - \omega_2]$	$-\frac{45}{64} A^9 d f \sin[3 \omega_1 - 2 \omega_2]$	$-\frac{45}{64} A^9 d f \sin[4 \omega_1 - 3 \omega_2]$
5 th	$ \begin{aligned} & + \frac{25}{8} A^5 e \sin[2 \omega_1 - \omega_2] \\ & + \frac{3375 A^{15} e f^2 \sin[2 \omega_1 - \omega_2]}{4096} \end{aligned} $	$+\frac{5}{8} A^5 e \sin[3 \omega_1 - 2 \omega_2]$	
6 th	$ \begin{aligned} & - \frac{1125}{512} A^{11} f^2 \sin[2 \omega_1 - \omega_2] \\ & - \frac{50625 A^{21} f^4 \sin[2 \omega_1 - \omega_2]}{262144} \end{aligned} $	$ \begin{aligned} & - \frac{1125}{256} A^{11} f^2 \sin[3 \omega_1 - 2 \omega_2] \\ & - \frac{151875 A^{21} f^4 \sin[3 \omega_1 - 2 \omega_2]}{262144} \end{aligned} $	$ \begin{aligned} & - \frac{1125}{256} A^{11} f^2 \sin[4 \omega_1 - 3 \omega_2] \\ & - \frac{151875 A^{21} f^4 \sin[4 \omega_1 - 3 \omega_2]}{262144} \end{aligned} $
7 th	$ \begin{aligned} & + \frac{735}{64} A^7 g \sin[2 \omega_1 - \omega_2] \\ & + \frac{118125 A^{17} f^2 g \sin[2 \omega_1 - \omega_2]}{16384} \\ & + \frac{15946875 A^{27} f^4 g \sin[2 \omega_1 - \omega_2]}{33554432} \end{aligned} $	$ \begin{aligned} & + \frac{245}{64} A^7 g \sin[3 \omega_1 - 2 \omega_2] \\ & + \frac{23625 A^{17} f^2 g \sin[3 \omega_1 - 2 \omega_2]}{16384} \end{aligned} $	$ \begin{aligned} & + \frac{35}{64} A^7 g \sin[4 \omega_1 - 3 \omega_2] \\ & + \frac{4725 A^{17} f^2 g \sin[4 \omega_1 - 3 \omega_2]}{65536} \end{aligned} $

$2 \omega_1 + 4 \omega_2$ from 6 th $+\frac{15}{32} A^6 f \cos[2 \omega_1 + 4 \omega_2]$	IMD3	IMD5	IMD7
2 nd			
3 rd	$+\frac{3}{4} A^3 c \sin[2 \omega_1 - \omega_2]$		
4 th			
5 th	$ \begin{aligned} & + \frac{25}{8} A^5 e \sin[2 \omega_1 - \omega_2] \\ & + \frac{3375 A^{15} e f^2 \sin[2 \omega_1 - \omega_2]}{4096} \end{aligned} $	$+\frac{5}{8} A^5 e \sin[3 \omega_1 - 2 \omega_2]$	
6 th	$-\frac{45}{512} A^{11} f^2 \sin[2 \omega_1 - \omega_2]$		

Appendix C: FULL RESULTS OF HARMONIC AND IMD FEEDBACK ANALYSIS

7 th	$+\frac{735}{64}A^7g\sin[2\omega_1-\omega_2]$ $+\frac{118125A^{17}f^2g\sin[2\omega_1-\omega_2]}{16384}$ $+\frac{15946875A^{27}f^4g\sin[2\omega_1-\omega_2]}{33554432}$	$+\frac{245}{64}A^7g\sin[3\omega_1-2\omega_2]$ $+\frac{23625A^{17}f^2g\sin[3\omega_1-2\omega_2]}{16384}$	$+\frac{35}{64}A^7g\sin[4\omega_1-3\omega_2]$
-----------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------	-----------------------------------------------

