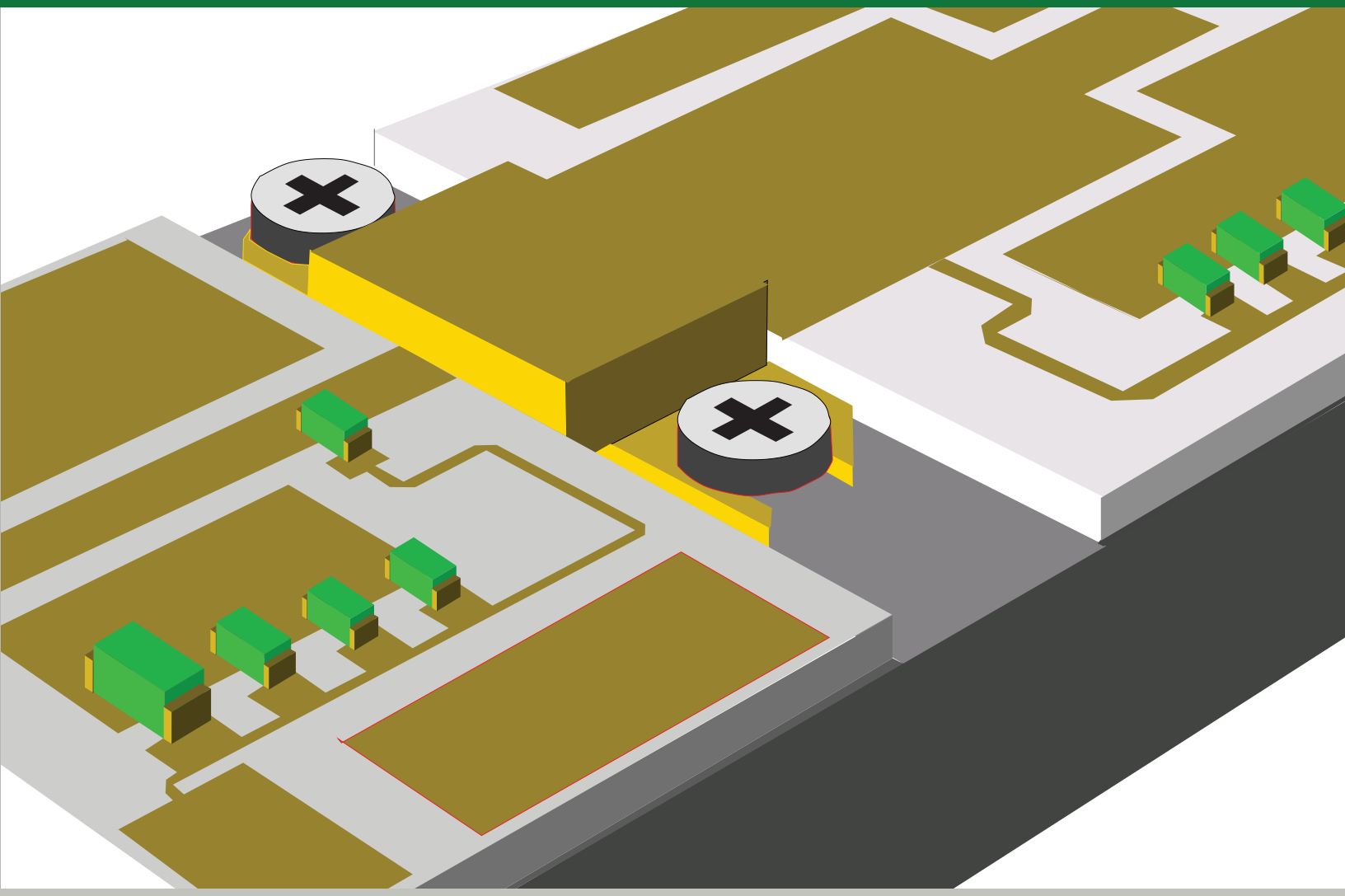
Volume 5

MICROWAVE AND RF DESIGN

AMPLIFIERS AND OSCILLATORS



Michael Steer

Third Edition

Microwave and RF Design
Amplifiers and Oscillators

Volume 5

Third Edition

Michael Steer

Microwave and RF Design
Amplifiers and Oscillators

Volume 5

Third Edition

Michael Steer

Citation: Steer, Michael. *Microwave and RF Design: Amplifiers and Oscillators*. Volume 5. (Third Edition), NC State University, 2019. doi: https//doi.org/10.5149/
9781469656991 Steer

ISBN 978-1-4696-5698-4 (paperback)
ISBN 978-1-4696-5699-1 (open access ebook)

Published by NC State University

NC STATE UNIVERSITY

Distributed by the University of North Carolina Press
www.uncpress.org

Printing: 1

To my Mother

Mary Elizabeth

Preface

The book series *Microwave and RF Design* is a comprehensive treatment of radio frequency (RF) and microwave design with a modern "systems-first" approach. A strong emphasis on design permeates the series with extensive case studies and design examples. Design is oriented towards cellular communications and microstrip design so that lessons learned can be applied to real-world design tasks. The books in the Microwave and RF Design series are:

- Microwave and RF Design: Radio Systems, Volume 1
- Microwave and RF Design: Transmission Lines, Volume 2
- Microwave and RF Design: Networks, Volume 3
- Microwave and RF Design: Modules, Volume 4
- Microwave and RF Design: Amplifiers and Oscillators, Volume 5

The length and format of each is suitable for automatic printing and binding.

Rationale

The central philosophy behind this series's popular approach is that the student or practicing engineer will develop a full appreciation for RF and microwave engineering and gain the practical skills to perform system-level design decisions. Now more than ever companies need engineers with an ingrained appreciation of systems and armed with the skills to make system decisions. One of the greatest challenges facing RF and microwave engineering is the increasing level of abstraction needed to create innovative microwave and RF systems. This book series is organized in such a way that the reader comes to understand the impact that system-level decisions have on component and subsystem design. At the same time, the capabilities of technologies, components, and subsystems impact system design. The book series is meticulously crafted to intertwine these themes.

Audience

The book series was originally developed for three courses at North Carolina State University. One is a final-year undergraduate class, another an introductory graduate class, and the third an advanced graduate class. Books in the series are used as supplementary texts in two other classes. There are extensive case studies, examples, and end of chapter problems ranging from straight-forward to in-depth problems requiring hours to solve. A companion book, *Fundamentals of Microwave and RF Design*, is more suitable for an undergraduate class yet there is a direct linkage between the material in this book and the series which can then be used as a career-long reference text. I believe it is completely understandable for senior-level students where a microwave/RF engineering course is offered. The book series is a comprehensive RF and microwave text and reference, with detailed index, appendices, and cross-references throughout. Practicing engineers will find the book series a valuable systems primer, a refresher as needed, and a

reference tool in the field. Additionally, it can serve as a valuable, accessible resource for those outside RF circuit engineering who need to understand how they can work with RF hardware engineers.

Organization

This book is a volume in a five volume series on RF and microwave design. The first volume in the series, *Microwave and RF Design: Radio Systems*, addresses radio systems mainly following the evolution of cellular radio. A central aspect of microwave engineering is distributed effects considered in the second volume of this book series, *Microwave and RF Design: Transmission Lines*. Here transmission lines are treated as supporting forward- and backward-traveling voltage and current waves and these are related to electromagnetic effects. The third volume, *Microwave and RF Design: Networks*, covers microwave network theory which is the theory that describes power flow and can be used with transmission line effects. Topics covered in *Microwave and RF Design: Modules*, focus on designing microwave circuits and systems using modules introducing a large number of different modules. Modules is just another term for a network but the implication is that is is packaged and often available off-the-shelf. Other topics that are important in system design using modules are considered including noise, distortion, and dynamic range. Most microwave and RF designers construct systems using modules developed by other engineers who specialize in developing the modules. Examples are filter and amplifier modules which once designed can be used in many different systems. Much of microwave design is about maximizing dynamic range, minimizing noise, and minimizing DC power consumption. The fifth volume in this series, *Microwave and RF Design: Amplifiers and Oscillators*, considers amplifier and oscillator design and develops the skills required to develop modules.

Volume 1: Microwave and RF Design: Radio Systems

The first book of the series covers RF systems. It describes system concepts and provides comprehensive knowledge of RF and microwave systems. The emphasis is on understanding how systems are crafted from many different technologies and concepts. The reader gains valuable insight into how different technologies can be traded off in meeting system requirements. I do not believe this systems presentation is available anywhere else in such a compact form.

Volume 2: Microwave and RF Design: Transmission Lines

This book begins with a chapter on transmission line theory and introduces the concepts of forward- and backward-traveling waves. Many examples are included of advanced techniques for analyzing and designing transmission line networks. This is followed by a chapter on planar transmission lines with microstrip lines primarily used in design examples. Design examples illustrate some of the less quantifiable design decisions that must be made. The next chapter describes frequency-dependent transmission line effects and describes the design choices that must be taken to avoid multimoding. The final chapter in this volume addresses coupled-lines. It is shown how to design coupled-line networks that exploit this distributed effect to realize novel circuit functionality and how to design networks that minimize negative effects. The modern treatment of transmission lines in this volume emphasizes planar circuit design and the practical aspects of designing

around unwanted effects. Detailed design of a directional coupler is used to illustrate the use of coupled lines. Network equivalents of coupled lines are introduced as fundamental building blocks that are used later in the synthesis of coupled-line filters. The text, examples, and problems introduce the often hidden design requirements of designing to mitigate parasitic effects and unwanted modes of operation.

Volume 3: Microwave and RF Design: Networks

Volume 3 focuses on microwave networks with descriptions based on S parameters and $ABCD$ matrices, and the representation of reflection and transmission information on polar plots called Smith charts. Microwave measurement and calibration technology are examined. A sampling of the wide variety of microwave elements based on transmission lines is presented. It is shown how many of these have lumped-element equivalents and how lumped elements and transmission lines can be combined as a compromise between the high performance of transmission line structures and the compactness of lumped elements. This volume concludes with an in-depth treatment of matching for maximum power transfer. Both lumped-element and distributed-element matching are presented.

Volume 4: Microwave and RF Design: Modules

Volume 4 focuses on the design of systems based on microwave modules. The book considers the wide variety of RF modules including amplifiers, local oscillators, switches, circulators, isolators, phase detectors, frequency multipliers and dividers, phase-locked loops, and direct digital synthesizers. The use of modules has become increasingly important in RF and microwave engineering. A wide variety of passive and active modules are available and high-performance systems can be realized cost effectively and with stellar performance by using off-the-shelf modules interconnected using planar transmission lines. Module vendors are encouraged by the market to develop competitive modules that can be used in a wide variety of applications. The great majority of RF and microwave engineers either develop modules or use modules to realize RF systems. Systems must also be concerned with noise and distortion, including distortion that originates in supposedly linear elements. Something as simple as a termination can produce distortion called passive intermodulation distortion. Design techniques are presented for designing cascaded systems while managing noise and distortion. Filters are also modules and general filter theory is covered and the design of parallel coupled line filters is presented in detail. Filter design is presented as a mixture of art and science. This mix, and the thought processes involved, are emphasized through the design of a filter integrated throughout this chapter.

Volume 5: Microwave and RF Design: Amplifiers and Oscillators

The fifth volume presents the design of amplifiers and oscillators in a way that enables state-of-the-art designs to be developed. Detailed strategies for amplifiers and voltage-controlled oscillators are presented. Design of competitive microwave amplifiers and oscillators are particularly challenging as many trade-offs are required in design, and the design decisions cannot be reduced to a formulaic flow. Very detailed case studies are presented and while some may seem quite complicated, they parallel the level of sophistication required to develop competitive designs.

Case Studies

A key feature of this book series is the use of real world case studies of leading edge designs. Some of the case studies are designs done in my research group to demonstrate design techniques resulting in leading performance. The case studies and the persons responsible for helping to develop them are as follows.

1. Software defined radio transmitter.
2. High dynamic range down converter design. This case study was developed with Alan Victor.
3. Design of a third-order Chebyshev combline filter. This case study was developed with Wael Fathelbab.
4. Design of a bandstop filter. This case study was developed with Wael Fathelbab.
5. Tunable Resonator with a varactor diode stack. This case study was developed with Alan Victor.
6. Analysis of a 15 GHz Receiver. This case study was developed with Alan Victor.
7. Transceiver Architecture. This case study was developed with Alan Victor.
8. Narrowband linear amplifier design. This case study was developed with Dane Collins and National Instruments Corporation.
9. Wideband Amplifier Design. This case study was developed with Dane Collins and National Instruments Corporation.
10. Distributed biasing of differential amplifiers. This case study was developed with Wael Fathelbab.
11. Analysis of a distributed amplifier. This case study was developed with Ratan Bhatia, Jason Gerber, Tony Kwan, and Rowan Gilmore.
12. Design of a WiMAX power amplifier. This case study was developed with Dane Collins and National Instruments Corporation.
13. Reflection oscillator. This case study was developed with Dane Collins and National Instruments Corporation.
14. Design of a C-Band VCO. This case study was developed with Alan Victor.
15. Oscillator phase noise analysis. This case study was developed with Dane Collins and National Instruments Corporation.

Many of these case studies are available as captioned YouTube videos and qualified instructors can request higher resolution videos from the author.

Course Structures

Based on the adoption of the first and second editions at universities, several different university courses have been developed using various parts of what was originally one very large book. The book supports teaching two or three classes with courses varying by the selection of volumes and chapters. A standard microwave class following the format of earlier microwave texts can be taught using the second and third volumes. Such a course will benefit from the strong practical design flavor and modern treatment of measurement technology, Smith charts, and matching networks. Transmission line propagation and design is presented in the context of microstrip technology providing an immediately useful skill. The subtleties of multimoding are also presented in the context of microstrip lines. In such

a class the first volume on microwave systems can be assigned for self-learning.

Another approach is to teach a course that focuses on transmission line effects including parallel coupled-line filters and module design. Such a class would focus on Volumes 2, 3 and 4. A filter design course would focus on using Volume 4 on module design. A course on amplifier and oscillator design would use Volume 5. This course is supported by a large number of case studies that present design concepts that would otherwise be difficult to put into the flow of the textbook.

Another option suited to an undergraduate or introductory graduate class is to teach a class that enables engineers to develop RF and microwave systems. This class uses portions of Volumes 2, 3 and 4. This class then omits detailed filter, amplifier, and oscillator design.

The fundamental philosophy behind the book series is that the broader impact of the material should be presented first. Systems should be discussed up front and not left as an afterthought for the final chapter of a textbook, the last lecture of the semester, or the last course of a curriculum.

The book series is written so that all electrical engineers can gain an appreciation of RF and microwave hardware engineering. The body of the text can be covered without strong reliance on this electromagnetic theory, but it is there for those who desire it for teaching or reader review. The book is rich with detailed information and also serves as a technical reference.

The Systems Engineer

Systems are developed beginning with fuzzy requirements for components and subsystems. Just as system requirements provide impetus to develop new base technologies, the development of new technologies provides new capabilities that drive innovation and new systems. The new capabilities may arise from developments made in support of other systems. Sometimes serendipity leads to the new capabilities. Creating innovative microwave and RF systems that address market needs or provide for new opportunities is the most exciting challenge in RF design. The engineers who can conceptualize and architect new RF systems are in great demand. This book began as an effort to train RF systems engineers and as an RF systems resource for practicing engineers. Many RF systems engineers began their careers when systems were simple. Today, appreciating a system requires higher levels of abstraction than in the past, but it also requires detailed knowledge or the ability to access detailed knowledge and expertise. So what makes a systems engineer? There is not a simple answer, but many partial answers. We know that system engineers have great technical confidence and broad appreciation for technologies. They are both broad in their knowledge of a large swath of technologies and also deep in knowledge of a few areas, sometimes called the “T” model. One book or course will not make a systems engineer. It is clear that there must be a diverse set of experiences. This book series fulfills the role of fostering both high-level abstraction of RF engineering and also detailed design skills to realize effective RF and microwave modules. My hope is that this book will provide the necessary background for the next generation of RF systems engineers by stressing system principles immediately, followed by core RF technologies. Core technologies are thereby covered within the context of the systems in which they are used.

Supplementary Materials

Supplementary materials available to qualified instructors adopting the book include PowerPoint slides and solutions to the end-of-chapter problems. Requests should be directed to the author. Access to downloads of the books, additional material and YouTube videos of many case studies are available at `https://www.lib.ncsu.edu/do/open-education`

Acknowledgments

Writing this book has been a large task and I am indebted to the many people who helped along the way. First I want to thank the more than 1200 electrical engineering graduate students who used drafts and the first two editions at NC State. I thank the many instructors and students who have provided feedback. I particularly thank Dr. Wael Fathelbab, a filter expert, who co-wrote an early version of the filter chapter. Professor Andreas Cangellaris helped in developing the early structure of the book. Many people have reviewed the book and provided suggestions. I thank input on the structure of the manuscript: Professors Mark Wharton and Nuno Carvalho of Universidade de Aveiro, Professors Ed Delp and Saul Gelfand of Purdue University, Professor Lynn Carpenter of Pennsylvania State University, Professor Grant Ellis of the Universiti Teknologi Petronas, Professor Islam Eshrah of Cairo University, Professor Mohammad Essaaidi and Dr. Otman Aghzout of Abdelmalek Essaadi Univeristy, Professor Jianguo Ma of Guangdong University of Technology, Dr. Jayesh Nath of Apple, Mr. Sony Rowland of the U.S. Navy, and Dr. Jonathan Wilkerson of Lawrence Livermore National Laboratories, Dr. Josh Wetherington of Vadum, Dr. Glen Garner of Vadum, and Mr. Justin Lowry who graduated from North Carolina State University.

Many people helped in producing this book. In the first edition I was assisted by Ms. Claire Sideri, Ms. Susan Manning, and Mr. Robert Lawless who assisted in layout and production. The publisher, task master, and chief coordinator, Mr. Dudley Kay, provided focus and tremendous assistance in developing the first and second editions of the book, collecting feedback from many instructors and reviewers. I thank the Institution of Engineering and Technology, who acquired the original publisher, for returning the copyright to me. This open access book was facilitated by John McLeod and Samuel Dalzell of the University of North Carolina Press, and by Micah Vandergrift and William Cross of NC State University Libraries. The open access ebooks are host by NC State University Libraries.

The book was produced using LaTeX and open access fonts, line art was drawn using xfig and inkscape, and images were edited in gimp. So thanks to the many volunteers who developed these packages.

My family, Mary, Cormac, Fiona, and Killian, gracefully put up with my absence for innumerable nights and weekends, many more than I could have ever imagined. I truly thank them. I also thank my academic sponsor, Dr. Ross Lampe, Jr., whose support of the university and its mission enabled me to pursue high risk and high reward endeavors including this book.

Michael Steer
North Carolina State University
Raleigh, North Carolina
mbs@ncsu.edu

List of Trademarks

Contents

Preface v

1 Introduction to Active RF and Microwave Circuits 1
1.1 Introduction to Amplifiers and Oscillators 1
1.2 Book Outline 3
1.3 Transistor Technology 3
1.3.1 BJT and HBT Fundamentals 4
1.3.2 MOSFET Fundamentals 6
1.3.3 MESFET, HEMT, and JFET Fundamentals 10
1.4 References 13
1.A Active Device Models 14
1.A.1 Level 3 MOSFET Model 14
1.A.2 Materka–Kacprzak MESFET Model 19
1.A.3 Gummel–Poon: Bipolar Junction Transistor Model . . 20

2 Linear Amplifiers 25
2.1 Introduction 25
2.2 Linear Amplifier Design Strategies 26
2.3 Amplifier Gain Definitions 26
2.3.1 Gain in Terms of Scattering Parameters 30
2.3.2 Design Using Gain Metrics 34
2.3.3 Gain Circles 35
2.4 Amplifier Efficiency 37
2.5 Class A, AB, B, and C Amplifiers 38
2.5.1 Class A Amplifier 40
2.5.2 Amplifier Efficiency 42
2.6 Amplifier Stability 44
2.6.1 Two-Port Stability Analysis 45
2.6.2 Unconditional Stability: Two-Port Stability Circles . . 46
2.6.3 Rollet's Stability Criterion — k-factor 50
2.6.4 Edwards–Sinsky Stability Criterion — μ-factor 52
2.6.5 Nyquist Stability Criterion 53
2.6.6 Summary 55
2.7 Amplifier Noise 55
2.8 Trading Off Gain, Noise, and Stability in Amplifier Design . . 56
2.9 Case Study: Narrowband Linear Amplifier Design 57
2.9.1 Bias Circuit Topology 58
2.9.2 Stability Considerations 58
2.9.3 Output Matching Network Design 58
2.9.4 Input Matching Network Design 61
2.9.5 Bias Network Design 63
2.10 Summary 63
2.11 References 64

2.12 Exercises . . . 65
2.12.1 Exercises By Section . . . 68
2.12.2 Answers to Selected Exercises . . . 68

3 Wideband Amplifiers . . . 69
3.1 Introduction . . . 69
3.1.1 Wideband Amplifier Design Strategies . . . 70
3.2 Distributed Amplifiers . . . 71
3.3 Case Study: Analysis of a Distributed Amplifier . . . 72
3.4 Negative Image Amplifier Design . . . 74
3.5 Case Study: Wideband Amplifier Design . . . 77
3.5.1 Transistor Properties . . . 77
3.5.2 Negative Image Design . . . 81
3.5.3 Final Design . . . 84
3.6 Differential Amplifiers . . . 86
3.6.1 Fully and Pseudo-Differential Amplifiers . . . 86
3.6.2 Even, Common, Odd, and Differential Modes . . . 91
3.6.3 Asymmetrical Loading . . . 92
3.6.4 Hybrids and Differential Amplifiers . . . 94
3.7 Case Study: Distributed Biasing of a Pseudo-Differential Amplifier . . . 96
3.8 Amplifiers and RFICs . . . 97
3.9 Summary . . . 101
3.10 References . . . 102
3.11 Exercises . . . 103
3.11.1 Exercises By Section . . . 107
3.11.2 Answers to Selected Exercises . . . 107

4 Power Amplifiers . . . 109
4.1 Introduction . . . 109
4.2 Simulation of Nonlinear Microwave Circuits . . . 110
4.2.1 Harmonic Balance Analysis of RF Circuits . . . 111
4.2.2 Example: Harmonic Balance Analysis of a Simple Circuit 112
4.2.3 User's Guide to Using Harmonic Balance Analysis . . . 115
4.2.4 Periodic Steady-State Simulation of RF Circuits . . . 116
4.3 Switching Amplifiers, Classes D, E, and F . . . 116
4.3.1 Dynamic Waveforms . . . 117
4.3.2 Conduction Angle . . . 118
4.3.3 Class D . . . 119
4.3.4 Class E . . . 120
4.3.5 Class F . . . 121
4.3.6 Inverted Amplifiers . . . 122
4.3.7 Summary . . . 122
4.4 Distortion and Digitally Modulated Signals . . . 122
4.4.1 PMEPR and Probability Density Function . . . 122
4.4.2 Design Guidelines . . . 124
4.5 Loadpull . . . 127
4.6 Case Study: Design of a WiMAX Power Amplifier . . . 128
4.6.1 Input and Output Matching Networks . . . 128
4.6.2 Load-Pull . . . 134
4.6.3 Two-Tone Characterization . . . 135

4.6.4 Summary . . . 136
4.7 Linearization . . . 136
4.7.1 Predistortion . . . 136
4.7.2 Feed-Forward Linearization . . . 137
4.7.3 Summary . . . 137
4.8 Advanced Power Amplifier Architectures . . . 138
4.8.1 Doherty Amplifier . . . 138
4.8.2 Envelope Tracking Amplifier . . . 139
4.8.3 LINC Amplifier . . . 139
4.8.4 LITMUS Amplifier . . . 140
4.8.5 Summary . . . 142
4.9 MMIC Power Amplifiers . . . 142
4.10 RFIC Power Amplifiers . . . 144
4.10.1 Distortion in a MOSFET Enhancement-Depletion Amplifier Stage . . . 144
4.10.2 Distortion in the Ultralinear MOS Connection . . . 147
4.10.3 RFIC Power Amplifiers with Minimal Distortion . . . 148
4.11 Summary . . . 151
4.12 References . . . 152
4.13 Exercises . . . 155
4.13.1 Exercises By Section . . . 158
4.13.2 Answers to Selected Exercises . . . 158

5 Oscillators . . . 159
5.1 Introduction . . . 159
5.2 Oscillator Theory . . . 160
5.2.1 Theory of Oscillation . . . 161
5.2.2 Basic Oscillator Configurations . . . 161
5.3 Reflection Oscillators . . . 165
5.3.1 Kurokawa Oscillation Condition . . . 165
5.3.2 Reflection Oscillator Design Approach . . . 166
5.3.3 Summary . . . 167
5.4 Case Study: Reflection Oscillator . . . 167
5.4.1 Design Procedure . . . 167
5.4.2 Summary . . . 171
5.5 Voltage-Controlled Oscillator (VCO) . . . 172
5.5.1 Design Procedure . . . 172
5.5.2 Managing Multioscillation and Phase Noise . . . 174
5.5.3 Negative Resistance Oscillator . . . 175
5.5.4 Summary . . . 176
5.6 Case Study: Design of a C-Band VCO . . . 176
5.6.1 Design Philosophy and Topology . . . 176
5.6.2 Design Strategy . . . 181
5.6.3 Oscillator Start-Up Considerations . . . 181
5.6.4 Initial Design . . . 182
5.6.5 Avoiding Multiple Oscillations Through Reflection Coefficient Shaping . . . 184
5.6.6 VCO Performance . . . 189
5.6.7 Summary . . . 190
5.7 Negative Transconductance Differential Oscillator . . . 190
5.8 Advanced Discussion of Oscillator Noise . . . 194

5.8.1 Observations of Oscillator Noise in the Frequency Domain . . . 194
5.8.2 Observations of Oscillator Noise in the Time Domain . 197
5.8.3 Excess Oscillator Noise: The Leeson Effect and Flicker Noise . . . 198
5.8.4 Excess Oscillator Noise: Linear Time-Variant Model . . 199
5.8.5 Excess Oscillator Noise: Chaotic Maps and Flicker Noise 201
5.8.6 Summary . . . 204
5.9 Case Study: Oscillator Phase Noise Analysis . . . 206
5.10 Summary . . . 209
5.11 References . . . 210
5.12 Exercises . . . 212
5.12.1 Exercises by Section . . . 216
5.12.2 Answers to Selected Exercises . . . 216

Index . . . 217

CHAPTER 1

Introduction to Active RF and Microwave Circuits

1.1 Introduction to Amplifiers and Oscillators 1
1.2 Book Outline 3
1.3 Transistor Technology 3
1.4 References 13
1.A Appendix: Active Device Models 14

1.1 Introduction to Amplifiers and Oscillators

Design of microwave amplifiers and oscillators is the most challenging of microwave designs determining the performance and DC power consumption of microwave systems. Most of the challenges arise because of the capacitive parasitics of microwave transistors and also because some types of transistors, such as silicon transistors, have quite low intrinsic power gain. Packaged transistors which are used in hybrid design and design using modules have the additional complexity of package inductance as well as additional capacitance from the package. With amplifiers one of the great challenges is achieving wide bandwidth so that the same amplifier can be used for multiple frequency bands. For example, with cellular communications it is desirable if one amplifier could be used for multiple cellular bands. However most of the time a cellular system handset must have different transmit and receive amplifiers for each of the cellular bands. The transistor's capacitive, and if packaged inductive, parasitics determine the minimum Q and thus maximum bandwidth. The matching required to interface the input and the of an amplifier output of transistors can only further reduce bandwidth.

Microwave amplifier design generally uses the topology shown in Figure 1-1 with the transistor biased in a high-gain region and the input and output matching networks used to provide good power transfer at the input and output of the transistors. The DC bias control circuit is fairly standard; it does not involve any microwave constraints other than the need to block RF currents from the bias circuit. The lowpass filters (in the bias circuits) can have one of several forms and are often integrated into the input and output matching networks. Synthesis of the input and output matching networks (and occasionally a feedback network required for stability and broadband operation) is the primary objective of any amplifier design.

Design of linear microwave amplifiers where narrowband operation is

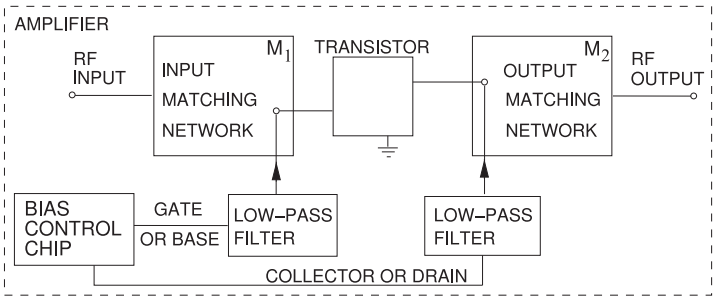


Figure 1-1: Block diagram of an RF amplifier including biasing networks.

sufficient is considered in Chapter 2 where the topology shown in Figure 1-1 is nearly always followed. The input and output matching networks limit the bandwidth of the amplifier and are ideally lossless. This chapter develops the skills required to trade off gain, noise, and stability. These trade-offs are required with all types of microwave amplifier design. The chapter presents a case study of narrowband linear amplifier design.

Chapter 3 presents strategies for designing a wideband amplifier and again the topology shown in Figure 1-1 is usually followed. Wideband is still limited as usually the best that can be achieved for an efficient amplifier is a bandwidth of only half-an-octave such as a bandwidth from 2 to 3 GHz. Sometimes there is inductive and capacitive feedback around the transistor to compensate for the inherent gain roll-off with respect to frequency of transistors, especially FETs. The chapter includes a case study on the design of a wideband amplifier. A case study is a good way to present design methods as it enables design decisions to be discussed. Rarely can design decisions be reduced to a formulaic flow. One of the important trade-offs is trading off gain and noise performance and in the case study it will be seen how this can be done graphically. A distributed amplifier is one type of amplifiers that has a very wide bandwidth, perhaps 2–4 octaves, e.g. 2 to 4 GHz or 2 to 16 GHz, but has an efficiency of only a few percent. The topology differs significantly from the input and output matching network–based topology of Figure 1-1. The distributed amplifier achieves wide bandwidth by incorporating the parasitics of multiple transistors in a transmission line where the input and output capacitances of transistors augment the capacitances in the LC model of an actual transmission line. A case study is presented that analyses a distributed amplifier. Efficient biasing of a wideband amplifier can be a challenge and a final case study presents a technique for distributed biasing of a differential amplifier.

The fourth chapter considers power amplifiers where the emphasis is on producing large powers at high efficiency and bandwidth is sacrificed. Usually at high powers efficiencies are achieved by engineering the transistor's current and voltage waveforms by manipulating the impedances presented at harmonics. This is the source of the low bandwidth. A case study of the design of a WiMAX power amplifier is undertaken.

The final chapter of this book considers the design of microwave oscillators. Microwave oscillator design is particularly challenging. Oscillators consume considerable DC power and are a competitive differentiator. There are two quite different classes of design, one for oscillators that are fixed in frequency and one for the much more useful voltage-controlled oscillator which has variable frequency. Case studies are presented for each of these two types of oscillators.

The remainder of this current chapter describes transistor technology and an appendix describes particular models of transistors that are used in circuit simulators. Today's simulators use transistor models that are very sophisticated compared to those models described in the appendix, however they are not amenable to developing a designer's intuition. The simpler models considered in the appendix, state-of-the-art models from 20 and 30 years ago, provide the desired intuition for a designer.

1.2 Book Outline

This book is the fifth volume in a series on microwave and RF design. The first volume in the series addresses radio systems [1] mainly following the evolution of cellular radio. A central aspect of microwave engineering is distributed effects considered in the second volume of his book series [2]. Here transmission lines are treated as supporting forward- and backward-traveling voltage and current waves and these are related to electromagnetic effects. The third volume [3] covers microwave network theory which is the theory that describes power flow and can be used with transmission line effects. Topics covered in this volume include scattering parameters, Smith charts, and matching networks that enable maximum power transfer. The fourth volume [4] focuses on designing microwave circuits and systems using modules introducing a large number of different modules. Modules is just another term for a network but the implication is that is is packaged and often available off-the-shelf. Other topics in this chapter that are important in system design using modules are considered including noise, distortion, and dynamic range. Most microwave and RF designers construct systems using modules developed by other engineers who specialize in developing the modules. Examples are filter and amplifier chip modules which once designed can be used in many different systems. Much of microwave design is about maximizing dynamic range, minimizing noise, and minimizing DC power consumption.

The books in the Microwave and RF Design series are:

- Microwave and RF Design: Radio Systems
- Microwave and RF Design: Transmission Lines
- Microwave and RF Design: Networks
- Microwave and RF Design: Modules
- Microwave and RF Design: Amplifiers and Oscillators

1.3 Transistor Technology

Transistors are semiconductor devices with three (and sometimes more) terminals. The third terminal enables output current to be controlled by a relatively small and low-power input signal. In amplifiers, transistors are used to achieve current gain, voltage gain, or power gain. Most often power gain is the objective in RF and microwave design. Most transistors are fabricated using silicon (Si) or **compound semiconductors** such as **gallium-arsenide** (**GaAs**), **indium phosphide** (**InP**), or **gallium-nitride** (**GaN**). The overwhelming trend is to use silicon technology because of the much higher integration density that is possible, with compound semiconductor technology used only when it provides a unique advantage such as high power, superior noise performance, or high efficiency. Germanium is used

as a dopant in silicon and then silicon is referred to as silicon germanium but usually germanium is in a very small proportion to silicon so SiGe as described here is silicon with a dopant. With comparable concentrations of silicon and germanium SiGe is a compound semiconductor and this is used as a compound semiconductor at times.

There are three fundamental types of microwave transistors [5, 6]: **bipolar junction transistors**, (**BJTs**); **junction field effect transistors**, (**JFETs**); and **insulated gate FETs**, (**IGFETs**), with the **metal-oxide-semiconductor FETs**, (**MOSFETs**), being the most common type of IGFET. The schematics and terminal definitions of the three fundamental types of transistors are shown in Figure 1-2. The three fundamental types of transistors are considered in the following subsections.

1.3.1 BJT and HBT Fundamentals

A bipolar transistor has three semiconductor regions called the collector (C), base (B), and emitter (E), as shown in the BJT cross section of Figure 1-3(a). An npn BJT has n-type semiconductor at the emitter and collector, and p-type semiconductor forms the base. In this transistor, the positive sense of current flow is from the collector through the base to the emitter (see Figure 1-3(a)) and the dominant carriers in the p-type base region are electrons, and so this is called a **minority carrier device**. The collector current is dependent on the number of carriers injected into the base region from the base terminal. In a pnp BJT the collector, base, and emitter are p-type, n-type, and p-type, respectively, and the majority carriers in the base are holes. Current flow is then from the emitter through the base to the collector. If the base region is thin and the emitter is doped at a higher concentration than the base, then the collector current, I_C, is much greater than I_B, with $I_C = \beta_F I_B$, where β_F is called the forward current gain and commonly has a value of several hundred. The key to high performance is a thin base region.

When realized in silicon, a bipolar transistor is called a bipolar junction transistor, (BJT); and in compound semiconductor technology it is a **heterostructure bipolar transistor**, (**HBT**). In a silicon germanium (SiGe) BJT transistor, germanium is normally used to increase the hole and electron mobility and the device is not regarded as a compound semiconductor transistor.

The fundamental operation of a BJT transistor was described by Gummel and Poon [7] using equations that are now implemented in circuit simulators and known as the Gummel–Poon model. The circuit schematic of the Gummel–Poon model is shown in Figure 1-3(c). It is the basis for more

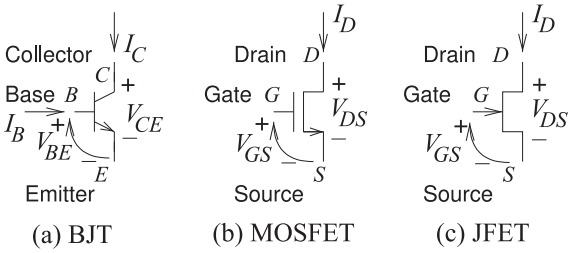
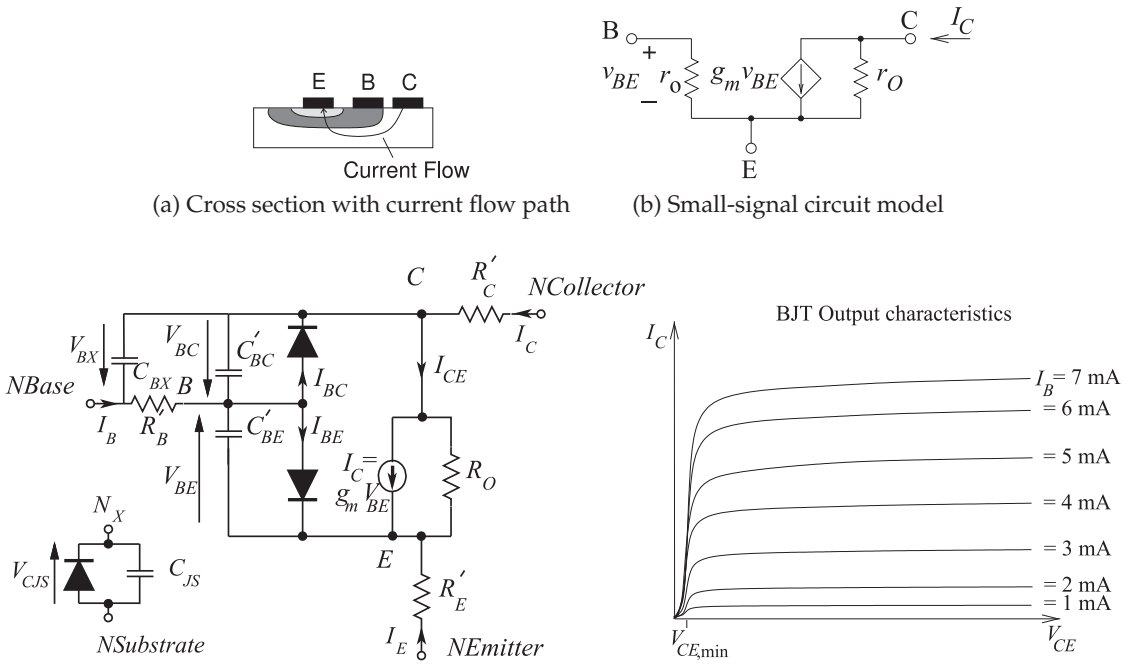


Figure 1-2: Transistor schematics: (a) pnp bipolar transistor with B for the base terminal, C for the collector terminal, and E for the emitter terminal; (b) n-type MOSFET (nMOS); and (c) n-type JFET (nJFET) with G for the gate terminal, D for the drain terminal, and S for the source terminal. The schematic symbol for a BJT is used for HBTs; and the schematic symbol for a JFET is used for MESFETs, HEMTS, and pHEMTs.



(a) Cross section with current flow path (b) Small-signal circuit model

(c) Gummel–Poon model schematic (d) Output i-v characteristic

Figure 1-3: BJT details.

sophisticated BJT and HBT models that capture parasitic and other second-order effects. Both hole and electron charge carriers are involved in current conduction, hence the term bipolar. The Gummel–Poon model is described in Section 1.A.3 and the fundamental operation is described by Equations (1.72)–(1.82). Summarizing, the base-emitter current is

$$I_{BE} = I_{BF}/\beta_F + I_{LE} \tag{1.1}$$

and the base-collector current is

$$I_{BC} = I_{BR}/\beta_R + I_{LC}, \tag{1.2}$$

where β_R is the reverse current gain. The collector-emitter current is

$$I_{CE} = I_{BF} - I_{BR}/K_{QB}. \tag{1.3}$$

The forward diffusion current is

$$I_{BF} = I_S \left(\mathrm{e}^{V_{BE}/(N_F V_{TH})} - 1 \right), \tag{1.4}$$

the nonideal base-emitter current is

$$I_{LE} = I_{SE} \left(\mathrm{e}^{V_{BE}/(N_E V_{TH})} - 1 \right), \tag{1.5}$$

the reverse diffusion current is

$$I_{BR} = I_S \left(\mathrm{e}^{V_{BC}/(N_R V_{TH})} - 1 \right), \tag{1.6}$$

the nonideal base-collector current is

$$I_{LC} = I_{SC}\left(\mathrm{e}^{V_{BC}/(N_C V_{TH})} - 1\right), \tag{1.7}$$

and the base charge factor is

$$K_{QB} = \frac{1}{2}\left[1 - \frac{V_{BC}}{V_{AF}} - \frac{V_{BE}}{V_{AB}}\right]^{-1}\left[1 + \sqrt{1 + 4\left(\frac{I_{BF}}{I_{KF}} + \frac{I_{BR}}{I_{KR}}\right)}\right]. \tag{1.8}$$

Thus the conductive current flowing into the base is

$$I_B = I_{BE} + I_{BC}, \tag{1.9}$$

the conductive current flowing into the collector is

$$I_C = I_{CE} - I_{BC}, \tag{1.10}$$

and the conductive current flowing into the emitter is

$$I_E = I_{BE} + I_{CE}. \tag{1.11}$$

The forward current gain, β_F, is much greater than the reverse current gain, β_R, and the nonideal base-emitter and base-collector currents are small. Equations (1.1)–(1.11) can then be reduced so that the base current is approximately

$$I_B = \frac{I_S}{\beta_F}\left(\mathrm{e}^{V_{BE}/(N_F V_{TH})} - 1\right), \tag{1.12}$$

the conductive current flowing into the collector is

$$I_C = \beta_F I_B, \tag{1.13}$$

and the conductive current flowing into the emitter is

$$I_E = I_B + I_C. \tag{1.14}$$

From Equations (1.12) and (1.13) it is seen that the fundamental operation of a BJT is as a voltage-controlled current source. This leads to the small-signal circuit model of a BJT, biased in its fundamental mode of operation, shown in Figure 1-3(b).

The schematic symbols used for BJTs are shown in Table 1-1 with the arrow pointing to the n-type semiconductor. The BJT symbol is also the symbol for a HBT.

1.3.2 MOSFET Fundamentals

There are several types of FETs, with the MOSFET being the most common. With all FETs there is a channel between two terminals, the source and drain, and an applied field produced by a voltage at a third terminal, the gate, controls the cross section of the channel and the number of carriers in the channel. Hence the gate voltage controls the current flow between the drain and the source. With some FETs, the channel does not exist until a gate field is applied and pulls carriers from the bulk into the channel, and this

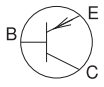
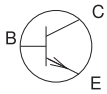
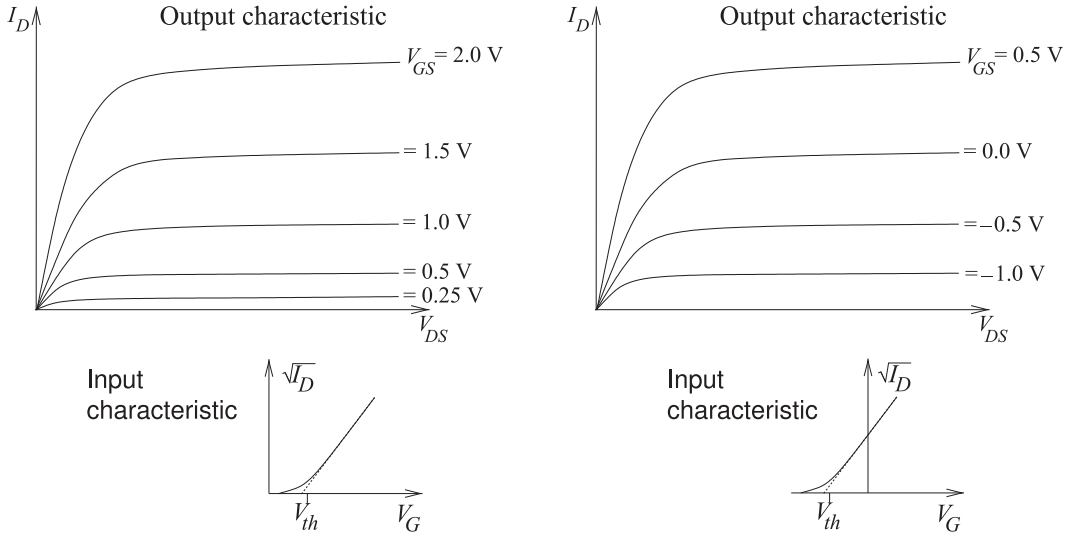
Transistor	IEEE symbol	Commonly used symbol
BJT, pnp		
BJT, npn		

Table 1-1: IEEE standard schematic symbols for bipolar junction transistors (BJTs and HBTs) [8] and commonly used symbols in layouts [9]. The letters indicate terminals: B (base), C (collector), E (emitter). These symbols are used for silicon BJTs and compound semiconductor HBTs.



(a) Enhancement MOSFET (b) Depletion MOSFET

Figure 1-4: Current-voltage characteristics of depletion- and enhancement-mode MOSFETs.

is called an **enhancement-mode** FET. The input and output characteristics of the enhancement-mode FET are shown in Figure 1-4(a). With some MOSFETs with a particular doping profile, carriers are in the channel even without an applied field and a gate voltage either enhances the cross section of the channel or closes it off. Most often the gate voltage is used to reduce current conduction, and this type of FET is called a **depletion-mode** FET. The input and output characteristics of the depletion-mode FET are shown in Figure 1-4(b). The enhancement-mode MOSFET is much more common than the depletion-mode.

The enhancement MOSFET is a relatively simple device to fabricate and is the smallest of the semiconductor transistors. It is the preferred technology for high-density integration. The three-dimensional view and cross sections of a MOSFET are shown in Figure 1-5(a–c). The cross-section of an nMOS transistor is shown in Figure 1-5(b) where there is a p-type substrate and an n-type channel is created when there is sufficient voltage at the gate. As well as the source and drain connection, there is a fourth terminal call the body connection denoted as U in Figure 1-5(b) but B is also used and this can be confused with the base of a BJT transistor. The body is typically

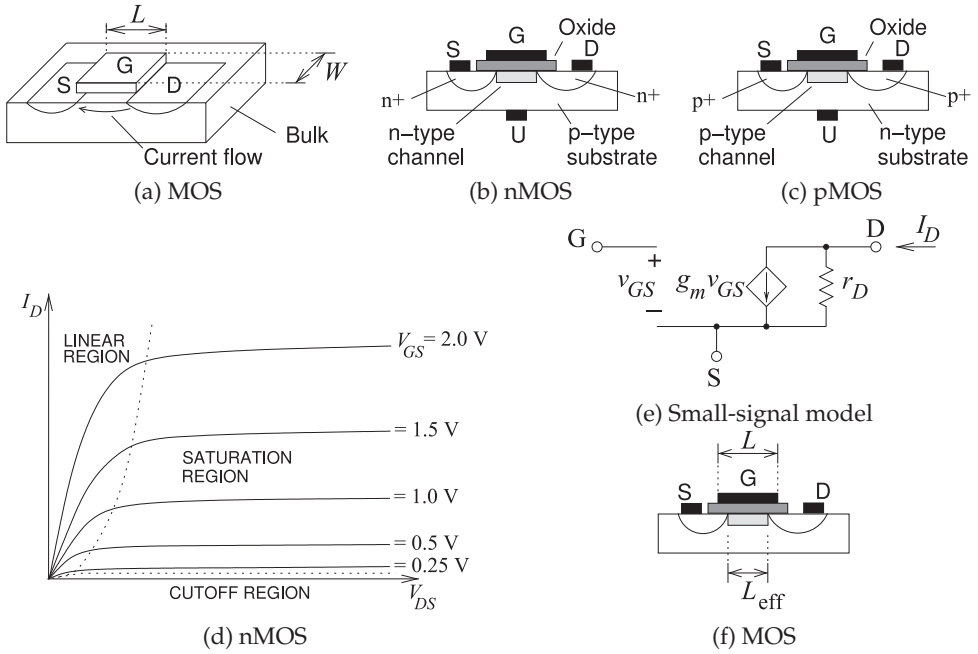


Figure 1-5: MOSFET details: (a) three-dimensional view of a MOSFET; (b) cross section of an nMOS transistor with metal or **polysilicon** contacts indicated by the black blocks; (c)the corresponding cross section of a pMOS transistor; (d) current-voltage characteristics of an enhancement-mode MOSFET; (e) circuit model of fundamental operation; and (f) cross section showing the effective gate length, L_{EFF}. The **linear region** is sometimes (but less often) called the **triode region** because of similarity to the characteristics of the triode vacuum tube device. Similarly the **saturation region** is sometimes called the **pentode region**.

connected to the most negative voltage in the circuit so that the substrate-to-channel interface is a reverse-biased diode. A similar situation occurs with the pMOS transistor with the cross-section of Figure 1-5(c). Now there is a p-type channel and an n-type substrate so that the body (U) must typically be connected to the most positive voltage in the circuit to ensure a reverse-biased junction between the substrate and the channel.

A MOSFET has metal or polysilicon (a reasonable conductor [5, 6, 10]) connections at the drain (D), source (S) and gate (G). The MOSFET is nearly always silicon, but possibly (**GaN!MOSFET** [11, 12]. The source and drain connections are highly doped (n+ for nMOS and p+ for pMOS) semiconductor regions providing a good ohmic contact rather than forming a Schottky barrier.[1] The gate is not in direct contact with the semiconductor, but separated by a thin layer of oxide. With no voltage applied at the gate, there are no carriers below the gate oxide that can conduct current between the source and drain. A gate voltage is necessary to draw carriers to the channel region, forming a conducting channel. That is, a voltage applied to the gate creates an electric field that induces electrons (the n carriers for

[1] A Schottky barrier occurs at the abrupt interface between a metal and a doped semiconductor.

an nMOSFET) to form a conducting channel immediately under the oxide.[2] This process is called inversion. The length of the channel is denoted L_{eff} (the effective gate length), which is less than the actual gate length L as the highly doped source and drain regions must extend under the gate to ensure good contact to the induced channel. This is indicated in Figure 1-5(e). The number of carriers in the channel is controlled by the gate voltage. A higher frequency of operation is obtained by reducing L_{eff}.

Three distinct regions of operation, identified in Figure 1-5(d), are recognized for a MOSFET. In the linear region the drain-source current, I_{DS}, continues to increase as the drain-source voltage, V_{DS}, increases. I_{DS} depends on both the drain-source and gate-source voltage, V_{DS} and V_{GS}, so the linear region is sometimes exploited in mixers. In the saturation region, I_{DS} is almost independent of V_{DS} and almost solely controlled by V_{GS}. MOSFET amplifiers operate in the saturation region. The cutoff region is when there is negligible drain current, and a FET is particularly effective at shutting off conduction and so makes a good voltage-controlled switch.

In initial design the mode of fundamental operation must be intuitively understood and simple models and equations are needed. In contrast, a circuit simulator requires a detailed model capturing subtle physical effects. A model of a MOSFET that can be used in a circuit simulator is presented in Section 1.A.1. The model presented is known as the Level 3 MOSFET model and captures the fundamental operation of MOSFETs as well as capacitive parasitic effects. Models are developed using physical insight into semiconductor operation. All semiconductor device models, not just MOSFETs, require extensive fitting to measured data and have limited accuracy. Consequently the design, fabrication, and test cycle are critically important to realizing transistor circuits.

In the saturation region (see Figure 1-5(d)) the fundamental operation of a MOSFET is described by Equation (1.47), which is repeated here:

$$I_{DS} = \frac{W_{\text{eff}}}{L_{\text{eff}}}\,\mu_{\text{eff}}\,C_{ox}\left[(V_{GS}-V_{th})-1+\frac{F_B}{2}\,V_{d\text{sat}}\right]V_{d\text{sat}}. \tag{1.15}$$

Here C_{ox} is the capacitance of the gate oxide, W_{eff} is the effective gate width, which is the gate width W modified by fringing and related effects, V_{th} is the **threshold voltage**, and μ_{eff} is the effective mobility[3] of the carriers in the channel (electrons for an nMOSFET and holes for a pMOSFET). $V_{d\text{sat}}$ is the drain saturation voltage and is the drain source voltage at which the device enters the saturation region from the linear region. F_B is due to the charge in the bulk semiconductor (below the channel) on which the gate-induced electric field terminates. L_{eff} is the effective gate length and this is modulated by the drain-source voltage so that [5, 6, 10]

$$L_{\text{eff}} = \frac{L}{1+\lambda V_{DS}}. \tag{1.16}$$

Accounting for channel length modulation, described by Equation (1.16),

[2] The discussion is similar for a pMOSFET, but with holes (p-type carriers) forming the channel.

[3] **Mobility**, μ, is the proportionality of the velocity of carriers to the applied electric field, $v_d = \mu E$, where v_d is the average drift velocity of carriers and E is the applied electric field. Mobility has the units $\text{m}^2/(\text{V}\cdot\text{s})$, i.e. $\text{m}^2\cdot\text{V}^{-1}\cdot\text{s}^{-1}$.

and simplifying [5, 6, 10], Equation (1.15) becomes

$$I_{DS} = \frac{W}{L}\frac{\mu_{\text{eff}} C_{ox}}{2}\left(V_{GS} - V_{th}\right)^2\left(1 + \lambda V_{DS}\right). \tag{1.17}$$

This equation embodies the fundamental operation needed in developing initial designs. The key is that the MOSFET can be modeled (at least in the saturation region) as a voltage-controlled current source as shown in the model of Figure 1-5(e). The transconductance, g_m (in saturation), is obtained by differentiating Equation (1.17) so that (ignoring channel length modulation)

$$g_m = \frac{\partial I_{DS}}{\partial V_{GS}} = \frac{W}{L}\mu_{\text{eff}} C_{ox}\left(V_{GS} - V_{th}\right). \tag{1.18}$$

This can also be written as

$$g_m = \sqrt{\frac{W}{L} 2\mu_{\text{eff}} C_{ox} I_{DS}}. \tag{1.19}$$

Generally the gate length L is fixed at the minimum supported by a particular process, as this provides the highest frequency of operation. However, both L and W can be selected to control the current, I_{DS}. For example, if V_{GS} is fixed, then the MOSFET acts as a current source, with the value of the current adjusted in design by setting L and W provided that there is sufficient V_{DS}.

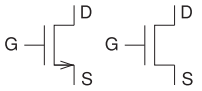
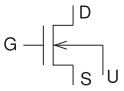
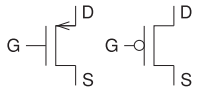
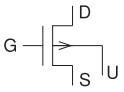
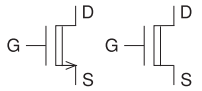
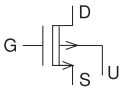
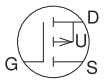
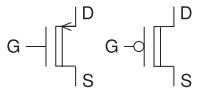
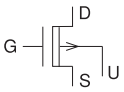
The current-voltage characteristics shown in Figure 1-5(d) are those of an enhancement-mode MOSFET, which requires the simplest processing. Applying a gate-source voltage enhances the channel and increases I_{DS}. With additional processing [5, 6, 10, 13] a depletion-mode MOSFET can be fabricated so that the channel exists even without an applied gate voltage. The same equations are used to describe operation with the threshold voltage changed. I_{DS} increases as the gate-source voltage increases, and it reduces as the gate-voltage becomes negative. The contrast between enhancement-mode and depletion-mode MOSFETs is illustrated in Figure 1-4.

The voltage of the bulk semiconductor affects the operation of a MOSFET and is a fourth terminal controlling drain-source conduction, but has a much smaller effect than the gate does. Most often the bulk is connected electrically to the most negative voltage in a circuit for an nMOSFET and to the most positive voltage for a pMOSFET. The standard schematic symbols of MOSFETs are shown in Table 1-2.

1.3.3 MESFET, HEMT, and JFET Fundamentals

The MESFETand HEMT are types of JFETs fabricated using compound semiconductors, with JFET most commonly referring to silicon devices only. The cross section of a JFET is shown in Figure 1-6(a), where the depth (cross section) of the conducting channel is varied by the thickness of the depletion region of a reverse-biased junction. With the silicon JFET, the voltage applied to the gate terminal changes the amount of reverse bias and hence the depletion region thickness. Increased reverse bias reduces the cross section of the current-carrying channel. Thus a JFET looks like a variable conductance. The controlling field of the FET is created at the reverse-biased pn junction at

Table 1-2: IEEE standard schematic symbols for MOSFET transistors [8] and symbols more commonly used in schematics [9]. The MOSFET symbols are for enhancement- and depletion-mode transistors. The letters indicate terminals: G (gate), D (drain), S (source), U (bulk). The three-terminal nMOSFET symbol is most often used when the bulk is connected to the most negative connection in the circuit, and the three-terminal pMOSFET symbol is used when the bulk is tied to V_{DD} (the most positive connection).

Transistor	IEEE symbol	Commonly used symbol (3 terminal)	Commonly used symbol (4 terminal)
FET, nMOS, depletion			
FET, pMOS, depletion			
FET, nMOS, enhancement			
FET, pMOS, enhancement			

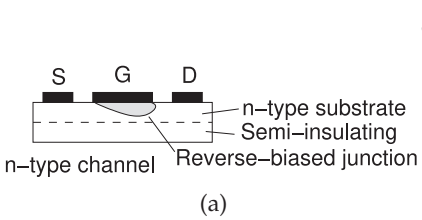
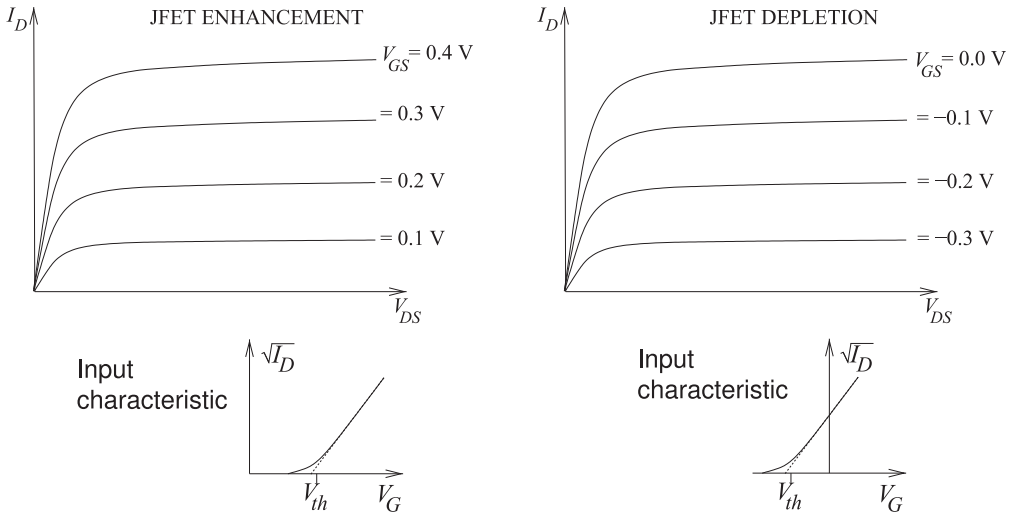


Figure 1-6: JFET details: (a) cross section; and (b) circuit model of fundamental operation.

the gate terminal. The term JFET most commonly refers to a silicon junction FET. With compound semiconductors such as GaAs, the pn junction of a silicon JFET is replaced by a Schottky barrier junction and the transistor is called a **metal-epitaxy-semiconductor FET** (**MESFET**). A device similar to the MESFET is the **high electron mobility transistor** (**HEMT**), where the field is established at the junction of two compound semiconductor materials having different band gaps, called a heterojunction. The channel is formed at the heterojunction. The HEMT is also called the **heterostructure FET** (**HFET**). A MESFET with a graded junction is called a **modulation-doped FET** (**MODFET**). A **pseudomorphic HEMT** (**pHEMT**) has an extremely thin layer establishing the channel so that the crystal structure stretches and a very high bandgap is established. Enhancement-mode and depletion-mode JFETs are contrasted in Figure 1-7.

The Materka-Kacprzak transistor model was developed for GaAs MESFET transistors [14] but is used to model silicon JFET and HEMT transistors as well. The model is described in Section 1.A.2 and the fundamental operation is described by Equation (1.64), which is repeated here without the area



(a) Enhancement mode (b) Depletion mode

Figure 1-7: Current-voltage characteristics of depletion-mode and enhancement-mode JFETs.

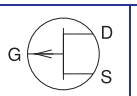
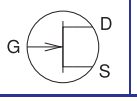
multiplier:

$$I_{DS} = I_{DSS}\left[1 + S_S\frac{V_{DS}}{I_{DSS}}\right]\left[1 - \frac{V_{GS}(t-\tau)}{V_{P0} + \gamma V_{DS}}\right]^{(E + K_E V_{GS}(t-\tau))} \times \tanh\left[\frac{S_L V_{DS}}{I_{DSS}(1 - K_G V_{GS}(t-\tau))}\right]. \quad (1.20)$$

Here I_{DSS} is the drain saturation current, and this, along with all quantities in Equation (1.20) other than V_{DS}, V_{GS}, and I_{DS} are constants and specified as inputs by the user. Equation (1.20) indicates that the fundamental operation of a JFET is that of a voltage-controlled current source. Thus the small-signal circuit model of fundamental operation is as shown in Figure 1-6(b).

The schematic symbols used for the MESFET, HEMT, and JFET are shown in Table 1-3. The only MESFET type used, however, is the n-type, as the p-type MESFET has poor performance due to the low mobility of holes.

Table 1-3: IEEE standard schematic symbols for JFETs (MESFET, HEMT, JFET) [8] and symbols more commonly used in schematics. The letters indicate terminals: G (gate), D (drain), S (source).

Transistor	IEEE symbol	Commonly used symbol
FET, pJFET		
FET, nJFET, MESFET, HEMT		

1.4 References

[1] M. Steer, *Microwave and RF Design, Radio Systems*, 3rd ed. North Carolina State University, 2019.

[2] ——, *Microwave and RF Design, Transmission Lines*, 3rd ed. North Carolina State University, 2019.

[3] ——, *Microwave and RF Design, Networks*, 3rd ed. North Carolina State University, 2019.

[4] ——, *Microwave and RF Design, Modules*, 3rd ed. North Carolina State University, 2019.

[5] B. B. Streetman and S. Banerjee, *Solid State Electronic Devices*, 6th ed. Prentice Hall, 2006.

[6] S. Sze and K. Ng, *Physics of Semiconductor Devices*, 3rd ed. John Wiley & Sons, 2007.

[7] H. Gummel and H. Poon, "An integral charge control model of bipolar transistors," *Bell Syst. Tech. J.*, vol. 49, pp. 827–852, may–jun 1970.

[8] IEEE Standard 315-1975, Graphic Symbols for Electrical and Electronics Diagrams (Including Reference Designation Letters), Adopted Sept. 1975, Reaffirmed Dec. 1993. Approved by American National Standards Institute, Jan. 1989. Approved adopted for mandatory use, Department of Defense, United States of America, Oct. 1975. Approved by Canadian Standards Institute, Oct. 1975.

[9] R. Baker, *CMOS Circuit Design, Layout, and Simulation*, 2nd ed. Wiley-Interscience, IEEE Press, 2008.

[10] D. Schroder, *Semiconductor Material and Device Characterization*. IEEE Press and Wiley, 2006.

[11] M. Johnson, D. Barlage, and D. Braddock, "Prospect for III-nitride heterojunction MOSFET structures and devices," in *Materials Research Society Proc.*, 2004.

[12] C. Roff, P. McGovern, J. Benedikt, P. Tasker, M. Johnson, D. Barlage, W. Sutton, and D. Braddock, "Pulsed-iv and RF waveform measurements of unique high-k dielectric GaN MOSFETs," in *IEEE Int Conf. on Microwaves, Communications, Antennas and Electronic Systems, 2008 (COMCAS 2008)*, May 2008, pp. 1–4.

[13] Y. Bito, N. Iwata, and M. Tomita, "64% efficiency enhancement-mode power heterojunction FET for 3.5 V Li-ion battery operated personal digital cellular phones," in *1998 IEEE MTT-S Int. Microwave Symp. Dig.*, Jun. 1998, pp. 439–442.

[14] A. Materka and T. Kacprzak, "Computer calculation of large-signal GaAs FET amplifier characteristics," *IEEE Trans. on Microwave Theory and Techniques*, vol. 33, no. 2, pp. 129–135, Feb. 1985.

[15] C. Enz and E. Vittoz, *Charge-Based MOS Transistor Modeling: the EKV Model for Low-Power and RF IC Design*. Wiley, 2006.

[16] W. Grabinski, B. Nauwelaers, and D. Schreurs, Eds., *Transistor Level Modeling for Analog/RF IC Design*. Springer, 2006.

[17] D. Foty, *MOSFET Modeling With SPICE: Principles and Practice*. Prentice-Hall, 1997.

[18] W. Liu, *MOSFET Models for Spice Simulation, Including BSIM3v3 and BSIM4*. John Wiley & Sons, 2001.

[19] P. Yang, B. Epler, and P. Chatterjee, "An investigation of the charge conservation problem for MOSFET circuit simulation," *IEEE J. of Solid-State Circuits*, vol. 18, no. 1, pp. 128–138, Feb. 1983.

[20] I. Angelov, H. Zirath, and N. Rosman, "A new empirical nonlinear model for HEMT and MESFET devices," *IEEE Trans. on Microwave Theory and Techniques*, vol. 40, no. 12, pp. 2258–2266, Dec. 1992.

[21] R. Pengelly, *Microwave Field-Effect Transistors: Theory, Design, and Applications*. Noble, 1994.

[22] F. Schwierz and J. Liou, *Modern Microwave Transistors: Theory, Design, and Performance*. Wiley, 2003.

[23] M. Rudolph, *Introduction to Modeling HBTs*. Artech House, 2006.

[24] J. Yuan, *SiGe, GaAs, and InP Heterojunction Biopolar Transistors*. Wiley, 1999.

Appendix

1.A Active Device Models

1.A.1 Level 3 MOSFET Model 14
1.A.2 Materka–Kacprzak MESFET and HEMT Model 19
1.A.3 Gummel–Poon: Bipolar Junction Transistor Model 20

This appendix presents the model parameters of the three most common transistor types used in microwave designs. These models are available in nearly all circuit simulators. Transistor models implement device equations that have been developed from physical insight with necessary simplifications required for implementation in a simulator. The purpose of presenting these models is so that the basic physical description of operation can be examined.

1.A.1 Level 3 MOSFET Model

The level 3 MOSFET model is the model of a silicon MOSFET transistor and is one of a large number of different MOSFET models that are used [15–18]. MOSFETs are the most complicated transistor to model, as their operation relies on attracting carriers into the channel under the gate in a process called inversion. The MOS level 3 model here uses the charge-conserving Yang–Chatterjee model [19] for modeling charge and capacitance. For many years the level 3 MOSFET model was implemented in circuit simulators but did not conserve charge. An example of errors that can exist in device models.

The model parameters listed in Table 1-4 can be specified by the circuit designer.

Table 1-4: Level 3 MOSFET model parameters.

Name	Description	Units	Default
gamma	Bulk threshold parameter	$V^{0.5}$	0

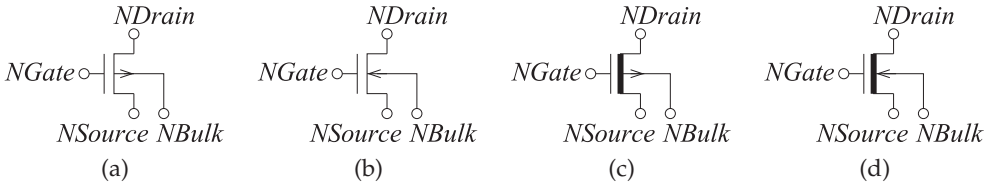


Figure 1-8: MOSFET types: (a) enhancement-mode p type; (b) enhancement-mode p type; (c) depletion-mode n type; (d) depletion-mode n type;

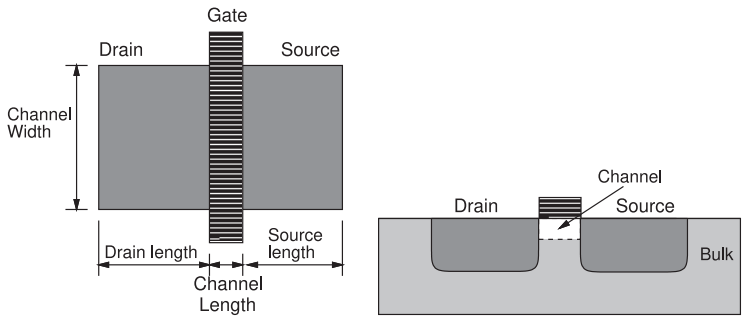


Figure 1-9: Physical layout of a MOSFET transistor.

Name	Description	Units	Default
kp	Transconductance parameter	A/V^2	0.000021
l	Device length	m	0.000002
w	Device width	m	0.00005
ld	Lateral diffusion length	m	0
wd	Lateral diffusion width	m	0
nsub	Substrate doping	cm^{-3}	0
phi	Surface inversion potential	V	0.6
tox	Oxide thickness	m	1×10^{-7}
u0	Surface mobility	$cm^2/V\text{-}s$	600
vt0	Zero bias threshold voltage	V	0
kappa	Saturation field factor	m	0.2
t	Device temperature	degrees	300.15
tnom	Nominal temperature	degrees	300.15
nfs	Fast surface state density	cm^{-2}	0
eta	Static feedback on threshold voltage		0
theta	Mobility modulation	$1/V$	0
tpg	Gate material type		0
nss	Surface state density	cm^{-2}	0
vmax	Maximum carrier drift velocity	m/sec	0
xj	Metallurgical junction depth		0
delta	Width effect on threshold voltage		0

Device Equations

The device equations here are specifically for a p-type MOSFET. There are sign changes required to get the appropriate current directions for an n-type MOSFET. The subscript D refers to the drain, S refers to the source, and G refers to the gate. The constants used are

$$\begin{aligned} q &= 1.6021918 \times 10^{-19}\ (\mathrm{As}), & k &= 1.3806226 \times 10^{-23}\ (\mathrm{J/K}), \\ \epsilon_0 &= 8.85421487 \times 10^{-12}\ (\mathrm{F/m}), & \epsilon_s &= 11.7\,\epsilon_0. \end{aligned}$$

All parameters used are indicated in `THIS` font.

$$E_g = 1.16 - \frac{7.02 \times 10^{-4}\,\mathtt{T}^2}{\mathtt{T} + 1108}\ (\mathrm{V}) \qquad C_{ox} = \frac{\epsilon_0\, 3.9}{\mathtt{TOX}}\ (\mathrm{F}) \tag{1.21}$$

$$L_{\text{eff}} = \mathtt{L} - 2\,\mathtt{LD} \qquad W_{\text{eff}} = \mathtt{W} - 2\,\mathtt{WD} \tag{1.22}$$

Depletion layer width coefficient:

$$X_d = \sqrt{\frac{2\,\epsilon_s}{q\,\mathtt{NSUB}\,10^6}}. \tag{1.23}$$

Built in voltage:

$$V_{bi} = \mathtt{VT0} - \mathtt{GAMMA}\sqrt{\mathtt{PHI}}. \tag{1.24}$$

Square root of substrate voltage:

$$\begin{aligned} V_{BS} \le 0 \quad &\Longrightarrow \quad SqV_{BS} = \sqrt{\mathtt{PHI} - V_{BS}} \\ V_{BS} > 0 \quad &\Longrightarrow \quad SqV_{BS} = \sqrt{\frac{\mathtt{PHI}}{1 + \frac{0.5}{\mathtt{PHI}}\,V_{BS}(1 + \frac{0.75}{\mathtt{PHI}}V_{BS})}}. \end{aligned} \tag{1.25}$$

Short-channel effect correction factor:
In a short-channel device, the device threshold voltage tends to be lower since part of the depletion charge in the bulk terminates the electric fields at the source and drain. The value of this correction factor is determined by the metallurgical depth, XJ.

$$c_0 = 0.0631353 \tag{1.26}$$
$$c_1 = 0.8013292 \tag{1.27}$$
$$c_2 = -0.01110777 \tag{1.28}$$
$$T_1 = \mathtt{XJ}\,(c_0 + c_1\,X_d\,SqV_{BS} + c_2\,(X_d\,SqV_{BS})^2) \tag{1.29}$$
$$F_s = 1 - \frac{\mathtt{LD} + T_1}{L_{\text{eff}}}\sqrt{1 - \left(\frac{X_d\,SqV_{BS}}{\mathtt{XJ} + X_d\,SqV_{BS}}\right)^2}. \tag{1.30}$$

Narrow-channel effect correlation factor:
The edge effects in a narrow channel cause the depletion charge to extend beyond the width of the channel. This has the effect of increasing the threshold voltage:

$$F_n = \frac{\pi\,\epsilon_s\,\mathtt{DELTA}}{2\,C_{ox}\,W_{\text{eff}}}. \tag{1.31}$$

Static feedback coefficient:
The threshold voltage lowers because the charge under the gate terminal depleted by the drain junction field increases with V_{DS}. This effect is drain-induced barrier lowering (DIBL):

$$\sigma = \frac{8.14 \times 10^{-22}\,\mathtt{ETA}}{{L_{\text{eff}}}^3\,C_{ox}}. \tag{1.32}$$

Threshold voltage:

$$V_{th} = V_{bi} - \sigma\,V_{DS} + \mathtt{GAMMA}\,SqV_{BS}\,F_s\, + F_n\,{SqV_{BS}}^2. \tag{1.33}$$

Subthreshold operation:
This variable is invoked depending on the value of the parameter NFS and is used only when in the subthreshold mode:

$$X_n = 1 + \frac{q\,\mathtt{NFS}\,10^4}{C_{ox}} + \frac{F_n}{2} + \frac{\mathtt{GAMMA}}{2}\,\frac{F_s}{SqV_{BS}}. \tag{1.34}$$

Modified threshold voltage:
This variable defines the limit between weak and strong inversion:

$$\begin{array}{ll} \mathtt{NFS} > 0 & \Longrightarrow \quad V_{on} = V_{th} + \frac{kT}{q}\,X_n \\ \mathtt{NFS} \le 0 & \Longrightarrow \quad V_{on} = V_{th}. \end{array} \tag{1.35}$$

Subthreshold gate voltage:

$$V_{gsx} = \text{MAX}(V_{GS}, V_{on}). \tag{1.36}$$

Surface mobility:

$$\mu_s = \frac{\mathtt{U0}10^{-4}}{1 + \mathtt{THETA}(V_{gsx} - V_{th})}. \tag{1.37}$$

Saturation voltage:
Calculation of this voltage requires many steps. The effective mobility is calculated as

$$\mu_{\text{eff}} = \mu_s \, F_{\text{drain}}, \tag{1.38}$$

where

$$F_{\text{drain}} = \left(1 + \frac{\mu_s \, V_{DS}}{\texttt{VMAX}\, L_{\text{eff}}}\right)^{-1} \tag{1.39}$$

$$\beta = \frac{W_{\text{eff}}}{L_{\text{eff}}} \, \mu_{\text{eff}} \, C_{ox}. \tag{1.40}$$

The Taylor expansion of bulk charge is

$$F_B = \frac{\texttt{GAMMA}}{4} \frac{F_s}{SqV_{BS}} + 2 \, F_n. \tag{1.41}$$

The standard value of the saturation voltage is calculated as

$$V_{\text{sat}} = \frac{V_{gsx} - V_{th}}{1 + F_B}. \tag{1.42}$$

The final value of the saturation voltage depends on the parameter VMAX:

$$\begin{array}{ll} \texttt{VMAX} = 0 & \Longrightarrow \quad V_{d\text{sat}} = V_{\text{sat}} \\ \texttt{VMAX} > 0 & \Longrightarrow \quad V_{d\text{sat}} = V_{\text{sat}} + V_c - \sqrt{V_{\text{sat}}^2 + V_c^2}, \quad V_c = \texttt{VMAX}\, L_{\text{eff}}/\mu_s. \end{array} \tag{1.43}$$

Velocity saturation drain voltage:
This ensures that the drain voltage does not exceed the saturation voltage:

$$V_{dsx} = \text{MIN}(V_{DS}, V_{d\text{sat}}). \tag{1.44}$$

Drain current:
Linear region:

$$I_{DS} = \beta \, \frac{\mu_s}{\texttt{U0}\; 10^{-4}} \, F_{\text{drain}} \, (V_{gsx} - V_{th} - \frac{1 + F_B}{2} \, V_{dsx}) \, V_{dsx}. \tag{1.45}$$

Saturation region:

$$I_{DS} = \beta \left[(V_{GS} - V_{th}) - \frac{1 + F_B}{2} \, V_{d\text{sat}}\right] V_{d\text{sat}}. \tag{1.46}$$

Using Equation (1.40), this becomes

$$I_{DS} = \frac{W_{\text{eff}}}{L_{\text{eff}}} \, \mu_{\text{eff}} \, C_{ox} \left[(V_{GS} - V_{th}) - \frac{1 + F_B}{2} \, V_{d\text{sat}}\right] V_{d\text{sat}}. \tag{1.47}$$

Cutoff region:

$$I_{DS} = 0. \tag{1.48}$$

Channel length modulation:
As V_{DS} increases beyond $V_{d\text{sat}}$, the point where the carrier velocity begins to saturate moves toward the source. This is modeled by the term $\triangle_\ell$:

$$\triangle_\ell = X_d \sqrt{\frac{X_d^2 \, E_p^2}{4} + \texttt{KAPPA}\, (V_{DS} - V_{d\text{sat}})} - \frac{E_p \, X_d^2}{2}, \tag{1.49}$$

where E_p is the lateral field at pinch-off and is given by

$$E_p = \frac{\texttt{VMAX}}{\mu_s\,(1 - F_{\rm drain})} \tag{1.50}$$

The drain current is multiplied by a correction factor, $l_{\rm fact}$. This factor prevents the denominator $(L_{\rm eff} - \triangle_\ell)$ from going to zero:

$$\begin{array}{lll} \triangle_\ell \le 0.5\,L_{\rm eff} & \Longrightarrow & l_{\rm fact} = \frac{L_{\rm eff}}{L_{\rm eff} - \triangle_\ell} \\ \triangle_\ell > 0.5\,L_{\rm eff} & \Longrightarrow & l_{\rm fact} = \frac{4\triangle_\ell}{L_{\rm eff}}. \end{array} \tag{1.51}$$

The corrected value of the drain-source current is

$$I_{DS\rm new} = I_{DS}\,l_{\rm fact}. \tag{1.52}$$

Subthreshold operation:
For subthreshold operation, if the fast surface density parameter NFS is specified and $V_{GS} \le V_{\rm on}$, then the final value of the drain-source current is given by

$$I_{DS\rm final} = I_{DSnew}\,{\rm e}^{\frac{kt}{q}\,\frac{V_{GS} - V_{\rm on}}{X_n}}. \tag{1.53}$$

Yang–Chatterjee charge model [19]
This model ensures continuity of the charges and capacitances throughout different regions of operation. The intermediate quantities are

$$V_{FB} = V_{to} - \texttt{GAMMA}\,\sqrt{\texttt{PHI}} - \texttt{PHI} \tag{1.54}$$

and

$$C_o = C_{ox}\,W_{\rm eff}\,L_{\rm eff}. \tag{1.55}$$

Accumulation region, $V_{GS} \le V_{FB} + V_{BS}$:

$$Q_d = 0, \quad Q_s = 0, \quad Q_b = -C_o\,(V_{GS} - V_{FB} - V_{BS}). \tag{1.56}$$

Cutoff region, $V_{FB} + V_{BS} < V_{GS} \le V_{th}$:

$$Q_d = 0, \quad Q_s = 0, \quad Q_b = -C_o\,\frac{\texttt{GAMMA}^2}{2}\,\{-1 + \sqrt{1 + \frac{4\,(V_{GS} - V_{FB} - V_{BS})}{\texttt{GAMMA}^2}}\}. \tag{1.57}$$

Saturation region, $V_{th} < V_{GS} \le V_{DS} + V_{th}$:

$$Q_d = 0, \quad Q_s = -\tfrac{2}{3}\,C_o\,(V_{GS} - V_{th}), \quad Q_b = C_o\,(V_{FB}\,\texttt{PHI} - V_{th}). \tag{1.58}$$

Linear region, $V_{GS} > V_{DS} + V_{th}$:

$$Q_d = -C_o\,\left[\frac{V_{DS}^2}{8\,(V_{GS} - V_{th} - \frac{1}{2}V_{DS})} + \frac{V_{GS} - V_{th}}{2} - \frac{3}{4}\,V_{DS}\right] \tag{1.59}$$

$$Q_s = -C_o\,\left[\frac{V_{DS}^2}{24\,(V_{GS} - V_{th} - \frac{1}{2}V_{DS})} + \frac{V_{GS} - V_{th}}{2} + \frac{1}{4}\,V_{DS}\right] \tag{1.60}$$

$$Q_b = C_o\,(V_{FB}\,\texttt{PHI} - V_{th}). \tag{1.61}$$

The final currents at the transistor nodes are given by

$$I_d = I_{DS\rm final} + \frac{dQ_d}{dt} \qquad I_g = \frac{dQ_g}{dt} \qquad I_s = -I_{DS\rm final} + \frac{dQ_s}{dt}. \tag{1.62}$$

1.A.2 Materka–Kacprzak MESFET Model

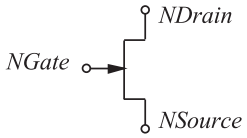


Figure 1-10: MESFET element.

The Materka–Kacprzak transistor model was developed for GaAs MESFET transistors [14] but is used to model silicon JFETs and compound semiconductor HEMT transistors as well. It is based on physical interpretation of a transistor with a junction-based gate. There are a number of other models [20–22], but the Materka–Kacprzak model is representative of JFETs.

Table 1-5: Materka–Kacprzak model parameters

Name	Description	Units	Default
AFAB	Slope factor of breakdown current ($AFAB$)	$1/V$	0.0
AFAG	Slope factor of gate conduction current ($AFAG$)	$1/V$	38.696
AREA	Area multiplier ($AREA$)	-	1.0
C10	Gate source Schottky barrier capacitance for (C_{10})	F	0.0
CFO	Gate drain feedback capacitance for (C_{F0})	F	0.0
CLS	Constant parasitic component of gate-source capacitance (C_{LS})	F	0.0
E	Constant part of power law parameter (E)	-	2.0
GAMA	Voltage slope parameter of pinch-off voltage (γ)	$1/V$	0.0
IDSS	Drain saturation current for (I_{DSS})	A	0.1
IG0	Saturation current of gate-source Schottky barrier (I_{G0})	A	0.0
K1	Slope parameter of gate-source capacitance (K_1)	$1/V$	1.25
KE	Dependence of power law on V_{GS} , (K_E)	$1/V$	0.0
KF	Slope parameter of gate-drain feedback capacitance (K_F)	$1/V$	1.25
KG	Drain dependence on V_{GS} in the linear region, (K_G)	$1/V$	0.0
KR	Slope factor of intrinsic channel resistance (K_R)	$1/V$	0.0
RI	Intrinsic channel resistance for (R_I)	Ω	0.0
SL	Slope of the drain characteristic in the saturated region, (S_L)	S	0.15
SS	Slope of the drain characteristic in the saturated region (S_S)	S	0.0
T	Channel transit-time delay (τ)	s	0.0
VBC	Breakdown voltage (V_{BC})	V	10^{10}
VP0	Pinch-off voltage for (V_{P0})	V	-2.0

The physical constants used in the model evaluation are

k	the Boltzmann constant	$1.3806226\,10^{-23}$ J/K
q	electronic charge	$1.6021918\,10^{-19}$ C

Standard calculations:

$$V_{\mathrm{TH}} = (kT)/q, \tag{1.63}$$

where T is the analysis temperature. Also

V_{DS} is the intrinsic drain source voltage,
V_{GS} is the intrinsic gate source voltage, and
V_{GD} is the intrinsic gate drain voltage.

Device Equations

Current characteristics:

$$I_{DS} = \text{Area} I_{DSS}\left[1 + S_S\frac{V_{DS}}{I_{DSS}}\right]\left[1 - \frac{V_{GS}(t-\tau)}{V_{P0} + \gamma V_{DS}}\right]^{(E + K_E V_{GS}(t-\tau))} \times \tanh\left[\frac{S_L V_{DS}}{I_{DSS}(1 - K_G V_{GS}(t-\tau))}\right] \tag{1.64}$$

$$I_{GS} = \text{Area} I_{G0}\left[\mathrm{e}^{A_{FAG}V_{GS}} - 1\right] - I_{B0}\left[e^{-A_{FAB}(V_{GS} + V_{BC})}\right] \tag{1.65}$$

$$I_{GD} = \text{Area} I_{G0}\left[\mathrm{e}^{A_{FAG}V_{GD}} - 1\right] - I_{B0}\left[e^{-A_{FAB}(V_{GD} + V_{BC})}\right] \tag{1.66}$$

$$R_I = \begin{cases} R_{10}(1 - K_R V_{GS})/\text{Area} & K_R V_{GS} < 1.0 \\ 0 & K_R V_{GS} \geq 1.0\,. \end{cases} \tag{1.67}$$

Capacitance:

C_{LVL} = 1 (default) for the standard Materka–Kacprzak capacitance model described below is used. The Materka–Kacprzak capacitances are

$$C'_{DS} = C_{DS} \tag{1.68}$$

$$C'_{GS} = \begin{cases} \left[C_{10}(1 - K_1 V_{GS})^{M_{GS}} + C_{1S}\right] & K_1 V_{GS} < F_{CC} \\ \left[C_{10}(1 - F_{CC})^{M_{GS}} + C_{1S}\right] & K_1 V_{GS} \geq F_{CC} \end{cases} \tag{1.69}$$

$$C'_{GD} = \begin{cases} \text{Area}\left[C_{F0}(1 - K_1 V_1)^{M_{GD}}\right] & K_1 V_1 < F_{CC} \\ \text{Area}\left[C_{F0}(1 - F_{CC})^{M_{GD}}\right] & K_1 V_1 \geq F_{CC} \end{cases}\,. \tag{1.70}$$

1.A.3 Gummel–Poon: Bipolar Junction Transistor Model

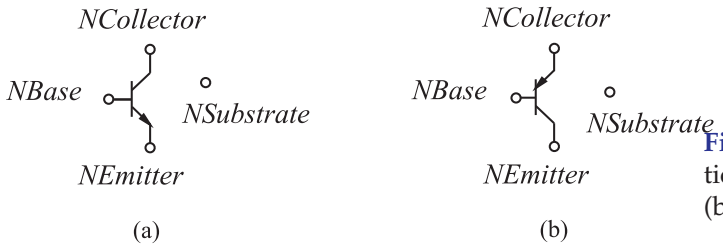


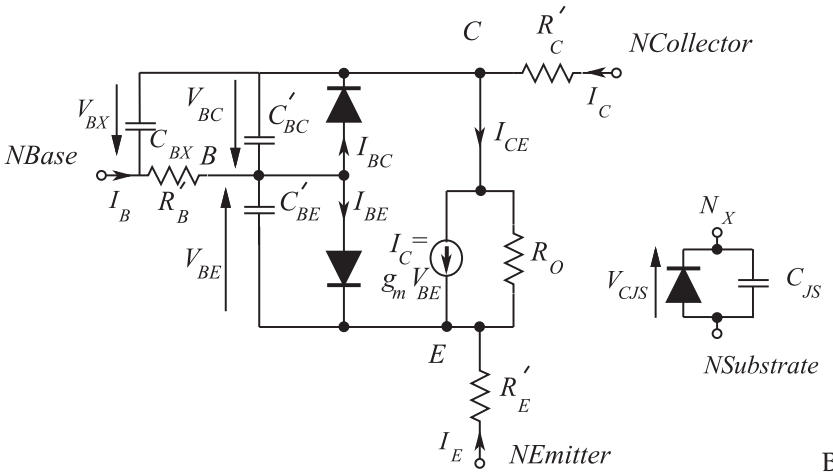
Figure 1-11: Q — bipolar junction transistor: (a) npn transistor; (b) pnp transistor.

Bipolar transistor models are based on the Gummel–Poon model [7] described here. The key feature of the model is that it captures the dependence of the forward and reverse current gain on current. In essence, the BJT model is a current-controlled current source. The Gummel–Poon model and its derivatives are used to model silicon BJTs and compound semiconductor HBTs [23, 24].

Table 1-6: Gummel–Poon BJT model parameters

Name	Description	Units	Default
AREA	Current multiplier		1.0
BF	Ideal maximum forward beta (B_F)		100.0
BR	Ideal maximum reverse beta (B_R)		1.0
C2	Base-emitter leakage saturation coefficient		I_{SE}/I_S
C4	Base-collector leakage saturation coefficient		(I_{SC}/I_S)
CJC	Base collector zero bias p-n capacitance (C_{JC})	F	0.0

Name	Description	Units	Default
CJE	Base emitter zero bias p-n capacitance (C_{JE})	F	0.0
EG	Bandgap voltage (E_G)	eV	1.11
FC	Forward bias depletion capacitor coefficient (F_C)		0.5
IKF	Corner of forward beta high-current roll-off (I_{KF})	A	10^{-10}
IKR	Corner for reverse-beta high current roll off (I_{KR})		10^{-10}
IS	Transport saturation current (I_S)	A	10^{-16}
ISC	Base collector leakage saturation current (I_{SC})	A	0.0
ISE	Base-emitter leakage saturation current (I_{SE})	A	0.0
IRB	Current at which RB falls to half of R_{BM} (I_{RB})	A	10^{-10}
ITF	Transit time dependency on IC (I_{TF})	A	0.0
MJC	Base collector p-n grading factor (M_{JC})		0.33
MJE	Base emitter p-n grading factor (M_{JE})		0.33
NC	Base-collector leakage emission coefficient (N_C)		2.0
NE	Base-emitter leakage emission coefficient (N_E)		1.5
NF	Forward current emission coefficient (N_F)		1.0
NR	Reverse current emission coefficient (N_R)		1.0
RB	Zero bias base resistance (R_B)	Ω	0.0
RBM	Minimum base resistance (R_{BM})	Ω	R_B
RE	Emitter ohmic resistance (R_E)	Ω	0.0
RC	Collector ohmic resistance (R_C)	Ω	0.0
T	Operating Temperature T	K	300
TF	Ideal forward transit time (T_S)	secs	0.0
TNOM	Nominal temperature (T_{NOM})	K	300
TR	Ideal reverse transit time (T_R)	S	0.0
TRB1	RB temperature coefficient (linear) (T_{RB1})		0.0
TRB2	RB temperature coefficient (quadratic) (T_{RB2})		0.0
TRC1	RC temperature coefficient (linear) (T_{RC1})		0.0
TRC2	RC temperature coefficient (linear) (T_{RC2})		0.0
TRE1	RE temperature coefficient (linear) (T_{RE1})		0.0
TRE2	RE temperature coefficient (quadratic) (T_{RE2})		0.0
TRM1	RBM temperature coefficient (linear) (T_{RM1})		0.0
TRM2	RBM temperature coefficient (quadratic) (T_{RM2})		0.0
VA	Alternative keyword for VAF (V_A)	V	10^{-10}
VAF	Forward early voltage (V_{AF})	V	10^{-10}
VAR	Reverse early voltage (V_{AR})		10^{-10}
VB	alternative keyword for VAR (V_B)		10^{-10}
VJC	Base collector built in potential (V_{JC})	V	0.75
VJE	Base emitter built in potential (V_{JE})	V	0.75
VTF	Transit time dependency on VBC (V_{TF})	V	10^{-10}
XCJC	Fraction of CBC connected internal to RB (X_{CJC})		1.0
XTB	Forward and reverse beta temperature coefficient (X_{TB})		0.0
XTF	Transit time bias dependence coefficient (X_{TF})		0.0
XTI	IS temperature effect exponent (X_{TI})		3.0



BJT model schematic.

Standard Calculations

The physical constants used in the model evaluation are

k	the Boltzmann constant	$1.3806226\,10^{-23}$ J/K
q	electronic charge	$1.6021918\,10^{-19}$ C

Absolute temperatures (in kelvin, K) are used. The thermal voltage is

$$V_{\rm TH}(T_{\rm NOM}) = k\,T_{\rm NOM}/q. \tag{1.71}$$

Current characteristics:
The base-emitter current is

$$I_{BE} = I_{BF}/\beta_F + I_{LE}. \tag{1.72}$$

the base-collector current is

$$I_{BC} = I_{BR}/\beta_R + I_{LC}. \tag{1.73}$$

The collector-emitter current is

$$I_{CE} = I_{BF} - I_{BR}/K_{QB}, \tag{1.74}$$

where the forward diffusion current is

$$I_{BF} = I_S\left(\mathrm{e}^{V_{BE}/(N_F V_{\rm TH})} - 1\right). \tag{1.75}$$

The nonideal base-emitter current is

$$I_{LE} = I_{SE}\left(\mathrm{e}^{V_{BE}/(N_E V_{\rm TH})} - 1\right). \tag{1.76}$$

The reverse diffusion current is

$$I_{BR} = I_S\left(\mathrm{e}^{V_{BC}/(N_R V_{\rm TH})} - 1\right). \tag{1.77}$$

The nonideal base-collector current is

$$I_{LC} = I_{SC}\left(\mathrm{e}^{V_{BC}/(N_C V_{\rm TH})} - 1\right). \tag{1.78}$$

The base charge factor is

$$K_{QB} = 1/2\left[1 - V_{BC}/V_{AF} - V_{BE}/V_{AB}\right]^{-1}\left(1 + \sqrt{1 + 4\left(I_{BF}/I_{KF} + I_{BR}/I_{KR}\right)}\right). \quad (1.79)$$

Thus the conductive current flowing into the base is

$$I_B = I_{BE} + I_{BC}, \quad (1.80)$$

the conductive current flowing into the collector is

$$I_C = I_{CE} - I_{BC}, \quad (1.81)$$

and the conductive current flowing into the emitter is

$$I_E = I_{BE} + I_{CE}. \quad (1.82)$$

Capacitances

$C_{BE} = \text{Area}(C_{BE\tau} + C_{BEJ})$, where the base-emitter transit time or diffusion capacitance is

$$C_{BE\tau} = \tau_{F,\text{EFF}}\left(\partial I_{BF}/\partial V_{BE}\right) \quad (1.83)$$

and the effective base transit time is empirically modified to account for base punchout, space-charge limited current flow, quasi-saturation, and lateral spreading, which tend to increase τ_F:

$$\tau_{F,\text{EFF}} = \tau_F\left[1 + X_{TF}(3x^2 - 2x^3)\text{e}^{(V_{BC}/(1.44V_{TF})}\right], \quad (1.84)$$

and $x = I_{BF}/(I_{BF} + \text{Area}I_{TF})$.