$3\omega_1-3\omega_2$ from 6 th $-\frac{5}{8}A^6f\cos[3\omega_1-3\omega_2]$	IMD3	IMD5	IMD7
2 nd		$-80A^7bf\sin[3\omega_1-2\omega_2]$	$-80A^7bf\sin[4\omega_1-3\omega_2]$
3 rd	$+\frac{3}{4}A^3c\sin[2\omega_1-\omega_2]$		
4 th	$-\frac{15}{16}A^9df\sin[2\omega_1-\omega_2]$	$-\frac{45}{16}A^9df\sin[3\omega_1-2\omega_2]$ $-\frac{375A^{19}df^3\sin[3\omega_1-2\omega_2]}{1024}$	$-\frac{45}{16}A^9df\sin[4\omega_1-3\omega_2]$ $-\frac{375A^{19}df^3\sin[4\omega_1-3\omega_2]}{1024}$
5 th	$+\frac{25}{8}A^5e\sin[2\omega_1-\omega_2]$ $+\frac{375}{256}A^{15}ef^2\sin[2\omega_1-\omega_2]$	$+\frac{5}{8}A^5e\sin[3\omega_1-2\omega_2]$	
6 th	$-\frac{375}{64}A^{11}f^2\sin[2\omega_1-\omega_2]$ $-\frac{5625A^{21}f^4\sin[2\omega_1-\omega_2]}{4096}$	$-\frac{375}{32}A^{11}f^2\sin[3\omega_1-2\omega_2]$ $-\frac{16875A^{21}f^4\sin[3\omega_1-2\omega_2]}{4096}$ $-\frac{46875A^{31}f^6\sin[3\omega_1-2\omega_2]}{262144}$	$-\frac{375}{32}A^{11}f^2\sin[4\omega_1-3\omega_2]$ $-\frac{16875A^{21}f^4\sin[4\omega_1-3\omega_2]}{4096}$ $-\frac{46875A^{31}f^6\sin[4\omega_1-3\omega_2]}{262144}$
7 th	$+\frac{735}{64}A^7g\sin[2\omega_1-\omega_2]$ $+\frac{13125A^{17}f^2g\sin[2\omega_1-\omega_2]}{1024}$ $+\frac{196875A^{27}f^4g\sin[2\omega_1-\omega_2]}{131072}$	$+\frac{245}{64}A^7g\sin[3\omega_1-2\omega_2]$ $+\frac{2625A^{17}f^2g\sin[3\omega_1-2\omega_2]}{1024}$	$+\frac{35}{64}A^7g\sin[4\omega_1-3\omega_2]$ $+\frac{2625A^{17}f^2g\sin[4\omega_1-3\omega_2]}{2048}$

$3\omega_1+3\omega_2$ from 6 th $+\frac{5}{8}A^6f\cos[3\omega_1+3\omega_2]$	IMD3	IMD5	IMD7
2 nd			
3 rd	$+\frac{3}{4}A^3c\sin[2\omega_1-\omega_2]$		
4 th			
5 th	$+\frac{25}{8}A^5e\sin[2\omega_1-\omega_2]$ $+\frac{375}{256}A^{15}ef^2\sin[2\omega_1-\omega_2]$	$+\frac{5}{8}A^5e\sin[3\omega_1-2\omega_2]$	
6 th	$-\frac{75}{128}A^{11}f^2\sin[2\omega_1-\omega_2]$	$-\frac{15}{128}A^{11}f^2\sin[3\omega_1-2\omega_2]$	
7 th	$+\frac{735}{64}A^7g\sin[2\omega_1-\omega_2]$ $+\frac{13125A^{17}f^2g\sin[2\omega_1-\omega_2]}{1024}$ $+\frac{196875A^{27}f^4g\sin[2\omega_1-\omega_2]}{131072}$	$+\frac{245}{64}A^7g\sin[3\omega_1-2\omega_2]$ $+\frac{2625A^{17}f^2g\sin[3\omega_1-2\omega_2]}{1024}$	$+\frac{35}{64}A^7g\sin[4\omega_1-3\omega_2]$

$4\omega_1-2\omega_2$ from 6 th $+\frac{15}{32}A^6f\cos[4\omega_1-2\omega_2]$	IMD3	IMD5	IMD7
2 nd		$-960A^7bf\sin[3\omega_1-2\omega_2]$	$-960A^7bf\sin[4\omega_1-3\omega_2]$
3 rd	$+\frac{3}{4}A^3c\sin[2\omega_1-\omega_2]$		
4 th	$-\frac{45}{64}A^9df\sin[2\omega_1-\omega_2]$	$-\frac{135}{64}A^9df\sin[3\omega_1-2\omega_2]$ $-\frac{10125A^{19}df^3\sin[3\omega_1-2\omega_2]}{65536}$	$-\frac{135}{64}A^9df\sin[4\omega_1-3\omega_2]$ $-\frac{10125A^{19}df^3\sin[4\omega_1-3\omega_2]}{65536}$
5 th	$+\frac{25}{8}A^5e\sin[2\omega_1-\omega_2]$ $+\frac{3375A^{15}ef^2\sin[2\omega_1-\omega_2]}{4096}$	$+\frac{5}{8}A^5e\sin[3\omega_1-2\omega_2]$	