The base-emitter junction (depletion) capacitance is

$$C_{BEJ} = \begin{cases} C_{JE}\left(1 - V_{BE}/V_{JE}\right)^{-M_{JE}} & V_{BE} \le F_C V_{JE} \\ C_{JE}\left(1 - F_C\right)^{-(1+M_{JE})}\left(1 - F_C(1+M_{JE}) + M_{JE}V_{BE}/V_{JE}\right) & V_{BE} > F_C V_{JE}. \end{cases} \quad (1.85)$$

The base-collector capacitance is $C_{BC} = \text{Area}(C_{BC\tau} + X_{CJC}C_{BCJ})$, where the base-collector transit time or diffusion capacitance is

$$C_{BC\tau} = \tau_R \partial I_{BR}/\partial V_{BC}. \quad (1.86)$$

The base-collector junction (depletion) capacitance is

$$C_{BCJ} = \begin{cases} C_{JC}\left(1 - V_{BC}/V_{JC}\right)^{-M_{JC}} & V_{BC} \le F_C V_{JC} \\ C_{JC}\left(1 - F_C\right)^{-(1+M_{JC})}\left(1 - F_C(1+M_{JC}) + M_{JC}V_{BC}/V_{JC}\right) & V_{BC} > F_C V_{JC}. \end{cases} \quad (1.87)$$

The capacitance between the extrinsic base and the intrinsic collector is

$$C_{BX} = \begin{cases} \text{Area}(1 - X_{CJC})C_{JC}\left(1 - V_{BX}/V_{JC}\right)^{-M_{JC}} & V_{BX} \le F_C V_{JC} \\ (1 - X_{CJC})C_{JC}\left(1 - F_C\right)^{-(1+M_{JC})} & V_{BX} > F_C V_{JC} \\ \quad \times \left(1 - F_C(1+M_{JC}) + M_{JC}V_{BX}/V_{JC}\right) & \end{cases}. \quad (1.88)$$

The substrate junction capacitance is

$$C_{JS} = \begin{cases} \text{Area}C_{JS}\left(1 - V_{CJS}/V_{JS}\right)^{-M_{JS}} & V_{CJS} \le 0 \\ \text{Area}C_{JS}\left(1 + M_{JS}V_{CJS}/V_{JS}\right) & V_{CJS} > 0. \end{cases} \quad (1.89)$$

CHAPTER 2

Linear Amplifiers

2.1 Introduction 25
2.2 Linear Amplifier Design Strategies 26
2.3 Amplifier Gain Definitions 26
2.4 Amplifier Efficiency 37
2.5 Class A, AB, B, and C Amplifiers 38
2.6 Amplifier Stability 44
2.7 Amplifier Noise 55
2.8 Trading Off Gain, Noise, and Stability in Amplifier Design 56
2.9 Case Study: Narrowband Linear Amplifier Design 57
2.10 Summary 63
2.11 References 64
2.12 Exercises 65

2.1 Introduction

Amplifiers increase the power of an RF signal by converting DC power to AC power. Amplifiers can be optimized for low noise, moderate to high gain, high efficiency, low distortion, or specific output power. At the same time stability must be assured, which is a problem with feedback due to parasitics and the internal feedback of transistors. However, it is not possible to optimize all of the parameters simultaneously. This has led to several amplifier design strategies and amplifier topologies trading off design complexity and performance. In this chapter the major active devices and linear amplifiers based on them are examined. A critical common aspect is minimizing noise, maximizing the efficiency of power conversion to RF, and ensuring stability.

A common characteristic of linear amplifier design, the subject of this chapter, is that the operation of the amplifier at RF is similar to its operation at low frequencies. The design strategy is based on small signal design and it is sufficient to use the small-signal S parameters of a device at the chosen operating point. In contrast, the design of power amplifiers considered in a future chapter is more complicated, as the amplifier operation is dependent on the signal level and design must use the large signal model of transistors and use nonlinear simulators.

2.2 Linear Amplifier Design Strategies

Linear amplifier design requires that the transistor(s) be biased in a high-gain region and that input and output matching networks be used to provide good power transfer at the input and output of the transistor stages. This circuit arrangement is shown in Figure 2-1. The DC bias control circuit is fairly standard; it does not involve any microwave constraints. The lowpass filters (in the bias circuits) can have one of several forms and are often integrated into the input and output matching networks. Synthesis of the input and output matching networks (and occasionally a feedback network required for stability and broadband operation) is the primary objective of any amplifier design.

RF transistors used to amplify small signals should have high maximum available gain and low noise characteristics. For transistors used in transmitters, where the efficient generation of power is critical, it is important to linear amplifier design that the transistor characteristics be close to linear in the central region of the output current-voltage characteristics so that distortion is minimized. The ultimate limit on output power is determined by the breakdown voltage at high drain-source voltages and also by the maximum current density that can be supported. Finally, for efficient amplification of large signals, the knee voltage (where the current-voltage curves bend over and starts to flatten) should be low.

Manufacturers of discrete transistors and amplifier modules provide substantial information, including S parameters and, in some cases, reference designs. An extract from the datasheet of a pHEMT transistor is shown in Figure 2-2. The intended application is provided and the device structure has been optimized for the application.

Design examples presented in the next few sections will use the pHEMT transistor documented in Figure 2-2. This discrete transistor is described as a low-noise, high-frequency, packaged pHEMT that can be used in amplifiers operating at up to 18 GHz. It shares a common characteristic of FET devices in that S_{21} is highest at low frequencies and the feedback parameter, S_{12}, is lowest at low frequencies. This means that gain is harder to achieve at higher frequencies and the higher-level feedback means that stability is often a problem at higher frequencies. However, the loop gain described by $S_{21}S_{12}$ is large at low frequencies so stability is also a problem at low frequencies.

2.3 Amplifier Gain Definitions

As with all circuit design, a few figures of merit (FOMs) are used to guide design. The most important metric in amplifier design is the gain

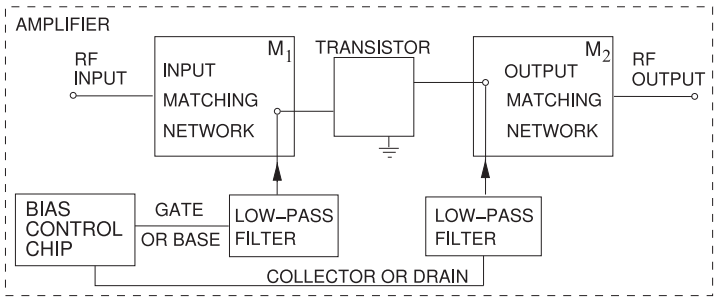


Figure 2-1: Block diagram of an RF amplifier including biasing networks.

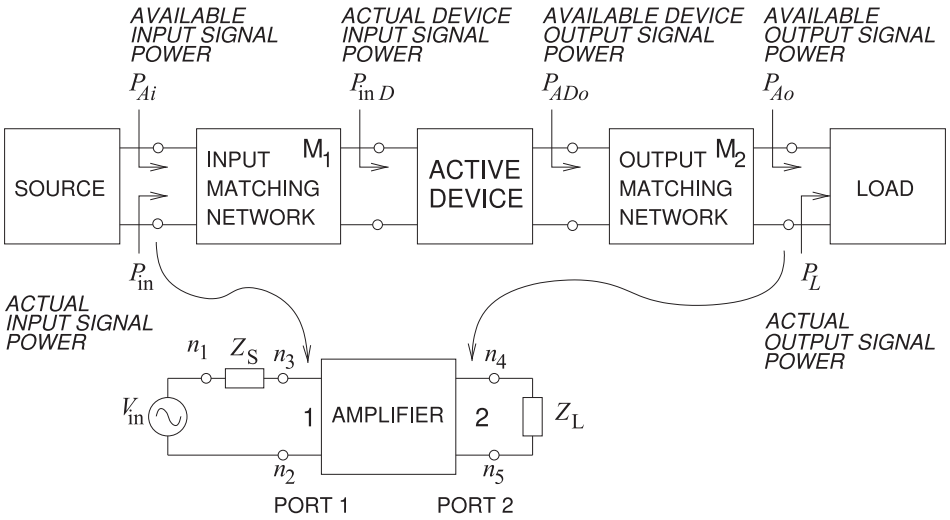
Data Sheet Extract.

Transistor technology:	Depletion-mode pHEMT.
Model:	FPD6836P70 from QORVO, Inc.
Description:	Low-noise, high-frequency packaged pHEMT.
	Optimized for low-noise, high-frequency applications.
Synopsis:	22 dBm output power (P1dB).
	15 dB power gain (G1dB) at 5.8 GHz, Usable gain to 18 GHz.
	0.8 dB noise figure at 5.8 GHz, 32 dBm output IP3 at 5.8 GHz.
	45% power-added efficiency at 5.8 GHz.
	Usable gain to 18 GHz.

Frequency (GHz)	$\|S_{11}\|$	$\angle S_{11}$ degrees	$\|S_{21}\|$	$\angle S_{21}$ degrees	$\|S_{12}\|$	$\angle S_{12}$ degrees	$\|S_{22}\|$	$\angle S_{22}$ degrees
0.500	0.976	−20.9	11.395	161.5	0.011	78.3	0.635	−11.5
1.000	0.925	−41.3	10.729	145.1	0.021	67.8	0.614	−22.2
2.000	0.796	−78.2	8.842	116.7	0.034	51.4	0.553	−37.9
3.000	0.694	−106.8	7.180	94.5	0.041	40.4	0.506	−48.9
4.000	0.614	−127.3	6.002	76.7	0.044	33.9	0.475	−57.7
5.000	0.555	−147.0	5.249	60.3	0.048	28.4	0.453	−66.4
6.000	0.511	−170.2	4.729	43.7	0.052	23.3	0.438	−76.0
7.000	0.493	163.9	4.261	26.8	0.057	14.0	0.391	−87.6
8.000	0.486	140.4	3.784	11.2	0.057	6.4	0.340	−99.1
9.000	0.473	122.5	3.448	−2.4	0.059	5.2	0.332	−109.6
10.000	0.488	103.4	3.339	−17.3	0.073	0.9	0.355	−124.8
11.000	0.539	79.8	3.166	−35.0	0.086	−10.1	0.349	−145.6
12.000	0.626	60.8	2.877	−51.9	0.095	−21.4	0.307	−169.6
13.000	0.685	47.6	2.604	−68.2	0.100	−32.5	0.295	165.3
14.000	0.724	36.2	2.392	−83.8	0.106	−43.3	0.312	142.7
15.000	0.787	20.9	2.225	−99.7	0.109	−55.1	0.320	125.4
16.000	0.818	5.2	2.067	−116.6	0.112	−68.4	0.340	103.9
17.000	0.831	−9.6	1.855	−134.4	0.108	−83.5	0.373	76.1
18.000	0.852	−19.5	1.603	−148.6	0.103	−94.2	0.406	54.7
19.000	0.815	−20.5	1.440	−159.3	0.102	−103.0	0.449	43.1
20.000	0.780	−26.8	1.382	−171.2	0.106	−113.5	0.460	37.9
21.000	0.779	−46.8	1.333	171.2	0.109	−130.7	0.438	31.4
22.000	0.786	−62.1	1.195	152.0	0.110	−148.4	0.417	6.0
23.000	0.774	−70.1	1.073	137.2	0.108	−162.4	0.428	−16.5
24.000	0.744	−81.7	1.025	123.5	0.112	−175.2	0.433	−29.0
25.000	0.704	−90.9	1.061	107.3	0.132	170.0	0.396	−46.5
26.000	0.677	−111.1	1.065	85.8	0.148	147.8	0.298	−71.0

Figure 2-2: Scattering parameters of an enhancement mode pHEMT transistor biased at $V_{DS} = 5$ V, $I_D = 55$ mA, $V_{GS} = -0.42$ V. Extract from the data sheet of the FPD6836P70 discrete transistor [1].

of the overall amplifier. There are a surprisingly large number of different definitions of gain that are useful at different stages in the design process. Each provides information about the performance of an amplifier and using the full set enables design to be approached in a systematic way. The FOMs are used to describe the performance of an amplifier, to develop an understanding of the active device, to compare different active devices, and,



Power	Description
P_{in}	Actual input power delivered to the amplifier.
P_{Ai}	Available input power from the source. $P_{\text{in}} \leq P_{Ai}$. If M_1 provides conjugate matching as seen from the source, then $P_{\text{in}} = P_{Ai}$.
$P_{\text{in}D}$	Actual device input power delivered to the active device. $P_{\text{in}D} \leq P_{\text{in}}$. If M_1 is lossless, $P_{\text{in}D} = P_{\text{in}}$.
P_{ADo}	Available device output power of the active device.
P_{Ao}	Available amplifier output power. $P_{Ao} \leq P_{ADo}$. If M_2 is lossless, $P_{Ao} = P_{ADo}$.
P_L	Actual output power delivered to load. Amplifier output power. $P_L \leq P_{Ao}$. If M_2 is lossless and provides conjugate matching, $P_L = P_{Ao} = P_{ADo}$.

Figure 2-3: Parameters used in defining gain measures. The input and output matching networks are lossless so that the actual device input signal power, $P_{\text{in}D}$, is the power delivered by the source. Similarly the actual output signal power delivered to the load, P_L, is the power delivered by the active device (the transistor including biasing network).

coupled with experience, to formulate an idea of how difficult a design will be.

The quantities used in the various gain definitions are defined in Figure 2-3. The power delivered to the amplifier is P_{in}, and this is equal to the available input power from the source if the source is conjugately matched to the input matching network. The power delivered to the active device, $P_{\text{in}D}$, is equal to the amplifier input power, P_{in}, if the input matching network is lossless. The available output power from the active device, P_{Ao}, is the actual device output power, P_o, delivered to the output matching network if the output of the active device is conjugately matched. This power is also delivered to the load as P_L if M_2 is lossless. In summary,

$P_{\text{in}} = P_{Ai}$, if the generator is conjugately matched
$P_{\text{in}} = P_{\text{in}D}$, if M_1 is lossless
$P_o = P_{Ao}$, if the output of active device is conjugately matched
$P_L = P_o$, if M_2 is lossless.

These power definitions refer to different circuit conditions. This enables a number of different gain definitions to be developed that relate to different stages in the development of an amplifier and indicate the ultimate performance achievable from an amplifier. The basic gain definitions are

- ***System gain***:

$$G = \frac{P_L}{P_{\text{in}}}. \tag{2.1}$$

  The system gain is the power actually delivered to the load relative to the input power delivered by the source. This gain is sometimes called the **actual power gain**.

- ***Power gain***:

$$G_P = \frac{P_L}{P_{\text{in},D}}. \tag{2.2}$$

  This gain is G, but with the loss of M_1 removed.

- ***Transducer gain***:

$$G_T = \frac{P_L}{P_{Ai}}. \tag{2.3}$$

  This is the ratio of the power delivered to the load divided by the power available from the source. This is the gain that really matters, the power actually delivered to the load relative to the power available from the source.

- ***Available gain***:

$$G_A = \frac{P_{Ao}}{P_{Ai}}. \tag{2.4}$$

  The transducer gain is the power available to the load relative to the input power available from the source. This gain is G_T with optimum M_2. That is, G_A is the system gain G with lossless M_1 and M_2 both optimized for maximum power transfer.

These gains are measures that can be used to characterize the performance of an amplifier but do not guide design. The development of guidelines begins by developing expressions for gains using the device's S parameters and then considering gain under idealized conditions such as optimum matching networks or with the device adjusted using feedback so that it is effectively unilateral.

EXAMPLE 2.1 Amplifier Gain

A source that drives an amplifier has an available output power of 1 mW. However, the load of the amplifier is mismatched so that the load reflection coefficient is 1 dB. The power actually delivered to the load of the amplifier is 1 W. Is it possible to determine the system gain? If so, what is it in decibels?

Solution:

There are many gain definitions so the information provided must be examined. The power delivered to the load is $P_L = 1$ W, and the available input power $P_{Ai} = 1$ mW. Examining the gains defined in Equations (2.1)–(2.4), Equation (2.3) yields

$$G_T = \frac{P_L}{P_{Ai}} = \frac{1\text{W}}{1\text{ mW}} = 1000 = 30\text{ dB}. \tag{2.5}$$

No other gain can be determined, i.e., the system gain cannot be determined as the actual input power is unknown.

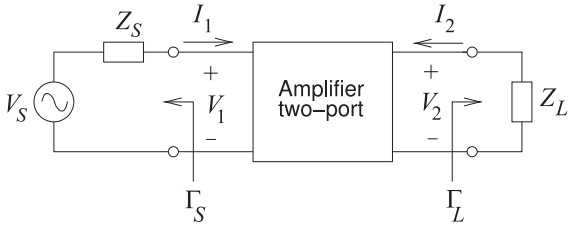
2.3.1 Gain in Terms of Scattering Parameters

This section develops the generalized scattering parameters of an amplifier and leads to expressions for gain in terms of the S parameters of the active device.

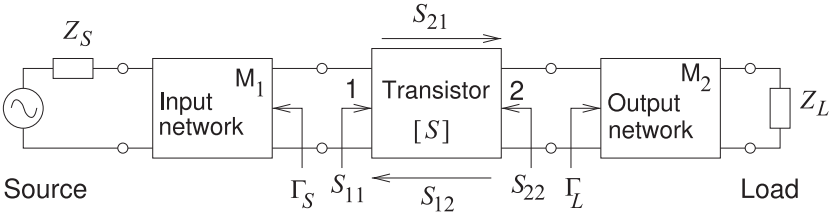
A linear amplifier can be represented as a two-port with a Thevenin equivalent source at Port 1 and a load at Port 2, as shown in Figure 2-4(a). This section illustrates the usefulness of generalized S parameters in working with power flow in systems with different system impedances at the ports. Let **S** be the normalized scattering matrix of the two-port, with Z_0 being the normalizing real characteristic impedance:

$$\mathbf{S} = [S] = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}. \tag{2.6}$$

The development in this section uses the normalized S parameters of the amplifier in Figure 2-4(a). The aim is to develop an expression for the



(a) An amplifier



(b) Active device with input and output matching networks

Figure 2-4: Two-port network with source and load used in defining unilateral gain measures.

unilateral transducer gain and for the maximum unilateral transducer gain. The unilateral transducer gain is restricted to the amplifier (Figure 2-4(a)), and the maximum unilateral transducer gain can use the S parameters of either the transistor (Figure 2-4(b)) or the amplifier (Figure 2-4(a)).

The generalized scattering matrix of the amplifier will be normalized to the source impedance, Z_S, and load impedance, Z_L, shown in Figure 2-4. First, the reflection coefficients Γ_S and Γ_L (normalized to Z_0) at the source and load are found using

$$\Gamma_S = \frac{Z_S - Z_0}{Z_S + Z_0} \quad \text{and} \quad \Gamma_L = \frac{Z_L - Z_0}{Z_L + Z_0}. \tag{2.7}$$

From Equation (2.132) of [2], the generalized scattering parameters (with Port 1 normalized to Z_S and Port 2 normalized to Z_L) are

$${}^G\mathbf{S} = (\mathbf{D}^*)^{-1}\,(\mathbf{S} - \mathbf{\Gamma}^*)(\mathbf{U} - \mathbf{\Gamma S})^{-1}\mathbf{D}, \tag{2.8}$$

where

$$\mathbf{\Gamma} = \begin{bmatrix} \Gamma_S & 0 \\ 0 & \Gamma_L \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} D_{11} & 0 \\ 0 & D_{22} \end{bmatrix}, \tag{2.9}$$

$$D_{11} = \frac{(1-\Gamma_S)\sqrt{1-|\Gamma_S|^2}}{|1-\Gamma_S^*|} \quad \text{and} \quad D_{22} = \frac{(1-\Gamma_L)\sqrt{1-|\Gamma_L|^2}}{|1-\Gamma_L^*|}. \tag{2.10}$$

Following tedious algebraic manipulations, the following expressions are obtained:

$${}^GS_{11} = \frac{1}{W}\frac{1-\Gamma_S}{1-\Gamma_S^*}[(S_{11} - \Gamma_S^*)(1-\Gamma_L S_{22}) + S_{12}S_{21}\Gamma_L] \tag{2.11}$$

$${}^GS_{12} = \frac{1}{W}\frac{(1-\Gamma_S)(1-\Gamma_L)}{|1-\Gamma_S||1-\Gamma_L|}S_{12}[(1-|\Gamma_S|^2)(1-|\Gamma_L|^2)]^{\frac{1}{2}} \tag{2.12}$$

$${}^GS_{21} = \frac{1}{W}\frac{(1-\Gamma_S)(1-\Gamma_L)}{|1-\Gamma_S||1-\Gamma_L|}S_{21}[(1-|\Gamma_S|^2)(1-|\Gamma_L|^2)]^{\frac{1}{2}} \tag{2.13}$$

$${}^GS_{22} = \frac{1}{W}\frac{1-\Gamma_L}{1-\Gamma_L^*}[(S_{22} - \Gamma_L^*)(1-\Gamma_S S_{11}) + S_{12}S_{21}\Gamma_S], \tag{2.14}$$

where

$$W = (1-\Gamma_S S_{11})(1-\Gamma_L S_{22}) - S_{12}S_{21}\Gamma_S\Gamma_L \tag{2.15}$$

and

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = {}^G\mathbf{S}\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}. \tag{2.16}$$

Here a and b are the root power waves defined in Figure 2-17 of [2].

A number of useful observations can be made. First, in a matched condition where $\Gamma_S = \Gamma_L = 0$ (i.e., $Z_S = Z_L = Z_0$), ${}^G\mathbf{S} = \mathbf{S}$. Second, for a reciprocal two-port network with $S_{12} = S_{21}$, the generalized scattering parameters are also reciprocal (i.e., ${}^GS_{12} = {}^GS_{21}$). An amplifier is not reciprocal, however, and for a good amplifier, ${}^GS_{12}$ is approximately zero and $\left|{}^GS_{21}\right|$ is greater than one, indicating power gain.

Now the transducer power gain, G_T, can be expressed in terms of device S parameters. G_T (Equation (2.3)) is defined as the ratio of the average power, P_L, delivered to the load Z_L, and the maximum input power available from the generator, P_{Ai}, that is,

$$G_T = \frac{P_L}{P_{Ai}}, \tag{2.17}$$

where the available power from the generator is

$$P_{Ai} = \frac{1}{8}\frac{|V_S|^2}{\Re\{Z_S\}} = \frac{1}{2}|a_1|^2 \tag{2.18}$$

and the power delivered to the load is

$$P_L = \frac{1}{2}\Re\{Z_L\}|-I_2|^2 = \frac{1}{2}|b_2|^2. \tag{2.19}$$

These quantities are defined in terms of the root power waves a_1 and b_2 and these are related by the generalized S parameters. Thus the transducer gain

$$G_T = \left|\frac{b_2}{a_1}\right|^2 = |^G S_{21}|^2, \tag{2.20}$$

and so, using Equation (2.13),

$$G_T = |S_{21}|^2 \frac{(1-|\Gamma_S|^2)(1-|\Gamma_L|^2)}{|(1-\Gamma_S S_{11})(1-\Gamma_L S_{22}) - \Gamma_S \Gamma_L S_{12} S_{21}|^2}. \tag{2.21}$$

This combines the inherent voltage gain of the device (S_{21}) with the effect of load and source mismatches through Γ_L and Γ_S. This expression simplifies under particular circumstances that will now be considered.

If the source and load impedances are both equal to the system impedance, then the transducer gain becomes

$$G_T|_{Z_0} = |S_{21}|^2. \tag{2.22}$$

This is the transducer gain without matching networks, so that $\Gamma_S = 0 = \Gamma_L$. This is nearly always much lower than what can be achieved using matching networks. For one, an active device has significant input and output reactances at microwave frequencies that need to be tuned out.

For a unilateral two-port, $S_{12} = 0$ (and this is a reasonable approximation for many amplifiers, as there is little feedback from the output to the input). Then the transducer gain becomes the **unilateral transducer gain**, G_{TU}:

$$G_{TU} = |S_{21}|^2 \left(\frac{1-|\Gamma_S|^2}{|1-\Gamma_S S_{11}|^2}\right)\left(\frac{1-|\Gamma_L|^2}{|1-\Gamma_L S_{22}|^2}\right). \tag{2.23}$$

This is often referred to as just the **unilateral gain**.

From the last expression it can be seen that by choosing $\Gamma_S = (S_{11})^*$ and $\Gamma_L = (S_{22})^*$ (i.e., the complex conjugates), G_{TU} achieves its maximum value, the **maximum unilateral transducer gain**:

$$G_{TU\text{max}} = |S_{21}|^2 \left(\frac{1}{1-|S_{11}|^2}\right)\left(\frac{1}{1-|S_{22}|^2}\right). \tag{2.24}$$

Frequency (GHz)	$G_{TU\max}$ (dB)	Frequency (GHz)	$G_{TU\max}$ (dB)
0.5	36.62	14	11.25
1	31.07	15	11.61
2	24.88	16	11.64
3	21.26	17	11.11
4	18.73	18	10.50
5	16.00	19	8.89
6	15.74	20	7.91
7	14.52	21	7.48
8	13.26	22	6.55
9	12.36	23	5.46
10	12.24	24	4.62
11	12.07	25	4.22
12	11.77	26	3.61
13	11.46		

Table 2-1: Maximum unilateral transducer gain, $G_{TU\max}$, of the pHEMT transistor documented in Figure 2-2.

Note that up to now the S parameters have been those of the transistor, see Figure 2-4(b). So $G_{TU\max}$ is the maximum unilateral transducer gain available from the active device. This is a good measure of the maximum power gain readily obtained from the device. However, with feedback (consider the general amplifier configuration of Figure 2-17) the effective $S_{12} \neq 0$, and any gain can be achieved, even oscillation. Also, higher gain is obtained at the expense of reduced bandwidth. As a general design guideline, the closer the gain specified for an amplifier is to $G_{TU\max}$, the more challenging the design task.

The maximum unilateral transducer gain, $G_{TU\max}$, of the pHEMT transistor described in Figure 2-2 in shown in Table 2-1. The maximum unilateral transducer gain is largest at low frequencies and monotonically reduces as frequency increases. This means that it is difficult to design a broadband amplifier using a pHEMT (and this is true with most FETs). It is also a challenge to ensure stable amplification at low frequencies and matching networks must be chosen to suppress low-frequency gain.

The power gain with the input conjugately matched is

$$G_P = \frac{|S_{21}|^2(1-|\Gamma_L|^2)}{|(1-S_{22}\Gamma_L)(1-|S_{11}|^2)}, \tag{2.25}$$

and with $\Gamma_L = 0$,

$$G_P|_{\Gamma_L=0} = \frac{|S_{21}|^2}{(1-|S_{11}|^2)}. \tag{2.26}$$

The available power gain with the output conjugately matched is

$$G_A = \frac{|S_{21}|^2(1-|\Gamma_S|^2)}{|1-S_{11}\Gamma_S|^2(1-|S_{22}|^2)}, \tag{2.27}$$

and with $\Gamma_S = 0$,

$$G_A|_{\Gamma_S=0} = \frac{|S_{21}|^2}{1-|S_{22}|^2}. \tag{2.28}$$

The maximum available power gain, G_{MA}, equal to both G_A and G_T with optimum M_1 and M_2 is

$$G_{\text{MA}} = \left|\frac{S_{21}}{S_{12}}\right| \left(k - \sqrt{k^2 - 1}\right), \tag{2.29}$$

where the **Rollet's stability factor** [3, 4]

$$k = \left(\frac{1 - |S_{11}|^2 - |S_{22}|^2 + |\Delta|^2}{2|S_{12}|\,|S_{21}|}\right) \quad \text{and} \quad \Delta = S_{11}S_{22} - S_{12}S_{21}. \tag{2.30}$$

Rollett developed the expressions for G_{MA} and k in terms of z, y, and h parameters [3, 4]. The expressions in Equations (2.29) and (2.30) are developed using the equivalences given in Table 2-2 of [2].

G_{MA} is only defined when $k \geq 1$ indicating that if $k < 1$ the amplifier can be expected to oscillate if both M_1 and M_2 are optimized for maximum power transfer. Provided that $k \geq 1$, G_{MA} is the maximum gain that can be achieved without using feedback from the output of the active device to its input. It is still possible to obtain stable gain when $k < 1$ but it is necessary to use either nonoptimum M_1 or M_2, or feedback across the active device.

Another important gain metric is the maximum stable gain G_{MS} that can be achieved. Of course this is just G_{MA} if the amplifier is unconditionally stable. If $k < 1$, Rollett showed that the amplifier is potentially unstable (this will be discussed in greater detail in Section 2.6) [3, 4]. Then G_{MS} will be less than G_{MA}. Rollett ensured stability of the amplifier by putting shunt admittances at the input and output ports of the transistor so that for the augmented transistor two-port $k = 1$. Then the maximum stable power gain, G_{MS}, is G_{MA} in Equation 2.29 but with k set to 1:

$$G_{\text{MS}} = \left|\frac{S_{21}}{S_{12}}\right|. \tag{2.31}$$

From Table 2-2 of [2] it is seen that $S_{21}/S_{12} = z_{21}/z_{12} = y_{21}/y_{12} = h_{21}/h_{12}$ so that the surprising result is that the maximum stable gain, a power gain, is the ratio of the forward parameter to the reverse parameter and not the square of the ratio, and those parameters could be S, z, y, or h parameters.

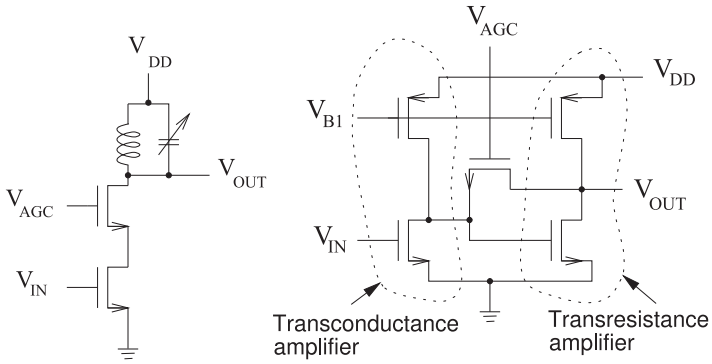
Experience indicates that G_{MS} is the practical limit to the stable gain that can be achieved with moderate design effort. (Note that the amplifier could be stable without being conditionally stable which is the criterion used in developing G_{MS}, and schemes other than using shunt admittances could be used to achieve unconditional stability). G_{MS} is interpreted as the maximum gain that can be achieved while ensuring that the amplifier is unconditionally stable.

The final useful gain metric is the unilateral power gain, U. This is the maximum available gain, G_{MA}, with feedback across the active device adjusted so that the effective device feedback parameter $S_{12} = 0$. When $U = 1$, the devices go from being active to being passive [5]:

$$U = \left(\frac{|S_{21}/S_{12} - 1|^2}{2k|S_{21}/S_{12}| - 2\Re(S_{21}/S_{12})}\right). \tag{2.32}$$

2.3.2 Design Using Gain Metrics

$G_{TU,\max}$, G_{MS}, G_{MA}, and U are used by designers as measures of the ultimate performance of a device and to guide design. They are important FOMs



(a) Tuned amplifier (b) Wideband amplifier

Figure 2-5: Low-noise amplifiers. V_{AGC} is the automatic gain control voltage and sets the gain of the amplifiers.

as they are defined with extreme conditions. Of these metrics only G_{MA} relates to the unaltered device S parameters. G_{MA} is the maximum gain that can be obtained with optimum input and output matching networks. If the amplifier is potentially unstable with optimum M_1 and M_2, then G_{MA} is undefined.

$G_{TU,\max}$ is the gain when the effective S_{12} is set to zero (perhaps using feedback) but instead of optimum M_1 and M_2, Γ_S is set to S_{11}^* and Γ_L is set to S_{22}^*. (S_{11} and S_{22} are those of the active device and are unaffected by feedback.) With $S_{12} = 0$, the choice for Γ_L is the same as using an optimum M_2. However the choice for Γ_S (i.e. $\Gamma_S = S_{11}^*$) is not the same as using an optimum M_1. Choosing an optimum M_1 would lead to a higher system gain (i.e., $G > G_{TU,\max}$). Experience is that $G_{TU,\max}$ can be achieved with little design effort. It is used primarily in initial choice of the active device.

G_{MS} and U are the maximum available gains for two different specific although artificial conditions. G_{MS} is the maximum available gain with k set 1, perhaps using feedback, but otherwise not changing the device's S parameters. Design experience is that G_{MS} is the stable gain that can be achieved with moderate design effort. U is the maximum available gain with S_{12} set to 1, perhaps using feedback, but otherwise not changing the device's S parameters. The design experience is that U indicates the highest frequency at which gain can be achieved and this is when $U = 1$.

$G_{TU,\max}$, G_{MS}, G_{MA}, and U are tabulated in Table 2-2 for the pHEMT transistor documented in Figure 2-2. As will be shown in Section 2.6.3, the amplifier is unconditionally stable from 5 to 11 GHz and above 22 GHz. Outside those ranges, matching networks can be chosen so that the amplifier could oscillate and then G_{MA} is not defined. G_{MS} is usually taken as the highest gain that can be easily achieved. However, higher gains can be achieved with more attention to stability, but then usually amplification is available only over a very narrow bandwidth.

Once a design is completed, the only gain that matters is the transducer gain, G_T, which is the ratio of the power delivered to a load to the power available from the source.

2.3.3 Gain Circles

The expressions for gains developed in Section 2.3.1 were in terms of absolute values of complex numbers. It is therefore possible to present gains at a particular frequency using circles on the complex reflection coefficient

Table 2-2: Device gain metrics for the pHEMT transistor in Figure 2-2.

Freq. (GHz)	Max. unilateral transducer gain $G_{TU,\max}$ (db)	Max. available power gain G_{MA} (db)	Max. stable power gain G_{MS} (db)	Unilateral power gain U (db)
0.5	36.6	–	30.1	41.9
1	31.1	–	27.1	39.4
2	24.9	–	24.2	34.4
3	21.3	–	22.4	29.7
4	18.7	–	21.3	25.6
5	17.0	18.6	20.4	23.6
6	15.7	17.0	19.6	22.6
7	14.5	15.5	18.7	20.8
8	13.6	14.0	18.2	17.9
9	12.4	13.0	17.7	16.4
10	12.2	13.2	16.6	17.8
11	12.1	13.7	15.7	19.5
12	11.8	–	14.8	20.5
13	11.5	–	14.2	20.5
14	11.2	–	13.5	21.2
15	11.6	–	13.1	27.6
16	11.6	–	12.7	24.2
17	11.1	–	12.3	21.4
18	10.5	–	11.9	17.0
19	8.88	–	11.5	12.9
20	7.91	–	11.2	11.4
21	7.48	–	10.9	11.0
22	6.55	8.42	10.4	9.22
23	5.46	6.62	9.97	7.43
24	4.62	5.55	9.62	6.22
25	4.23	5.17	9.05	5.73
26	3.61	4.27	8.57	4.79

plane [6]. The mathematics behind this are developed in Section 1.A.13 of [7].

Two of the more useful gains to use in making trade-offs are the maximum available gain G_{MA} (Equation (2.29)) and the maximum stable gain G_{MS} (Equation (2.31)). G_{MA} is the available G_A and G_T with optimum M_1 and M_2, but only has a finite value when the transistor is unconditionally stable. Microwave circuit simulators use another gain measure to handle this situation. If a transistor is conditionally stable then resistive loading is used to ensure unconditional stability. So the maximum available gain calculated in microwave simulators is G_{MA} if it has a finite value but is G_{MS} otherwise. Introducing G_{MAX} to describe this gain:

$$G_{\text{MAX}} = \begin{cases} G_{\text{MA}} = \left|\dfrac{S_{21}}{S_{12}}\right| \left(k - \sqrt{k^2-1}\right) & \text{if } k \geq 1 \\ G_{\text{MS}} = \left|\dfrac{S_{21}}{S_{12}}\right| & \text{if } k < 1 \end{cases} . \tag{2.33}$$

Here k is Rollet's stability factor (see Equation (2.30).

The discussion can now turn to defining gain circles. Gain circles plot the locus of the available power gain, G_A, as defined in Equation (2.27), on the input reflection coefficient, Γ_S, plane. G_A is a function of the magnitude of complex numbers and so circles can be defined in the complex plane. Figure

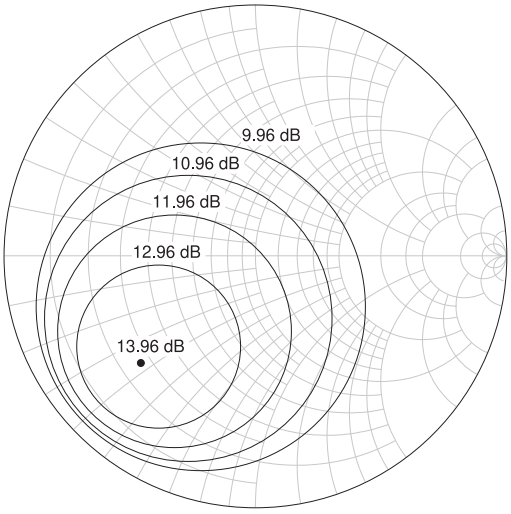


Figure 2-6: Gain circles of the transistor in Figure 2-2 at 8 GHz plotted on the Γ_S plane.

2-6 plots the G_A gain circles for the pHEMT transistor at 8 GHz on the input reflection coefficient plane. The center of the family of circles is G_{MAX} and this point defines the Γ_S value required to achieve G_{MAX}. The other circles moving out plot the locus of Γ_S for reductions of available power gain, G_A, in 1 dB steps below G_{MAX}. That is, a circle defines the values of Γ_S that will yield a specific G_A.

2.4 Amplifier Efficiency

There are several ways of expressing amplifier efficiency depending on how the RF input power is treated. One measure of efficiency of a circuit is the useful output power divided by the input power, and considers the contribution of the RF input power. This measure of efficiency is called **power-added efficiency** (**PAE**). At RF and microwave frequencies, the most common definition of PAE used with power amplifiers focuses on the additional RF power divided by the DC input power. Thus

$$\eta_{\text{PAE}} = \frac{P_{\text{RF,out}} - P_{\text{RF,in}}}{P_{\text{DC}}}. \tag{2.34}$$

There is another definition of PAE, but it is less commonly used at RF and microwave frequencies [8]. However, this alternative definition can be used with any two-port network. RF and microwave circuit designers refer to this as the **total power-added efficiency** and at lower frequencies it is called the **transmit chain efficiency**. This efficiency is denoted η_{total} and is defined as

$$\eta_{\text{total}} = \frac{P_{\text{RF,out}}}{P_{\text{DC}} + P_{\text{RF,in}}}. \tag{2.35}$$

RF engineers also refer to this as the **overall amplifier efficiency**:

$$\eta_{\text{overall}} = \eta_{\text{total}} = \frac{P_{\text{RF,out}}}{P_{\text{DC}} + P_{\text{RF,in}}}. \tag{2.36}$$

A final definition is the **average amplifier efficiency**, η_{avg} [9]. This metric takes into account the time-varying level of a modulated communications

Table 2-3: Comparison of efficiency metrics for an amplifier producing 1 W RF output power and consuming 2 W of DC power with various power gains.

Power gain (dB)	η_{TOTAL}	η_{PAE}	η_D
3	40%	25%	50%
6	44%	37%	50%
10	48%	45%	50%
15	49%	48%	50%
20	50%	50%	50%
40	50%	50%	50%

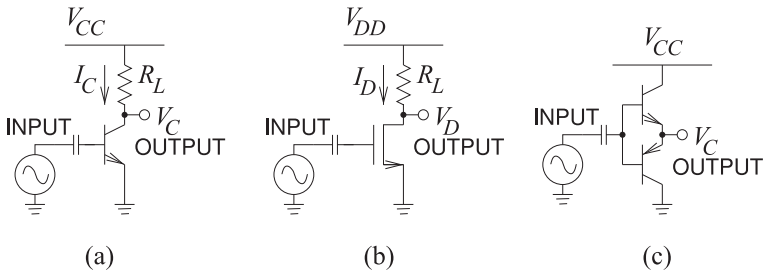


Figure 2-7: Class A single-ended resistively biased amplifiers: (a) BJT transistor with B for base terminal, C for collector terminal, and E for emitter terminal; (b) MOSFET transistor with G for gate terminal, D for drain terminal, and S for source terminal; and (c) Class B or Class C push-pull amplifier.

signal and is the ratio of the average RF output power to the average DC input power:

$$\eta_{\mathrm{avg}} = \frac{P_{\mathrm{RF,out,avg}}}{P_{\mathrm{DC,avg}}}. \tag{2.37}$$

For high-gain amplifiers, $P_{\mathrm{RF,in}} \ll P_{\mathrm{DC}}$, and all of the efficiencies become approximately equivalent and is simply called the **efficiency**, η, of the amplifier:

$$\eta = \frac{P_{\mathrm{RF,out}}}{P_{\mathrm{DC}}} \approx \eta_{\mathrm{PAE}} \approx \eta_{\mathrm{total}} \approx \eta_{\mathrm{avg}} \approx \eta_{\mathrm{overall}} \quad \text{(high gain)}. \tag{2.38}$$

When the primary input DC power is fed to the drain of a FET, the term **drain efficiency**, η_D, is used:

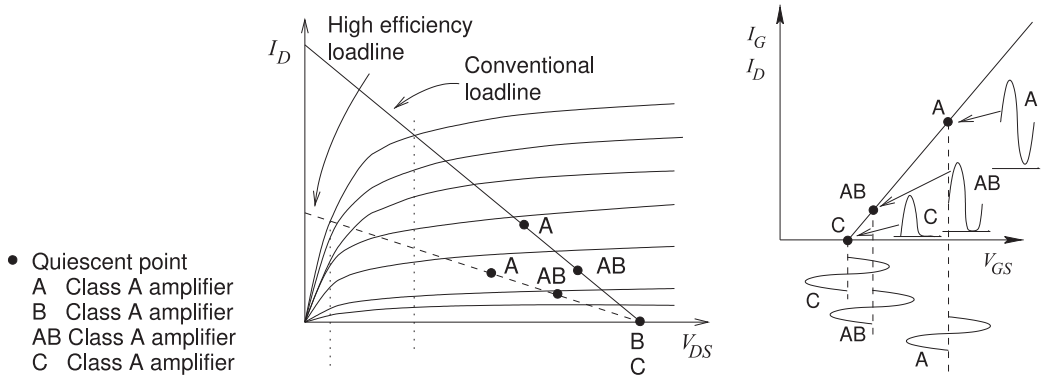
$$\eta_D = \frac{P_{\mathrm{RF,out}}}{P_{\mathrm{DC}}}. \tag{2.39}$$

This term is also used when the device is a BJT or HBT, although the term **collector efficiency** would be more appropriate.

The efficiency metrics are compared in Table 2-3 for an amplifier with 1 W RF output power. The first amplifier has a power gain of 3 dB, which is commonly the gain of the final amplifier stage producing the maximum output power available from a particular transistor technology.

2.5 Class A, AB, B, and C Amplifiers

Transistor amplifiers use several different biasing strategies. The strategies are identified as classes of amplifiers ranging from Class A to Class G. In this section Class A, AB, B, and C amplifiers are considered and these have the basic topologies of Figure 2-7, where input and output matching networks have been omitted. Class A–C amplifiers have the same impedance presented to the output of the amplifier at the operating frequency and at



(a) Output characteristic (b) Input characteristic

Figure 2-8: Current-voltage characteristics of a transistor used in an amplifier showing the quiescent points of various amplifier classes.

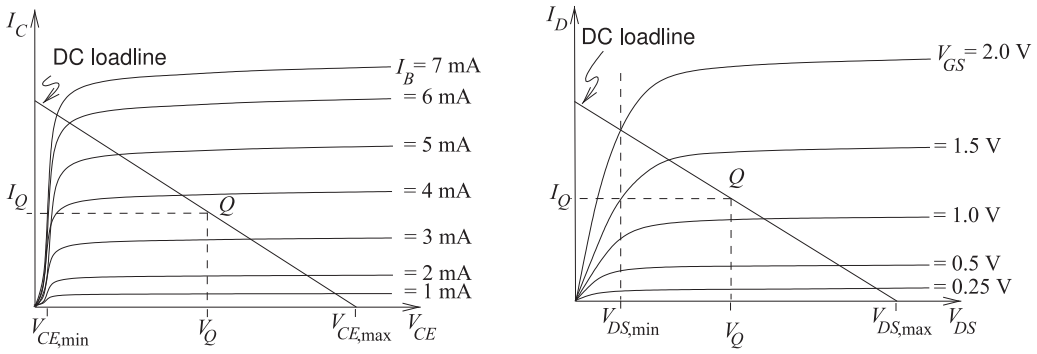
harmonics. Figure 2-8(a) is the output characteristic of a transistor and shows the distinguishing quiescent points for the various classes of amplifier. The input characteristics are shown in Figure 2-8(b), where the input (I_G) and output (I_D) current waveforms are shown for a sinusoidal input waveform (V_{GS}).

Amplifier design consists of both design for low-power linear operation, requiring maximum power transfer at the input and output of the amplifier, and a trade-off of acceptable distortion and efficiency. In practice, a certain level of distortion must be tolerated, and what is acceptable is embedded in the specifications of the various wireless systems.

For low distortion, the peaks of the RF signal must be amplified linearly, however, the DC power consumed depends on the amplifier class. With Class A amplifiers, the DC power must be sufficient to provide low-level distortion of the largest RF signal so that the DC power is proportional to the peak AC power.

The situation is similar for Class AB amplifiers, with the difference being that the intent is to accept some distortion of the peak signal so that the relationship between peak power and DC power still exists, but the direct proportionality no longer holds. For Class C and higher class amplifiers, the DC power is mostly proportional to the average RF power. So for Class A and Class AB amplifiers, the average operating point must be "backed off" to allow for manageable distortion of the peaks of a signal, with the level of back-off required being proportional to the PMEPR of the modulated signal. For Class C and higher classes, the back-off required comes from experience and experimentation. The characteristics of the signal also determine how much distortion can be tolerated.

The PMEPR of the signal is an indication of the type and amount of distortion that can be tolerated. The PMEPR of the two-tone signal is 3 dB, and digitally modulated signals can have PMEPRs ranging from 0 dB to 20 dB or more. A signal with a higher PMEPR results in lower efficiency, as more back-off is required. Putting this another way, the DC bias must be set so that there is minimum distortion when the signal is at its peak, but



(a) Bipolar amplifier (b) FET amplifier

Figure 2-9: Current-voltage characteristics of transistor amplifiers shown with a Class A amplifier loadline: (a) bipolar amplifier; and (b) FET amplifier.

the average RF power produced can be much less than the peak RF power. (The average RF power is approximately an amount PMEPR below the peak RF power.) Thus for a high-PMEPR signal, generally a higher DC power is required to produce the same RF power. This is especially true for Class A amplifiers.

2.5.1 Class A Amplifier

The Class A amplifier has limited efficiency because there is always substantial quiescent current flowing whether or not RF current is flowing. Higher-order classes of amplifiers achieve higher efficiency, but distort the RF signal. The current and voltage loci of Class A, B, AB, and C amplifiers have a similar trajectory on the output current-voltage characteristics of a transistor. The output characteristics of a transistor are shown in Figure 2-8(a), showing what is called the linear or DC loadline and the bias points for the various amplifier classes. The loadline is the locus of the DC current and voltage as the DC input voltage is varied.

With the Class A amplifier, the transistor is biased in the middle of the transistor characteristics, where the response has the highest linearity. That is, when the gate voltage varies due to an applied signal, the output voltage and current variations are nearly linearly proportional to the applied input. The drawback is that there is always considerable DC current flowing, even when the input signal is very small. That is, there is DC power consumption whether or not RF power is being generated at the output of the transistor. This is not of concern if small RF signals are to be amplified, as then a small transistor can be chosen so that the DC current levels are small. It is a problem if an amplifier must handle both large and small signals.

The Class A amplifier is defined by its ability to amplify small to medium and even large signals with minimal distortion. This is achieved by biasing a transistor in the middle of its I-V (i.e., current-voltage) characteristics. Figure 2-9 shows the I-V characteristics of the bipolar and FET transistors shown in Figure 2-7, together with the DC loadline. The loadline is the locus of the output current and voltage. For the Class A amplifiers in Figure 2-7(a and b)

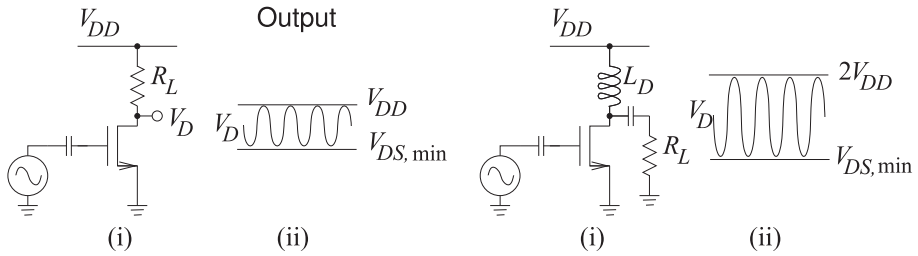


Figure 2-10: Single-ended Class A MOSFET amplifiers: (i) schematic; and (ii) drain voltage waveforms.

the loadlines are defined by

$$I_C = (V_{CC} - V_{CE})/R_L \quad \text{and} \quad I_D = (V_{DD} - V_{DS})/R_L \tag{2.40}$$

respectively. These are called single-ended amplifiers, as the input and output voltages are referred to ground. An amplifier using a bipolar transistor (either a BJT or an HBT), as shown in Figure 2-7(a), has the output characteristics shown in Figure 2-9(a). Here the output voltage of the bipolar amplifier is V_{CE}, and this swings from a maximum value of $V_{CE,\max}$ to a minimum of $V_{CE,\min}$. For a bipolar transistor $V_{CE,\min}$ is approximately 0.2 V, while $V_{CE,\max}$ for a resistively biased circuit is just the supply voltage, V_{CC}. The quiescent or bias point is shown with collector-emitter voltage V_Q and quiescent current I_Q. For a Class A amplifier, the quiescent point is just the bias point, and this is in the middle of the output voltage swing and the slope of the loadline is established by the load resistor, R_L.

The output I-V characteristic of a FET amplifier is shown in Figure 2-9(b). The notable difference between these characteristics and those of the bipolar transistor is that the curves are less abrupt at low output voltage (i.e., low V_{DS}). This results in the FET amplifier's minimum output voltage, $V_{DS,\min}$, being larger than that of a BJT-based amplifier, $V_{CE,\min}$. For a typical RF FET amplifier, $V_{DS,\min}$ is 0.5 V.

The bipolar and FET amplifiers of Figure 2-7 use resistive biasing so that the maximum output voltage swing is limited. As well, the bias resistor is also the load resistor. Various alternative topologies have been developed yielding a range of output voltage swings. The common variations are shown in Figure 2-10 for a FET amplifier. Figure 2-10(a) is a resistively biased Class A amplifier with the output voltage swing between $V_{DS,\min}$ and V_{DD}. The quiescent drain-source voltage is halfway between these extremes. The load, R_L, also provides correct biasing. A more efficient Class A amplifier uses inductive biasing, as shown in Figure 2-10(b). Bias current is now provided via the drain inductor, and the load, R_L, is not part of the bias circuit. With the inductively loaded Class A amplifier, the quiescent voltage is V_{DD} and the output voltage swing is between $V_{DS,\min}$ and $2V_{DD}$, slightly more than twice the voltage swing of the resistively loaded amplifier.

A Class A amplifier can be designed using the S parameters of the transistor in a specific configuration. Ideally the effect of bias circuitry would be included in the S parameters of the transistor, but bias circuit

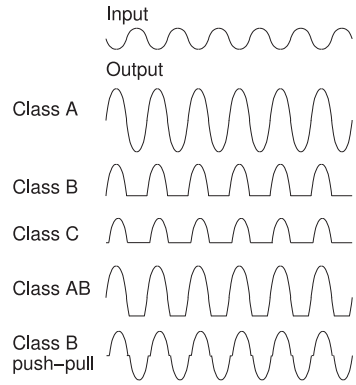


Figure 2-11: Input and output waveforms for various classes of amplifier.

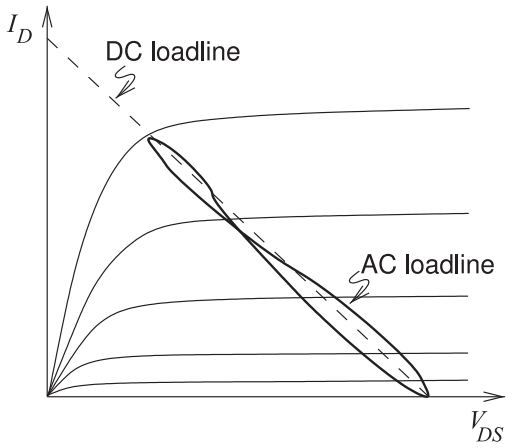


Figure 2-12: DC and RF loadlines of Class A, B, and C amplifiers. The AC loadline is also called the dynamic loadline.

design attempts to present RF open or short circuits as required to minimize impact on RF performance. Generally design begins with the transistor's S parameters and assuming no bias circuit impact.

2.5.2 Amplifier Efficiency

Since the Class A amplifier is always drawing DC current, its efficiency is near zero when the input signal is very small. The maximum efficiency of Class A amplifiers is 25% if resistive biasing is used and 50% when inductive biasing is used. Efficiency is improved by reducing the DC power, and this is achieved by moving the bias point further down the DC loadline, as in the Class B, AB, and C amplifiers shown in Figure 2-8. Reducing the bias results in signal distortion for large RF signals. This can be seen in the various output waveforms shown in Figure 2-11. Class A amplifiers have the highest linearity and Class B and C amplifiers result in considerable distortion. As a compromise, Class AB amplifiers are used in many cellular applications, although Class C amplifiers are used with constant envelope modulation schemes, as in GSM. Nearly all small-signal amplifiers are Class A.

The effect of parasitic capacitances and delay effects (such as those due to the time it takes carriers to move across a base for a BJT or under the gate for a FET) result in the current-voltage locus for RF signals differing from the DC situation. This effect is captured by the **dynamic or AC loadline**, which is shown in Figure 2-12.

The Class A amplifier is a low-efficiency, but highly linear class. The other

amplifier classes have higher efficiencies but varying degrees of distortion, as seen in Figure 2-11. The output of the Class B amplifier contains an amplified version of only half of the input signal but draws just a small leakage current when no signal is applied. With the Class C amplifier there must be some positive RF input signal before there is an output: there is more distortion but no current flows, not even leakage current, when there is no RF input signal. The Class AB amplifier is a compromise between Class A and Class B amplifiers. Less DC current flows than with Class A when there is negligible input signal, and the distortion is less than with Class B. Filtering, often provided by matching networks, eliminates harmonics from the output of the amplifier, but in-band distortion of finite bandwidth signals remains.

Class C amplifiers are biased so that there is almost no drain-source (or collector-emitter) current when no RF signal is applied, so the output waveform has considerable distortion, as shown in Figure 2-11. This distortion is important only if there is information in the amplitude of the signal. FM, PM, and to a lesser extent GMSK schemes result in signals with constant (or for GMSK near constant) RF envelopes, thus there is no information contained in the amplitude of the signal. Therefore errors introduced into the amplitude of a signal are of lesser significance and efficient saturating mode amplifiers such as a Class C amplifier can be used. In contrast, PSK and QAM modulation schemes do not result in signals with constant RF envelopes and so some information is contained in the amplitude of the RF signal. For these modulation techniques, reasonably linear amplifiers are required.

The Class A amplifier presents input and output impedances that are almost independent of the level of the signal. However, a Class B, AB, or C amplifier presents input and output impedances that vary depending on the level of the RF signal. Thus design requires more care, as the chances of instability are higher and it is more likely that an oscillation condition will be met. Also, Class B, AB, and C amplifiers are generally not used in broadband applications or at high frequencies (say above 20 GHz) mainly because of the problem of maintaining stability. Class A amplifiers are then the preferred solution as design is simpler and the amplifier is more tolerant of parasitic effects and variations.

EXAMPLE 2.2 Efficiency of a Class A Amplifier

Determine the efficiency of a Class A FET amplifier with resistive biasing using the FET characteristics shown in Figure 2-9(b).

Solution:

For maximum output voltage swing, the quiescent point should be halfway between the maximum and minimum drain source voltages,

$$V_O = V_{DS,\min} + \left(\frac{V_{DS,\max} - V_{DS,\min}}{2}\right) = \left(\frac{V_{DD} + V_{DS,\min}}{2}\right), \tag{2.41}$$

since $V_{DS,\max} = V_{DD}$. The quiescent (DC) current through R_L is

$$I_Q = \frac{V_{DD} - V_O}{R_L} = \left(V_{DD} - \frac{V_{DD} + V_{DS,\min}}{2}\right)\frac{1}{R_L} = \frac{V_{DD} - V_{DS,\min}}{2R_L}, \tag{2.42}$$

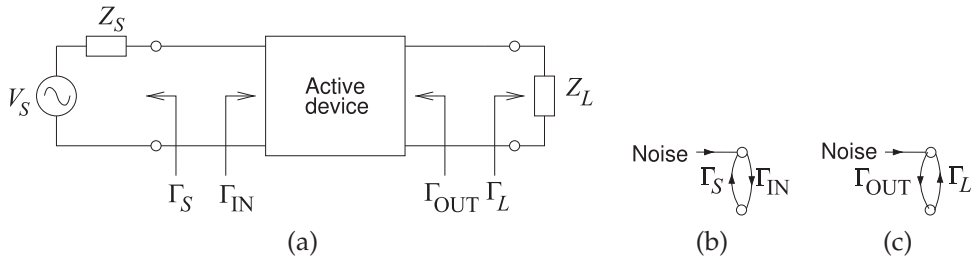


Figure 2-13: A two-port network with inputs at the source and load used in defining stability measures: (a) network; (b) input signal flow graph; and (c) output signal flow graph.

and the quiescent (DC) power consumed is

$$P_{DC} = V_{DD} I_Q = V_{DD} \left(\frac{V_{DD} - V_{DS,\min}}{2R_L} \right) = \frac{V_{DD}^2}{2R_L} \left(1 - \frac{V_{DS,\min}}{V_{DD}} \right). \tag{2.43}$$

The peak voltage of the AC output is

$$V_p = (V_{DD} - V_{DS,\min})/2, \tag{2.44}$$

so that the RF output power in the load is

$$P_{RF,\text{out}} = \frac{1}{2} \left(\frac{V_{DD} - V_{DS,\min}}{2} \right)^2 \frac{1}{R_L} = \frac{V_{DD}^2}{8R_L} \left(1 - \frac{V_{DS,\min}}{V_{DD}} \right)^2. \tag{2.45}$$

Thus the efficiency of the amplifier is

$$\eta = \frac{P_{RF,\text{out}}}{P_{DC}} = \frac{1}{4} \left(1 - \frac{V_{DS,\min}}{V_{DD}} \right). \tag{2.46}$$

If the minimum drain voltage is ignored ($V_{DS,\min} = 0$), then $\eta = 1/4 = 25\%$. This is the maximum efficiency of a resistively biased Class A amplifier.

2.6 Amplifier Stability

The potential exists for an amplifier to be unstable and oscillate. Generally this is not a constant oscillation with fixed amplitude and frequency, but a chaotic response. Oscillator design is not as simple as designing an unstable amplifier, if only one could be so lucky. There is not a simple metric that will indicate whether an amplifier will be stable or not. In the worst case a transistor will oscillate no matter what is done to the external circuitry [10]. A manufacturer could not sell such a transistor and it would not be a useful component in a monolithic integrated process. So it is not surprising that stable amplifiers can be designed using available transistors. For stability analysis, the active device in a linear amplifier can often be treated as a two-port (see Figure 2-13(a)).

Stability considerations affect the maximum (stable) gain that can be achieved and the ease of design. If the maximum stable gain is too low then another transistor needs to be selected. If the specified gain is close to the maximum stable gain, then design will be challenging, especially for broader

bandwidths. So design effort increases with simpler (and hence cheaper) transistors. Experience is the best guide to making this tradeoff.

There are many ways of looking at stability. In the time-domain instability is manifested as the growth of signals over time independent of the level of the input signal. However it is most efficient to analyze and design RF and microwave circuits in the frequency domain and so stability must be assessed in the frequency domain as well. The many frequency domain techniques available to assess stability vary by the level of coverage and ease by which they can be applied. The simplest and most easily applied technique is based on two-port analysis, which leads to stability metrics based on two-port S parameters and to the concept of regions (called circles) of stability on a Smith chart and a more general technique is Nyquist stability analysis. These are considered below.

2.6.1 Two-Port Stability Analysis

Stability analysis should be applied to the innermost two-port containing the active device and any feedback networks. If the innermost two-port (including any feedback) is unconditionally stable, then the amplifier in which it is embedded will be stable. Thus there is a natural selection process so that available transistors tend to not suffer from internal instabilities.

Oscillation will initiate if signals reflected at the input port increase in amplitude as the signal reflects first from the source, Γ_S, and then from the input port. That is, if

$$|\Gamma_S \Gamma_{\rm IN}| > 1, \tag{2.47}$$

the amplifier will be potentially unstable at the input. Whether or not it is actually unstable will depend on the phases of Γ_S and $\Gamma_{\rm IN}$. This situation is shown in the signal flow graph of Figure 2-13(b), where oscillation is initiated by noise. Similarly oscillation will occur if multiple reflections between the output and the load build in amplitude. That is, if

$$|\Gamma_L \Gamma_{\rm OUT}| > 1, \tag{2.48}$$

with the oscillation initiated by noise as shown in Figure 2-13(c). Now

$$|\Gamma_{\rm IN}| = \left| S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L} \right| \tag{2.49}$$

and

$$|\Gamma_{\rm OUT}| = \left| S_{22} + \frac{S_{12}S_{21}\Gamma_S}{1 - S_{11}\Gamma_S} \right|. \tag{2.50}$$

Combining Equations (2.47)–(2.50), the amplifier will be unstable if

$$|\Gamma_S \Gamma_{\rm IN}| = \left| \Gamma_S S_{11} + \frac{S_{12}S_{21}\Gamma_S\Gamma_L}{1 - S_{22}\Gamma_L} \right| > 1 \tag{2.51}$$

or

$$|\Gamma_L \Gamma_{\rm OUT}| = \left| \Gamma_L S_{22} + \frac{S_{12}S_{21}\Gamma_S\Gamma_L}{1 - S_{11}\Gamma_S} \right| > 1. \tag{2.52}$$

The coupling of Γ_S and Γ_L makes it difficult to independently design the input and output matching networks. It is much more convenient to consider the **unconditionally stable** situation whereby the input is stable no matter

what the load and output matching network present, and the output is stable no matter what the source and input matching network present. As a first stage in design, a linear amplifier is designed for unconditional stability. The design space is larger if the more rigorous test for stability, embodied in Equations (2.51) and (2.52), is used to determine stability. The advantage (i.e., a larger design space), however, is often small.

If the source and load are passive, then $|\Gamma_S| < 1$ and $|\Gamma_L| < 1$ so that oscillations will build up if

$$|\Gamma_{\rm IN}| > 1 \quad \text{and} \quad |\Gamma_{\rm OUT}| > 1. \tag{2.53}$$

For guaranteed stability for all passive source and load terminations (i.e., unconditional stability), then

$$|\Gamma_{\rm IN}| < 1 \quad \text{and} \quad |\Gamma_{\rm OUT}| < 1. \tag{2.54}$$

Amplifiers are often realized as stages whereby one amplifier stage feeds another. This complicates stability analysis, as it is possible for Γ_S and Γ_L to be more than unity. If Γ_S and Γ_L are both less than one for multiple amplifier stages, then the amplifier stability being described here can be used.

For the stability criteria to be used in design they must be put in terms of the scattering parameters of the active device. There are two suitable stability criteria commonly used, the k-factor and the μ-factor, which will now be considered.

2.6.2 Unconditional Stability: Two-Port Stability Circles

The input reflection coefficient of an active device is determined by the S parameters of the device and the load:

$$\Gamma_{\rm IN} = S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L}, \tag{2.55}$$

and so for stability (for $|\Gamma_S| \leq 1$)

$$|\Gamma_{\rm IN}| = \left| S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L} \right| < 1. \tag{2.56}$$

Similarly, considering the output of the active device, for stability (with $|\Gamma_L| \leq 1$),

$$|\Gamma_{\rm OUT}| = \left| S_{22} + \frac{S_{12}S_{21}\Gamma_S}{1 - S_{11}\Gamma_S} \right| < 1. \tag{2.57}$$

Equations (2.56) and (2.57) must hold for all

$$|\Gamma_S| \leq 1 \quad \text{and} \quad |\Gamma_L| \leq 1. \tag{2.58}$$

When the active is unilateral, $S_{12} = 0$, and Equations (2.56) and (2.57) simplify to the requirement that $|S_{11}| < 1$ and $|S_{22}| < 1$. Otherwise, given a device, there will be a limit on the values of Γ_S and Γ_L that will ensure stability. The stability criteria are in terms of the magnitudes of complex numbers, and this is known to specify circles in the complex plane. The following development will lead to the center and radius defining the

stability circles that define the boundaries between stable and potentially unstable regions.

For $|\Gamma_{\mathrm{IN}}| = 1$, the output stability circle is defined by

$$\left| S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L} \right| = 1. \tag{2.59}$$

The development that follows puts this into the standard form for a circle, that is, in the form of

$$|\Gamma_L - c| = r, \tag{2.60}$$

where c is a complex number and defines the center of the circle on a reflection coefficient plot, and r is a real number and is the radius of the circle. This circle defines the boundary between stable and potentially unstable values of Γ_L. Now Equation (2.59) can be rewritten as

$$|S_{11} - (S_{11}S_{22} - S_{12}S_{21})\Gamma_L| = |1 - S_{22}\Gamma_L|, \tag{2.61}$$

which includes the determinant, Δ, of the scattering parameter matrix. With

$$\Delta = S_{11}S_{22} - S_{12}S_{21}, \tag{2.62}$$

Equation (2.61) becomes

$$|S_{11} - \Delta\Gamma_L| = |1 - S_{22}\Gamma_L|. \tag{2.63}$$

Removing the absolute signs by multiplying each side by its complex conjugate and then rearranging,

$$(S_{11} - \Delta\Gamma_L)(S_{11} - \Delta\Gamma_L)^* = (1 - S_{22}\Gamma_L)(1 - S_{22}\Gamma_L)^* \tag{2.64}$$

$$S_{11}S_{11}^* + \Delta\Delta^*\Gamma_L\Gamma_L^* - (\Delta\Gamma_L S_{11}^* + \Delta^*\Gamma_L^* S_{11}) = 1 + S_{22}S_{22}^*\Gamma_L\Gamma_L^* - (S_{22}\Gamma_L + S_{22}^*\Gamma_L^*) \tag{2.65}$$

$$\left(|S_{22}|^2 - |\Delta|^2\right)\Gamma_L\Gamma_L^* - (S_{22} - \Delta S_{11}^*)\Gamma_L - (S_{22}^* - \Delta^* S_{11})\Gamma_L^* = |S_{11}|^2 - 1 \tag{2.66}$$

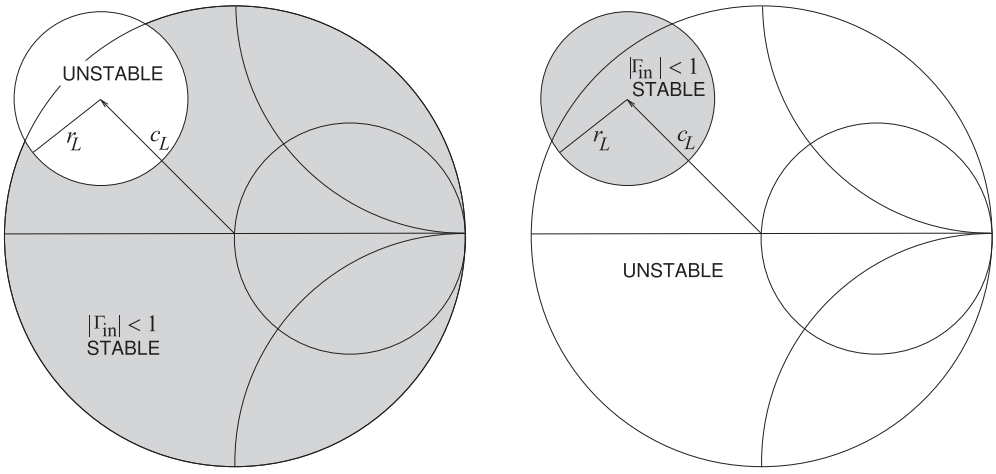
$$\Gamma_L\Gamma_L^* - \frac{(S_{22} - \Delta S_{11}^*)\Gamma_L}{|S_{22}|^2 - |\Delta|^2} - \frac{(S_{22} - \Delta S_{11}^*)^*\Gamma_L^*}{|S_{22}|^2 - |\Delta|^2} = \frac{|S_{11}|^2 - 1}{|S_{22}|^2 - |\Delta|^2}. \tag{2.67}$$

Adding the same term to both sides,

$$\Gamma_L\Gamma_L^* - \frac{(S_{22} - \Delta S_{11}^*)\Gamma_L}{|S_{22}|^2 - |\Delta|^2} - \frac{(S_{22} - \Delta S_{11}^*)^*\Gamma_L^*}{|S_{22}|^2 - |\Delta|^2} + \frac{(S_{22} - \Delta S_{11}^*)(S_{22} - \Delta S_{11}^*)^*}{\left(|S_{22}|^2 - |\Delta|^2\right)^2} = \frac{|S_{11}|^2 - 1}{|S_{22}|^2 - |\Delta|^2} + \frac{(S_{22} - \Delta S_{11}^*)(S_{22} - \Delta S_{11}^*)^*}{\left(|S_{22}|^2 - |\Delta|^2\right)^2} \tag{2.68}$$

and collecting terms leads to the equation for a circle:

$$\left| \Gamma_L - \frac{S_{22} - \Delta S_{11}^*}{|S_{22}|^2 - |\Delta|^2} \right| = \left| \frac{S_{12}S_{21}}{|S_{22}|^2 - |\Delta|^2} \right|. \tag{2.69}$$



(a) $|S_{11}| < 1$ (b) $|S_{11}| > 1$

Figure 2-14: Output stability circles on the Γ_L plane. The shaded regions denote the values of Γ_L that will result in unconditional stability at the input indicated by $|\Gamma_{\text{in}}| < 1$.

This defines a circle, called the output stability circle, in the Γ_L plane with center c_L and radius r_L (the development of this is given in Section 1.A.13) of [7]:

$$\text{center}: \; c_L = \frac{(S_{22} - \Delta S_{11}^*)^*}{|S_{22}|^2 - |\Delta|^2} \quad (2.70) \qquad \text{radius}: \; r_L = \left| \frac{S_{12}S_{21}}{|S_{22}|^2 - |\Delta|^2} \right|. \quad (2.71)$$

This circle is plotted on a Smith chart in Figure 2-14 for the two conditions $|S_{11}| < 1$ and $|S_{11}| > 1$. When $|S_{11}| < 1$ the shaded region in Figure 2-14(a) indicates unconditional stability. That is, as long as Γ_L is chosen to lie in the shaded region, the input reflection coefficient, Γ_{in}, will be less than one. It does not matter what the source impedance is, as long as it is passive there will not be oscillation due to multiple reflections between the input of the amplifier and the source.

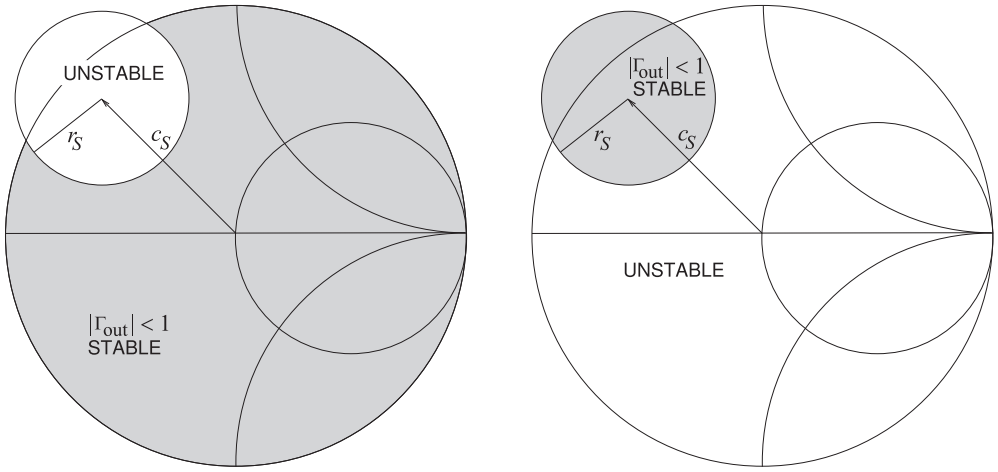
Similarly an input stability circle can be defined for Γ_S, where

$$\text{center}: \; c_S = \frac{(S_{11} - \Delta S_{22}^*)^*}{|S_{11}|^2 - |\Delta|^2} \quad (2.72) \qquad \text{radius}: \; r_S = \left| \frac{S_{12}S_{21}}{|S_{11}|^2 - |\Delta|^2} \right|. \quad (2.73)$$

The interpretation of the input stability circles, shown in Figure 2-15, is similar to that for the output stability circles.

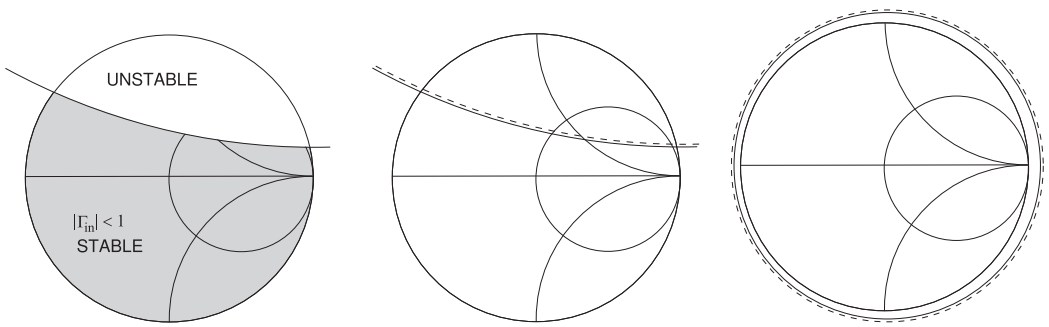
The stability criterion provided by the input and output stability circles is very conservative. For example, the input stability circle (for Γ_S) indicates the value of Γ_S that will ensure stability no matter what passive load is presented. Thus the stability circles here are called unconditional stability circles. However, an amplifier can be stable for loads (or source impedances) other than those that ensure unconditional stability. The use of stability circles provides a good first pass in design of the matching networks between the actual source and load and the amplifier. The stability circles will change with frequency, and so ensuring stability requires a broad frequency view. This is considered in the linear amplifier design case study in Section 2.9.

Shading is commonly used in publications to indicate the stable and



(a) $|S_{22}| < 1$ (b) $|S_{22}| > 1$

Figure 2-15: Input stability circles on the Γ_S plane. The shaded regions denote the values of Γ_S that will result in unconditional stability at the output indicated by $|\Gamma_{\text{out}}| < 1$.



(a) (b) (c)

Figure 2-16: Stability circles: (a) using the absence of shading to indicate the potentially unstable region; (b) using a dashed line to indicate the same potentially unstable region; and (c) stability circle of an unconditionally stable (different) two-port.

potentially unstable regions on a Smith chart. An alternative commonly used by RF and microwave computer-aided design programs is to use a dashed line to indicate the side of the stability circle that is potentially unstable (see Figure 2-16). Figure 2-16(a) uses shading to indicate the potentially unstable region of the Smith chart while the stability circle in Figure 2-16(b) indicates the potentially unstable region using a dashed line. The stability circle in Figure 2-16(c) identifies a two-port that is stable for all passive terminations.

If an amplifier is unconditionally stable, amplifier design is considerably simplified. Matching network design then needs to be concerned about the impact of matching networks on stability. Detailed stability analyses are presented in [11] and [12]. Note that an amplifier can be stable even if it

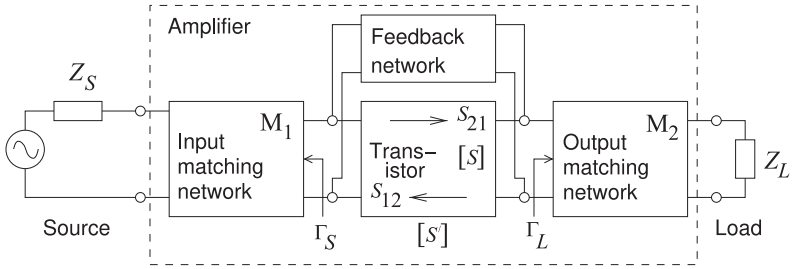


Figure 2-17: General amplifier configuration.

is not unconditionally stable.

2.6.3 Rollet's Stability Criterion — k-factor

Rollet rhymes with wallet.

The k-factor method, also known as **Rollet's stability criterion** [3, 4] is the most commonly used stability metric. It is based on the input and output reflection coefficients of an active device. The most general amplifier is depicted in Figure 2-17. Here the active device has scattering parameters

$$\mathbf{S} = [S] = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \tag{2.74}$$

and, as it will be used a lot, define the determinant $\Delta = S_{11}S_{22} - S_{12}S_{21}$. The overall amplifier has scattering parameters $\mathbf{S}'$. Γ_S is the generator reflection coefficient and Γ_L is the reflection coefficient of the load. For unconditional stability, $|S'_{11}| < 1$ and $|S'_{22}| < 1$. If $|S'_{11}| > 1$ or $|S'_{22}| > 1$, then there is a negative resistance and oscillation could possibly occur. Unconditional stability is defined as when $|S'_{11}| < 1$ for all passive loads Γ_L (i.e., $|\Gamma_L| \leq 1$) and $|S'_{22}| < 1$ for all passive source impedances Γ_S (i.e., $|\Gamma_S| \leq 1$). These inequalities are the same as saying the real part of the input and output impedances of the amplifier are positive resistances. These requirements also describe the unit circle on a Smith chart. Beginning with these definitions and ignoring the feedback network, a stability factor, k, can be defined as

$$k = \left(\frac{1 - |S_{11}|^2 - |S_{22}|^2 + |\Delta|^2}{2|S_{12}|\,|S_{21}|} \right), \tag{2.75}$$

where $k > 1$ is required (but not sufficient) for unconditional stability. This is the Rollet stability condition. What is done here is rolling two unconditional stability requirements (on $|S'_{11}|$ and $|S'_{22}|$) into one. If $k \leq 1$, the amplifier may not be unstable, but extra care is required when a load is presented to the amplifier. If $k > 1$, the design is relatively straightforward, but if k is near 1 or $k \leq 1$, design will be troublesome. The closer design is to the limits of operation of a device, the more likely k will be near or less than 1. For example, the limit could be the maximum stable gain at the intended frequency of operation.

It has been shown that unconditional stability is assured if

$$k = \left(\frac{1 - |S_{11}|^2 - |S_{22}|^2 + |\Delta|^2}{2|S_{12}|\,|S_{21}|} \right) > 1, \tag{2.76}$$

Table 2-4: Rollet's stability factor, k-factor, $|\Delta|$, and stability circle parameters of the pHEMT transistor (in Figure 2-2). For the active device to be unconditionally stable two conditions must be met: $k > 1$ and $|\Delta| < 1$. Frequencies at which the device is unconditionally stable (5–11 and 22–26 GHz) are in bold.

Freq. (GHz)	k > 1	$\|\Delta\|$ < 1	C_L	R_L	C_S	R_S
0.5	0.15178	0.62757	$1.0388 + j13.5176$	13.370	$0.92820 + j0.50682$	0.22434
1	0.24895	0.58720	$0.78679 + j7.26968$	6.9988	$0.69258 + j0.96669$	0.44107
2	0.46535	0.46821	$0.73554 + j3.96784$	3.4718	$-0.087892 + j1.481129$	0.72546
3	0.67865	0.37045	$0.63608 + j3.17780$	2.4779	$-0.86485 + j1.46382$	0.85475
4	0.91943	0.29706	$0.52381 + j2.82057$	1.9223	$-1.4232 + j1.2217$	0.91458
5	1.0838	0.24365	$0.32018 + j2.76164$	1.7276	$-1.90410 + j0.77291$	1.0132
6	1.1839	0.18765	$0.065264 + j2.679153$	1.5700	$-2.182123 + j0.024217$	1.0885
7	1.2846	0.14113	$-0.39857 + j2.97836$	1.8266	$-2.06262 - j0.85258$	1.0885
8	1.5225	0.092502	$-0.91418 + j3.21866$	2.0150	$-1.5996 - j1.4914$	0.94750
9	1.6448	0.056060	$-1.2526 + j3.0479$	1.8998	$-1.1418 - j1.8922$	0.92223
10	1.3151	0.072750	$-1.7681 + j2.6944$	2.0189	$-0.54178 - j2.13439$	1.0468
11	1.1043	0.11703	$-2.6666 + j2.4074$	2.5186	$0.24025 - j2.02042$	0.98357
12	0.98784	0.16159	$-4.4423 + j2.2978$	4.0111	$0.73109 - j1.58120$	0.74724
13	0.92131	0.18991	$-5.9502 + j1.0596$	5.1100	$0.96118 - j1.24308$	0.60117
14	0.84098	0.21905	$-5.95467 - j0.75389$	5.1368	$1.10689 - j0.97667$	0.53246
15	0.69555	0.24320	$-6.0313 - j1.9645$	5.6068	$1.20251 - j0.58617$	0.43291
16	0.63420	0.27993	$-5.7310 - j3.8333$	6.2171	$1.26271 - j0.23704$	0.39187
17	0.68792	0.32454	$-3.5163 - j5.6496$	5.9268	$1.257206 + j0.087208$	0.34232
18	0.72764	0.36197	$-1.1258 - j5.5386$	4.8824	$1.17955 + j0.29932$	0.27755
19	0.89194	0.35755	$0.72551 - j2.82689$	1.9913	$1.20751 + j0.32462$	0.27383
20	0.97085	0.32318	$1.1030 - j2.0753$	1.3671	$1.19540 + j0.46897$	0.29068
21	0.97475	0.28626	$1.3859 - j1.8451$	1.3221	$0.94075 + j0.85514$	0.27681
22	1.1014	0.28501	$2.1696 - j1.1961$	1.4187	$0.66834 + j1.07377$	0.24499
23	1.3123	0.29397	$2.35683 - j0.15064$	1.1976	$0.51539 + j1.17421$	0.22605
24	1.4714	0.28083	$2.23754 + j0.46967$	1.0569	$0.28045 + j1.30061$	0.24185
25	1.4273	0.22850	$2.2473 + j1.2506$	1.3389	$0.079906 + j1.412442$	0.31586
26	1.5308	0.17235	$2.6309 + j3.0590$	2.6671	$-0.45588 + j1.43292$	0.36774

combined with any one of the following auxiliary conditions [13–19]:

$$B_1 = 1 + |S_{11}|^2 - |S_{22}|^2 - |\Delta|^2 > 0 \tag{2.77}$$

$$B_2 = 1 - |S_{11}|^2 + |S_{22}|^2 - |\Delta|^2 > 0 \tag{2.78}$$

$$|\Delta| = |S_{11}S_{22} - S_{12}S_{21}| < 1 \tag{2.79}$$

$$C_1 = 1 - |S_{11}|^2 - |S_{12}S_{21}| > 0 \tag{2.80}$$

$$C_2 = 1 - |S_{22}|^2 - |S_{12}S_{21}| > 0. \tag{2.81}$$

The conditions in Equations (2.77)–(2.81) are not independent, and it can be shown that one implies the others if $k > 1$ [13].

Rollet's stability criteria, k and $|\Delta|$, are tabulated in Table 2-4 for the pHEMT described in Figure 2-2. The device is unconditionally stable at the frequencies 5–11 GHz and 22–26 GHz. It is seen that the device could be potentially unstable at frequencies below 5 GHz and from 12 to 21 GHz. At these frequencies stability circles need to be used in designing matching networks.

Table 2-5: Edwards–Sinsky stability parameters for the pHEMT transistor documented in Figure 2-2. For stability, $\mu > 1$. The frequencies at which the device is unconditionally stable are in bold. The device is unconditionally stable at the frequencies 5–11 GHz and 22–26 GHz. The other columns refer to Equations (2.77)–(2.81) and are provided for completeness.

Freq. (GHz)	μ > 1	B_1 > 0	B_2 > 0	$\lvert\Delta\rvert$ < 1	C_1 > 0	C_2 > 0
0.5	0.18785	1.1555	0.056799	0.62757	−0.077921	0.47143
1	0.31338	1.1338	0.17657	0.58720	−0.080934	0.39770
2	0.56362	1.1086	0.45297	0.46821	0.065756	0.39356
3	0.76297	1.0884	0.63717	0.37045	0.22398	0.44958
4	0.94651	1.0631	0.76039	0.29706	0.35892	0.51029
5	1.0526	1.0434	0.83782	0.24365	0.44002	0.54284
6	1.1100	1.0341	0.89551	0.18765	0.49297	0.56225
7	1.1783	1.0703	0.88992	0.14113	0.51407	0.60424
8	1.3310	1.1120	0.87085	0.092502	0.54812	0.66871
9	1.3954	1.1104	0.88335	0.056060	0.57284	0.68634
10	1.2039	1.1068	0.88259	0.072750	0.51811	0.63023
11	1.0739	1.1550	0.81758	0.11703	0.43720	0.60592
12	0.99026	1.2715	0.67626	0.16159	0.33481	0.63244
13	0.93383	1.3461	0.58173	0.18991	0.27037	0.65257
14	0.86542	1.3788	0.52518	0.21905	0.22227	0.64910
15	0.73637	1.4578	0.42389	0.24320	0.13811	0.65507
16	0.67769	1.4752	0.36811	0.27993	0.099372	0.65290
17	0.72762	1.4461	0.34324	0.32454	0.10910	0.66053
18	0.76942	1.4300	0.30791	0.36197	0.10899	0.67006
19	0.92718	1.3348	0.40953	0.35755	0.18890	0.65152
20	0.98311	1.2924	0.49876	0.32318	0.24511	0.64191
21	0.98558	1.3331	0.50306	0.28626	0.24786	0.66286
22	1.0587	1.3627	0.47486	0.28501	0.25075	0.69466
23	1.1641	1.3295	0.49769	0.29397	0.28504	0.70093
24	1.2294	1.2872	0.55509	0.28083	0.33166	0.69771
25	1.2330	1.2866	0.60899	0.22850	0.36433	0.70313
26	1.3676	1.3398	0.60077	0.17235	0.38405	0.75358

2.6.4 Edwards–Sinsky Stability Criterion — μ-factor

Rollet's stability criterion, Equations (2.76)–(2.81), ensures unconditional stability but it does not provide a relative measure of stability. That is, the k factor cannot be used to determine how close a particular design is to the edge of stability. There is no ability to compare the relative stability of different designs. Edwards and Sinsky [13] developed a test that can be used to compare the relative stability of different designs. This is called the μ-factor stability criterion, with unconditional stability having

$$\mu = \frac{1 - |S_{11}|^2}{|S_{22} - S_{11}^* \Delta| + |S_{21} S_{12}|} > 1. \tag{2.82}$$

An important result is that a larger value of μ indicates greater stability. The μ factor is a single quantity that provides a sufficient and necessary condition for unconditional stability. That is, it does not matter what passive source and load are presented (i.e. if $|\Gamma_S| \leq 1$ and $|\Gamma_L| \leq 1$) then the amplifier will be stable if $\mu > 1$. This contrasts with Rollet's stability criterion in which two conditions must be met.

Edwards–Sinsky stability parameters for the pHEMT transistor (documented in Figure 2-2) are shown in Table 2-5. The unconditionally stable frequencies of operation are 5–11 GHz and 22–26 GHz. These are the same unconditionally stable frequencies determined by using Rollet's stability criterion (in Table 2-4). The additional information available with the Edwards–

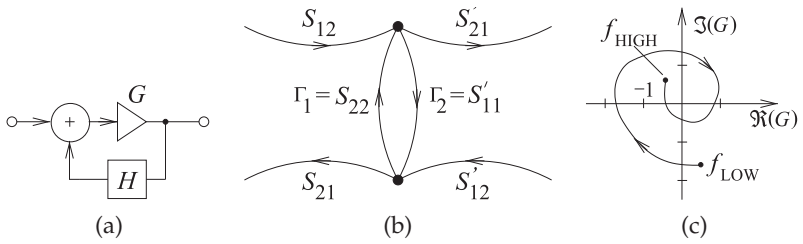


Figure 2-18: Nyquist stability analysis: (a) feedback amplifier; (b) a loop with reflection coefficients Γ_1 and Γ_2; and (c) a Nyquist stability plot of loop gain $G = -\Gamma_1\Gamma_2$ (with H set to 1).

Sinsky stability criterion is that μ indicates the relative stability. In Table 2-5, the transistor is unconditionally stable in the 5–11 GHz range, and in this range the device is most stable between 8 and 10 GHz. At the high end, above 22 GHz, the device is increasingly more stable. This can be expected to continue as the device capacitive parasitics short out the device. (The result of the low impedance of capacitors at high frequencies is that smaller RF voltages will be generated for the same drive level.) So this transistor will make a very good 8–10 GHz amplifier provided that appropriate matching networks are chosen to ensure stability below 5 GHz and above 11 GHz. Note, however, that the amplifier can be used up to about 20 GHz, but with extra care in design. Above 20 GHz the maximum unilateral transducer gain is too small to be useful (see Table 2-1, where $G_{TU,\max}$ is tabulated for this transistor).

The Edwards–Sinsky stability factor, μ, is also called a **geometric stability factor**. It is the distance from the center of the Smith chart (i.e., the origin of the S parameter polar plot) to the nearest potentially unstable point on the input source plane. That is, it is the shortest distance from the origin to the input stability circle, where a negative μ indicates that the stability circle encompasses the origin of the Smith chart.

Edwards and Sinsky also defined a dual parameter, μ' [13]:

$$\mu' = \frac{1 - |S_{22}|^2}{|S_{11} - S_{22}^* \Delta| + |S_{21} S_{12}|}. \tag{2.83}$$

This is the distance from the center of the Smith chart (i.e., the origin of the S parameter polar plot) to the nearest potentially unstable point on the output load plane. That is, it is the shortest distance from the origin to the output stability circle. If $\mu > 1$ (indicating the possibility of two-port instability), then $\mu' > 1$ as well. Thus to determine whether or not a two-port is unconditionally stable, it is only necessary to consider one of them. Also, μ and μ' are sometimes referred to as the input and output geometric stability factors respectively, and sometimes simply as μ_1 (MU1) and μ_2 (MU2), respectively. Calculating both μ and μ' is useful, as this enables the relative stability at the input and the output to be examined. Table 2-6 lists μ and μ' for the pHEMT transistor considered previously.

2.6.5 Nyquist Stability Criterion

The Nyquist stability criterion is the most complete way of analyzing stability. It is based on the analysis of the feedback system shown in Figure 2-18(a) [20–23]. The closed-loop transfer function is

$$T = \frac{G}{1 + GH}. \tag{2.84}$$

Table 2-6: Input and output Edwards–Sinsky stability parameters for the pHEMT transistor documented in Figure 2-2. For stability, $\mu > 1$ and $\mu' > 1$. The device is unconditionally stable at the frequencies 5–11 GHz and 22–26 GHz.

Freq. (GHz)	μ (MU1) (input)	μ' (MU2) (output)
0.5	0.18785	0.8332
1	0.31338	0.74811
2	0.56362	0.75828
3	0.76297	0.84547
4	0.94651	0.96112
5	1.05257	1.04174
6	1.10997	1.09373
7	1.17828	1.14339
8	1.33101	1.23947
9	1.39544	1.28779
10	1.20387	1.15528
11	1.07392	1.05108
12	0.99026	0.99479
13	0.93383	0.97018
14	0.86542	0.94371
15	0.73637	0.90486
16	0.67769	0.89289
17	0.72762	0.91790
18	0.76942	0.93939
19	0.92718	0.97655
20	0.98311	0.99342
21	0.98558	0.99451
22	1.05874	1.01979
23	1.16408	1.05629
24	1.22944	1.08865
25	1.23297	1.09884
26	1.36763	1.13596

The feedback system is unstable if the open-loop transfer function $GH = -1$. Usually in stability analysis H is considered to be 1 so that the system is unstable if the open-loop transfer function $G = -1$. Relating this to microwave circuits, every loop in a signal flow graph is considered. Generally it is sufficient to consider all possible loops containing one pair of nodes, as shown in Figure 2-18(b). It would be best to select the pair of nodes at the input or output of an active device, as the active device is surely going to be involved in instability. There is of course a chance that the critical pair of nodes will be missed, which is more likely to happen in a multistage amplifier. In that situation several pairs of nodes should be considered individually. Returning to a single pair of nodes, as shown in Figure 2-18(b), here the open-loop frequency-domain transfer function is

$$G = \Gamma_1 \Gamma_2. \tag{2.85}$$

The Nyquist stability criterion is that the system is unstable if the open-loop transfer function, G, plotted on the complex plain encircles the -1 point in a clockwise rotation with increasing frequency. The details behind this are provided in nearly every book on linear control systems (e.g. [20–23]).

To ensure stability the Nyquist plot, as Figure 2-18(c) is called, should be plotted for every loop in a system. However, experience is a good guide and only a few loops need to be considered in practice. There are also metrics such as the S probe parameter (sometimes called the G probe) used in microwave circuit simulators that provide a good estimate of whether the

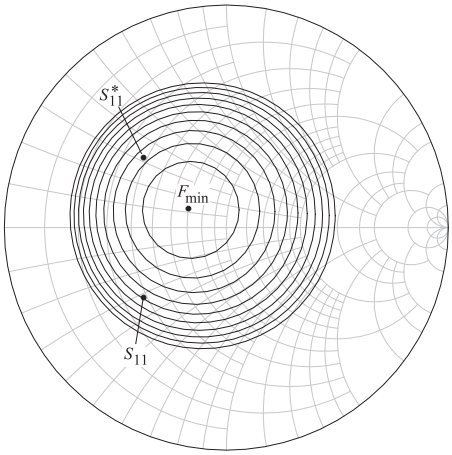


Figure 2-19: Noise figure circles at 8 GHz, plotted on the input Smith chart (i.e. on the Γ_S Smith chart) for the pHEMT transistor documented in Figure 2-2. $F_{\min}$ is 1.04 dB and the circles are in 0.1 dB circles so that the outer circle is the locus of Γ_S that will produce a noise figure of 2.04 dB.

Nyquist plot will encircle the point [24]. The utility of this is that a two-port SPROBE or GPROBE element can be inserted into a circuit and used to provide a measure of Nyquist stability by considering the open-loop gain based on reflection coefficients looking from each port.

A similar technique to applying the Nyquist stability criterion is to create a Bode plot [25]. However, this is not easy to do with many microwave circuits and not often used.

2.6.6 Summary

This section presented criteria that can be used in determining the stability of transistor amplifiers and of active devices. Stability tests should be applied to the innermost two-port, specifically the active device, to provide an enhanced confidence in design. Two criteria in the forms of FOMs were presented for unconditional stability: the k-factor and μ-factor. The k-factor, part of Rollet's unconditional stability criterion, is a test of whether a device is unconditionally stable or not. It does not provide a relative measure of stability. The μ-factor, derived by Edwards and Sinsky, is a factor that indicates the relative stability of a network. Stable amplifiers can be designed even if the amplifier is not unconditionally stable. Stability circles aid in the design of such amplifiers. Stability is an extensive topic and the reader is directed to in-depth stability analysis by Suárez and Quéré [12] for further information.

Two graphical techniques were also presented. Stability circles can be used in design to enable graphically based trade-off of stability, noise, gain, and bandwidth information displayed on a Smith chart. It is surprising how well the trade-off can be made by a designer.