Appendix C: FULL RESULTS OF HARMONIC AND IMD FEEDBACK ANALYSIS

6 th	$-\frac{1125}{256} A^{11} f^2 \sin[2 \omega_1 - \omega_2]$ $-\frac{151875}{262144} A^{21} f^4 \sin[2 \omega_1 - \omega_2]$	$-\frac{1125}{128} A^{11} f^2 \sin[3 \omega_1 - 2 \omega_2]$ $-\frac{455625}{262144} A^{21} f^4 \sin[3 \omega_1 - 2 \omega_2]$ $-\frac{11390625}{268435456} A^{31} f^6 \sin[3 \omega_1 - 2 \omega_2]$	$-\frac{1125}{128} A^{11} f^2 \sin[4 \omega_1 - 3 \omega_2]$ $-\frac{455625}{262144} A^{21} f^4 \sin[4 \omega_1 - 3 \omega_2]$ $-\frac{11390625}{268435456} A^{31} f^6 \sin[4 \omega_1 - 3 \omega_2]$
7 th	$+\frac{735}{64} A^7 g \sin[2 \omega_1 - \omega_2]$ $+\frac{118125}{16384} A^{17} f^2 g \sin[2 \omega_1 - \omega_2]$ $+\frac{15946875}{33554432} A^{27} f^4 g \sin[2 \omega_1 - \omega_2]$	$+\frac{245}{64} A^7 g \sin[3 \omega_1 - 2 \omega_2]$ $+\frac{23625}{16384} A^{17} f^2 g \sin[3 \omega_1 - 2 \omega_2]$	$+\frac{35}{64} A^7 g \sin[4 \omega_1 - 3 \omega_2]$ $+\frac{23625}{65536} A^{17} f^2 g \sin[4 \omega_1 - 3 \omega_2]$

$4 \omega_1 + 2 \omega_2$ from 6 th $+\frac{15}{32} A^6 f \cos[4 \omega_1 + 2 \omega_2]$	IMD3	IMD5	IMD7
2 nd			
3 rd	$+\frac{3}{4} A^3 c \sin[2 \omega_1 - \omega_2]$		
4 th			
5 th	$+\frac{25}{8} A^5 e \sin[2 \omega_1 - \omega_2]$ $+\frac{3375}{4096} A^{15} e f^2 \sin[2 \omega_1 - \omega_2]$	$+\frac{5}{8} A^5 e \sin[3 \omega_1 - 2 \omega_2]$	
6 th	$-\frac{225}{256} A^{11} f^2 \sin[2 \omega_1 - \omega_2]$	$-\frac{225}{512} A^{11} f^2 \sin[3 \omega_1 - 2 \omega_2]$	$-\frac{45}{512} A^{11} f^2 \sin[4 \omega_1 - 3 \omega_2]$
7 th	$+\frac{735}{64} A^7 g \sin[2 \omega_1 - \omega_2]$ $+\frac{118125}{16384} A^{17} f^2 g \sin[2 \omega_1 - \omega_2]$ $+\frac{15946875}{33554432} A^{27} f^4 g \sin[2 \omega_1 - \omega_2]$	$+\frac{245}{64} A^7 g \sin[3 \omega_1 - 2 \omega_2]$ $+\frac{23625}{16384} A^{17} f^2 g \sin[3 \omega_1 - 2 \omega_2]$	$+\frac{35}{64} A^7 g \sin[4 \omega_1 - 3 \omega_2]$