2.7 Amplifier Noise

Section 4.3.6 of [26] presented a discussion of noise in amplifiers. The final result was an expression for the noise figure of an amplifier given the noise parameters of the transistor. This expression was in terms of the magnitudes of complex numbers, which, as has been seen, leads to circles when plotted on the complex plane. Thus the noise figure of an amplifier can be pictured as noise figure circles on a Smith chart. Figure 2-19 shows the noise figure

circles at 8 GHz, plotted on the input Smith chart for the pHEMT transistor documented in Figure 2-2. The center of the (almost) concentric circles is $F_{\min}$ and the noise figure circles show the impact of source mismatch on the noise figure. The loading conditions have little impact on the noise figure of an amplifier as long as the output matching network is lossless and the gain of the amplifier is high. These circles prove useful in making design trade-offs.

2.8 Trading Off Gain, Noise, and Stability in Amplifier Design

The design of small signal amplifiers requires a trade-off of gain, stability, and noise figure. This can be further complicated by the bandwidth requirement. So first consider the trade-offs in narrowband amplifier design with input and output matching networks and with no feedback network between the output and input of a transistor. The amplifier consists of three cascaded two-ports, one of which (the transistor two-port) can produce a potentially unstable situation. Whether the amplifier is stable or not depends on the impedance seen looking from the transistor into the output matching network and the impedance seen from the transistor looking into the input matching network. Even in narrowband amplifier design, the designer must be concerned about stability out of band. The amplifier must also be stable no matter what the values of the load or source impedances. This is because in nearly every situation there is a severe penalty if the amplifier becomes unstable and oscillates. The instrument or device in which the amplifier is embedded certainly will not work, but in the case a communication system, the entire communication system could be corrupted and system operation not restored until the offending device is tracked down and turned off.

The impedance presented to the output of the transistor is not the load impedance, it will be modified by the output matching network and by losses in cabling and filters (if any), between the output of the amplifier and the load. As far as stability is concerned, this loss helps as it reduces the range of effective loads presented to the amplifier. Ignoring loss, the amplifier load could be a short circuit, an open circuit, a match, or any combination. Thus on a Smith chart the load could be anywhere. If the output matching network is lossless and of any topology, then the impedance presented to the output of the transistor could be anywhere on the Smith chart. Now if constraints are placed on the matching network, then the region of the effective load on the Smith chart could be constrained, but this involves a more sophisticated design and is something only very experienced amplifier designers would exploit.

This discussion is designed to provide convincing evidence that the amplifier should be designed for unconditional stability. So if the load can be any value, the stability circle on the input plane defines the values of Γ_S (the reflection coefficient looking into the input matching network from the transistor) that result in unconditional stability. The matching network must be designed to provide a Γ_S that ensures stability no matter what happens to the load. Only in extreme circumstances, say at very high frequencies (where design becomes very difficult) or where there is considerable loss, say in subsequent filtering, would an experienced designer consider designing an amplifier that is not unconditionally stable. Even then design would begin with the unconditionally stable situation and morph into the potentially less

stable situation.

The stability discussion above concerns designing for stability no matter what happens to the load. The discussion is similar for designing the output matching network no matter what happens to the source. So there is an apparent flaw in the stability argument. In designing the input and output matching networks for unconditional stability, the procedure described above considers the load being corrupted on its own independent of whether the source is corrupted (e.g. by the failure of a previous stage). If something goes wrong at both the source and the load, then the amplifier could be unstable even after best efforts have been undertaken in design. It is unlikely that both the load and source would be compromised simultaneously.

Now consider the trade-off between gain and the noise figure. This could be a difficult trade-off, but a simple design procedure has been adopted. If the amplifier is the first stage in a cascade of amplifiers, then the preferred choice is that the emphasis for the first stage is to design it for the minimum noise figure and ensure that at least some gain is obtained. In subsequent stages the emphasis is on gain and the noise figure is given little consideration. This trade-off is based on Friis's formula for the noise figure of cascaded systems, which indicates that if the gain of the first stage is high, then the noise figure of the first stage dominates the system noise figure. A better overall trade-off can be achieved using the optimizer provided in a microwave design tool. However, the manually directed design must be done first or the optimization problem is too difficult.

With wideband designs of a half-octave bandwidth amplifier, the additional problem of achieving stability and the minimum noise figure, or stability and maximum gain, over the frequency band is a further complicating factor. Here the inventive aspect of design is developing a matching network that has the desired frequency response.

Further complicating this is that efficiency and distortion must be considered as well. Even with a small signal, distortion is a concern, as a design goal is minimizing DC power consumed. This is because reducing distortion usually results in increased DC power consumption.

2.9 Case Study: Narrowband Linear Amplifier Design

The design procedure for linear amplifiers is well developed and the strategy forms the basis for all amplifier design. An amplifier has three major components: an input matching network, an active device, and an output matching network (see Figure 2-20). There are a number of design choices to be made and these will be illustrated by considering the design of an amplifier for maximum gain using the discrete pHEMT transistor examined previously (see Figure 2-2). The design specifications are

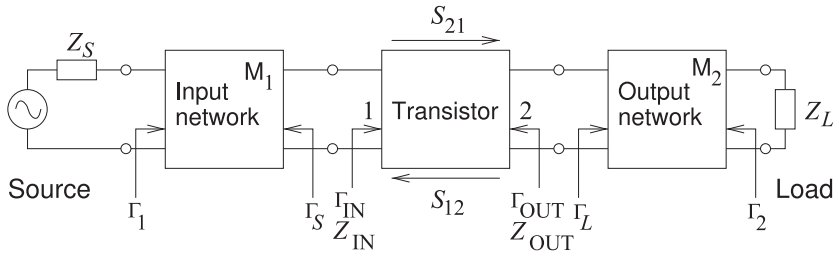


Figure 2-20: Linear amplifier comprising input and output matching networks and an active device in a specific configuration forming a two-port.

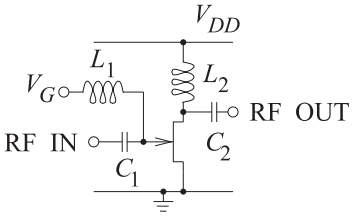


Figure 2-21: Bias configuration for pHEMT amplifier. L_1 and L_2 are RF chokes and large enough to block RF. C_1 and C_2 are DC blocking capacitors that block DC but allow RF to pass with negligible impedance.

Gain:	maximum gain at 8 GHz
Topology:	three two-ports (input and output matching networks, and the active device)
Stability:	broadband stability
Bandwidth:	maximum that can be achieved using two-element matching networks
Source impedance:	$Z_S = 50\ \Omega$
Load impedance:	$Z_L = 50\ \Omega$

2.9.1 Bias Circuit Topology

The first design step is to choose a biasing configuration, and this is directly related to the output voltage swing supported. The inductively biased configuration in Figure 2-21 will be used. Here L_1 and L_2 are RF chokes and large enough to block RF. C_1 and C_2 are DC blocking capacitors that block DC but allow RF to pass with negligible impedance. The input matching network is attached to the RF IN terminal and the output matching network is attached to the RF OUT terminal. V_{DD} is the supply voltage and V_G is the DC gate voltage typically provided by an analog integrated circuit available in conjunction with most RF designs. An additional design objective is to absorb the biasing circuit into the matching networks.

2.9.2 Stability Considerations

It is not sufficient to consider a single frequency in design, as stability must be ensured at low and high frequencies. The stability factor of the active device was given in Table 2-5. This indicates that the device is unconditionally stable from 5 to 11 GHz and from 22 to 26 GHz. At the high-frequency end, the gain of the device reduces rapidly with increasing frequency as the capacitive parasitics begin dominating transmission through the device. Therefore it is reasonable to assume that the device is unconditionally stable above 22 GHz.

Design nearly always commences with the output matching network. The first design step is to choose a matching network that will provide the appropriate impedances to ensure stability below 5 GHz and above 11 GHz. To do this the stability circles must be considered, as the device is only conditionally stable at these frequencies. The center and radius of the input and output stability circles are listed in Table 2-4. These are plotted in Figure 2-22 for selected frequencies.

2.9.3 Output Matching Network Design

The output stability circle at 1 GHz (see Figure 2-22(a)) indicates that for stable, low-frequency amplification, the output matching network, as seen from the transistor, could look like a short circuit, a matched load, or a capacitor at low frequencies. Figure 2-22(c), the output stability circle at

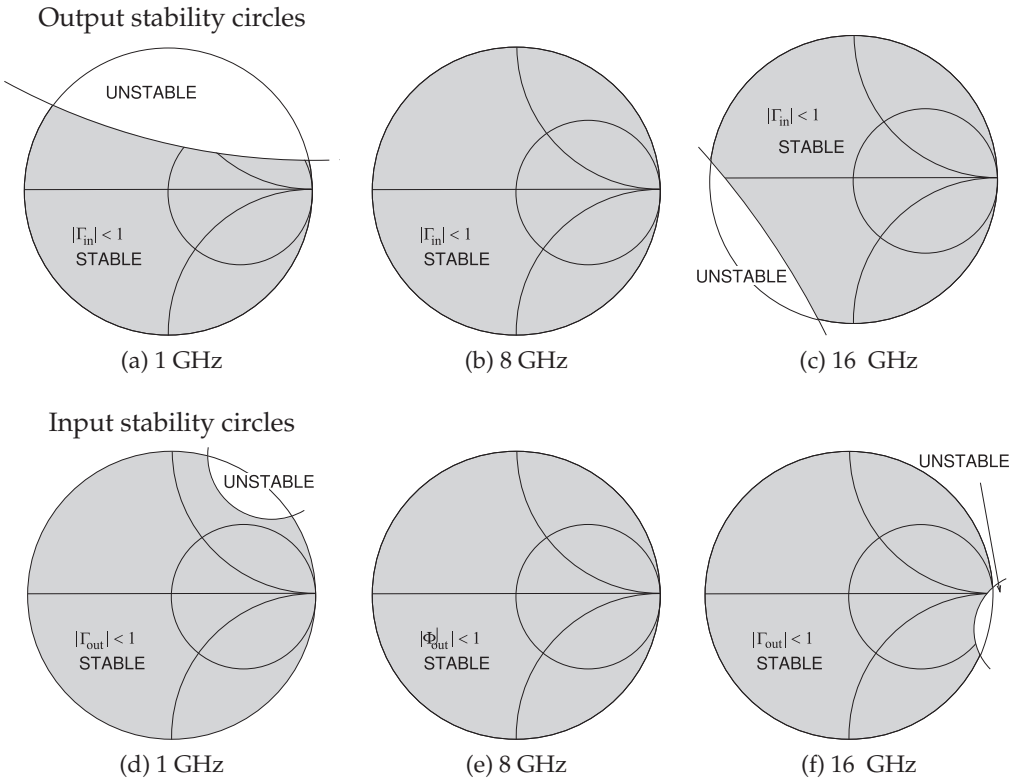


Figure 2-22: Input and output stability circles on the complex reflection coefficient planes for $|S_{11}| < 1$ and $|S_{22}| < 1$.

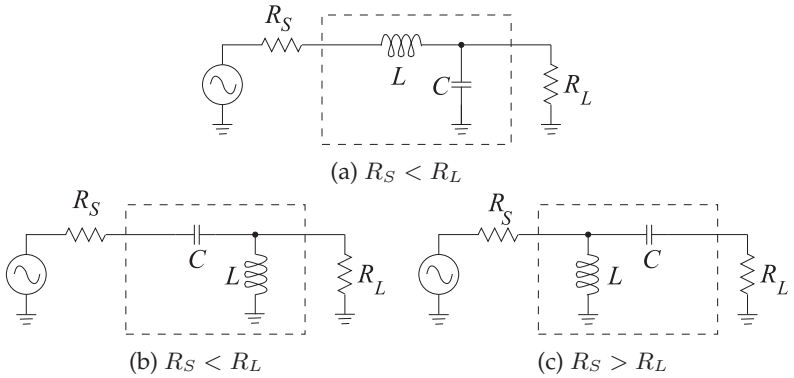


Figure 2-23: Output matching network candidates required for out-of-band stability. The active device is on the left.

16 GHz, indicates that for stable, high-frequency amplification the output matching network, as seen from the transistor, could look like an open circuit or a matched load at high frequencies. As expected, the output stability circle at 8 GHz (see Figure 2-22(b)) indicates unconditional stability. Examining the two-element matching networks in Figure 6-7 of [2], there are three candidate output matching networks that are shown in Figure 2-23.

From Figure 2-2, the device S parameters at 8 GHz are as follows:

$$S_{11} = 0.486\angle 140.4^\circ \qquad S_{21} = 3.784\angle 11.2^\circ$$
$$S_{12} = 0.057\angle 6.4^\circ \qquad S_{22} = 0.340\angle -99.1^\circ.$$

To start, ignore S_{12} so that $\Gamma_{\rm OUT} = S_{22} = 0.340\angle -99.1$. There is little choice here as $\Gamma_{\rm OUT}$ depends on the input matching network that has not yet been designed. It would have been possible to begin with the input matching network and make this approximation for $\Gamma_{\rm IN}$. However, experience is that the error introduced by starting with the output matching network is less. Once the output matching network has been designed, $\Gamma_{\rm IN}$ can be calculated without approximation. A thorough design would complete the first pass of the design and then make one more pass without the approximation that ignores S_{12}. Now, with $\Gamma_{\rm OUT} = S_{22}$, the output impedance of the active device is

$$\begin{aligned} Z_{\rm OUT} &= Z_0\frac{1+\Gamma_{\rm OUT}}{1-\Gamma_{\rm OUT}} = Z_0\frac{1+S_{22}}{1-S_{22}} \\ &= (50\ \Omega)\frac{1+(0.340\angle -99.1^\circ)}{1-(0.340\angle -99.1^\circ)} = (50\ \Omega)\frac{1+(-0.053774-\jmath 0.335721)}{1-(-0.053774-\jmath 0.335721)} \\ &= 36.153-\jmath 27.447\ \Omega, \end{aligned} \tag{2.86}$$

or $Y_{\rm OUT} = 1/Z_{\rm OUT} = 0.017547+\jmath 0.013322\,$S. The output of the active device looks like a 56.99 Ω resistor in parallel with a capacitor with a reactance of $-75.064\,\Omega$. So taking into account the bias objectives and the available output matching networks in Figure 2-23, the matching network topology of Figure 2-23(c) will be used where the source in the matching network is the active device. So the output matching network problem is as shown in Figure 2-24. This choice enables the inductor to be used to apply bias.

The complete output matching problem is shown in Figure 2-24(a). In part, the parallel configuration of the active device output impedance was chosen, as this is closer to reality since there is a capacitance at the output of the transistor. Resonance, as shown in Figure 2-24(b), will be used to cancel the effect of the active device capacitance, so that the matching problem reduces to that shown in Figure 2-24(c). Using the procedure outlined in Section 6.4.2 of [2],

$$|Q_S| = |Q_P| = \sqrt{\frac{R_S}{R_L}-1} = \sqrt{\frac{56.99}{50}-1} = 0.3739 \tag{2.87}$$

$$Q_S = \left|\frac{X_S}{R_L}\right| = \left|\frac{X_S}{50\ \Omega}\right| = 0.3739 \text{ and } Q_P = \left|\frac{R_S}{X_P}\right| = \left|\frac{56.99\ \Omega}{X_P}\right| = 0.3739, \tag{2.88}$$

$$\text{so}\quad X_S = -18.70\ \Omega \quad \text{and} \quad X_P = 152.4\ \Omega. \tag{2.89}$$

Now $X_x = -75.064\,\Omega$, so $75.064\,\Omega$ must be added to X_P in parallel, and the reactance of L_o is $50.29\,\Omega$, thus (at 8 GHz)

$$L_o = 1.00\ \text{nH} \quad \text{and} \quad C_o = 1.064\ \text{pF}. \tag{2.90}$$

The final output matching network design is shown in Figure 2-24(d).

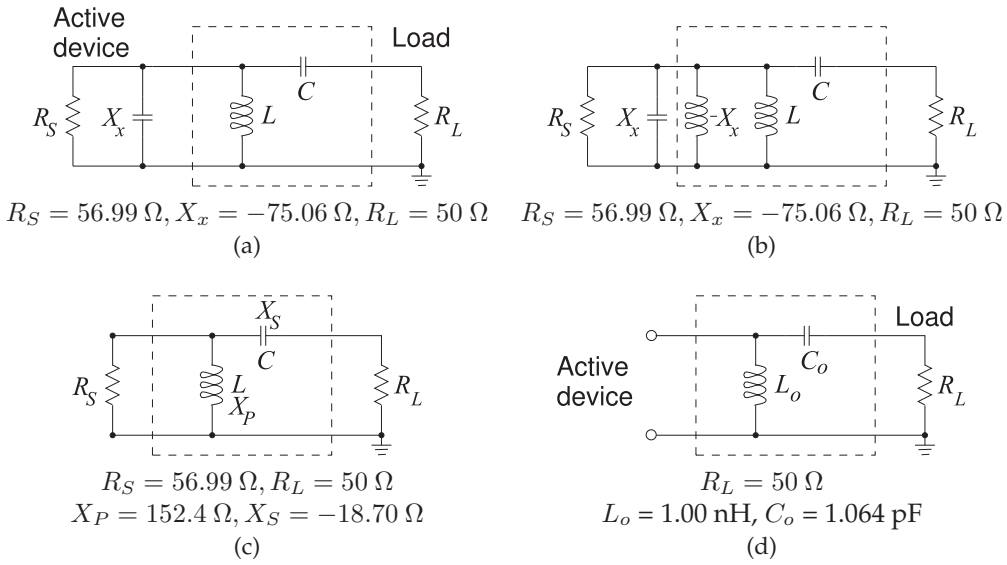


Figure 2-24: Steps in the design of the output matching network: (a) active device presents itself as a resistance in parallel with a capacitive reactance to the output matching network; (b) with inductor to resonate out active device reactance; (c) simplified matching network problem; and (d) final output matching network design.

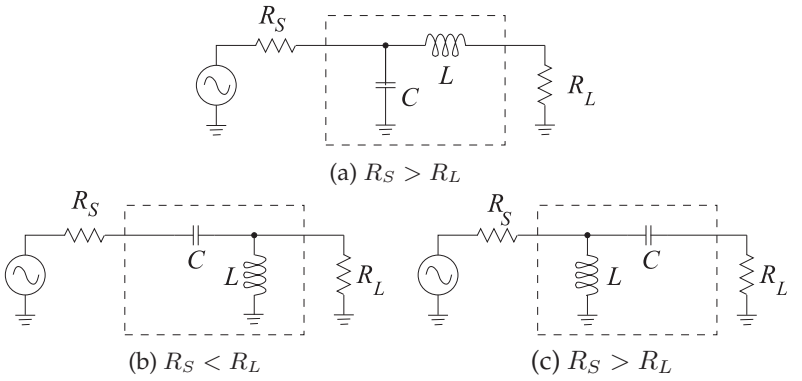


Figure 2-25: Input matching network candidates required for out-of-band stability. The active device is on the right.)

2.9.4 Input Matching Network Design

The input stability circle at 1 GHz (Figure 2-22(d)) indicates that at 1 GHz, the input matching network, as seen from the transistor, could look like a short circuit, a matched load, or a capacitor at low frequencies. Figure 2-22(f), the input stability circle at 16 GHz, indicates that the input matching network, as seen from the transistor, could look like a short circuit or a matched load at high frequencies. Examining the two-element matching networks in Figure 6-7 of [2], there are three candidate input matching networks as shown in Figure 2-25.

The reflection coefficient looking into the output matching network from the active device is $\Gamma_L = S_{22}^* = 0.340\angle 99.1°$ because of the design decision to ignore S_{12} for the output matching network. Now that the output matching

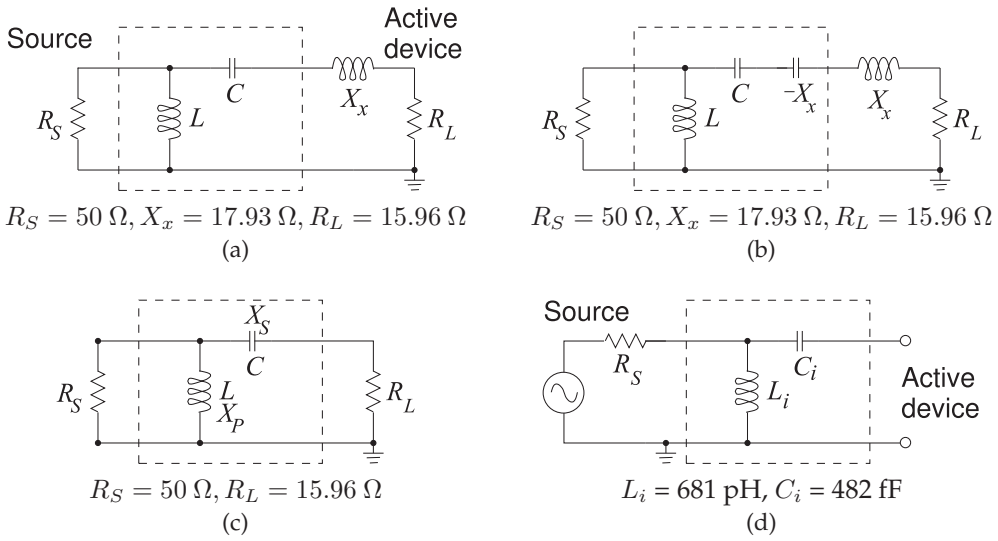


Figure 2-26: Steps in the design of the input matching network: (a) active device presents itself as a resistance in series with an inductive reactance to the input matching network; (b) with a capacitor to resonate out the active device reactance; (c) simplified matching network problem; and (d) final output matching network design.

network has been designed, the feedback parameter need no longer be ignored. So

$$\Gamma_{\text{IN}} = S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L} \tag{2.91}$$

$$= (0.486\angle 140.4^\circ) + \frac{(0.057\angle 6.4^\circ)(3.784\angle 11.2^\circ)(0.340\angle 99.1^\circ)}{1 - (0.340\angle -99.1^\circ)(0.340\angle 99.1^\circ)} \tag{2.92}$$

$$= -0.4117 + \jmath 0.3839. \tag{2.93}$$

That is, $Z_{\text{IN}} = 15.959 + \jmath 17.935\ \Omega$. So taking into account the bias objectives and the output matching networks shown in Figure 2-25, the matching network topology of Figure 2-25(c) will be used (where the load is the active device). The input matching network problem is as shown in Figure 2-26(c). (Biasing cannot be incorporated into this matching network.) Now

$$|Q_S| = |Q_P| = \sqrt{\frac{R_S}{R_L} - 1} = \sqrt{\frac{50}{15.959} - 1} = 1.4605 \tag{2.94}$$

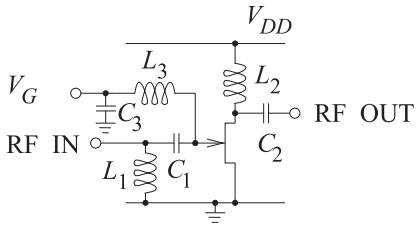
$$Q_S = \left|\frac{X_S}{R_L}\right| = \left|\frac{X_S}{15.959\ \Omega}\right| = 1.4605 \quad \text{and} \quad Q_P = \left|\frac{R_S}{X_P}\right| = \left|\frac{50\ \Omega}{X_P}\right| = 1.4605. \tag{2.95}$$

$$\text{So} \qquad X_S = -23.31\ \Omega \quad \text{and} \quad X_P = 34.23\ \Omega. \tag{2.96}$$

Since $X_x = 17.935\ \Omega$, $-17.935\ \Omega$ must be added to X_S, and the reactance of C_i is $41.24\ \Omega$, thus (at 8 GHz),

$$L_i = 681\ \text{pH} \quad \text{and} \quad C_i = 482\ \text{fF}. \tag{2.97}$$

The final input matching network design is shown in Figure 2-24(d).



$L_1 = L_i$ = 681 pH
$L_2 = L_o$ = 1.00 nH
L_3 = 10 nH
$C_1 = C_i$ = 482 fF
$C_2 = C_o$ = 1.064 pF
C_3 = 100 pF

Figure 2-27: Final amplifier schematic.

2.9.5 Bias Network Design

The final schematic of the linear amplifier design is shown in Figure 2-27. The output matching network, L_2 and C_2, enabled the biasing inductor to be replaced by L_2. So the output bias circuitry is absorbed into the output matching network. A similar result is not obtained with the input matching network, L_1 and C_1. A separate gate bias network is still required. L_3 should be a large enough value for it to act as an RF choke. A value of 10 nH provides a reactance of approximately 500 Ω at 8 GHz. The value of C_3 = 100 pF is chosen large enough to provide an RF short and stabilize the DC bias, V_G. This is a surprisingly simple circuit that provides maximum gain at 8 GHz, ensures out-of-band stability, and provides DC bias. Another design iteration with a more sophisticated input matching network may enable the separate bias inductor L_3 to be eliminated. As it is, the gate bias circuit further ensures stability at low frequencies, as then the gate tends to be shorted out. The amplifier has a calculated transducer gain of 13.2 dB, which can be compared to the gains reported in Table 2-2, where gain metrics were determined with S_{12} ignored.

Linear amplifier design for a specific gain is also possible. Now the errors involved in ignoring S_{12} during the design process are significant and a full bilateral treatment is required. This design approach is described in references [11, 27, 28].

2.10 Summary

This chapter addressed the design of narrowband amplifiers, but this forms the basis of wideband and power amplifier design to be considered in the following chapters. The bandwidth of an amplifier is dictated by the frequency-dependent characteristics of the active device and at microwave frequencies the device parasitic capacitances are usually significant. Without special broadbanding concepts, the narrowband approach covered in this chapter is good, usually, for amplifier designs with up to 5% bandwidth.

The basic topology used in amplifiers is an input matching network, an active device, and an output matching network. This arrangement is one of three cascaded two-ports. Sometimes an additional two-port is used in parallel with the active device to provide feedback and ensure stability or a flat gain response over frequency. With narrowband design the use of a feedback network is rarely required. The input matching network provides near-maximum power transfer from the system impedance to the usually higher input impedance of the active device. The near-maximum qualification is used since the requirements for maximum power transfer at the input conflict with the conditions for best noise performance. This originates because the active device has multiple partially correlated

physical noise sources and the input matching network affects how the correlated noise sources are combined so that it is possible to minimize their contributions. The active device is followed by an output matching network that matches the output impedance of the active device to the usually higher system impedance. In narrowband amplifier design the device capacitive parasitics are often incorporated into the matching networks. The topology of the input and output matching networks is chosen to ensure out-of-band stability and provide bias with minimum additional components.

Amplifier design is driven by metrics for the power gains at various optimum conditions, and the various gain metrics are used at various stages in design. The gain metrics are also used in choosing an active device and in estimating the design complexity that can be expected.

Amplifier design is a major endeavor and many books have been written about particular aspects of RF and microwave amplifier design. This chapter covered the main topics and also presented treatments that are broadly applied. The reader is directed to references [11, 14, 29–36] for specialized aspects of amplifier design. Numerous references are available for understanding, analyzing, and characterizing distortion in greater depth than covered here [37–48].

2.11 References

[1] QORVO, "FPD6836P70 data sheet, low noise high frequency packaged enhancement mode phemt transistor," http://www.qorvo.com.

[2] M. Steer, *Microwave and RF Design, Networks*, 3rd ed. North Carolina State University, 2019.

[3] J. Rollett, "Stability and power-gain invariants of linear twoports," *Circuit Theory, IRE Trans. on*, vol. 9, no. 1, pp. 29–32, Mar. 1962.

[4] ——, "Correction to stability and power-gain invariants of linear twoports," *IEEE Trans. on Circuit Theory*, vol. 10, no. 1, p. 107, Mar. 1963.

[5] M. Gupta, "Power gain in feedback amplifiers, a classic revisited," *IEEE Trans. on Microwave Theory and Techniques*, vol. 40, no. 5, pp. 864–879, May 1992.

[6] H. Fukui, "Available power gain, noise figure, and noise measure of two-ports and their graphical representations," *IEEE Trans. on Circuit Theory*, vol. 13, no. 2, pp. 137–142, Jun. 1966.

[7] M. Steer, *Microwave and RF Design, Transmission Lines*, 3rd ed. North Carolina State University, 2019.

[8] S. Lucyszyn, "Power-added efficiency errors with RF power amplifiers," *Int. J. of Electronics*, vol. 82, no. 3, pp. 303–312, Mar. 1997.

[9] F. Raab, P. Asbeck, S. Cripps, P. Kenington, Z. Popovic, N. Pothecary, J. Sevic, and N. Sokal, "Power amplifiers and transmitters for RF and microwave," *IEEE Trans. on Microwave Theory and Techniques*, vol. 50, no. 3, pp. 814–826, Mar. 2002.

[10] L. Esaki and R. Tsu, "Superlattice and negative differential conductivity in semiconductors," *IBM Journal of Research and Development*, vol. 14, no. 1, pp. 61–65, Jan. 1970.

[11] G. Gonzalez, *Microwave Transistor Amplifiers: Analysis and Design*, 2nd ed. Prentice Hall, 1997.

[12] A. Suárez and R. Quéré, *Stability Analysis of Nonlinear Microwave Circuits*. Artech House, 2003.

[13] M. Edwards and J. Sinsky, "A new criterion for linear 2-port stability using a single geometrically derived parameter," *IEEE Trans. on Microwave Theory and Techniques*, vol. 40, no. 12, pp. 2303–2311, Dec. 1992.

[14] T. Ha, *Solid State Microwave Amplifier Design*. Wiley, 1981.

[15] D. Woods, "Reappraisal of the unconditional stability criteria for active 2-port networks in terms of s parameters," *IEEE Trans. on Circuits and Systems*, vol. 23, no. 2, pp. 73–81, Feb. 1976.

[16] W. Ku, "Unilateral gain and stability criterion of active two-ports in terms of scattering parameters," *Proc. of the IEEE*, vol. 54, no. 11, pp. 1617–1618, Nov. 1966.

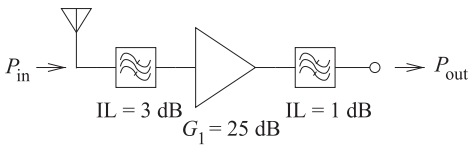
[17] D. Youla, "On the stability of linear systems," *IEEE Trans. on Circuit Theory*, vol. 10, no. 2, pp. 276–279, Jun. 1963.

[18] R. Meys, "Review and discussion of stability criteria for linear 2-ports," *IEEE Trans. on Circuits and Systems*, vol. 37, no. 11, pp. 1450–

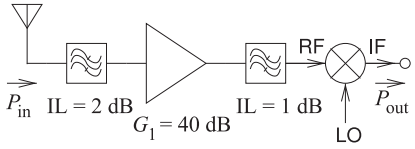
1452, Nov. 1990.
[19] G. Bodway, "Two port power flow analysis using generalized scattering parameters," *Microwave Journal*, pp. 61–69, May 1967.
[20] E. Faulkner, *Introduction to the Theory of Linear Systems*. Chapman & Hall, 1969.
[21] J. D'Azzo and C. Houpis, *Feedback Control System Analysis and Synthesis*. McGraw-Hill, 1960.
[22] A. Pippard, *Response & Stability*. Cambridge University Press, 1985.
[23] G. Franklin, *Feedback Control of Dynamic Systems*. Prentice Hall, 2002.
[24] K. Wang, M. Jones, and S. Nelson, "The s-probe-a new, cost-effective, 4-gamma method for evaluating multi-stage amplifier stability," in *1992 IEEE MTT-S Int. Microwave Symp. Dig.*, Jun. 1992, pp. 829–832.
[25] H. Bode, *Network analysis and feedback amplifier design*. Van Nostrand Company, 1951.
[26] M. Steer, *Microwave and RF Design, Modules*, 3rd ed. North Carolina State University, 2019.
[27] G. Vendelin, A. Pavio, and U. Rohde, *Microwave Circuit Design Using Linear and Nonlinear Techniques*. Wiley, 1990.
[28] I. Bahl and P. Bhartia, *Microwave Solid State Circuit Design*. John Wiley & Sons, 1988.
[29] A. Grebennikov, *RF and microwave power amplifier design*. McGraw-Hill, 2005.
[30] A. Shirvani, *Design and Control of RF Power Amplifiers*. Kluwer Academic, 2003.
[31] P. Kenington, *High-Linearity RF Amplifier Design*. Artech House, 2000.
[32] A. Grebennikov, *Switchmode RF Power Amplifiers*. Elsevier/Newnes, 2007.
[33] S. Cripps, *RF Power Amplifiers for Wireless Communications*. Artech House, 1999.
[34] ——, *Advanced techniques in RF Power Amplifiers Design*. Artech House, 2002.
[35] P. Reynaert, *RF Power Amplifiers for Mobile Communications*. Springer, 2006.
[36] M. Steer, J. Sevic, and B. Geller, "Editorial," *IEEE Trans. on Microwave Theory and Techniques*, vol. 49, no. 6, pp. 1145–1147, Jun. 2001.
[37] M. Steer and K. Gharaibeh, "Volterra modeling for analog and microwave circuits," in *Encyclopedia of RF and Microwave Engineering*. John Wiley, 2005, pp. 5507–5514.
[38] J. Sevic and M. Steer, "Analysis of GaAs MESFET spectrum regeneration driven by a $\pi/4$-DQPSK modulated source," in *1995 IEEE MTT-S Int. Microwave Symp. Dig.*, May 1995, pp. 1375–1378.
[39] J. Pedro and N. Carvalho, *Intermodulation Distortion in Microwave and Wireless Circuits*. Artech House, 2003.
[40] T. Turlington, *Behavioral Modeling of Nonlinear RF and Microwave Devices*. Artech House, 2000.
[41] S. Maas, *Nonlinear Microwave and RF Circuits*, 2nd ed. Artech House, 2003.
[42] J. Vuolevi and T. Rahkonen, *Distortion in RF Power Amplifiers*. Artech House, 2003.
[43] J. Hu, K. Gard, N. Carvalho, and M. Steer, "Time-frequency characterization of long-term memory in nonlinear power amplifiers," in *Microwave Symp. Dig., 2008 IEEE MTT-S Int.*, Jun. 2008, pp. 269–272.
[44] K. Gharaibeh and M. Steer, "Characterization of cross modulation in multichannel amplifiers using a statistically based behavioral modeling technique," *IEEE Trans. on Microwave Theory and Techniques*, vol. 51, no. 12, pp. 2434–2444, 2003.
[45] M. Steer, J. Bandler, and C. Snowden, "Computer-aided design of RF and microwave circuits and systems," *IEEE Trans. on Microwave Theory and Techniques*, vol. 50, no. 3, pp. 996–1005, Mar. 2002.
[46] H. Gutierrez, K. Gard, and M. Steer, "Nonlinear gain compression in microwave amplifiers using generalized power-series analysis and transformation of input statistics," *IEEE Trans. on Microwave Theory and Techniques*, vol. 48, no. 10, pp. 1774–1777, Oct. 2000.
[47] G. Rhyne and M. Steer, "Generalized power series analysis of intermodulation distortion in a MESFET amplifier simulation and experiment," in *1987 IEEE MTT-S Int. Microwave Symp. Dig.*, 1987, pp. 115–118.
[48] K. Gard, H. Gutierrez, and M. Steer, "Characterization of spectral regrowth in microwave amplifiers based on the nonlinear transformation of a complex gaussian process," *IEEE Trans. on Microwave Theory and Techniques*, vol. 47, no. 7, pp. 1059–1069, Jul. 1999.

2.12 Exercises

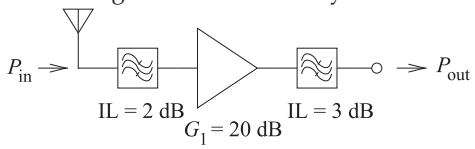
1. What is the gain of the following receiver system?



2. In the system below the mixer has a conversion loss of 10 dB. What is the gain of the receiver system?



3. What is the gain of the receiver system below?



4. A source that drives an amplifier has an available output power of 1 mW. The amplifier has been optimally matched in a $50\ \Omega$ system and then has a small signal gain of 20 dB. The amplifier load is now changed and the new load is mismatched with a VSWR of 1.5. What is the power delivered to the new load?

5. A MOSFET amplifier has the small signal S parameters $S_{11} = 0.3\angle 85^\circ$, $S_{12} = 0.05\angle 15^\circ$, $S_{21} = 2.5\angle 100^\circ$, and $S_{22} = 0.85\angle -50^\circ$ at 5.6 GHz.
   1. What is the maximum unilateral transducer gain?
   2. What is the maximum available power gain?
   3. What is the maximum stable power gain?
   4. What is the unilateral power gain?

6. An amplifier has a gain of 10 dB, an output power of 1 W, and a power-added efficiency of 25%.
   (a) What is the total efficiency of the amplifier as a percentage?
   (b) What is the efficiency of the amplifier as a percentage?

7. A Class A BJT amplifier has a collector bias voltage of 5 V and a collector bias current of 100 mA.
   (a) What is the efficiency of the amplifier if the RF output power is 1 mW?
   (b) What is the efficiency of the amplifier if the RF output power is 10 mW?
   (c) What is the efficiency of the amplifier if the RF output power is 100 mW?

8. A Class A MOS RF amplifier has a drain bias voltage of 20 V and a drain bias current of 1 A. If the output power of the amplifier is 5 W and the available input power is 1 W, what is the power-added efficiency of the amplifier?

9. A power amplifier with a gain of 10 dB draws 100 W of DC power and delivers 50 W of RF output power. What is the power-added efficiency of the amplifier?

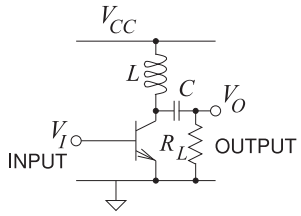
10. A FET power amplifier with a gain of 10 dB draws 100 W of DC power and delivers 50 W of RF output power. What is the drain efficiency of the amplifier?

11. Consider the design of a 15 GHz inductively biased Class A amplifier using a transistor with 50 Ω S parameters $S_{11} = 0.5\angle 45^\circ$, $S_{12} = 0.1\angle 0\circ$, $S_{21} = 2\angle 90^\circ$, and $S_{22} = 0.75\angle 45^\circ$.
   (a) If the input of the transistor is terminated in $50\ \Omega$, what is the impedance looking into the output of the transistor?
   (b) Design a two-element lumped-element output matching network for maximum power transfer from the output of the transistor into a $50\ \Omega$ load.

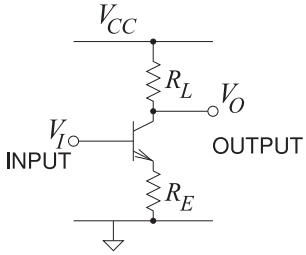
12. Consider the design of a 15 GHz inductively biased Class A amplifier using a transistor with 50 Ω S parameters $S_{11} = 0.96\angle 85^\circ$, $S_{12} = 0.056\angle 15^\circ$, $S_{21} = 2.56\angle 100^\circ$, and $S_{22} = 0.320\angle 54.6^\circ$.
   (a) If the input of the transistor is terminated in 50 Ω, what is the impedance looking into the output of the transistor?
   (b) Design a two-element lumped-element output matching network for maximum power transfer from the output of the transistor into a $50\ \Omega$ load.
   (c) As the first step in evaluating the power gain of the amplifier, determine which of the various gains defined for an amplifier is the power gain here. That is, several gains are defined in terms of S parameters and reflection coefficients (e.g., available gain, maximum stable gain, etc.). Which of these can be used to evaluate the power gain in this circumstance where there is not an input matching network, but there is an output matching network?
   (d) What is the power gain of the amplifier in decibels?

13. Consider the design of a 10 GHz inductively biased Class A amplifier using a transistor with 50 Ω S parameters $S_{11} = 0.9\angle 80^\circ$, $S_{12} = 0.06\angle 15^\circ$, $S_{21} = 2.5\angle 10^\circ$, and $S_{22} = 0.3\angle 45^\circ$.
   (a) If the input of the transistor is terminated in 55.5 Ω, what is the impedance looking into the output of the transistor?
   (b) Design a two-element lumped-element output matching network for maximum power transfer from the output of the transistor into a $50\ \Omega$ load.
   (c) What is the power gain of the amplifier in decibels?

14. The Class A BJT amplifier in the figure below has an RF choke providing collector current and acts as an open circuit at RF. The load, R_L, is driven through a capacitor, C, which is effectively a short circuit at RF. The maximum undistorted efficiency of this circuit is 50%. Derive this efficiency. Ignore the base-emitter voltage drop, $V_{CE,\min}$, and note that the maximum of V_O is $2V_{CC}$, allowing a voltage swing of $\pm V_{CC}$ around the collector quiescent operating voltage. [Parallels Example 2.2]



15. The Class A BJT amplifier in the figure below has a load, R_L, and a maximum undistorted efficiency of 25%. Derive the efficiency of this amplifier in terms of R_E and R_L. Assume that V_{CC} is much greater than V_{BE}. [Parallels Example 2.2]



16. Consider a Class C BJT amplifier with a resistive bias that is also the RF load. The supply voltage is 10 V.
    (a) Draw the loadline of the amplifier and indicate the loadline and bias point.
    (b) What is the collector bias current with no RF input signal?
    (c) With the RF input to the amplifier having a power of 10 mW, the RF output power is 100 mW, the quiescent collector-emitter voltage is 6 V, and the quiescent collector current is 20 mA. What is the power-added efficiency of the amplifier under these conditions?

17. Consider the design of a 15 GHz inductively biased Class A amplifier using the transistor in Figure 2-2 and with a $50\ \Omega$ source.
    1. What is the impedance presented at the output of the transistor?
    2. Design a two-element output matching network for maximum power transfer into a $50\ \Omega$ load.

18. A MOSFET amplifier has the small signal S parameters $S_{11} = 0.8\angle 90°$, $S_{12} = 0.05\angle 0°$, $S_{21} = 2.5\angle 0°$, and $S_{22} = 0.8\angle 0°$ at 5.6 GHz.
    (a) Calculate the radius and center of the input stability circle.
    (b) Draw conclusions from the plot of the input stability circle. That is, what restrictions are placed on the input matching network if the amplifier load is passive?

19. A MOSFET amplifier has the small signal S parameters $S_{11} = 0.9\angle 85°$, $S_{12} = 0.05\angle 15°$, $S_{21} = 2.5\angle 100°$, and $S_{22} = 0.85\angle -50°$ at 5.6 GHz.
    (a) Calculate the radius and center of the output stability circle.
    (b) Draw the output stability circle on a Smith chart.
    (c) Draw conclusions from the plot of the output stability circle. That is, what restrictions are placed on the output matching network?

20. A MOSFET amplifier has the small signal S parameters $S_{11} = 0.9\angle 85°$, $S_{12} = 0.025\angle 15°$, $S_{21} = 3\angle 100°$, and $S_{22} = 0.85\angle -50°$ at 2 GHz.
    (a) Calculate the radius and center of the input stability circle.
    (b) Draw conclusions from the plot of the input stability circle. That is, what restrictions are placed on the input matching network?

21. A MOSFET amplifier has the small signal S parameters $S_{11} = 0.9\angle 85°$, $S_{12} = 0.05\angle 15°$, $S_{21} = 2.5\angle 100°$, and $S_{22} = 0.85\angle -50°$ at 5.6 GHz.
    (a) What is the k-factor of Rollet's stability criterion?
    (b) What does the k-factor indicate about the stability of the transistor?
    (c) What is the μ-factor of the Edwards–Sinsky stability criterion?
    (d) What does the Edwards–Sinsky stability criterion indicate about the stability of the transistor?

22. Consider the design of a 15 GHz inductively biased Class A amplifier using the pHEMT transistor documented in Figure 2-2. Use the topology shown in Figure 2-20.
    (a) If the input of the transistor is terminated in $55.5\ \Omega$, what is the impedance looking into the output of the transistor?
    (b) Design a two-element output matching network for maximum power transfer into a $50\ \Omega$ load.

23. Consider the design of a 15 GHz inductively biased Class A amplifier using the pHEMT transistor documented in Figure 2-2. Use the topology shown in Figure 2-20.
    (a) If the input of the transistor is terminated in 150 Ω, what is the impedance looking into the output of the transistor?
    (b) Design a two-element output matching network for maximum power transfer into a 50 Ω load.
24. Consider the design of a 15 GHz inductively biased Class A amplifier using the pHEMT transistor documented in Figure 2-2. Use the topology shown in Figure 2-20.
    1. If the input of the transistor is terminated in 200 Ω, what is the impedance looking into the output of the transistor?
    2. Design a two-element output matching network for maximum power transfer into a 50 Ω load.
25. Design an amplifier for maximum stable gain using the discrete pHEMT transistor described in Figure 2-2. The design specifications are

Gain:	maximum gain at 24 GHz
Topology:	three two-ports (input and output matching networks, and the active device)
Stability:	broadband stability
Bandwidth:	maximum that can be achieved using two-element matching networks
Source Z:	$Z_S = 10\ \Omega$
Load Z:	$Z_L = 50\ \Omega$

26. Design an amplifier for maximum stable gain using the discrete pHEMT transistor described in Figure 2-2. The design specifications are

Gain:	maximum gain at 23 GHz
Topology:	three two-ports (input and output matching networks, and the active device)
Stability:	broadband stability
Bandwidth:	maximum that can be achieved using two-element matching networks
Source Z:	$Z_S = 50\ \Omega$
Load Z:	$Z_L = 50\ \Omega$

27. An inductively biased Class A HBT amplifier is biased with a collector-emitter quiescent voltage of 5 V and a quiescent collector-emitter current of 100 mA. When operated at the 1 dB gain compression point, the input RF power is 10 mW and the output power is 100 mW. Consider that the RF signal is a sinewave, and note that the quiescent collector-emitter voltage will be the supply rail voltage.
    (a) What is the quiescent DC power consumed? Express your answer in milliwatts.
    (b) What is the output power in dBm?
    (c) What is the efficiency of the amplifier? Note that the efficiency of a Class A amplifier can be more than 25% if distortion is tolerated.
    (d) What is the power-added efficiency of the amplifier?
    (e) If the input power is reduced by 10 dB so that the amplifier is no longer in compression, will the DC quiescent point change? Explain your answer.
    (f) If the input power is reduced 10 dB so that the amplifier is no longer in compression, what is the output power in dBm? Ignore any change in the quiescent point.
    (g) With 1 mW input power, what is the power-added efficiency of the amplifier if the quiescent point does not change?

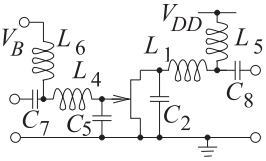
2.12.1 Exercises By Section

†challenging, ‡very challenging

§2.1 1, 2, 3
§2.3 4, 5†
§2.4 6, 7, 8, 9, 10
§2.5 11†, 12†, 13†, 14†, 15†, 16†, 17
§2.6 18†, 19†, 20, 21
§2.9 22†, 23†, 24†, 25‡, 26‡, 27†

2.12.2 Answers to Selected Exercises

1 1 dB
6 27.8%
5(d) 14.2 dB
10 50%
11(a) $436 - j105.6\ \Omega$
12(a) $61.3 - j35.6\ \Omega$
21(c) 0.563
25



$L_1 = 36.8$ fH, $C_2 = 36.8$ fF
$L_3 = 321$ pH, $C_4 = 27.4$ fF
L_5, L_6, C_7, C_8 are large
27 2.3

CHAPTER 3

Wideband Amplifiers

3.1 Introduction 69
3.2 Distributed Amplifiers 71
3.3 Case Study: Analysis of a Distributed Amplifier 72
3.4 Negative Image Amplifier Design 74
3.5 Case Study: Wideband Amplifier Design 77
3.6 Differential Amplifiers 86
3.7 Case Study: Distributed Biasing of Differential Amplifiers 96
3.8 Amplifiers and RFICs 97
3.9 Summary 101
3.10 References 102
3.11 Exercises 103

3.1 Introduction

Wideband amplifier design requires the synthesis of matching networks that provide a match over considerable bandwidths. The divisions between narrowband, wideband and ultra-wideband microwave amplifier design depend on operating frequency and the amplifier efficiency required. Generally, however, a microwave amplifier with a half-octave bandwidth, e.g. 2 to 3 GHz, is regarded as a wideband design.

The essential amplifier design problem is that at microwave frequencies the parasitic input and output capacitances of a transistor are significant and these must be canceled to achieve maximum power transfer. Synthesis of the input and output matching networks of a microwave amplifier at a single frequency leads to a narrowband amplifier with a bandwidth of perhaps 2–3%. At lower frequencies where the parasitic capacitances are less significant, the fractional bandwidth may be greater. An ideal response would be achieved if there were negative capacitors and typically resonance of lumped-elements can be used to at least partially present a negative capacitance-like characteristic over about a quarter-octave bandwidth.

The dominant reactive parasitics of a transistor are its input and output capacitances, but also the feedback capacitor between the collector/drain and base/gate becomes important at higher frequencies. Ignoring the feedback capacitance and thinking just about the input of the transistor, the input of the transistor is a capacitance sometimes in series (for a BJT)

and sometimes in parallel (for a FET) with the resistance describing the absorption of RF input power by the transistor. Ideal matching requires the synthesis of a negative capacitor (i.e., an element which has an inductive reactance that reduces with frequency). Simply using an inductor to provide matching provides a matching element whose impedance increases with frequency. The wideband input matching problem becomes essentially the synthesis of a terminated two-port network with an input impedance that has the required negative capacitance characteristic. This is not easy to achieve using lumped-elements alone.

This chapter presents three strategies for designing wideband linear amplifiers. One uses the image impedance method in which the required negative capacitance impedance is realized using a transmission line network. The next is a multistage distributed amplifier that incorporates the transistor capacitances into a transmission line. The third approach is akin to a parallel coupled-line filter design.

3.1.1 Wideband Amplifier Design Strategies

Generally a wideband amplifier has a half-octave bandwidth, e.g. 8 GHz to 12 GHz. Multiple objectives must be met in wideband amplifier design. Of course the gain must be flat over the specified bandwidth but it is also important to meet noise and stability objectives over the bandwidth. Of course the amplifier must be stable out-of-band as well. It is generally not possible to meet all of these objectives using computer optimization and it is necessary to simplify the design process. When computer optimization is used, it is done in stages and begins with a prototype design that is not too far away from the final design.

An ultra-wideband amplifier has a bandwidth of more than a half octave. There are two approaches to achieving ultra-wide bandwidth and both types of amplifiers have low efficiency. The first category of ultra-wideband amplifier is the distributed amplifiers which achieves multi-octave bandwidth by incorporating the parasitic capacitances of transistors in an artificial transmission line. The parasitic inductances are usually negligible but if not, they are incorporated in the artificial transmission line. In effect there is a multi-stage amplifier and each stage must be a Class A stage and thus have very low efficiency, think 5%. Ultra-wideband distributed amplifiers tend to be used in instrumentation. A non-aggressive Class A amplifier design is more likely to be stable. Distributed amplifier design is considered in Section 3.2 and a case study of a distributed amplifier in Section 3.3.

A second type of ultra-wideband amplifier is an operational amplifier with very high levels of feedback. In an operational amplifier the open loop amplifier (without feedback) has very high gain, but a gain that varies significantly with frequency. Then feedback is used, the loop is closed, to effectively throw away most of the gain to obtain an overall flatter gain over a wide bandwidth. This type of amplifier has very low efficiency and usually the gains available from microwave transistors are not high enough anyway. Even with the highest performing transistors, that is ones with very high S_{21} to S_{12} ratio, the transistors tend to be very expensive requiring finer lithography to achieve the required shorter gate. Microwave operational amplifiers really are not viable and so will not be considered further.

The highest bandwidth of a microwave amplifier that achieves flat gain across the band, has good efficiency, and meets noise and stability requirements is about half an octave. A straight-forward approach would seem to be to simultaneously design the input and output networks and employ computer optimization. This is complicated at microwave frequencies because feedback from the output to the input, i.e. S_{12}, is large. Design then becomes an optimization problem with multiple objectives and many parameters to adjust. An optimization-only approach rarely works. It is essential to simplify the problem and approach design in stages. The most successful wideband amplifier design technique is the negative image design method which begins by placing hypothetical negative capacitors in parallel with the input and output capacitances of a transistor. The procedure will be described in Section 3.4 and then a case study is presented in Section 3.5.

A final class of microwave amplifiers that achieves reasonably high bandwidths are the differential amplifiers used in RFICs. These are considered in Sections 3.6–3.8.

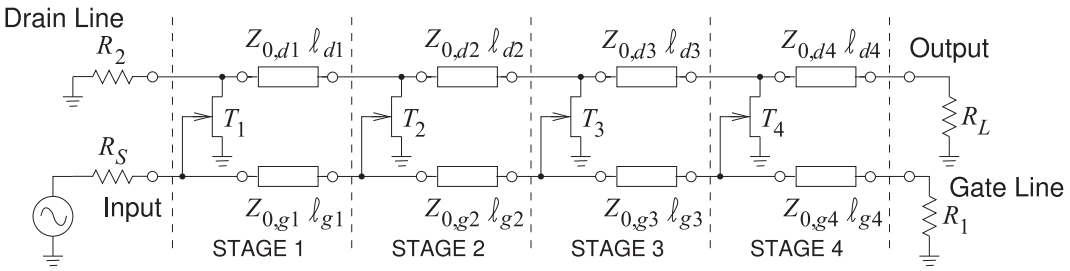
3.2 Distributed Amplifiers

Distributed amplifiers use the ability of transmission lines to combine the output of multiple transistor stages to realize an amplifier with a bandwidth of more than a decade [1, 2]. While the bandwidth is wider than that of the single-stage wideband amplifier discussed in the previous section, the efficiency is much lower.

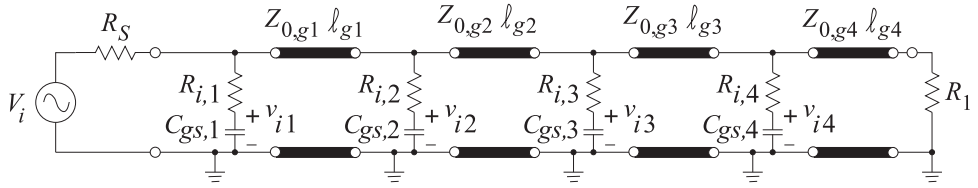
The topology of a four-stage distributed amplifier is shown in Figure 3-1(a). The distributed amplifier has two transmission lines, referred to as the gate line and drain line. Each stage includes an active device and two sections of transmission line, one being part of the gate line and the other being part of the drain line. The small-signal model of the input is shown in Figure 3-1(b) and that of the output in Figure 3-1(c). In the small signal model, the input of the transistor is modeled as a series resistance, $R_{i,n}$, and capacitance, $C_{gs,n}$, and these load the gate line. Ignoring $R_{i,n}$ for now, the small-signal input model is a transmission line that is loaded periodically by capacitors. This therefore appears as an artificial transmission line. If the line is terminated in an appropriate resistance, R_1, then no input signal is reflected at the end of the line segment. Proper design would result in very little power on the gate line after the final stage, as power is periodically coupled into the transistors. Design also ensures that there is a negligible backward-traveling wave on the input transmission line.

The small-signal model of the output drain line is similar, with a line periodically loaded by the output resistance and capacitance of the transistors. Now, however, there is a controlled current source that injects power onto the drain line. If there was only one stage, then the power delivered to the drain line would be split equally between forward- and backward-traveling components. However, here there are multiple stages, and the phase of the drain current injection changes along the line and current is preferably coupled into the forward-traveling wave. Still a termination resistor R_2 ensures that there is no backward-traveling wave on the drain line.

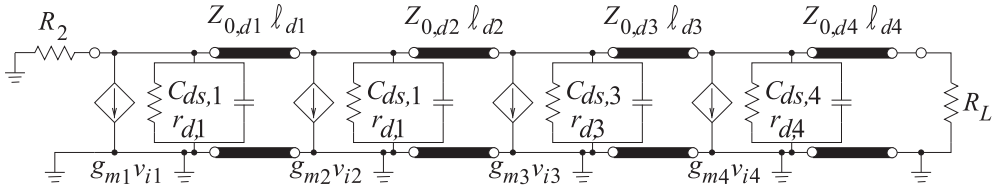
Power is periodically being tapped off of the gate line and an amplified signal is periodically inserted on the drain line. As a result, the transistors



(a) Distributed amplifier topology



(b) Small-signal input model (gate line)



(c) Small-signal output model (drain line)

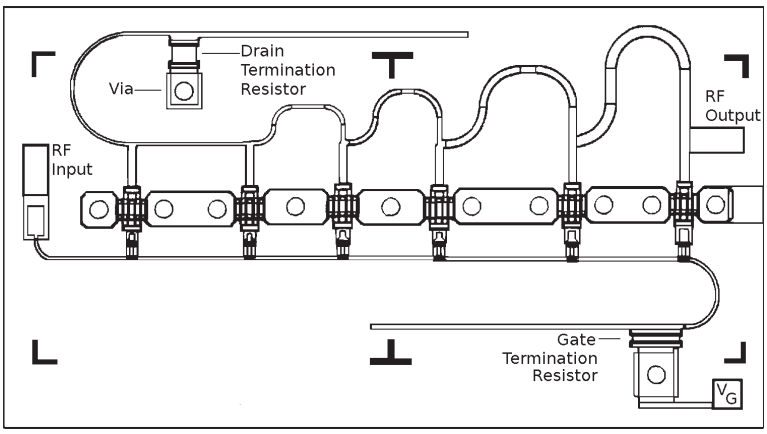
Figure 3-1: Distributed amplifier with four stages.

often are designed to increase in size along the length of the line. In this case the input and output capacitances of the transistors increase with each stage. Even if the transistors in each stage are of equal size, the characteristic impedances of the drain and gate transmission lines vary for each stage, and the lengths of the lines in the drain stage are not the same as the lengths of the lines in the gate line.

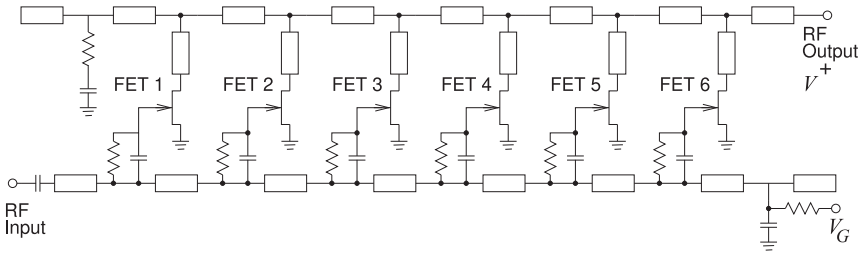
Distributed amplifiers can simplify stability constraints and enable amplification over multiple octaves. They also find application at millimeter-wave frequencies even when bandwidths of greater than one-half octave are not required [3]. At millimeter-wave frequencies parasitic capacitances are significant and these can be incorporated into the synthesis of the loaded transmission lines. Since the need to cancel parasitic capacitances is not required, it can be easier to achieve stable amplification.

3.3 Case Study: Analysis of a Distributed Amplifier

Figure 3-2(a) shows the layout of a monolithically integrated, large signal, distributed, GaAs power amplifier [4], model TGA8220 from Texas Instruments. The schematic of the amplifier is shown in Figure 3-2(b). This amplifier circuit is designed to deliver +25 dBm output power at 1 dB compression when operating from 2–18 GHz. Identical FETs are used in



(a) Layout



(b) Schematic

Figure 3-2: The TGA8220 MMIC distributed amplifier with six FETs (numbered 1–6 from the RF input). After [4], copyright *Microwaves & RF*, used with permission.

the six-stage amplifier and have a gate length $L = 0.5$ μm and gate width $W = 335$ μm. Models of bends, tee junctions, vias, bond wires, and FET parasitics must be used. After bias and RF sources are defined, the circuit can be simulated in a nonlinear microwave simulator [5].

The small signal gain and the output power and efficiency at 1 dB gain compression are shown in Figure 3-3 with V_{DS} = 8 V and 50% I_{DSS}. The circuit was designed with series-gate capacitors to increase the gate-line cut-off frequency and to tailor the gate voltage excitation to maximize output power. Further gain and power enhancements were achieved with the help of tapered drain lines and a large drain termination resistor. This enables a 1 dB compressed output power of +25 dBm with a power added efficiency of 10% to be obtained over the 2–18 GHz band. The small-signal gain of at least 19 dB also has a small positive slope with frequency.

Nonlinear circuit simulation can also be used to understand the intuitive operation of the nonlinear circuit. This is particularly important for complex circuits such as this one, which uses FETs that are increasingly loaded by distributing them in a systematic manner. One particularly useful form of display that aids in understanding is the I-V phase plane, which is the locus of the I-V trajectory of an individual device's operating characteristic. The dynamic loadline of each FET in the amplifier at 18 GHz for 1 dB gain

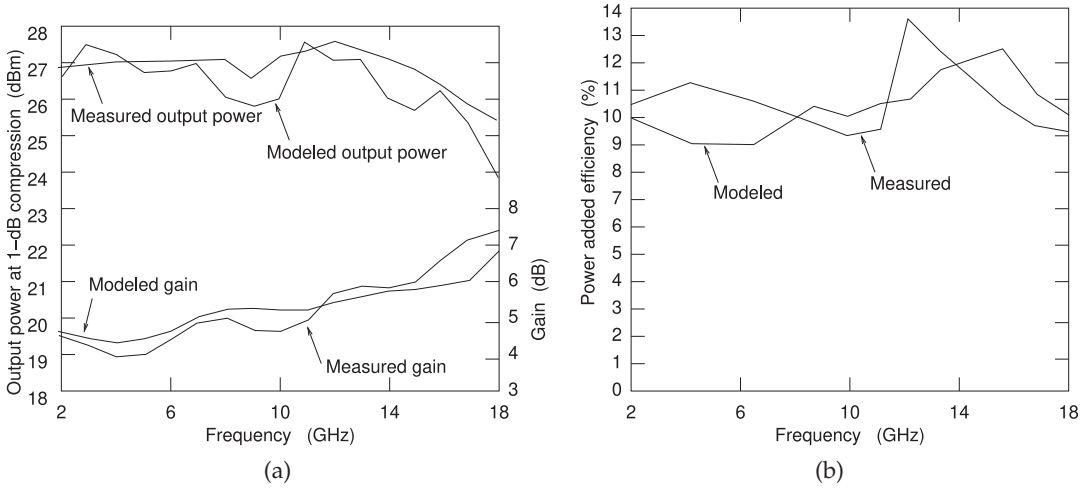


Figure 3-3: Gain, power, and efficiency of the TGA8220 MMIC distributed amplifier: (a) small-signal gain and output power at 1-dB compression; and (b) efficiency at 1-dB compression. After [4], copyright *Microwaves & RF*, used with permission.

compression at the output is given in Figure 3-4. The DC loadline (not shown) is a single line for each FET. The dynamic loadline opens up because of reactive loading. The match becomes better going from FET 1 to FET 6. Ideally the closed dynamic loadline would be achieved for all FETs, but there are many trade-offs required to achieve good gain, output power, and broad bandwidth. For efficiency, emphasis should be on maximizing the efficiency of the final stage, which operates at the highest power levels. The design here results in optimum output power across the entire bandwidth without unduly degrading the small-signal gain and input and output match for the remaining FETs in this amplifier. The use of series-gate capacitors and tapered drain lines is the mechanism that achieves this in this design [4]. A distributed amplifier delivers maximum output power and efficiency when each FET reaches gate and drain voltage limits for maximum power and efficiency simultaneously over a broad range of frequencies.

3.4 Negative Image Amplifier Design

Design using the negative image method is illustrated in the case study in Section 3.5 but here the philosophy behind the technique will be explained. The method breaks the stages the design in to much simpler steps.

In this section wideband amplifier design using the negative image method will be described for an amplifier having a single transistor. The idea can be applied to amplifier designs with multiple transistors. The basic model of a microwave transistor has shunt input and output capacitances and a feedback capacitance between the output and the input. A good amplifier design strategy would be to first place (ideal) negative capacitances in shunt with each of these capacitors and then attempt to synthesize a circuit that looks like a negative capacitor. This describes the essence of the negative image amplifier design method except that no attempt is made to directly

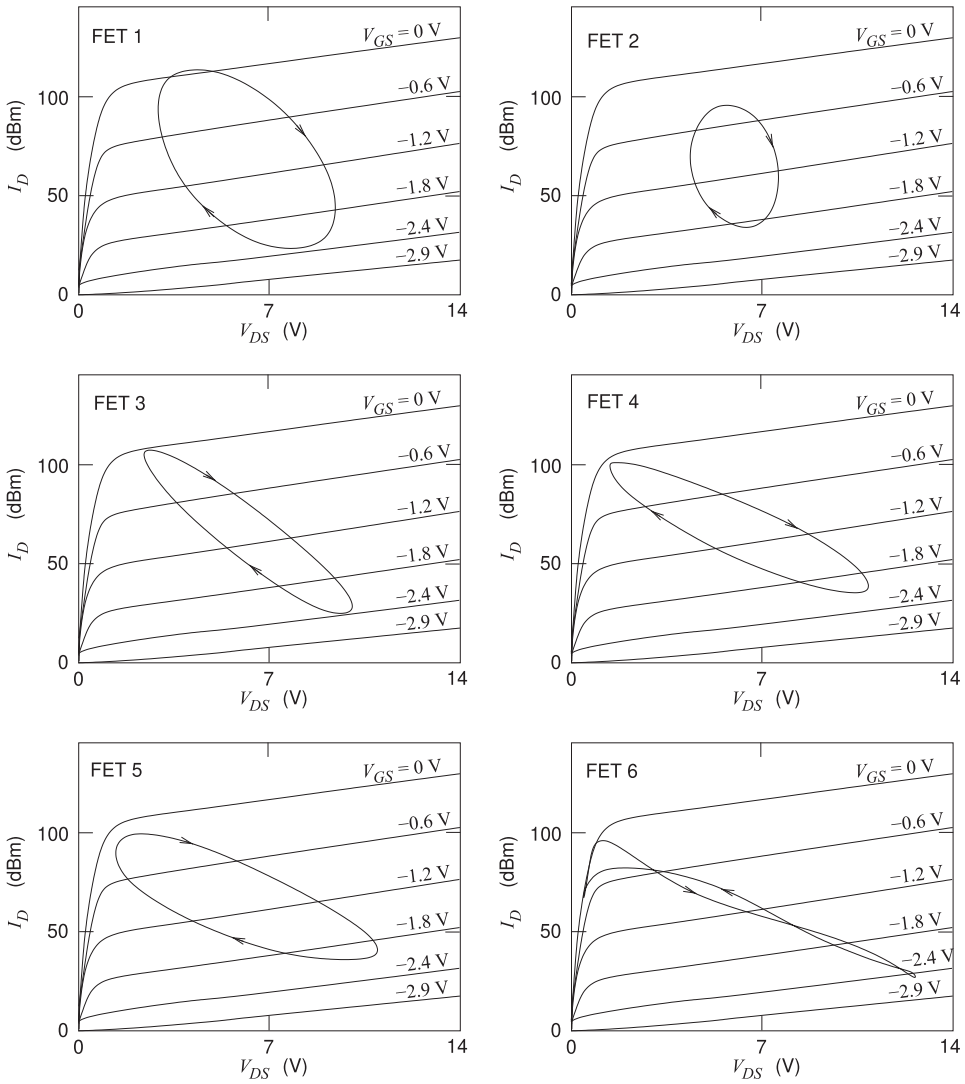


Figure 3-4: Dynamic loadlines of each FET in the TGA8220 MMIC distributed amplifier. After [4], copyright *Microwaves & RF*, used with permission.

cancel the transistor's feedback capacitor and instead the input and output matching networks are adjusted to account for it.

The negative image amplifier design method achieves high bandwidth by synthesizing input and output matching networks beginning with each network comprising a negative capacitance in parallel with the input and output, respectively, parasitic capacitances of the transistor. Then resistive loads, chosen for maximum power transfer, are put in parallel with the negative capacitances. This circuit is simulated with the actual model of the transistor and this is followed by an optimization step to optimize the broadband gain and noise responses of the amplifier while simultaneously ensuring stability of the amplifier. In this process the effect of the feedback

capacitance and the full complexity of the transistor are accounted for. The input and output matching networks are the negative image networks. The next step is to separately synthesize networks that provide the characteristics of the ideal input and output matching networks. This can only be done over a limited bandwidth but usually this is about half an octave. This process indicates the optimum characteristics of the input and output matching networks. The negative image networks can now be synthesized individually and once synthesized can be incorporated with the transistor model to obtain an overall circuit that can further be adjusted but perhaps only a few percent adjustment will be necessary.

The negative image method regularly achieves a half octave bandwidth. What is being done in the matching network design is using a topology that presents what looks like a negative capacitor and also the right impedance transformation (usually to 50 ohms). This is where invention comes into play. The impedance looking into the input (or output) of a transistor rotates with respect to frequency in the clockwise direction on a Smith chart. The complex conjugate impedance rotates in the counter-clockwise direction. The designer tries to develop a matching network that tracks the counter-rotating locus but with one capacitor and one inductor the impedance locus (with respect to frequency) looking into the matching network will rotate in the clockwise direction. The only simple circuit that will give you the right characteristics includes two or more transmission lines. The designer is using a topology that someone else discovered. It is not possible to synthesize the best network.

There are some limitations, both the input and output of the transistor will have some series inductance due to bondwires for discrete transistor parts and due to lengths of transmission line for on-chip transistors . The impedance of these inductances will be small compared to the parasitic capacitances and will really only matter if the input resistance of the transistor (in the case of the inductance at the input port) or the output resistance of the transistor (in the case of the inductance at the output port) is also small. Generally it is only necessary to compensate for the output inductance in the negative image model of the output network. and this is done by using use a negative inductance.

At this stage the amplifier design consists of input and output negative image matching networks which are quite simple and contain negative elements. The operation of the amplifier is optimized using these simple matching networks. The transistor is not unilateral so the input and output matching networks must be adjusted iteratively to get the optimum performance. The matching networks are so simple that manual tuning can be used in the circuit simulator..

With the simple matching networks designed, the challenge is now to realize the simple matching networks with real elements. L's and C's in a filter-like structure could possibly be used, but the result is rarely very good. The best results are obtained when transmission line structures are used. There is not a way to systematically synthesize these matching networks. The transmission line-based topologies that are used in the case study are inventions. There are very few other structures that work. The design problem is greatly simplified and design can focus on designing first the input matching network and then the output matching network without the transistor included.

3.5 Case Study: Wideband Amplifier Design

In this section an X-band wideband low-noise amplifier is designed.[1] The topology of the amplifier is shown in Figure 3-5, and this is the same topology used in narrowband amplifier design. The design specification is for a maximum noise figure (NF) of 1 dB and a gain of 14 ± 1 dB from 8 GHz to 12 GHz.

3.5.1 Transistor Properties

The transistor chosen is the packaged NEC NE32400A transistor and its parameters are given in Table 3-1 in what is called the Touchstone® format used by microwave simulators. The file begins with a number of comment lines (identified by '!') followed by the option line:

```
# <frequency unit> <parameter> <format> R <n>
```

where the <frequency unit> is GHz, <parameter> specifies the kind of network parameter data, and here S is scattering parameters. The <format>, MA, indicates that the data is in magnitude-angle(degrees) format, and the <n> term is 50, indicating that the S parameters are normalized to $50\ \Omega$. The line of data is ordered as f, $|S_{11}|$, $\angle S_{11}$, $|S_{21}|$, $\angle S_{21}$, $|S_{12}|$, $\angle S_{12}$, $|S_{22}|$, and $\angle S_{22}$. The second set of data contains the noise parameters and there are five entries for each frequency arranged as frequency (using the previously specified units), the minimum noise figure $\mathrm{NF}_{\min}$ (in dB), then $|\Gamma_{\mathrm{opt}}|$, $\angle\Gamma_{\mathrm{opt}}$, and $r_n/50$. These are the noise parameters used with two-port amplifiers as described in Section 4.3.6 of [6] with $\mathrm{NF}_{\min} = 10\log(F_{\min})$ and Γ_{opt} being the reflection coefficient of Γ_{opt} referenced to Y_0 (= 0.02 S here).

The S parameters of the transistor are plotted in Figure 3-6. S_{11}, S_{12}, and S_{22} are plotted on a Smith chart in Figure 3-6(a). However, S_{21} is greater than one and so this is plotted on a polar plot in Figure 3-6(b). All of the S parameters vary significantly with frequency. In Figure 3-6(a) the locus of S_{11} from 8 GHz to 12 GHz is highlighted and the segment is labeled A. Ideal matching would require that the reflection coefficient, Γ_S, looking into the lossless input matching network from the transistor be the complex conjugate of S_{11} (if Port 2 of the transistor is terminated in $50\ \Omega$).

[1] AWR Design Environment Project File: X_Band_LNA.emp

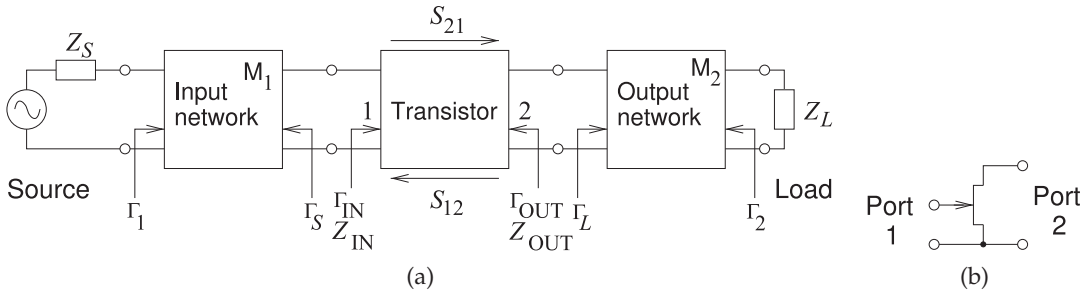
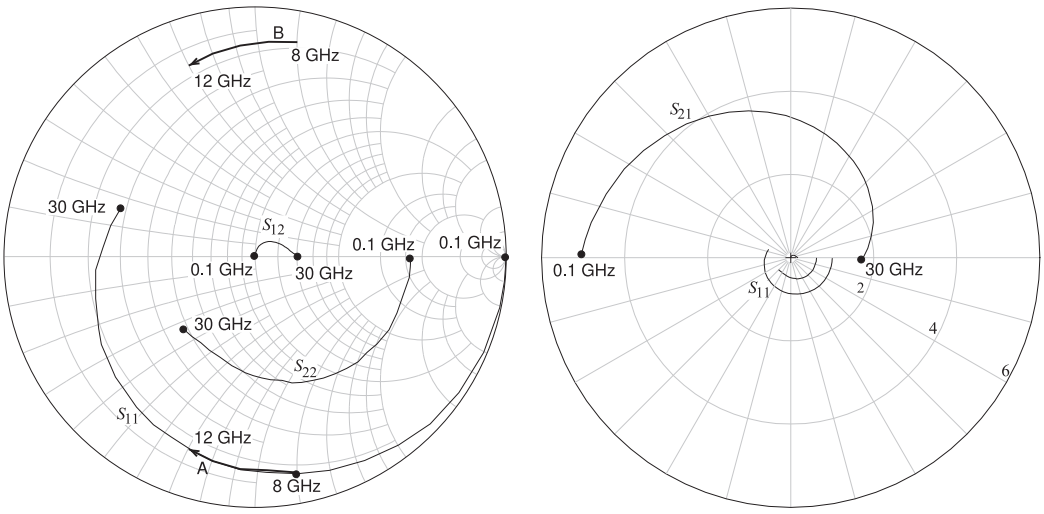


Figure 3-5: Wideband amplifier: (a) topology; and (b) port definition for transistor parameters.

Table 3-1: S parameter data file for a packaged NE32400A HJFET (heterojunction FET) transistor. `# GHZ S MA R 50` indicates frequency in GHz, S parameters in magnitude-angle(degrees) format, and reference to 50 Ω. The S parameter data are f(GHz), magnitude and angle of S_{11}, $|S_{21}|$, S_{12}, $|S_{22}|$. Noise data are f(GHz), $\mathrm{NF}_{\min}$ (dB), $|\Gamma_{\mathrm{opt}}|$, ang(Γ_{opt}) (in degrees), and $r_n/50$.

```
! FILENAME: N32400A.S2P VERSION: 5.0.
! NEC PART NUMBERS: NE32400 DATE:06/91
! BIAS CONDITIONS: VDS=2V, IDS=10mA
! NOTE: S-PARAMETERS INCLUDES BOND WIRES.
! GATE: TOTAL 2 WIRES, 1 PER BOND PAD, EACH WIRE 0.0132"(335um) LONG.
! DRAIN: TOTAL 2 WIRES, 1 PER BOND PAD, EACH WIRE 0.0094"(240um) LONG.
! SOURCE: TOTAL 4 WIRES, 2 PER SIDE, EACH WIRE 0.0070"(178um) LONG.
! WIRE: 0.0007"(17.8um) DIAMETER, GOLD
# GHZ S MA R 50
 0.1   .999    -1   5.04   179   .002   89   .62    -1
 0.2   .999    -3   5.02   178   .004   89   .62    -1
 0.5   .999    -6   4.97   175   .008   87   .62    -4
 1.0   .997   -12   4.88   170   .016   84   .62    -8
 2.0   .990   -23   4.70   161   .030   77   .61   -15
 3.0   .980   -34   4.54   152   .042   71   .61   -22
 4.0   .970   -44   4.38   144   .052   65   .61   -29
 5.0   .950   -53   4.22   136   .062   59   .60   -36
 6.0   .930   -62   4.08   128   .071   53   .59   -41
 7.0   .910   -71   3.93   120   .079   48   .59   -46
 8.0   .890   -79   3.80   113   .086   43   .58   -51
 9.0   .870   -87   3.67   106   .092   38   .57   -56
10.0   .860   -94   3.54    99   .099   34   .56   -61
11.0   .840  -102   3.42    92   .104   30   .55   -65
12.0   .820  -108   3.30    86   .109   27   .54   -70
13.0   .800  -115   3.19    80   .114   24   .53   -74
14.0   .790  -121   3.08    74   .119   21   .51   -78
15.0   .770  -128   2.97    68   .123   18   .50   -83
16.0   .750  -134   2.87    63   .127   16   .49   -87
17.0   .740  -139   2.77    57   .131   14   .48   -91
18.0   .720  -145   2.68    52   .135   12   .47   -95
19.0   .710  -150   2.59    47   .138   10   .46   -98
20.0   .690  -155   2.50    42   .142    8   .45  -102
22.0   .660  -165   2.32    32   .148    6   .43  -109
24.0   .640  -175   2.16    23   .153    4   .42  -116
26.0   .610   177   2.01    15   .159    3   .41  -122
28.0   .590   168   1.87     7   .163    1   .41  -128
30.0   .570   160   1.73    -1   .168    0   .41  -134
! NOISE PARAMETERS
! NOTE: NOISE PARAMETERS FOR 28 & 30 GHZ
! ARE EXTRAPOLATED, NOT MEASURED.
 1   0.30   .81    10   .39
 2   0.31   .79    17   .36
 4   0.33   .75    31   .33
 6   0.38   .72    45   .30
 8   0.43   .70    59   .27
10   0.50   .68    77   .24
12   0.60   .66    92   .22
14   0.71   .64   108   .19
16   0.85   .62   126   .18
18   1.00   .58   140   .15
20   1.20   .55   153   .13
22   1.50   .52   164   .11
24   1.80   .49   175   .10
26   2.10   .48  -176   .08
28   2.40   .46  -168   .07
30   2.80   .46  -160   .05
```



(a) Smith chart (b) Polar plot

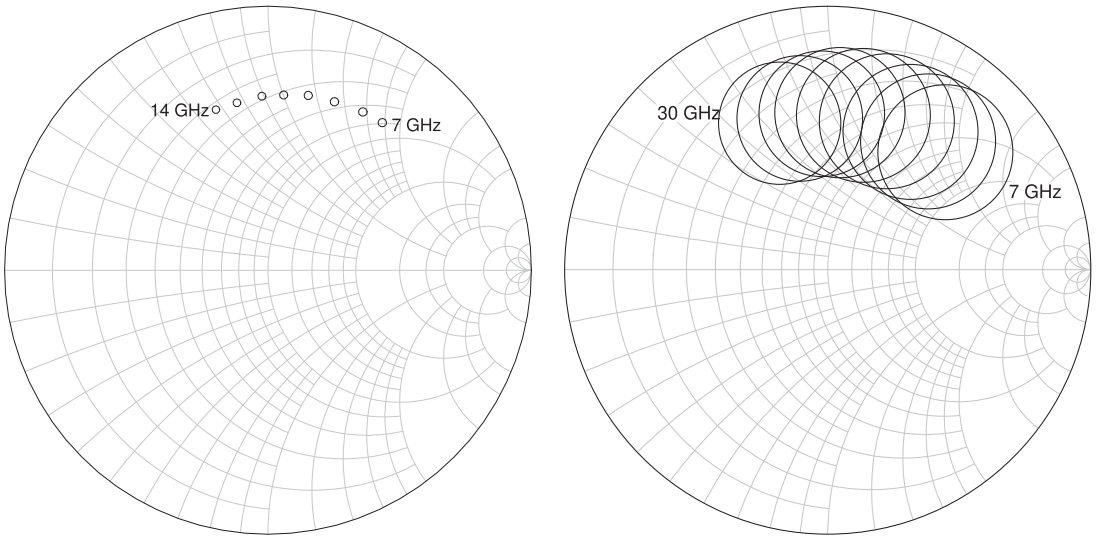
Figure 3-6: S parameters of the N32400A transistor. $|S_{21}|$ exceeds unity and is shown on a polar plot in (b) where 2, 4, and 6 indicate the magnitudes (radii) of constant $|\Gamma|$ circles.

Thus the locus of the optimum Γ_S is shown as segment B in Figure 3-6(a). Note that Γ_S rotates in the counterclockwise direction with increasing frequency. However, the input reflection coefficient of a simple matching network would rotate in the clockwise direction. Thus a reasonable match will only be achieved over a very narrow bandwidth. The output matching network situation is similar. The ideal matching network must have an input reflection coefficient that is counter to that of a simple network. So this illustrates the big difference between wideband and narrowband amplifier design. The matching networks must be designed to present the required complex conjugate impedance over a broad range of frequencies.

Another design task is simultaneously minimizing noise. The noise data of the transistor is plotted in Figure 3-7. Figure 3-7(a) plots the value of Γ_S $(= \Gamma_{\rm opt})$ required to achieve the minimum noise figure. The points are just $\Gamma_{\rm opt}$ at different frequencies. These points do not coincide with the Γ_S for optimum matching as shown in Figure 3-6(a). So a compromise is needed. This compromise step is guided by the noise figure circles. Figure 3-6(b) plots the noise figure circles when the noise figure is 0.25 dB higher than $\rm NF_{min}$. For example, at one frequency, if Γ_S is on the circle for that frequency, the noise figure will be 0.25 dB higher than $\rm NF_{min}$. If Γ_S is inside the circle the noise figure will be less than 0.25 dB above $\rm NF_{min}$.

A more complete set of noise figure circles at 10 GHz, the middle of the amplifier band, is shown in Figure 3-8. $\rm NF_{min}$ is 0.50 dB and the noise figure circles are in 0.1 dB steps.

Another consideration affecting the choice of matching networks is the stability of the amplifier. The input and output stability circles for the transistor are shown in Figure 3-9 starting at 2 GHz and up to 30 GHz. For the transistor to be unconditionally stable, Γ_S, must lie in the unconditionally stable region of the Smith chart. The stable regions are shown in Figure



(a) Minimum noise figure, $\mathrm{NF}_{\mathrm{min}}$ (b) 0.25 dB noise figure circles

Figure 3-7: Noise characteristic from 7 to 14 GHz in 1 GHz steps plot on the input (Γ_S) plane. $\mathrm{NF}_{\mathrm{min}}$ = 0.38 dB, 0.41 dB, 0.43 dB, 0.47 dB, 0.50 dB, 0.55 dB, 0.60 dB, 0.66 dB, and 0.71 dB from 7 to 14 GHz in 1 GHz steps. The noise figure on each circle is $\mathrm{NF}_{\mathrm{min}} + 0.25$ dB.

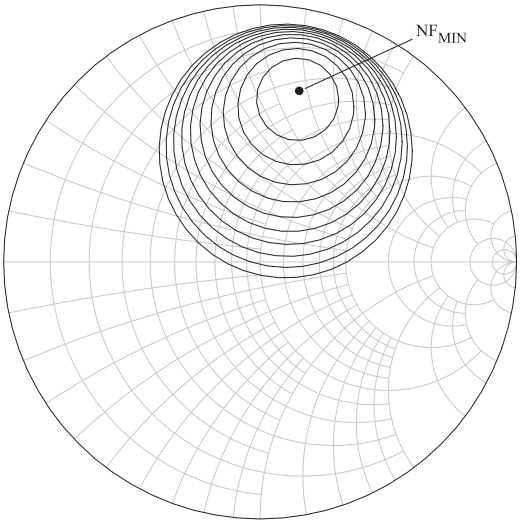
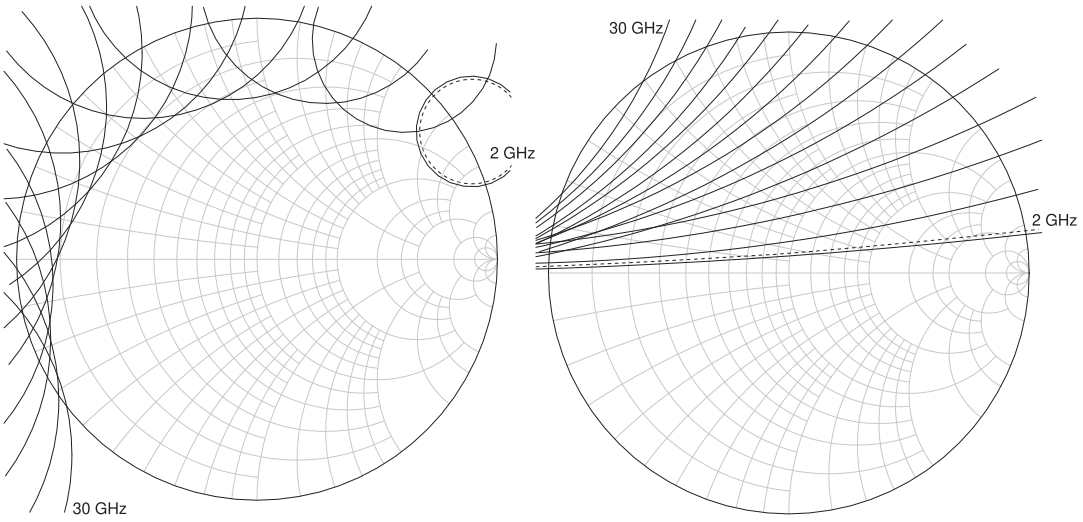


Figure 3-8: Noise figure circles at 10 GHz where $\mathrm{NF}_{\mathrm{min}} = 0.50$ dB. Circles have 0.1 dB steps so that the inner-most circle indicates the values of Γ_S that achieve NF = 0.6 dB.

3-9(a) at all frequencies. Similarly, Γ_L must lie in the stable region of the Smith chart in Figure 3-9(b) at all frequencies. The final consideration is the maximum available gain, G_{MAX}. The G_{MAX} circles are shown in Figure 3-10. At 8 GHz $G_{\mathrm{MAX}} = 16.6$ dB and it reduces to 14.8 dB at 12 GHz. This further complicates design as the gain of the final amplifier should be flat across the band and not monotonically decreasing.

So the complete design problem is to determine the matching network topology and then develop the input and output matching networks that meet all of the constraints implied by the stability circles, the noise figure



(a) Input stability circles (b) Output stability circles

Figure 3-9: Stability circles in 2 GHz steps starting at 2 GHz and continuing up to 30 GHz. The potentially unstable region is indicated by the dashed line on the 2 GHz circle.

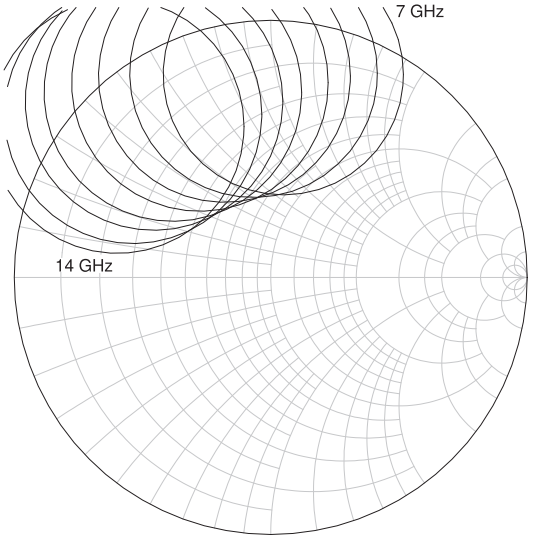


Figure 3-10: Maximum available gain, G_{MAX}, circles. $G_{\mathrm{MAX}} = 17.0$ dB, 16.5 dB, 16.0 dB, 15.5 dB, 15.2 dB, 14.8 dB, 14.5 dB, 14.1 dB at 7 to 14 GHz in 1 GHz steps.

circles, and the G_{MAX} circles.

3.5.2 Negative Image Design

A successful strategy for wideband design is the negative image amplifier design technique. The development begins by considering the fundamental input and output impedances of the transistor. The input of a transistor can be approximated as a resistance in series with a capacitance. The output of the transistor appears as a capacitance in parallel with a resistance. So a simple matching strategy is to consider an input matching network that presents a negative series capacitance (the image) to the

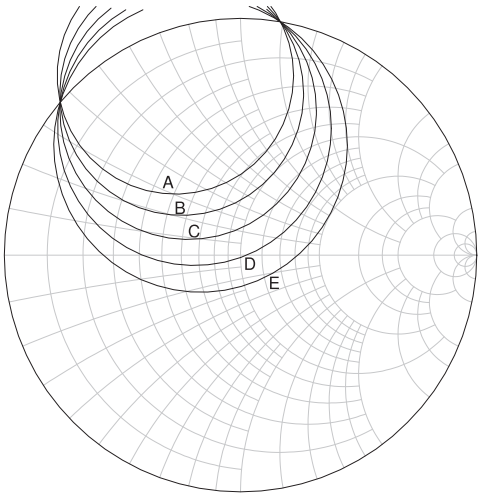
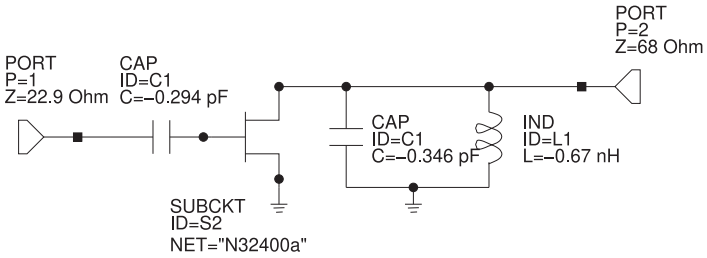
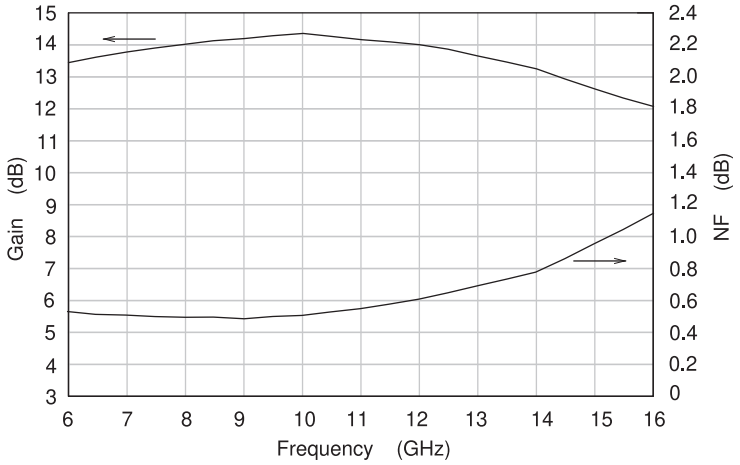


Figure 3-11: Maximum available gain, G_{MAX}, circles at 10 GHz in 1 dB steps. The central circle has G_{MAX} =15.5 dB.



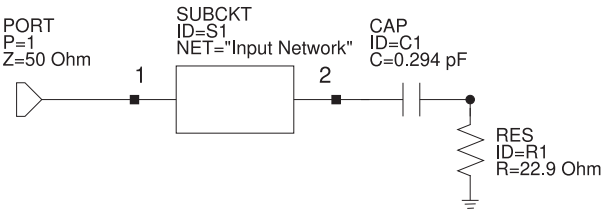
(a) Image amplifier schematic.



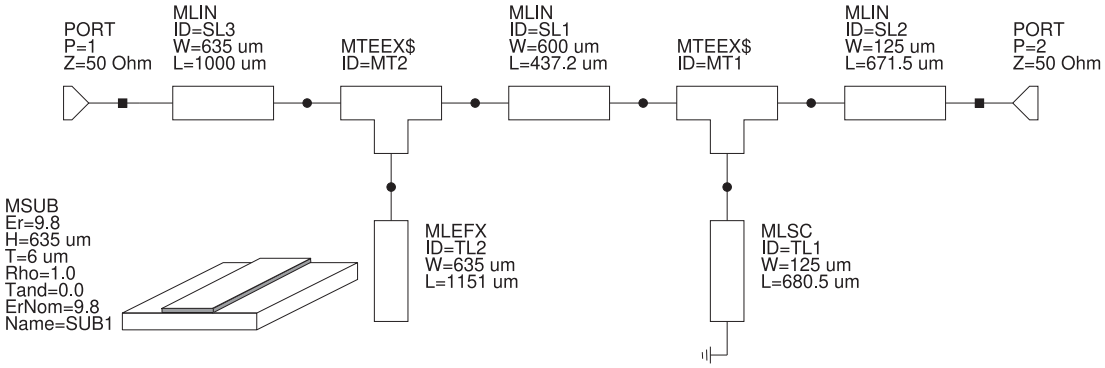
(b) Gain and noise figure of the image amplifier.

Figure 3-12: Amplifier using negative image model.

input of the transistor and a negative shunt capacitance to the output of the transistor. Such an amplifier is shown in Figure 3-12(a). The output matching network also includes a negative shunt inductance that cancels the bondwire inductance of the packaged transistor. The input and output port impedances were adjusted to achieve the required gain. The values of the input and output port impedances as well as of the three reactive elements can be optimized or adjusted using the manual tuning feature in most microwave simulators. Tuning is a useful feature that provides considerable



(a) Input image model



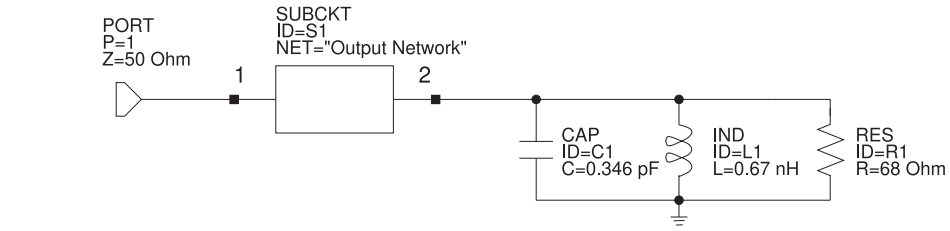
(b) Microstrip schematic of input network

Figure 3-13: Microstrip realization of the input matching network using a microstrip substrate, MSUB. The layout begins with a port element PORT 1 (P=1) with a reference impedance of $50\ \Omega$. The MLIN element is a microstrip line with width W and length L; the MTEEX$ element models a microstrip tee and supports the shunt connection of the MLEFX element, which is an open-circuited microstrip line with end effects modeled; another line and tee connects a shorted stub, the MLSC element; then another line; and finally a second PORT element.