$5 \omega_1 + \omega_2$ from 6 th $+\frac{3}{16} A^6 f \cos[5 \omega_1 + \omega_2]$	IMD3	IMD5	IMD7
2 nd			
3 rd	$+\frac{3}{4} A^3 c \sin[2 \omega_1 - \omega_2]$		
4 th			
5 th	$+\frac{25}{8} A^5 e \sin[2 \omega_1 - \omega_2]$ $+\frac{135}{1024} A^{15} e f^2 \sin[2 \omega_1 - \omega_2]$	$+\frac{5}{8} A^5 e \sin[3 \omega_1 - 2 \omega_2]$	
6 th	$-\frac{45}{128} A^{11} f^2 \sin[2 \omega_1 - \omega_2]$	$-\frac{45}{128} A^{11} f^2 \sin[3 \omega_1 - 2 \omega_2]$	$-\frac{45}{256} A^{11} f^2 \sin[4 \omega_1 - 3 \omega_2]$
7 th	$+\frac{735}{64} A^7 g \sin[2 \omega_1 - \omega_2]$ $+\frac{4725}{4096} A^{17} f^2 g \sin[2 \omega_1 - \omega_2]$ $+\frac{25515}{2097152} A^{27} f^4 g \sin[2 \omega_1 - \omega_2]$	$+\frac{245}{64} A^7 g \sin[3 \omega_1 - 2 \omega_2]$ $+\frac{945}{4096} A^{17} f^2 g \sin[3 \omega_1 - 2 \omega_2]$	$+\frac{35}{64} A^7 g \sin[4 \omega_1 - 3 \omega_2]$

$5 \omega_1 - \omega_2$ from 6 th $-\frac{3}{16} A^6 f \cos[5 \omega_1 - \omega_2]$	IMD3	IMD5	IMD7
2 nd			
3 rd	$+\frac{3}{4} A^3 c \sin[2 \omega_1 - \omega_2]$		
4 th	$-\frac{3}{32} A^9 d f \sin[2 \omega_1 - \omega_2]$	$-\frac{9}{32} A^9 d f \sin[3 \omega_1 - 2 \omega_2]$	$-\frac{9}{32} A^9 d f \sin[4 \omega_1 - 3 \omega_2]$
5 th	$+\frac{25}{8} A^5 e \sin[2 \omega_1 - \omega_2]$ $+\frac{135}{1024} A^{15} e f^2 \sin[2 \omega_1 - \omega_2]$	$+\frac{5}{8} A^5 e \sin[3 \omega_1 - 2 \omega_2]$	
6 th	$-\frac{225}{256} A^{11} f^2 \sin[2 \omega_1 - \omega_2]$ $-\frac{405}{32768} A^{21} f^4 \sin[2 \omega_1 - \omega_2]$	$-\frac{225}{128} A^{11} f^2 \sin[3 \omega_1 - 2 \omega_2]$ $-\frac{1215}{32768} A^{21} f^4 \sin[3 \omega_1 - 2 \omega_2]$	$-\frac{225}{128} A^{11} f^2 \sin[4 \omega_1 - 3 \omega_2]$ $-\frac{1215}{32768} A^{21} f^4 \sin[4 \omega_1 - 3 \omega_2]$

Appendix C: FULL RESULTS OF HARMONIC AND IMD FEEDBACK ANALYSIS

7 th	$+ \frac{735}{64} A^7 g \sin[2 \omega_1 - \omega_2]$ $+ \frac{4725 A^{17} f^2 g \sin[2 \omega_1 - \omega_2]}{4096}$ $+ \frac{25515 A^{27} f^4 g \sin[2 \omega_1 - \omega_2]}{2097152}$	$+ \frac{245}{64} A^7 g \sin[3 \omega_1 - 2 \omega_2]$ $+ \frac{945 A^{17} f^2 g \sin[3 \omega_1 - 2 \omega_2]}{4096}$	$+ \frac{35}{64} A^7 g \sin[4 \omega_1 - 3 \omega_2]$
-----------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------