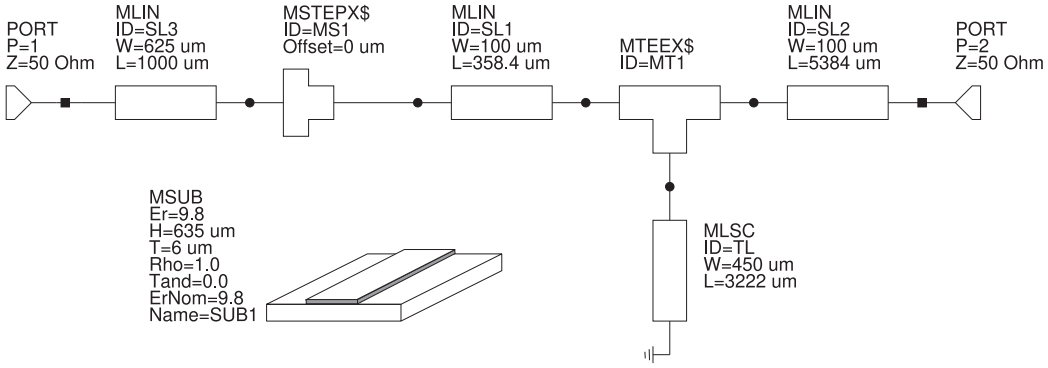
insight. The trade-offs are not always easy to make without using the image amplifier technique.

The noise and gain of the image amplifier of Figure 3-12(a) are shown in Figure 3-12(b). While the gain and noise figure do not meet the specification (13 dB gain and NF $<$ 1 dB), they are very close and it can be expected that optimization will achieve the required design.

At this stage the topologies of the input and output matching networks need to be selected since negative inductances and capacitors are not available. There are several ways this can be done. One way is to develop transfer function descriptions of the impedances of the input (and output) network seen from the transistor. The impedance functions can then be synthesized and developed as if they were filters. This can be a long process, but sometimes it is the only way to meet demanding specifications. Very often the topology can be adopted from an earlier design or from a design reported in a publication. The topology chosen here for the input network is shown in Figure 3-13(b). (The theory behind this topology is described in Section 7.7 of [6].) The parameters of the input matching network are optimized using the input image model shown in Figure 3-13(a). However, the sign of the negative capacitance in the input circuit of Figure 3-12(a) is changed. The series connection of the 0.294 pF capacitor in series with



(a) Output image model



(b) Output network

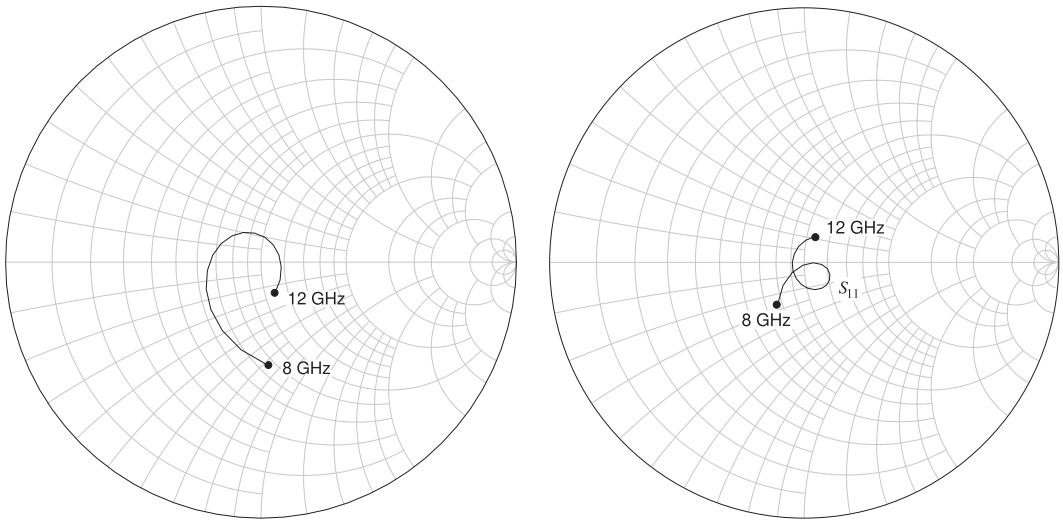
Figure 3-14: Realization of the output network.

the 22.9 Ω resistor approximates the input impedance of the (terminated) transistor. More specifically, it is the impedance that will result in the gain and noise profile in Figure 3-12(b). The goal in realizing the input matching network is to minimize the input reflection coefficient (normalized to 50 Ω) at the input port. A similar approach is used in developing the output matching network, shown in Figure 3-14. Following optimization, the input reflection coefficients at Port 1 of the matching networks terminated in the image networks are shown in Figure 3-15.

3.5.3 Final Design

The input and output matching networks are connected to the transistor as shown in Figure 3-16 and then the complete amplifier is optimized. The parameters of the optimized input and output matching networks using the complete transistor model are given in Figures 3-13(b) and 3-14(b), and the complete layout is shown in Figure 3-17. The gain and noise performance of the completed design is given in Figure 3-18. This wideband amplifier topology achieves a bandwidth up to one-half octave (e.g. 8–12 GHz).

In some cases, although not required here, it is necessary to introduce feedback from the output to the input of the transistor. This is often a simple circuit such as a cascade of a capacitor, an inductor, and a transmission line. The feedback network provides frequency-dependent feedback that flattens the gain response. A resistor can be also included in the feedback network to manage stability.



(a) Γ_{in} of the input network, in Figure 3-13 (b) Γ_{in} of the output network, in Figure 3-14

Figure 3-15: Reflection coefficient looking into Port 1 of the input and output matching networks terminated in the corrected image network.

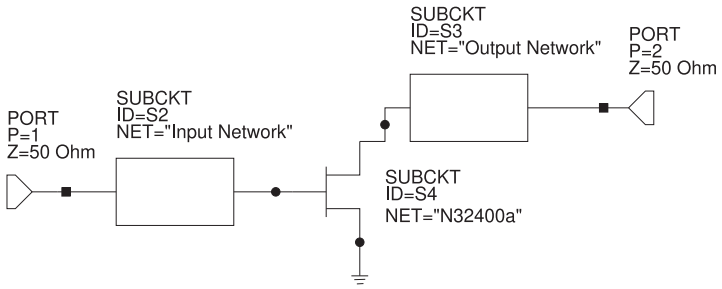


Figure 3-16: Final amplifier.

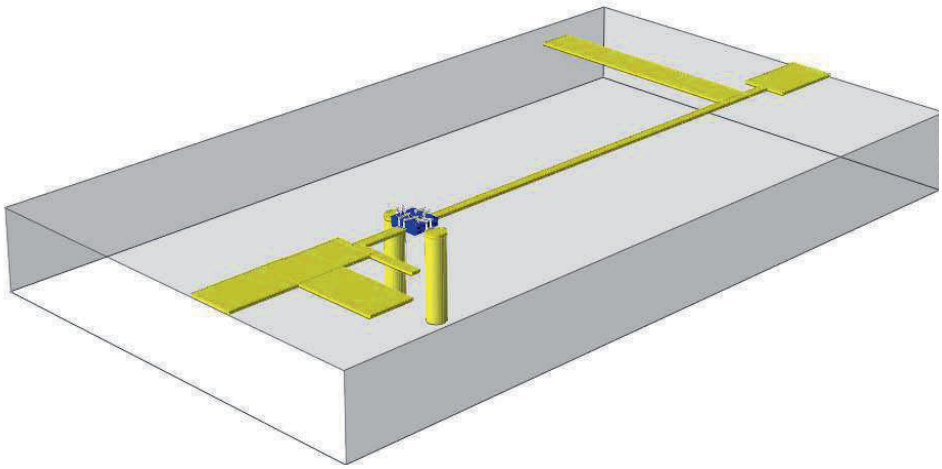


Figure 3-17: Final layout of the X-band LNA. The input is on the left and the output on the right.

3.6 Differential Amplifiers

A significant change in RF and microwave engineering has been the increasing importance of differential circuits such as the amplifier in Figure 3-19(a). In part this is because they are conveniently implemented in silicon technology. It is also a result of the use of monolithic integration and the maturity of semiconductor technologies resulting in repeatable RF active components. Differential amplifiers are the preferred amplifier topology with RFICs. Since substrate noise is common to all nodes of a differential amplifier, there is little differential substrate noise signal. Also, differential circuits lend themselves to current-mode biasing which is preferred on-chip. The defining characteristic of a differential amplifier is that there are paired signal paths that are differential. These amplifiers are also (but less commonly) called **balanced amplifiers**. When the RF signal on one side of the differential path is positive, the RF signal on the other side is negative.

3.6.1 Fully and Pseudo-Differential Amplifiers

Figure 3-19(a) shows a **fully differential amplifier** (**FDA**) with resistive biasing in the drain legs. As well as providing biasing current, the resistors load the amplifier. The supply voltage of an RFIC can be quite low (a few volts or less), so choosing circuit topologies that provide for large voltage swings is important, particularly for a driver amplifier being the last

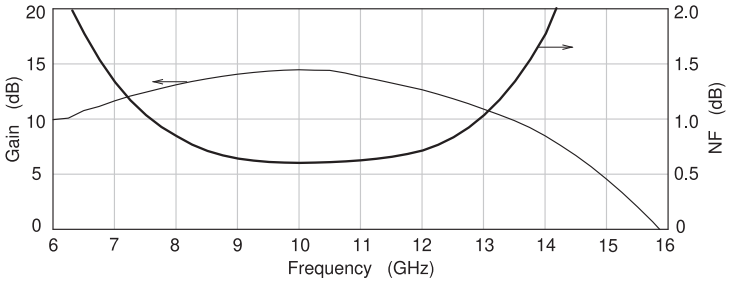


Figure 3-18: Final amplifier performance.

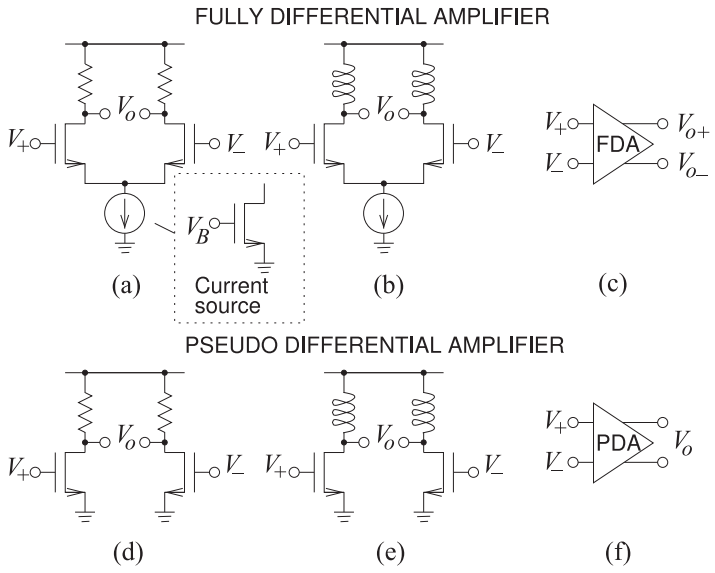


Figure 3-19: Differential amplifiers: (a) fully differential amplifier (FDA); (b) FDA with inductive biasing; (c) schematic representation; (d) pseudo-differential amplifier (PDA); (e) PDA with inductive biasing; and (f) schematic representation. The inset in (a) shows the implementation of current source as a single enhancement-mode MOSFET with a bias voltage, V_B, at the gate.

amplifier stage in a transmitter RFIC that drives a following power amplifier. Differential topologies lead to an almost doubling of the output voltage swing compared to the output voltage swing of a single-ended amplifier. An FDA includes a common current source that can be implemented quite simply using a single FET, as shown in the inset in Figure 3-19(a). Here the gate-source voltage is the bias voltage, V_B, which, from Figure 2-9(b), establishes a nearly constant drain current as long as there is sufficient drain-source voltage across the current source transistor.

Common-Mode Rejection

One of the attributes that makes FDAs attractive is that they are relatively immune to substrate noise (noise in the substrate produced by other circuits). The signal applied to the inputs of a differential amplifier have differential- and common-mode components. Referring to the differential amplifier in Figure 3-19(c), the differential-mode input signal is

$$V_{id} = V_+ - V_- \tag{3.1}$$

and the common-mode input signal is

$$V_{ic} = \tfrac{1}{2}(V_+ + V_-). \tag{3.2}$$

Similarly the differential- and common-mode output signals are

$$V_{od} = V_{o+} - V_{o-} \quad \text{and} \quad V_{oc} = \tfrac{1}{2}(V_{o+} + V_{o-}), \tag{3.3}$$

respectively. The differential-mode voltage gain is

$$A_d = \frac{V_{od}}{V_{id}} \tag{3.4}$$

and the common-mode gain is

$$A_c = \frac{V_{oc}}{V_{ic}}. \tag{3.5}$$

For good noise immunity, the common-mode gain should be low and the differential-mode gain should be high. The figure of merit that describes this is the **common-mode rejection ratio** (**CMRR**):

$$\text{CMRR} = \frac{A_d}{A_c}, \tag{3.6}$$

and the larger this is, the better. CMRR is usually expressed in decibels, and since CMRR is a voltage gain ratio, CMRR in decibels is $20\log(A_d/A_c)$. The current source at the source of the differential pair of the FDA has a large effect on the CMRR by suppressing the output common-mode voltage. Then the current source results in a large CMRR. Without the current source, the CMRR of the FDA of Figure 3-19(a) would be one.

Output Voltage Swing

Single-ended amplifiers were discussed in Section 2.5.1, where it was shown that inductive biasing enables higher output voltage swings than possible

with resistive biasing. A similar enhancement can be obtained with a differential amplifier. The inductively biased FDA of Figure 3-19(b) has a higher voltage swing than the resistively biased FDA of Figure 3-19(a). More can be achieved, however. The current sources at the common source point of the FDAs in Figure 3-19(a and b) limit the voltage swing, as there is a minimum drain-source voltage drop required across the current source for it to maintain constant current. When larger output voltage swings are required, the current source is eliminated and the resulting amplifier is called a **pseudo-differential amplifier** (PDA), as shown in Figure 3-19(d). Again, inductive biasing (see Figure 3-19(e)) almost doubles the possible voltage swing. The performance cost with the PDA circuit is that the CMRR is one.

The output voltage waveforms for single-ended and differential amplifiers with and without inductive biasing are shown in Figure 3-20. The inductively biased PDA, shown in Figure 3-20(d), has an output voltage swing that is about 4 times the voltage swing (or about 16 times the power into the same load) of the single-ended resistively biased Class A amplifier in Figure 3-20(a) (the actual factors depend on the supply voltage and $V_{DS,\min}$).

EXAMPLE 3.1 Calculation of Common-Mode Rejection Ratio

Determine the CMRR of the FET differential amplifier shown in Figure 3-21(a).

Solution:

The strategy for solving this problem is to develop the common-mode and differential-mode equivalent circuits and solve for the gain of each. The first step is to develop the small-signal model shown in Figure 3-21(b). The differential input signal is V_{id} and the common input signal is V_{ic} so that the input voltage signals are

$$V_{i+} = \tfrac{1}{2}\left(V_{id} + V_{ic}\right) \quad \text{and} \quad V_{i-} = -\tfrac{1}{2}\left(V_{id} + V_{ic}\right). \tag{3.7}$$

The expressions are similar for the output differential- and common-mode signals V_{od} and V_{oc}. This leads to the small signal differential-mode model of Figure 3-21(c) and the small signal common-mode model of Figure 3-21(d). From Figure 3-21(c), the output differential signal is

$$V_{od} = \frac{V_{id}}{2}\left[-g_m(r_d//R_L)\right] - \frac{-V_{id}}{2}\left[-g_m(r_d//R_L)\right] = \frac{-V_{id}g_m r_d R_L}{r_d + R_L}, \tag{3.8}$$

so the differential gain is

$$A_d = \frac{V_{od}}{V_{id}} = \frac{-g_m r_d R_L}{r_d + R_L}. \tag{3.9}$$

If, as usual, $r_d \gg R_L$, this becomes

$$A_d = -g_m R_L. \tag{3.10}$$

Focusing on the small signal common-mode model of Figure 3-21(d) yields the output common-mode signal. The two halves of the circuit are now identical. The sum of the currents at X is zero, so

$$\frac{V_S}{2R_S} + \frac{V_S - V_{oc}}{r_d} - g_m(V_{ic} - V_S) = 0, \tag{3.11}$$

and the sum of the currents at the output terminal is zero, so

$$\frac{V_{oc}}{R_L} + \frac{V_{oc} - V_S}{r_d} + g_m(V_{ic} - V_S) = 0. \tag{3.12}$$

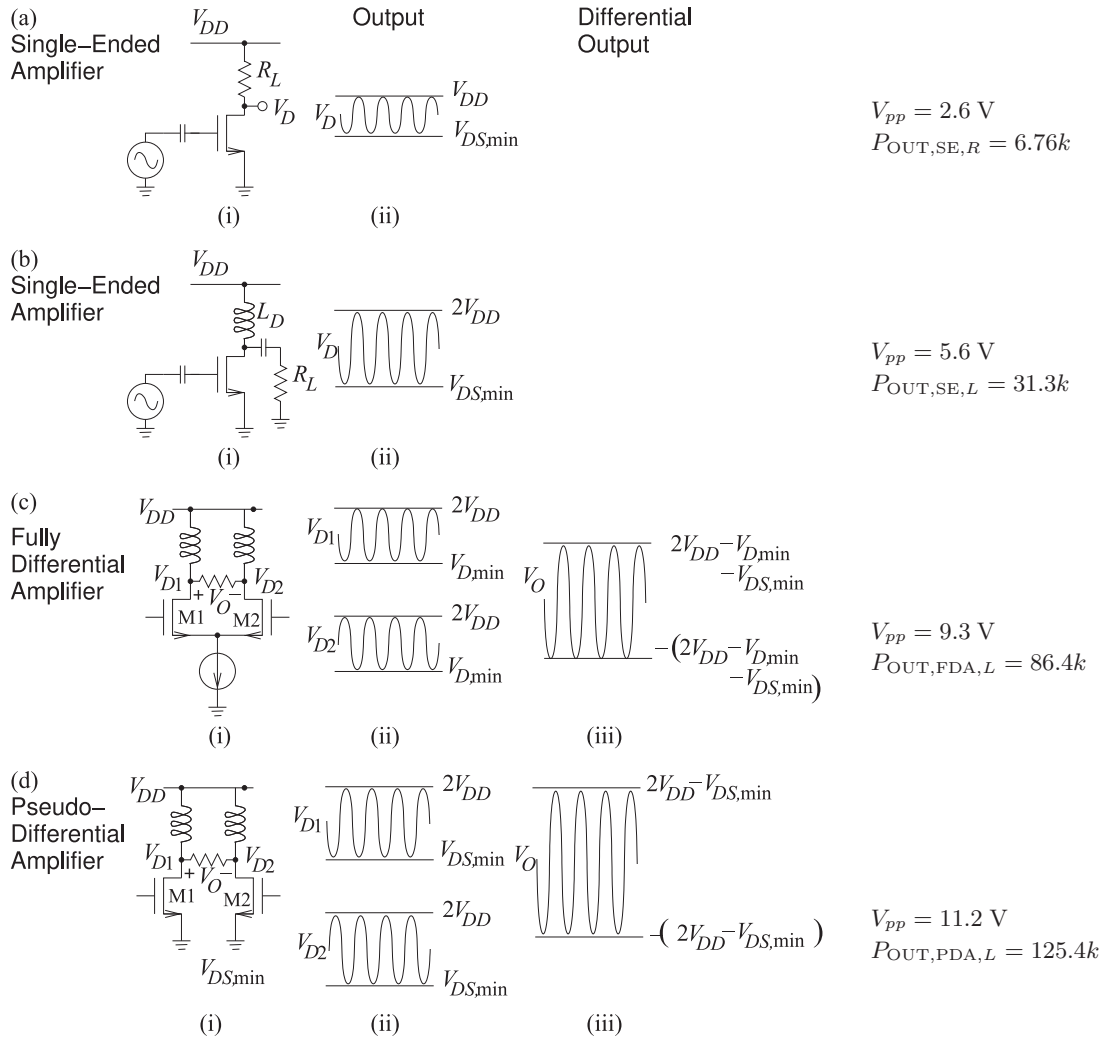


Figure 3-20: Class A MOSFET amplifiers with output voltage waveforms: (a) single-ended amplifier with resistive biasing; (b) single-ended amplifier with inductive biasing; (c) fully differential amplifier with inductive biasing; and (d) pseudo-differential amplifier. Schematic is shown in (i), drain voltage waveforms in (ii), and differential output in (iii). The final column gives the output voltage swing, V_{pp}, and the output power with $V_{DD} = 3$ V, $V_{D,\min} = 0.95$ V, $V_{DS,\min} = 0.4$ V, and k is a proportionality constant dependent on loading that is assumed to be the same for all amplifiers. $V_{D,\min}$ is the minimum voltage at the drain of the current-source MOSFET. The resistively biased single-ended amplifier has an output power proportional to 6.76 while the inductively biased PDA in (d) has an output power proportional to 125.4, 18.6 times larger.

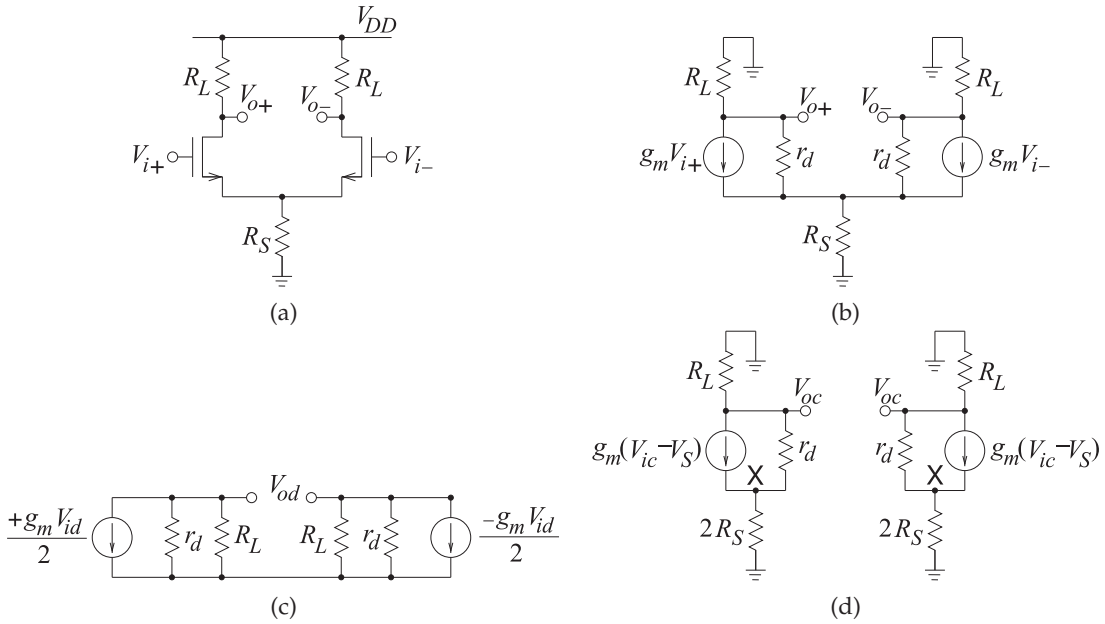


Figure 3-21: Differential amplifier: (a) schematic; (b) small signal model; (c) small signal model for calculating differential gain; and (d) small signal model for calculating common-mode gain.

Eliminating V_S from these equations leads to

$$V_{oc} = \frac{-g_m r_d R_L V_{ic}}{(1+g_m r_d)2R_S + r_d + R_L}, \tag{3.13}$$

so the common-mode gain is

$$A_c = \frac{V_{oc}}{V_{ic}} = \frac{-g_m r_d R_L}{(1+g_m r_d)2R_S + r_d + R_L}. \tag{3.14}$$

If, as usual, $r_d \gg R_L$, this becomes

$$A_c = \frac{-g_m R_L}{1+2g_m R_S}. \tag{3.15}$$

The CMRR (when $r_d \gg R_L$) is

$$\text{CMRR} = \frac{A_d}{A_c} = \frac{-g_m R_L(1+2g_m R_S)}{-g_m R_L} = (1+2g_m R_S). \tag{3.16}$$

So the CMRR depends on the value of R_S. If there is no resistor at the common-mode source point, as in a pseudo-differential amplifier, $R_S = 0$ and so

$$\text{CMRR}|_{R_S=0} = 1. \tag{3.17}$$

If there is an ideal current source at the source node, then R_S is effectively infinite, and so

$$\text{CMRR}|_{\text{current source}} = \infty. \tag{3.18}$$

3.6.2 Even, Common, Odd, and Differential Modes

The difference between even- and common-mode current, voltages and impedances comes down to bookkeeping as to how the voltages and currents are defined. The same is true for the odd- and differential-mode quantities. The reason both sets of definitions are used is because the even-/odd-mode set usage is preferred with transmission line structures and the common-/differential-mode set usage is preferred with complementary transistor circuits such as differential amplifiers.

Consider the amplifier shown in Figure 3-22(a) with two inputs and two outputs. The input and output voltages in the various modes are defined as follows:

Odd-mode input voltage, V_{io}, and current, I_{io} (with the second subscript indicating the mode):

$$V_{io} = \tfrac{1}{2}\left(V_{i1} - V_{i2}\right), \quad \text{and} \quad I_{io} = \tfrac{1}{2}\left(I_{i1} - I_{i2}\right). \tag{3.19}$$

Differential-mode input voltage, V_{id}, and current, I_{id}:

$$V_{id} = \left(V_{i1} - V_{i2}\right), \quad \text{and} \quad I_{id} = \tfrac{1}{2}\left(I_{i1} - I_{i2}\right). \tag{3.20}$$

Even-mode input voltage, V_{ie}, and current, I_{ie}:

$$V_{ie} = \tfrac{1}{2}\left(V_{i1} + V_{i2}\right), \quad \text{and} \quad I_{ie} = \tfrac{1}{2}\left(I_{i1} + I_{i2}\right). \tag{3.21}$$

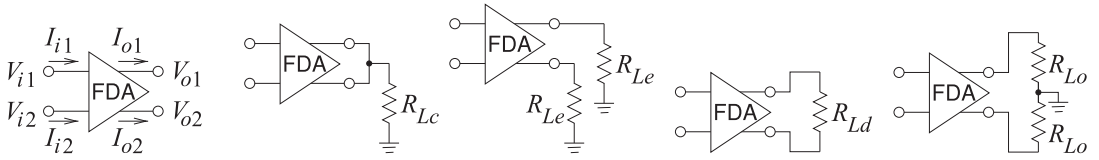
Common-mode input voltage, V_{ic}, and current, I_{ic}:

$$V_{ic} = \tfrac{1}{2}\left(V_{i1} + V_{i2}\right), \quad \text{and} \quad I_{ic} = \left(I_{i1} + I_{i2}\right). \tag{3.22}$$

Reversing the definitions, if there is no common-/even-mode input signal:

$$\begin{array}{lll} V_{i1} = \tfrac{1}{2}V_{id} = V_{io}, & & V_{i2} = -\tfrac{1}{2}V_{id} = -V_{io}, \\ I_{i1} = I_{id} = I_{io}, & \text{and} & I_{i2} = -I_{id} = -I_{io}. \end{array} \tag{3.23}$$

The output voltages and currents are similarly related. Figures 3-22(b)–3-22(e) show the conceptual even-, odd-, common- and differential- mode load definitions. In switching between definitions, say between the differential load and the odd-mode load, the actual resistor in the circuit does not change.

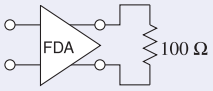


(a) Differential amplifier (b) Common-mode load (c) Even-mode load (d) Differential-mode load (e) Odd-mode load

Figure 3-22: Differential amplifiers and various loads. R_{Lc} is the common-mode load, R_{Le} is the even-mode load, R_{Ld} is the differential-mode load (often the term differential load is used), and R_{Lo} is the odd-mode load.

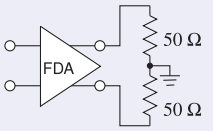
EXAMPLE 3.2 Odd-Mode Load

The differential amplifier to the right has a differential load of 100 Ω. What is the odd-mode load?



Solution:

The circuit is put into the odd-mode form to the right. Comparing this to the load definitions shown in Figure 3-21(d), it is seen that the odd-mode load impedance is 50 Ω.



3.6.3 Asymmetrical Loading

It is simple enough to determine the common and differential loads, or similarly the even- and odd-mode loads, when the loading of a differential amplifier is symmetrical. However when loading is asymmetrical, details of the driving differential amplifier are required to determine the coupling between the common and differential signals. The situation is similar to that with a terminated coupled line, see Section 5.7 of [7], where the Thevenin equivalent impedance of the source is required to determine the circuit conditions.

With an asymmetrical load there will be coupling between the even and odd modes. So even if the driving differential amplifier produces a differential output current and has zero common mode current, there could still be a common mode voltage. This is important as transistors operate as voltage-controlled current sources and many differential amplifiers are actually transconductance amplifiers as this gives the widest bandwidths, simplest biasing, and good noise immunity. The output stage of a differential amplifier appears as differential voltage-controlled current sources and in an RFIC adaptive mechanisms usually ensure that there is no common-mode current. But the differential current can induce a common-mode voltage which drives following stages. The design strategy then is to ensure that a differential amplifier produces minimal common-mode current, and loading is symmetrical. The following examples are illustrative.

EXAMPLE 3.3 Asymmetrical Loading of a Differential Amplifier

A differential amplifier has two output terminals with one of the outputs connected to a 60 Ω resistor and the other terminated in a 100 Ω resistor, as shown in Figure 3-23(a). The output stage of the differential amplifier is modeled as two controlled current sources I_{o1} and I_{o2} and the output common-mode current is zero.

(a) What is the common-mode voltage, V_c, at the load if the differential current is 1 mA? This problem will first be solved using the general loads shown in Figure 3-23(b). The voltages at the load are $V_{o1} = V_c + \frac{1}{2}V_d$ and $V_{o2} = V_c - \frac{1}{2}V_d$ where V_d is the differential output voltage. The output currents are $I_{o1} = \frac{1}{2}I_c + I_d = I_d$ and $I_{o2} = \frac{1}{2}I_c - I_d = -I_d$. Then

$$V_{o1} = V_c + \tfrac{1}{2}V_d = I_{o1}R_1 = +I_dR_1 \tag{3.24}$$

$$V_{o2} = V_c - \tfrac{1}{2}V_d = I_{o2}R_2 = -I_dR_2 \tag{3.25}$$

combining these

$$2V_c = I_d(R_1 - R_2), \quad \text{and so} \quad V_c = \tfrac{1}{2}(1\text{ mA}(60\ \Omega - 100\ \Omega) = 40\text{ mV}. \tag{3.26}$$

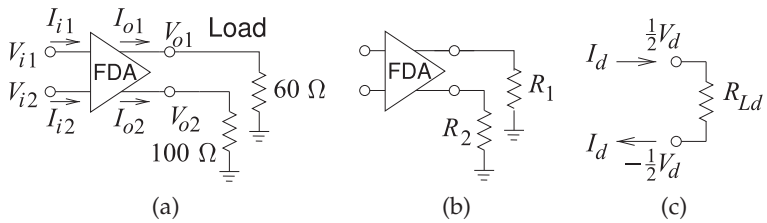


Figure 3-23: Terminated differential amplifier: (a) asymmetrical loading; (b) general representation of loading; and (c) definition of voltages and currents for the differential-mode load impedance R_{Ld}.

The common mode voltage at the output is 40 mV.

(b) What is the differential-mode load resistance, R_{Ld}?
The differential-mode load R_{Ld} is defined in Figure 3-23(d) so that $R_{Ld} = V_d/I_d$. Taking the difference of Equations (3.24) and (3.25) and eliminating V_c leads to

$$V_d = I_d(R_1 + R_2) + \tfrac{1}{2}I_c(R_1 - R_2) \quad \text{and} \quad R_{Ld} = \frac{V_d}{I_d} = (R_1 + R_2) = 160\ \Omega. \tag{3.27}$$

EXAMPLE 3.4 Differential- and Odd-Mode Loads

A differential amplifier is shown in Figure 3-24(a) with resistive loading. Find the differential- and odd-mode load resistances if the common-mode current is zero.

Solution:

Nodal analysis yields the circuit equations

$$I_1 = \frac{V_1}{10} + \frac{V_1 - V_2}{5} \quad \text{and} \quad I_2 = \frac{V_2}{20} + \frac{V_2 - V_1}{5}. \tag{3.28}$$

That is $$V_1 = (50I_1 + 40I_2)/7 \quad \text{and} \quad V_2 = (40I_1 + 60I_2)/7. \tag{3.29}$$

(a) What is the differential-mode load resistance R_{Ld}?
The differential-mode current $I_d = \frac{1}{2}(I_1 - I_2)$, so, since the common-mode current is zero, set $I_1 = I_d$ and $I_2 = -I_d$ and Equation (3.29) becomes

$$V_1 = \frac{10}{7}I_d, \quad \text{and} \quad V_2 = -\frac{20}{7}I_d. \tag{3.30}$$

The differential-mode voltage is $$V_d = (V_1 - V_2) = \frac{30}{7}I_d. \tag{3.31}$$

Thus $$R_{Ld} = \frac{V_d}{I_d} = \frac{30}{7} = 4.286\ \Omega. \tag{3.32}$$

(b) What is the odd-mode load resistance R_{Lo}?
The odd-mode current $I_o = \frac{1}{2}(I_1 - I_2)$, so set $I_1 = I_o$ and $I_2 = -I_o$ since the common-mode, and hence even-mode, current is zero and Equation (3.29) becomes

$$V_1 = \frac{10}{7}I_o \quad \text{and} \quad V_2 = \frac{-20}{7}I_o. \tag{3.33}$$

The odd-mode voltage is $$V_o = \tfrac{1}{2}(V_1 - V_2) = \frac{30}{14}I_o. \tag{3.34}$$

Thus $$R_{Lo} = \frac{V_o}{I_o} = \frac{30}{14} = 2.143\ \Omega. \tag{3.35}$$

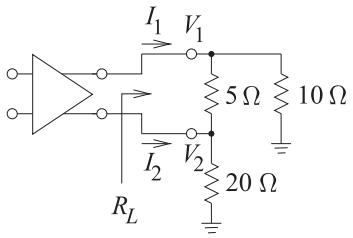


Figure 3-24: A differential amplifier with asymmetrical terminating resistors.

3.6.4 Hybrids and Differential Amplifiers

RFICs use both fully differential and pseudo-differential signal paths. If signal swing is not a concern, a fully-differential signal path is preferred, particularly because of its immunity to noise and its bias stability. The additional transistors involved in realizing a differential circuit (e.g., the current source) reduce the available voltage swing, and hence the power-handling capability. Pseudo-differential signaling uses, in effect, two parallel paths, each referred to ground, but of opposite polarity. The signal on one of the parallel paths is the mirror image of the signal on the other (i.e., the signal is not truly differential, which would imply that it was floating or independent of ground). Each of the parallel paths is unbalanced, but together their RF signal appears to be balanced, or pseudo-balanced. Another consideration is that in working with RFICs it is necessary to interface (unbalanced) microstrip circuits with the inputs and outputs of RFICs. The functionality here requires that signals be split and combined, and converted between balanced and unbalanced signals.

An FDA is shown in Figure 3-25(a). Both the input voltage, $V_i = V_+ - V_-$, and the output voltage, V_o, are differential. Figure 3-25(b and c) show a transformer being used to convert the differential output of the amplifier to an unbalanced signal that, for example, can be connected to a microstrip circuit. The output of many RFICs is pseudo-differential, as this signaling provides large voltage swings. A pseudo-differential amplifier is shown in Figure 3-25(d), but before dealing with manipulation of the signal path, first consider the hybrid on its own.

Figure 3-26(a) shows how two pseudo-balanced signals can be combined to yield a single balanced signal. This $180°$ hybrid function is realized by a **center-tapped transformer**. The signal at Terminal 2 is referenced to ground, and these two terminals are Port 2. The image component of the pseudo-differential signal is applied to Port 3, comprising Terminal 3 and ground. The balanced signal at Port 1 can be directly connected to a microstrip line that is, of course, unbalanced. Most implementations of hybrids at microwave frequencies have ports that are referenced to ground. This is emphasized in Figure 3-26(b), making it easier to see how a $180°$ hybrid can be used to combine a pseudo-differential signal, as shown in Figure 3-26(c). This pseudo-differential-to-unbalanced interface is shown in Figure 3-25(e–g) with a pseudo-differential amplifier.

Hybrids can be used at the input and output terminals of an RFIC pseudo-differential amplifier so that an unbalanced source can efficiently drive the RFIC and then the output of the RFIC can be converted to an unbalanced port to interface with unbalanced circuitry, such as filters and transmission lines. In the RFIC-based system in Figure 3-27, a $180°$ hybrid is first used as a **splitter** and then at the output as a **combiner**.

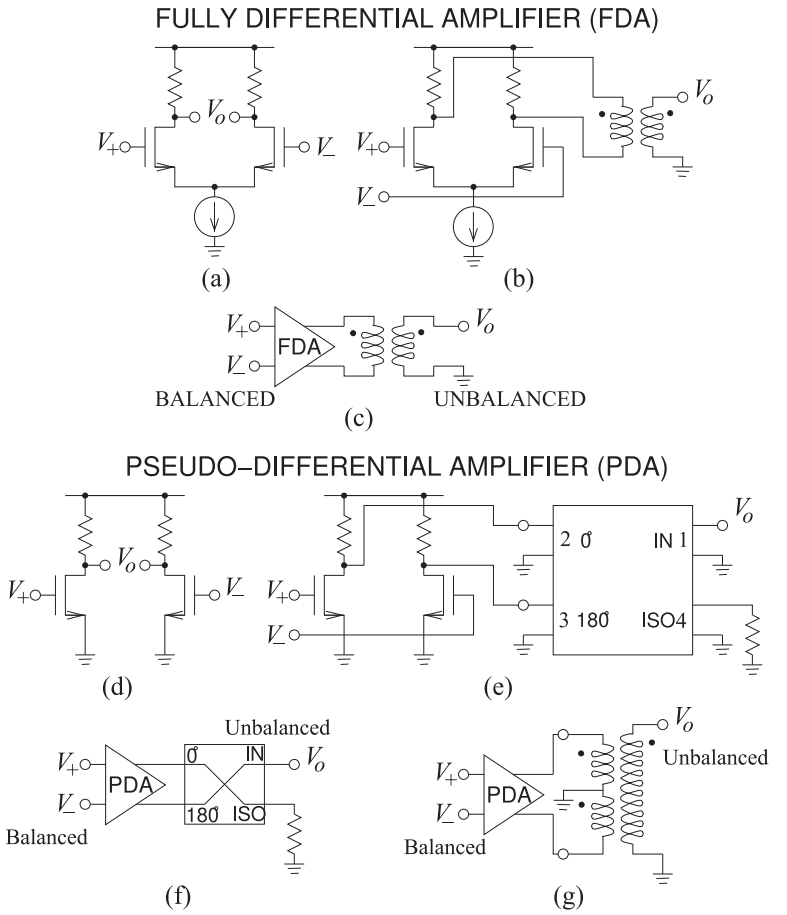


Figure 3-25: Configurations providing an unbalanced output from a differential amplifier: (a) FDA; (b) FDA configured with a balun; (c) schematic; (d) PDA; (e) PDA configured with a 180° hybrid to provide an unbalanced output; (f) schematic; and (g) PDA with a transformer connection yielding an unbalanced output.

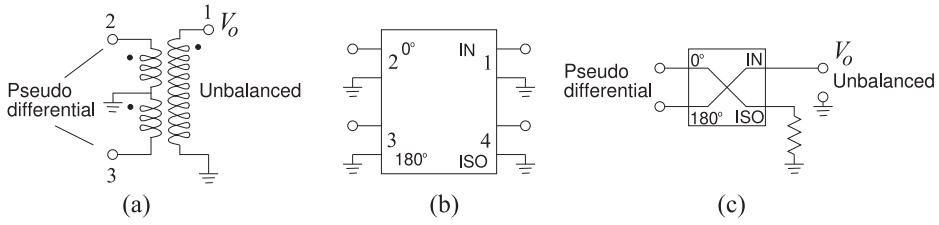


Figure 3-26: Equivalent representations of a 180° hybrid connected to provide an interface between a pseudo-differential balanced port and an unbalanced port: (a) a transformer configured as a 180° hybrid with pseudo-unbalanced-to-balanced configuration; (b) hybrid showing two terminal representation of ports; and (c) schematic of a 180° hybrid with the isolation port terminated in a matched load.

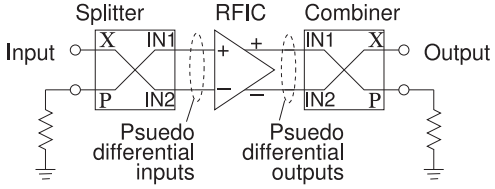


Figure 3-27: An RFIC with differential inputs and outputs driven by a 180° hybrid used as a splitter and followed by another 180° hybrid used as a combiner.

3.7 Case Study: Distributed Biasing of a Pseudo-Differential Amplifier

In this case study a broadband distributed balun-like section is presented as an alternative to inductor-biasing of a pseudo-differential amplifier (PDA). The distributed biasing circuit discriminates between differential- and common-mode signals, resulting in rejection of common-mode signals. A PDA is shown in Figure 3-28, where the inductors present high RF impedances to the transistors while providing low-impedance paths for bias currents. However, inductive biasing of pseudo-differential circuits presents the same environment to common- and differential-mode signals so that the CMRR is 1.

Differential amplifiers have large differential gain, A_d. At the same time it is desirable to minimize the common-mode gain, A_c, as the resulting high CMRR provides immunity to substrate-induced noise. Considering that each transistor has transconductance, g_m, and that even- and odd-mode impedances, $Z_{\rm EVEN}$ and $Z_{\rm ODD}$, are presented to the drains of the transistors, then the gains are approximately

$$A_d = g_m Z_{\rm ODD} \text{ and } A_c = g_m Z_{\rm EVEN}, \tag{3.36}$$

and so

$$\text{CMRR} = A_d/A_c = Z_{\rm ODD}/Z_{\rm EVEN}. \tag{3.37}$$

The desired amplifier characteristics are thus obtained by synthesizing the even- and odd-mode impedances.

First consider the inductively biased circuit in Figure 3-28. Modal analysis of the inductor biasing circuit results in the circuit model shown in Figure 3-29, from which the total even-mode impedance is

$$Z_{\rm EVEN}(s) = (sL + R_{DD}/2)\,//R_L, \tag{3.38}$$

where // indicates a parallel connection, and the total odd-mode impedance is

$$Z_{\rm ODD}(s) = sL//R_L. \tag{3.39}$$

Since R_{DD} is usually negligible and L is a bias or choke inductor so that sL is very large, $Z_{\rm ODD} \approx R_D \approx Z_{\rm EVEN}$ and so the CMRR is 1. However, a coupled-line network can provide different model impedances.

Now consider the Marchand balun-like structure in Figure 3-30 that replaces the drain bias circuit in Figure 3-28. The Marchand Balun structure

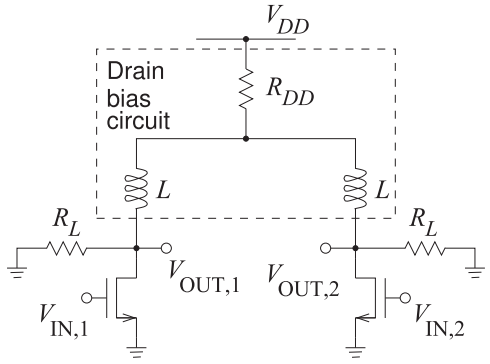


Figure 3-28: A PDA amplifier with bias inductors, L, at the drains, parasitic supply resistance, R_{DD}, and single-ended load impedance, R_L. L is also known as a choke inductor chosen so that its impedance is very large at the maximum operating frequency, i.e. $|sL| \gg R_L$.

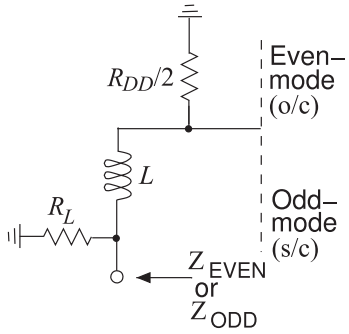


Figure 3-29: Modal subcircuits of the inductor-based biasing circuit of Figure 3-28, including single-ended load resistance R_L.

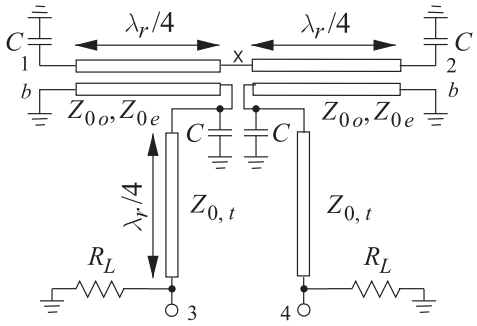


Figure 3-30: Marchand balun-like biasing circuit with single-ended load resistance R_L. External DC bias is applied at Ports b using decoupling capacitors to ensure RF ground.

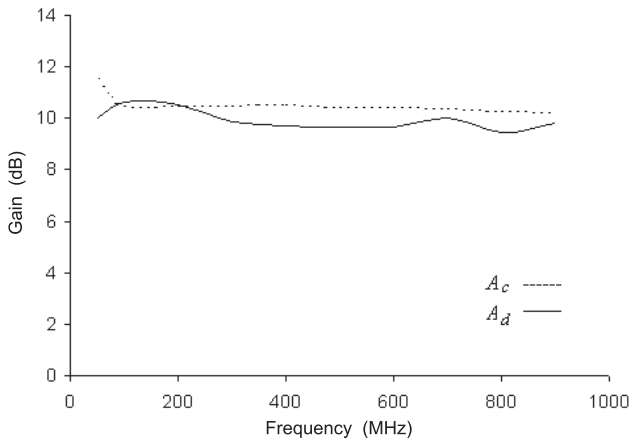
presents different impedances for common- and differential-mode signals. The synthesis of this biasing circuit is described in [8, 9] and follows a procedure similar to that for filter design. So high CMRR performance is the result of presenting different even- and odd-mode impedances to the active devices. The final results of the design are shown in Figure 3-31, first for inductive-biasing of the PDA and then for the coupled-line balun-like biasing circuit.

3.8 Amplifiers and RFICs

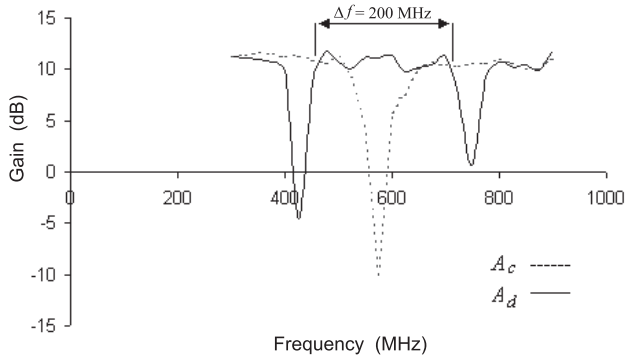
Silicon RFICs exploit the high-density integration possible with silicon MOSFET transistors. These transistors can be fabricated with high levels of repeatability so that the transistors in differential designs can be closely matched. As well, the inherent compatibility with digital circuits enables digital control of RF circuits. As far as amplifiers are concerned, there are a few commonly used basic circuits that use complementary MOSFETs (nMOS and pMOS), that is, CMOS transistors. MOSFET differential amplifiers have been presented throughout this chapter. Other common CMOS circuits are shown in Figure 3-32. The transistors in all the circuits described here operate in the saturation region.