$6 \omega \text{ from } 6^{\text{th}}$ $+ \frac{1}{32} A^6 f \cos[6 \omega_1]$ $+ \frac{1}{32} A^6 f \cos[6 \omega_2]$	IMD3	IMD5	IMD7
2 nd			
3 rd	$+ \frac{3}{4} A^3 c \sin[2 \omega_1 - \omega_2]$		
4 th			
5 th	$+ \frac{25}{8} A^5 e \sin[2 \omega_1 - \omega_2]$ $+ \frac{15 A^{15} e f^2 \sin[2 \omega_1 - \omega_2]}{2048}$	$+ \frac{5}{8} A^5 e \sin[3 \omega_1 - 2 \omega_2]$	
6 th	$- \frac{15}{512} A^{11} f^2 \sin[2 \omega_1 - \omega_2]$	$- \frac{15}{256} A^{11} f^2 \sin[3 \omega_1 - 2 \omega_2]$	$- \frac{15}{256} A^{11} f^2 \sin[4 \omega_1 - 3 \omega_2]$
7 th	$+ \frac{735}{64} A^7 g \sin[2 \omega_1 - \omega_2]$ $+ \frac{525 A^{17} f^2 g \sin[2 \omega_1 - \omega_2]}{8192}$ $+ \frac{945 A^{27} f^4 g \sin[2 \omega_1 - \omega_2]}{16777216}$	$+ \frac{245}{64} A^7 g \sin[3 \omega_1 - 2 \omega_2]$ $+ \frac{105 A^{17} f^2 g \sin[3 \omega_1 - 2 \omega_2]}{8192}$	$+ \frac{35}{64} A^7 g \sin[4 \omega_1 - 3 \omega_2]$ $+ \frac{105 A^{17} f^2 g \sin[4 \omega_1 - 3 \omega_2]}{16384}$

$\omega_1 - 5 \omega_2 \text{ from } 6^{\text{th}}$ $- \frac{3}{16} A^6 f \cos[\omega_1 - 5 \omega_2]$	IMD3	IMD5	IMD7
2 nd			
3 rd	$+ \frac{3}{4} A^3 c \sin[2 \omega_1 - \omega_2]$		
4 th			
5 th	$+ \frac{25}{8} A^5 e \sin[2 \omega_1 - \omega_2]$ $+ \frac{135 A^{15} e f^2 \sin[2 \omega_1 - \omega_2]}{1024}$	$+ \frac{5}{8} A^5 e \sin[3 \omega_1 - 2 \omega_2]$	
6 th	$- \frac{45}{256} A^{11} f^2 \sin[2 \omega_1 - \omega_2]$	$- \frac{45}{128} A^{11} f^2 \sin[3 \omega_1 - 2 \omega_2]$	$- \frac{45}{128} A^{11} f^2 \sin[4 \omega_1 - 3 \omega_2]$
7 th	$+ \frac{735}{64} A^7 g \sin[2 \omega_1 - \omega_2]$ $+ \frac{4725 A^{17} f^2 g \sin[2 \omega_1 - \omega_2]}{4096}$ $+ \frac{25515 A^{27} f^4 g \sin[2 \omega_1 - \omega_2]}{2097152}$	$+ \frac{245}{64} A^7 g \sin[3 \omega_1 - 2 \omega_2]$ $+ \frac{945 A^{17} f^2 g \sin[3 \omega_1 - 2 \omega_2]}{4096}$	$+ \frac{35}{64} A^7 g \sin[4 \omega_1 - 3 \omega_2]$

$\omega_1 + 5 \omega_2 \text{ from } 6^{\text{th}}$ $+ \frac{3}{16} A^6 f \cos[\omega_1 + 5 \omega_2]$	IMD3	IMD5	IMD7
2 nd			
3 rd	$+ \frac{3}{4} A^3 c \sin[2 \omega_1 - \omega_2]$		
4 th			
5 th	$+ \frac{25}{8} A^5 e \sin[2 \omega_1 - \omega_2]$ $+ \frac{135 A^{15} e f^2 \sin[2 \omega_1 - \omega_2]}{1024}$	$+ \frac{5}{8} A^5 e \sin[3 \omega_1 - 2 \omega_2]$	
6 th			
7 th	$+ \frac{735}{64} A^7 g \sin[2 \omega_1 - \omega_2]$ $+ \frac{4725 A^{17} f^2 g \sin[2 \omega_1 - \omega_2]}{4096}$ $+ \frac{25515 A^{27} f^4 g \sin[2 \omega_1 - \omega_2]}{2097152}$	$+ \frac{245}{64} A^7 g \sin[3 \omega_1 - 2 \omega_2]$ $+ \frac{945 A^{17} f^2 g \sin[3 \omega_1 - 2 \omega_2]}{4096}$	$+ \frac{35}{64} A^7 g \sin[4 \omega_1 - 3 \omega_2]$