A **cascode amplifier** is shown in Figure 3-32(a). There are two FETs, with the top FET acting as the drain load of the bottom FET. The gate of the top FET is held at ground so the voltage at the source of the top FET (and the drain of the bottom FET) is held at a nearly constant voltage. Thus the top FET presents a low-resistance load to the bottom FET. The voltage gain of the bottom FET is low, and this reduces the **Miller effect capacitance**, which is the effective input capacitance (being the gate-drain capacitance multiplied by the transistor voltage gain). The cascode topology increases the bandwidth of the circuit. Current gain, and hence power gain, is still realized by the bottom transistor. The voltage gain of the circuit depends on the resistance presented to the drain of the top transistor.

A variable gain **cascode amplifier** is shown in Figure 3-32(b). This is similar to the cascode amplifier of Figure 3-32(a), but now the voltage at



(a)



(b)

Figure 3-31: Measured common-mode gain, A_c, and differential-mode gain, A_d: (a) with inductor-based biasing circuit with lumped inductors of 0.75 μH; and (b) with balun-like biasing circuit without lumped capacitors.

the gate of the top transistor, V_{G1}, is selected so that a variable resistance is presented to the bottom transistor, thus the voltage gain of the circuit can be varied. This is a variable gain amplifier.

FET circuits are nearly always current biased, so circuits that realize current sources and current matching are particularly important. A single MOSFET can be used to realize a current source. If the gate-to-source voltage of a MOSFET is fixed, a near-constant current source is realized (see the drain-source current equation, Equation (1.17)). A differential amplifier circuit with a variable current source is shown in Figure 3-32(d). What is particularly interesting is that the transistor controlling the bias to the current source can be part of a digital circuit, enabling digital control of the analog circuit bias. The concept can be replicated by replacing M_1 by multiple transistors in parallel with each transistor having a binary signal at the gate. This is a fundamental component in the **digital control** of analog circuits, including RFICs. For example, modern RFICs incorporate digitally controlled trimming of RFICs to achieve, for example, enhanced IQ balance of quadrature modulators.

Another circuit that controls current in an RFIC is the **current mirror** shown in Figure 3-32(c). In this circuit, $I_1 = I_2$, as the gate-source voltages of transistors M_1 and M_2 are the same. The drain-gate connection of M_1 ensures

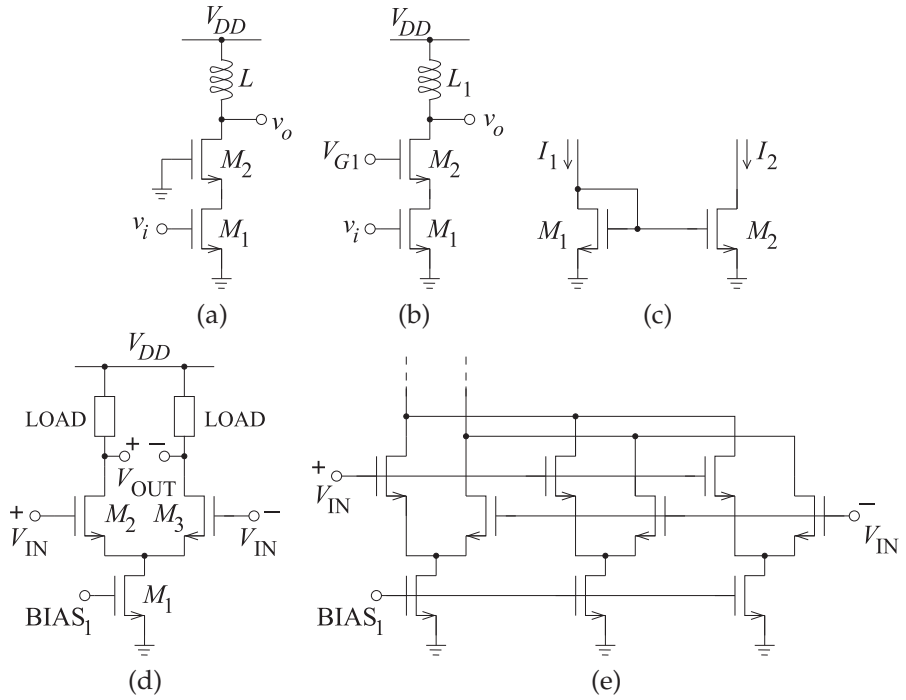


Figure 3-32: MOS analog circuits: (a) cascode amplifier; (b) variable gain cascode amplifier; (c) current mirror; (d) differential pair; (e) multi-tanh triplet implementation of a differential pair providing enhanced linearity.

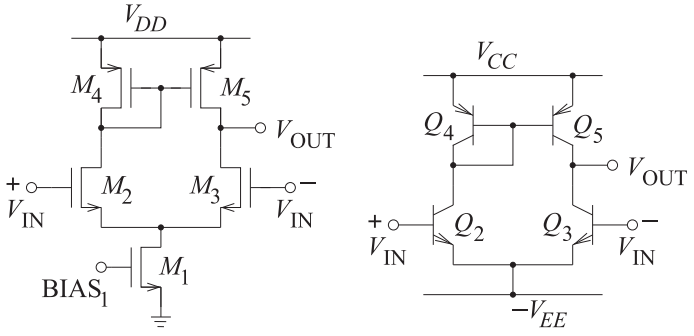


Figure 3-33: Differential pair circuits each with a current mirror.

that the gate-source voltage will be whatever is needed to support the current I_1 derived from the rest of the circuit.

A CMOS differential amplifier with a current mirror load is shown in Figure 3-33(a). In this configuration the current mirror presents a high differential impedance. If the load impedance presented to the terminal, labeled $V_{\rm OUT}$, is less than this load, then the current mirror-loaded differential amplifier realizes a single-ended output while having the essential functionality of a differential circuit to reject common-mode signals. There is a price to pay for this functionality. The circuit of Figure 3-33(a) has three drain-source voltage drops between the rails. This reduces the available

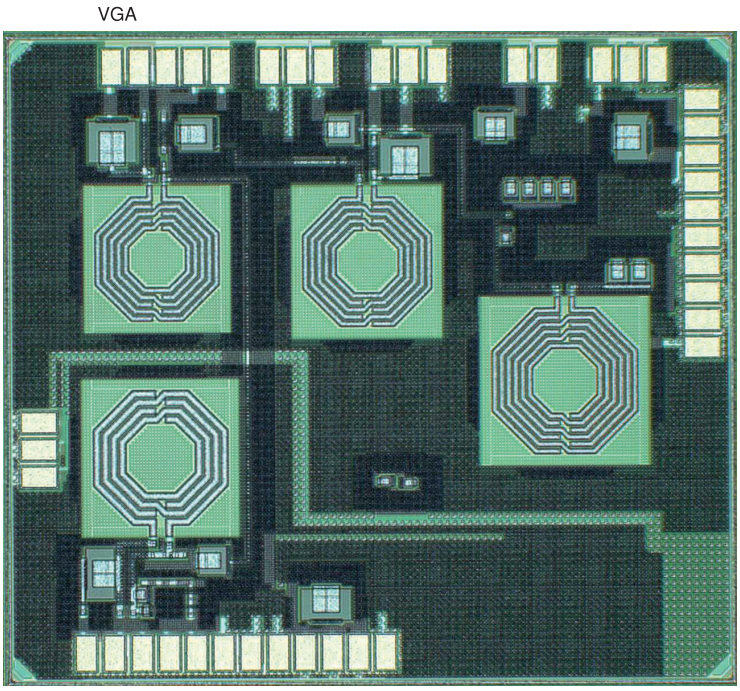


Figure 3-34: Photomicrograph of a 90 nm CMOS WCDMA transmitter showing (MIXER) the up-conversion mixer and quadrature LO divider; (VGA) the cascode variable gain amplifier; and the two-stage driver amplifier with 9.6 dBm output power. Die size is 1.1 mm $\times$ 1.4 mm. After Yang [21], and Yang and Gard [12]. Copyright K. Gard and X. Yang, used with permission.

voltage swing. This is one of the major problems encountered with RFICs, as the supply voltage is dictated by the relatively low supply voltages that can be supported in a process that is optimized for low-voltage digital circuits. A current mirror can also be realized using BJTs with the BJT-based current mirror-loaded differential pair shown in Figure 3-33(b) as an example.

Mathematically the simplified input-output characteristic of a MOSFET is essentially a quadratic (see Equation (1.17)). It is a challenge to take such fundamental algebraic models and derive the equations that describe the operation of a complete circuit; a challenge that must be addressed in the synthesis of a circuit with specified distortion and noise performance. It can be shown that the relationship between the drain current and the drain-source voltage has the form of a tanh function [10]. In terms of the transconductance, g_m, it appears as a quadratic-like function with a peak value at a drain voltage that is controlled by the W/L ratio. By putting several differential pairs in parallel, with each pair having staggered W/L ratios, a compound differential amplifier with enhanced linearity can be realized [11, 12]. This circuit is known as a multi-tanh differential pair. A triplet **multi-tanh differential pair** is shown in Figure 3-32(e). Detailed RFIC design involves the algebraic derivation of the required conditions. This network synthesis applied to RFICs is explored in numerous references [13–20] as well as a large number of papers on RFIC design. Synthesis to control distortion and noise is at the heart of RFIC design. Collecting novel circuit topologies, and the techniques to synthesize them (e.g. from conference and journal papers and patents), is an essential part of RFIC design; which is not that different from the process for all other forms of circuit design.

Figure 3-34 is a photomicrograph of a 90 nm WCDMA transmitter. The design consists of three blocks: an up-converting mixer (MIXER), a variable gain cascode amplifier (VGA), and a two-stage driver amplifier. The

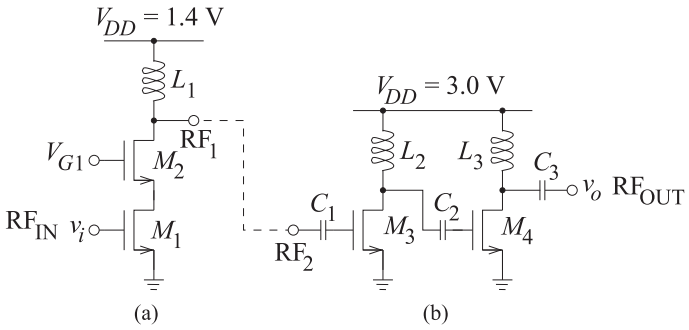


Figure 3-35: Three-stage amplifier of the WCDMA receiver shown in Figure 3-34: (a) a variable gain cascode amplifier; and (b) the two-stage driver amplifier. After Yang [21], and Yang and Gard [12]. Copyright K. Gard and X. Yang, used with permission.

schematic of the output amplifier stages is shown in Figure 3-35. The variable gain amplifier, the VGA block, is a cascode amplifier with variable biasing of the top FET in the cascode to realize a variable gain. Each of the amplifiers in the two-stage driver amplifier has a different V_{DD} so that the nonlinearities of the two stages can be designed to cancel, and thus the overall performance of the driver amplifier is linearized [12, 21].

3.9 Summary

The bandwidth of an amplifier is limited by both the frequency-dependent transconductance of the active device and by the device parasitics. The frequency-dependent transconductance is most significant with FETs as the basic operating control mechanism at the gate is the field produced at a capacitor. Thus the transconductance of a FET varies as the inverse of frequency. With unpackaged transistors, the device parasitics are capacitors. These are augmented by inductances and transmission line effects in packaged devices. Also the feedback capacitor between the collector/drain and base/gate becomes important at higher frequencies. Ignoring the feedback capacitance and thinking just about the input of the transistor, the input of the transistor is a capacitance sometimes in series and sometimes in parallel with a resistance, with the resistance describing the absorption of RF input power by the transistor.

There are two main approaches to wideband amplifier design. One is the synthesis of a matching network that provides a match over a bandwidth that is rarely more than one-half octave wide at microwave frequencies. One way of visualizing the design difficulty is to realize that a simple circuit comprising a resistor and a capacitor (this could represent the input and output equivalent circuits of an active deice) have a locus with respect to frequency that rotates clockwise on a Smith chart. Matching requires that the complex conjugate of the impedance be matched, so matching network design requires that a circuit be synthesized that has a counterclockwise rotation on the Smith chart. Such a circuit can be realized using a transmission line network, but matches over one-half octave of bandwidth are usually all that can be achieved. Another approach to wideband amplifier design is to use multiple active devices and incorporate the device parasitics into input and output transmission lines. This distributed amplifier approach can realize amplifier bandwidths of two or more octaves.

A third approach, which was considered through a case study, is to consider the design problem as a filter synthesis problem. Indeed, following the active device directly with a filter and bypassing the matching network

can broaden the amplifier bandwidth by eliminating the bandwidth limiting effect of matching to a specific system impedance. The necessary impedance transformation is performed in the filter. That is, the doubly terminated filter network has one impedance at the first port (e.g., the lower output impedance of the amplifier) and another impedance, say the system impedance, at the second port.

3.10 References

[1] R.-C. Liu, K.-L. Deng, and H. Wang, "A 0.6-22-GHz broadband cmos distributed amplifier," in *2003 IEEE Radio Frequency Integrated Circuits (RFIC) Symp.*, Jun. 2003, pp. 103–106.

[2] J. Beyer, S. Prasad, R. Becker, J. Nordman, and G. Hohenwarter, "MESFET distributed amplifier design guidelines," *IEEE Trans. on Microwave Theory and Techniques*, vol. 32, no. 3, pp. 268–275, Mar. 1984.

[3] M.-D. Tsai, H. Wang, J.-F. Kuan, and C.-S. Chang, "A 70 GHz cascaded multi-stage distributed amplifier in 90nm cmos technology," in *2005 IEEE Int. Solid-State Circuits Conf., 2005. Dig. of Technical Papers*, Feb. 2005, pp. 402–606.

[4] R. Bhatia, J. Gerber, and T. Kwan, "Analyze large signal distributed amps with nonlinear cae," *Microwaves and RF*, pp. 121–129, Nov. 1989.

[5] R. Gilmore and M. Steer, "Nonlinear circuit analysis using the method of harmonic balance—a review of the art: part i, introductory concepts," *Int. J. on Microwave and Millimeter Wave Computer Aided Engineering*, vol. 1, pp. 22–37, Jan. 1991.

[6] M. Steer, *Microwave and RF Design, Networks*, 3rd ed. North Carolina State University, 2019.

[7] ——, *Microwave and RF Design, Transmission Lines*, 3rd ed. North Carolina State University, 2019.

[8] W. Fathelbab and M. Steer, "Broadband network design," in *Multifunctional Adaptive Microwave Circuits and Systems*, M. Steer and W. Palmer, Eds., 2008, ch. 8.

[9] ——, "Distributed biasing of differential RF circuits," *IEEE Trans. on Microwave Theory and Techniques*, vol. 52, no. 5, pp. 1565–1572, May 2004.

[10] P. Gray, P. Hurst, S. Lewis, and R. Meyer, *Analysis and Design of Analog Integrated Circuits*, 4th ed. Wiley, 2001.

[11] B. Gilbert, "The multi-tanh principle: a tutorial overview," *IEEE J. of Solid-State Circuits*, vol. 33, no. 1, pp. 2–17, Jan. 1998.

[12] X. Yang, A. Davierwalla, D. Mann, and K. Gard, "A 90nm CMOS direct conversion transmitter for WCDMA," in *2007 IEEE Radio Frequency Integrated Circuits (RFIC) Symp.*, Jun. 2007, pp. 17–20.

[13] T. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*. Cambridge University Press, 2004.

[14] R. Baker, *CMOS Circuit Design, Layout, and Simulation*, 2nd ed. Wiley-Interscience, IEEE Press, 2008.

[15] A. Aktas and M. Ismail, *CMOS PLLs and VCOs for 4G wireless*. Kluwer, 2004.

[16] D. Pederson and K. Mayaram, *Analog Integrated Circuits for Communication: Principles, Simulation and Design*. Springer, 2008.

[17] B. Razavi, *Design of Analog CMOS Integrated Circuits*. McGraw-Hill, 2001.

[18] L. Dai and R. Harjani, *Design of High Performance CMOS Voltage-Controlled Oscillators*. Kluwer Academic Publishers, 2003.

[19] B. Leung, *VLSI for Wireless Communications*. Prentice Hall, 2002.

[20] M. Tiebout, *Low Power VCO Design in CMOS*. Springer, 2006.

[21] X. Yang, "90nm cmos transmitter design for WCDMA," Ph.D. dissertation, North Carolina State University, 2009.

3.11 Exercises

1. Consider a $Z_0 = 50\ \Omega$ transmission line of length $\lambda/10$ at 30 GHz.
   (a) Calculate the $ABCD$ parameters of the transmission line at 30 GHz?
   (b) With the transmission line shunted by 0.05 pF capacitors at each end, calculate the $ABCD$ parameters of the augmented transmission line.
   (c) At 30 GHz the augmented transmission line is equivalent to a single transmission line with characteristic impedance Z_{01} and length ℓ_1. What is ℓ_1 in terms of wavelengths?
   (d) What is Z_{01}?
2. A four-stage distributed FET amplifier as shown in Figure 3-1 has $R_S = R_L = 50\ \Omega$. If the capacitive and resistive loading of the transistors are ignored, what are the optimum values of R_1 and R_2? Provide your reasoning.
3. The four-stage distributed FET amplifier shown in Figure 3-1 has $R_S = 80\ \Omega$ and $R_L = 25\ \Omega$. If the capacitive and resistive loading of the transistors are ignored, what are the optimum values of R_1 and R_2? Provide your reasoning.
4. The input matching network of the wideband amplifier considered in Section 3.5 is shown in Figure 3-13(b). (Note that Port 2 of the input network is connected to the transistor.) Typically the complex conjugate of S_{22} of the input network would match the input reflection coefficient, Γ_{in}, of the transistor. Put your answers in magnitude-angle form.
   (a) Draw the input matching network showing where S_{22} is determined. Also draw the transistor terminated by the output matching network and indicate where Γ_{in} is calculated.
   (b) Use a microwave simulator to calculate the S_{22} of the input matching network at 8, 9, . . ., 12 GHz.
   (c) Determine S_{22}^* of the input matching network at 8, 9, . . ., 12 GHz.
   (d) Determine S_{11} of the transistor at 8, 9, . . ., 12 GHz.
   (e) Determine Γ_{in} of the transistor (terminated in the output matching network) at 8, 9, . . ., 12 GHz.
   (f) On a Smith chart plot S_{22}^* of the input network, S_{11} of the transistor, and Γ_{in} of the transistor.
   (g) Describe the input matching network condition for maximum power transfer used in narrowband amplifier design.
   (h) Discuss the mismatch of Γ_{in} of the transistor and S_{22}^* of the input matching network. Describe the effect that this has on the broadband response of the amplifier.
5. The output matching network of the wideband amplifier considered in Section 3.5 is shown in Figure 3-14(b). (Note that Port 2 of the input network is connected to the transistor.) Typically the complex conjugate of S_{22} of the output network would match the input reflection coefficient, Γ_{out}, of the transistor. Put your answers in magnitude-angle form.
   (a) Draw the output matching network showing where S_{22} is determined. Also draw the transistor terminated by the input matching network and indicate where Γ_{out} is calculated.
   (b) Use a microwave simulator to calculate the S_{22} of the output matching network at 8, 9, . . ., 12 GHz.
   (c) Determine S_{22}^* of the output matching network at 8, 9, . . ., 12 GHz.
   (d) Determine S_{22} of the transistor at 8, 9, . . ., 12 GHz.
   (e) Determine Γ_{out} of the transistor (terminated in the output matching network) at 8, 9, . . ., 12 GHz.
   (f) On a Smith chart plot S_{22}^* of the output network, S_{22} of the transistor, and Γ_{out} of the transistor.
   (g) Describe the output matching network condition for maximum power transfer used in narrowband amplifier design.
   (h) Discuss the mismatch of Γ_{out} of the transistor and S_{22}^* of the output matching network. Describe the effect that this has on the broadband response of the amplifier. You will want to consider the S_{21} of the transistor.
6. Plot the 50 Ω S_{11} and S_{22} parameters from 8 GHz to 12 GHz of the wideband amplifier considered in Section 3.5. It will be seen that the amplifier is not matched across the band. Discuss the reason why there is a mismatch even though the gain and noise figure of the amplifier, shown in Figure 3-18, are relatively flat from from 8 GHz to 12 GHz. Note that Port 1 is the input port of the amplifier and Port 2 is the output Port.
7. The output of a transistor is modeled as the shunt connection of a current source, a 20 Ω re-

sistor, a 0.35 pF capacitor, and a 0.7 nH inductor.

(a) What is the admittance of the transistor output at 8, 10, and 12 GHz?
(b) How does the susceptance vary with frequency?
(c) What is the shunt reactive element required to resonate the output admittance of the transistor at 8, 10, and 12 GHz?
(d) What are the equivalent inductances required to resonate the output admittance of the transistor at 8, 10, and 12 GHz?
(e) How does the inductance calculated in (d) vary with frequency?
(f) Describe a two-element circuit that has the characteristic identified in (e). (Note that this circuit would only be able to achieve the required characteristic over a smaller bandwidth than that required for a match from 8 GHz to 12 GHz.)

8. The output of a transistor is modeled as the shunt connection of a current source, a $68\ \Omega$ resistor, a 0.35 pF capacitor, and a 0.7 nH inductor.
   (a) What is the output admittance of the transistor at 8, 10 and 12 GHz?
   (b) How does the admittance vary with frequency?
   (c) Design a lumped-element matching network with two elements to match the transistor output at 10 GHz to a $50\ \Omega$ source.
   (d) Calculate the input admittance of the matching network, looking from the transistor, at 8, 10, and 12 GHz.
   (e) What is the input admittance of an ideal matching network, looking from the transistor, at 8, 10, and 12 GHz? Plot the actual and ideal admittance loci on a Smith chart using markers at 8, 10, and 12 GHz and indicating the direction of increasing frequency with arrows.

9. Consider a transistor having the S parameters shown in Table 3-1 and Figure 3-6(a). Ignore feedback effects and consider that the reflection coefficient looking into the output of the transistor is S_{22}^*.
   (a) Draw and describe the two-port input matching network problem with Port 1 at the output of the transistor and a $50\ \Omega$ termination at Port 2.
   (b) What is the ideal S_{11} of the input matching two-port at 8 GHz?
   (c) What is the ideal S_{11} of the input matching two-port at 10 GHz?
   (d) What is the ideal S_{11} of the input matching two-port at 12 GHz?
   (e) Plot the locus from 8 GHz to 12 GHz of S_{11} of the input matching two-port on a Smith chart.
   (f) Assume that the locus plotted in (e) from 8 GHz to 12 GHz can be realized using a lumped-element network. Comment on the difficulty of the design and the design approach.

10. At 10 GHz a capacitor, C_1, has a reactance of $-50\ \Omega$.
    (a) What is the impedance of C_1 at 8, 10, and 12 GHz?
    (b) How does the impedance of C_1 vary with frequency?
    (c) What is the inductance required to resonate the capacitance at 8, 10, and 12 GHz?
    (d) How does the inductance calculated in (b) vary with frequency?

11. The input of a transistor is modeled as a $20\ \Omega$ resistor in series with a 0.3 pF capacitor.
    (a) What is the impedance of the transistor input at 8, 10, and 12 GHz?
    (b) How does the impedance vary with frequency?
    (c) What is the series inductance required to resonate out the transistor capacitance at 8, 10, and 12 GHz?
    (d) Comment on whether a wideband match of a resistive source to the input of a transistor can be achieved using a frequency-independent inductor.

12. The input of a transistor is modeled as a $20\ \Omega$ resistor in series with a 0.3 pF capacitor. The transistor is part of an amplifier operating in a $50\ \Omega$ system.
    (a) Design a lumped-element matching network with two elements (inductors and/or capacitors) to match the transistor input at 10 GHz to a $50\ \Omega$ source.
    (b) Calculate the return loss (looking into the matching network from the source) at 8, 9, 10, 11, and 12 GHz.
    (c) Calculate the fraction of the available input power, expressed in decibels, delivered to the transistor at 8, 9, 10, 11, and 12 GHz and indicate the direction of increasing frequency with arrows.
    (d) Comment on the variation in amplifier gain solely due to mismatch at the transistor input.

13. Consider the input of a transistor having the S parameters shown in Table 3-1 and Figure 3-6(a). Ignore feedback effects so that for the active device $\Gamma_{\text{in}} = S_{11}$. Also an input matching

network terminated in 50 Ω at Port 1 and the active device at Port 2.

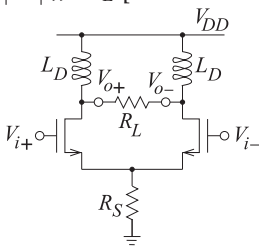
(a) What is the ideal 50 Ω S_{22} of the input matching network (i.e., seen from the transistor input) at 8 GHz?
(b) What is the ideal 50 Ω S_{22} of the input matching network (i.e., seen from the transistor input) at 10 GHz?
(c) What is the ideal 50 Ω S_{22} of the input matching network (i.e., seen from the transistor input) at 12 GHz?

14. Consider matching the input of a transistor having the S parameters shown in Table 3-1 and Figure 3-6(a). Ignore feedback effects and consider that the input reflection coefficient of the transistor $\Gamma_{\text{in}} = S_{11}$. Curve B in Figure 3-6(a) is the locus of the impedance looking into the matching network from the transistor. What two-element network has this locus? (One of the elements may be a resistor).

15. Consider synthesizing a two-port matching network terminated in a 50 Ω load and with an input reflection coefficient Γ_1 shown as Curve B in Figure 3-6(a). Draw and describe the two-port matching network problem.

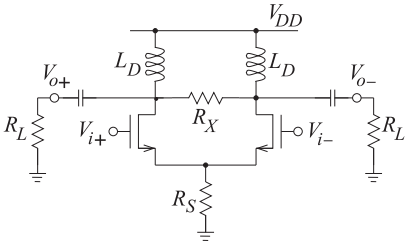
16. Consider synthesizing a two-port matching network terminated in a 50 Ω load and with an input reflection coefficient Γ_1 shown as Curve B in Figure 3-6(a). Can a broadband match be obtained using a two-element matching network? Explain your answer in terms of rotations on a Smith chart.

17. Consider the inductively biased differential Class A amplifier shown below. L_D is a choke inductor so $|sL| \gg R_L$.[Parallels Example 3.1.]



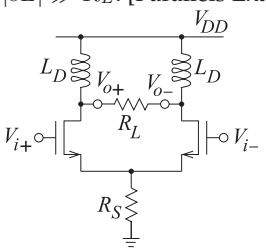
What is the CMRR when R_S = 20 kΩ, R_L = 10 kΩ, the transistor transconductance, g_m = 50 mS, and the drain-source resistance, r_d is 100 kΩ?

18. Consider the inductively biased differential Class A amplifier shown below. The capacitors can be treated as RF short circuits. L_D is a choke inductor so $|sL| \gg R_L$. [Parallels Example 3.1]



(a) Derive a symbolic expression for the CMRR of the amplifier assuming that the drain-source resistance of the transistors, r_0 or r_d, is much greater than both R_L and R_X, and so can be ignored.
(b) What is the CMRR when R_S = 10 kΩ, R_X = 30 kΩ, R_L = 10 kΩ, and the transistor transconductance, g_m is 10 mS.

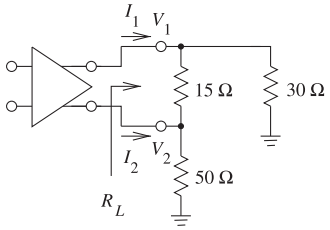
19. Consider the inductively-biased differential Class A amplifier shown below. L_D is a choke inductor so $|sL| \gg R_L$. [Parallels Example 3.1]



(a) Derive a symbolic expression for the differential mode gain of the amplifier.
(b) Derive a symbolic expression for the CMRR of the amplifier.
(c) What is CMRR when R_S = 10 kΩ, R_L = 10 kΩ, the transistor transconductance, g_m is 15 mS, and the transistors' drain-source resistance, r_d, is 100 kΩ?

20. A differential amplifier has a differential-mode gain of 20 dB and a common-mode gain of −3 dB.
(a) What is the the odd-mode gain?
(b) What is the the even-mode gain?

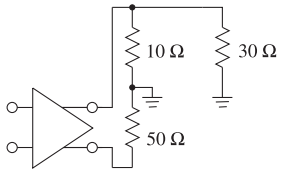
21. Consider the differential amplifier below. [Parallels Example 3.0]



(a) What is the differential load impedance?
(b) What is the odd-mode load impedance?

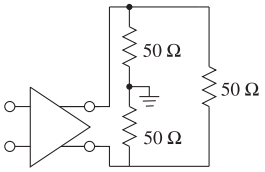
(c) What is the common-mode load impedance?
(d) What is the even-mode load impedance?

22. Consider the differential amplifier below. [Parallels Example 3.0]



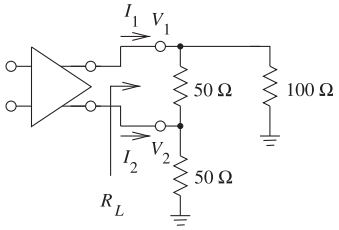
(a) What is the differential load impedance?
(b) What is the odd-mode load impedance?
(c) What is the common-mode load impedance?
(d) What is the even-mode load impedance?

23. Consider the differential amplifier below. [Parallels Example 3.0]



(a) What is the differential load impedance?
(b) What is the odd-mode load impedance?
(c) What is the common-mode load impedance?
(d) What is the even-mode load impedance?
(e) If the differential-mode gain of the amplifier is 20 dB and the common-mode gain is 2 dB, what is the odd-mode gain?

24. Consider the differential amplifier below. L_D is a choke inductor so $|sL| \gg R_L$. [Parallels Example 3.0]



(a) What is the differential load impedance?
(b) What is the odd-mode load impedance?
(c) What is the common-mode load impedance?
(d) What is the even-mode load impedance?

25. A pseudo-differential amplifier is shown in Figure 3-28. Distributed biasing of this amplifier (replacing the inductors and R_{DD}), presents a common-mode impedance of $5\ \Omega$ and a differential-mode impedance of $1\ \text{k}\Omega$ to the drain terminals of the transistors in the middle of the amplifier band. The transconductance of each transistor is $g_m = 1$ S, and the internal parasitics of the transistors can be ignored.

(a) Draw the common-mode amplifier schematic without the biasing elements. Include the common-mode load resistance R_{Lc}.
(b) Draw the odd-mode amplifier schematic without the biasing elements. Include the odd-mode load resistance R_{Lo}.
(c) Draw the even-mode amplifier schematic without the biasing elements. Include the odd-mode load resistance R_{Lo}.
(d) Draw the differential-mode amplifier schematic without the biasing elements. Include the common-mode load resistance R_{Lc}.
(e) What is the common-mode gain?
(f) What is the differential-mode gain?
(g) What is the common-mode rejection ratio in decibels?

26. A pseudo-differential amplifier is shown in Figure 3-28. Distributed biasing of this amplifier (replacing the inductors and R_{DD}), presents a common-mode impedance of $5\ \Omega$ and an odd-mode impedance of $1\ \text{k}\Omega$ to the drain terminals of the transistors in the middle of the amplifier band. The transconductance of each transistor is $g_m = 100$ mS and the internal parasitics of the transistors can be ignored.

(a) Draw the common-mode amplifier schematic without the biasing elements. Include the common-mode load resistance R_{Lc}.
(b) Draw the odd-mode amplifier schematic without the biasing elements. Include the odd-mode load resistance R_{Lo}.
(c) Draw the even-mode amplifier schematic without the biasing elements. Include the odd-mode load resistance R_{Lo}.
(d) Draw the differential-mode amplifier schematic without the biasing elements. Include the common-mode load resistance R_{Lc}.
(e) What is the even-mode impedance presented to the amplifier?
(f) What is the differential-mode impedance presented to the amplifier?
(g) What is the even-mode voltage gain?
(h) What is the differential-mode voltage gain?
(i) What is the common-mode voltage gain?
(j) What is the odd-mode voltage gain?
(k) What is the common-mode rejection ratio?

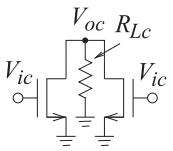
3.11.1 Exercises By Section

†challenging, ‡very challenging

§3.2 $1^{\dagger}, 2, 3$
§3.5 $4^{\ddagger}, 5^{\ddagger}, 6^{\dagger}, 7^{\dagger}, 8^{\dagger}, 9^{\dagger}$
§3.6 $10, 11, 12, 13, 14, 15, 16^{\dagger}, 17^{\dagger}, 18^{\dagger}, 19^{\dagger}, 20, 21, 22, 23^{\dagger}, 24^{\dagger}$
§3.7 $25^{\dagger}, 26^{\dagger}$

3.11.2 Answers to Selected Exercises

1(d) $34.7\ \Omega$
5(b) 121
10(d) $1/f^2$
12(c) 8 GHz, -2.77 dB
9 GHz, -0.57 dB
10 GHz, 0 dB
11 GHz, -0.33 dB
12 GHz, -1.03 dB
18(h) -100
20(b) -3 dB
21 $37.5\ \Omega$
25(a)



CHAPTER 4

Power Amplifiers

4.1 Introduction 109
4.2 Simulation of Nonlinear Microwave Circuits 110
4.3 Switching Amplifiers, Classes D, E, and F 116
4.4 Distortion and Digitally Modulated Signals 122
4.5 Loadpull 127
4.6 Case Study: Design of a WiMAX Power Amplifier 128
4.7 Linearization 136
4.8 Advanced Power Amplifiers 138
4.9 MMIC Power Amplifiers 142
4.10 RFIC Power Amplifiers 144
4.11 Summary 151
4.12 References 152
4.13 Exercises 155

4.1 Introduction

This chapter considers amplifiers that are designed to produce large RF powers and achieve high efficiency. The design of such amplifiers is more complicated than the design of small signal amplifiers, which can be designed using linear techniques. It is now necessary to use nonlinear simulation tools and laboratory optimization to develop designs that achieve high efficiency with acceptable distortion [1]. The design of power amplifiers involves the trade-off of distortion and amplifier efficiency. Before beginning this chapter the reader should be familiar with the metrics for nonlinear distortion described in Section 4.5 of [2].

Transistors used in microwave amplifiers are large, as can be seen in Figure 4-1. This transistor is targeted for use in WCDMA amplifiers operating between 2.11 and 2.17 GHz with an output power of 140 W at 1 dB gain compression. The transistor includes input and output matching networks in the package. This provision by the transistor vendor reduces the complexity of amplifier design.

4.2 Simulation of Nonlinear Microwave Circuits

The circuit simulation tools used to model RF circuits are linear circuit simulators, transient circuit simulators (e.g., Spice), and nonlinear steady-state simulators. Transient circuit simulators can model large signals in RF circuits, but those available to RF designers cannot be used to model circuits requiring high dynamic range simulation or to model circuits with sharp frequency responses such as circuits with high-order filters. These simulators can be very slow, or perhaps impossibly slow, in modeling circuits with narrowband signals such as a digitally modulated carrier. When it is important to capture nonlinear and frequency-response effects precisely, nonlinear steady-state simulators are preferred. The two main types of nonlinear steady-state simulators available to designers are harmonic balance (HB) simulators [3, 4] and Spice-like transient simulators modified to efficiently find the response of a circuit to periodic excitation [5] (so-called periodic steady-state (PSS) analysis).

Nonlinear steady-state simulators exploit the narrowband nature of most radio systems and the circuit waveforms that are essentially steady-state, although not necessarily periodic. Such waveforms are called quasi-periodic waveforms. As an example of the waveforms to be determined, consider the

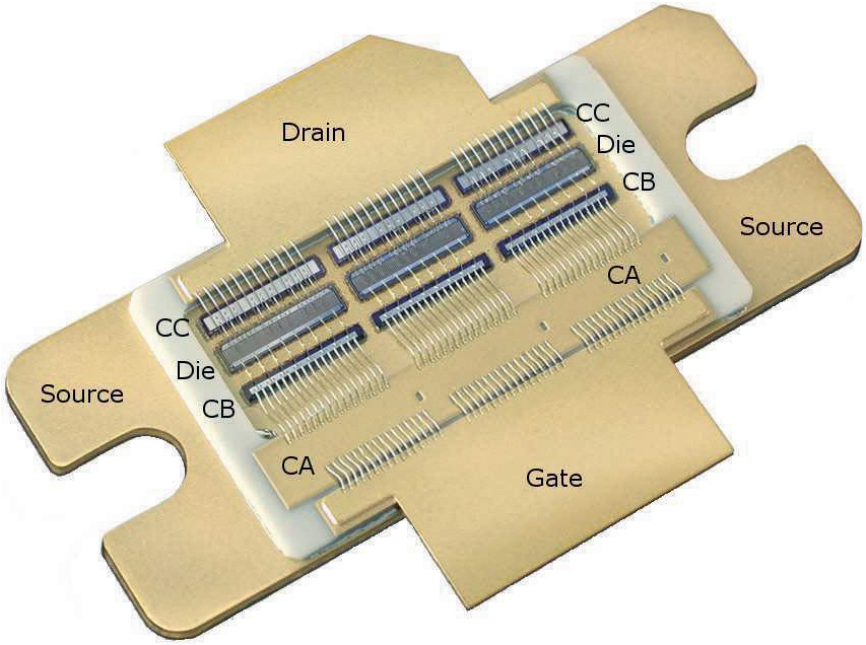


Figure 4-1: A 2.1 GHz silicon LDMOS transistor. The gate and drain tabs are 12.5 mm across. The three dies operate in parallel. The input matching network comprises a series inductance provided by bond wires, a shunt capacitor (Capacitor A, C_A), another series inductance from bond wires, and another shunt capacitor (C_B). The network is then connected to each gate finger of the transistor dies using short bond wires. The output matching network consists of a shunt capacitor (C_C) and series inductance. There are 189 bond wires. Used courtesy of Freescale Semiconductor Inc. Also see [1].

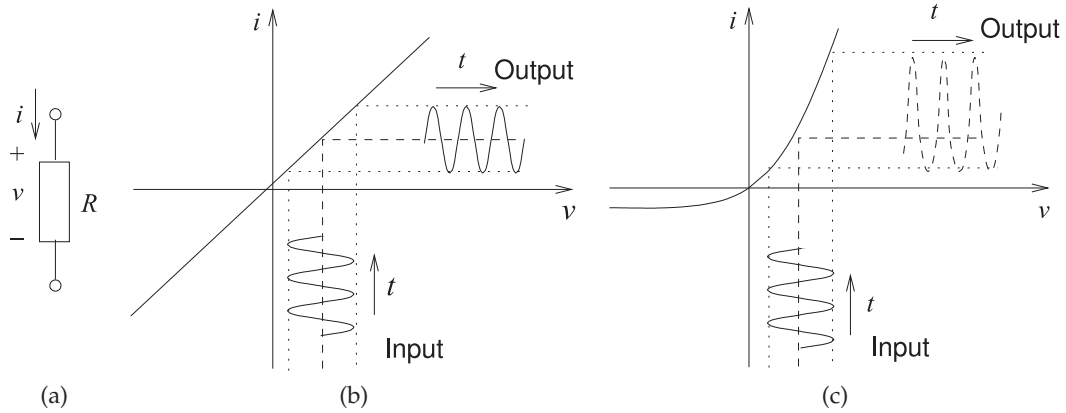


Figure 4-2: Current responses of a resistor: (a) resistor with passive convention defining voltage and current; (b) i-v characteristic of a linear resistor; and (c) i-v characteristic of a diode (a nonlinear resistor).

responses, shown in Figure 4-2, of linear and nonlinear resistors to an applied sinusoidal voltage.

Figure 4-2(b) shows the i-v characteristic of a linear resistor. With an applied sinusoidal voltage, the output current waveform of the linear resistor is also a sinusoid. If the applied voltage signal is a sum of sinusoids, then the output current is also a sum of sinusoids. The component of output current at each frequency only depends on the applied voltage component at that frequency. With a nonlinear resistor having the i-v characteristic in Figure 4-2(c), a large applied sinusoidal signal results in an output that is distorted and is a steady-state signal that has harmonics of the original signal. If the applied signal is the sum of two sinusoids of frequencies f_1 and f_2, then the output will be a steady-state signal with components having frequencies $mf_1 + nf_2$, where m and n are integers. The key feature here is that if a steady-state signal, a sum of sinusoids, is applied to a nonlinear circuit, then the output will also be a steady-state signal, a sum of sinusoids, but now each frequency component at the output is affected by every frequency component of the input signal. However, the simulation problem simplifies to finding the amplitudes and phases of the sinusoidal components rather than trying to determine the output waveform at a very large number of time points as done in a transient simulator. This is the basis of nonlinear steady-state simulation. In a narrowband radio, signals are very close to being sinusoidal with very slowly varying amplitude and phase [6].

4.2.1 Harmonic Balance Analysis of RF Circuits

With the harmonic balance method, the steady-state response of a nonlinear circuit is assumed to be a sum of sinusoids [3, 4]. This assumed form of the solution then allows simplification of the circuit equations, and simulation is used to determine the unknown coefficients: the magnitudes and phases of the sinusoids.

A harmonic-balance simulator can be many orders of magnitude more

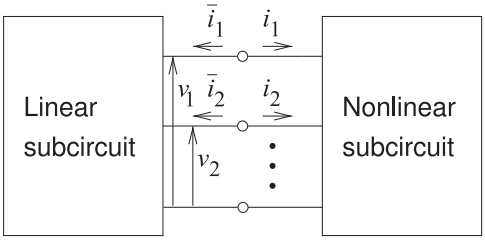


Figure 4-3: Analysis of a nonlinear circuit by the harmonic balance method partitions the circuit into linear and nonlinear subcircuits.

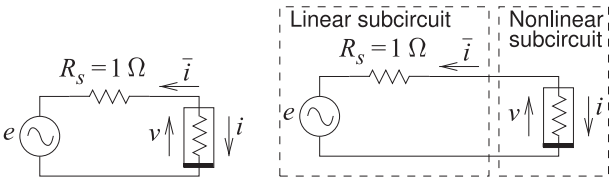


Figure 4-4: Example harmonic balance circuit: (a) circuit with nonlinear resistor; and (b) partitioned circuit. $e(t) = E\cos(\omega y)$.

efficient than a time-domain simulator, and lends itself well to analysis of narrowband circuits and to optimization. Another major advantage of the harmonic balance method is that linear circuits can be of practically any size, with no significant increase in overall simulation time.

4.2.2 Example: Harmonic Balance Analysis of a Simple Circuit

Internally a harmonic balance simulator partitions a circuit into two subcircuits, as shown in Figure 4-3. During analysis, the set of frequencies being considered is fixed and the linear block need only be calculated initially, and its admittance parameters stored and reused at every stage of the analysis. Newton's method (or similar iterative derivative-based minimization method) is used with harmonic balance to solve for the state of the circuit, that is, the amplitudes and phases of the voltage phasors at the interface.

As an example, consider the circuit in Figure 4-4(a) where the nonlinear resistor is described by

$$i(t) = v(t) + [v(t)]^2 . \tag{4.1}$$

The first step in analysis partitions the circuit into linear and nonlinear subcircuits, as shown in Figure 4-4(b). Harmonic balance simulation minimizes the frequency domain Kirchoff's current law error at the linear circuit-nonlinear circuit interface. Next the number of sinusoids (or tones) to be considered must be chosen. The choice in this example is to consider only the DC, fundamental at radian frequency ω, and second-harmonic tones. Then the voltage at the interface is

$$v(t) = V_0 + V_1\cos(\omega t) + V_2\cos(2\omega t). \tag{4.2}$$

Phase has been dropped, as this is a resistive circuit and all currents and voltages will have the same phase, nominally zero. Thus the unknowns are the amplitudes V_0, V_1, and V_2. With values of V_0, V_1, and V_2 assumed (and updated through iteration), the current flowing into the linear subcircuit can be calculated using the nodal admittance matrix of the linear subcircuit

yielding

$$\bar{i}(t) = \bar{I}_0 + \bar{I}_1 \cos(\omega t) + \bar{I}_2 \cos(2\omega t). \tag{4.3}$$

Similarly the nonlinear model of the element in the nonlinear subcircuit can be used to calculate the nonlinear currents:

$$i(t) = I_0 + I_1 \cos(\omega t) + I_2 \cos(2\omega t). \tag{4.4}$$

The linear subcircuit, shown in Figure 4-4(a), and the nonlinear subcircuit described by the quadratic model in Equation (4.1) result in the following circuit equations:

$$\begin{array}{lll} I_0 = V_0 + \frac{1}{2}V_2^2, & I_1 = V_1, & I_2 = \frac{1}{2}V_1^2 \\ \bar{I}_0 = V_0, & \bar{I}_1 = V_1 - E, & \text{and } \bar{I}_2 = V_2. \end{array} \tag{4.5}$$

Since only the DC, fundamental, and second-harmonic components are being considered, the KCL error, F, is

$$F = |f_0| + |f_1| + |f_2| \tag{4.6}$$

where the Kirchoff current error at DC, the fundamental, and the second harmonic are

$$f_0 = I_0 + \bar{I}_0, \quad f_1 = I_1 + \bar{I}_1, \quad \text{and} \quad f_2 = I_2 + \bar{I}_2, \tag{4.7}$$

respectively. Thus F is minimized, or alternatively the zeros of each sub-error, the f_ns, are found. The zeros can be found using a Newton–Raphson iterative technique to determine the voltages (V_0, V_1, and V_2) that yield the zeros of f_0, f_1, and f_2. Thus the $(i+1)$th analysis iteration is

$$^{i+1}\begin{bmatrix} V_0 \\ V_1 \\ V_2 \end{bmatrix} = {}^{i}\begin{bmatrix} V_0 \\ V_1 \\ V_2 \end{bmatrix} - \left[\mathbf{J}\left({}^{i}\begin{bmatrix} V_0 \\ V_1 \\ V_2 \end{bmatrix}\right)\right]^{-1} \times \begin{bmatrix} f_0({}^{i}[V_0, V_1, V_2]^{\mathrm{T}}) \\ f_1({}^{i}[V_0, V_1, V_2]^{\mathrm{T}}) \\ f_2({}^{i}[V_0, V_1, V_2]^{\mathrm{T}}) \end{bmatrix}. \tag{4.8}$$

The Jacobian, $\mathbf{J}$, is a matrix of derivatives of the f_ns with respect to the V_ns. Thus

$$\begin{aligned} &\mathbf{J}\left({}^{i}[V_0, V_1, V_2]^{\mathrm{T}}\right) \\ &= \begin{bmatrix} \dfrac{\partial f_0({}^{i}[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_0} & \dfrac{\partial f_0({}^{i}[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_1} & \dfrac{\partial f_0({}^{i}[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_2} \\ \dfrac{\partial f_1({}^{i}[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_0} & \dfrac{\partial f_1({}^{i}[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_1} & \dfrac{\partial f_1({}^{i}[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_2} \\ \dfrac{\partial f_2({}^{i}[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_0} & \dfrac{\partial f_2({}^{i}[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_1} & \dfrac{\partial f_2({}^{i}[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_2} \end{bmatrix}. \end{aligned} \tag{4.9}$$

Now each element of the Jacobian is affected by both the linear and nonlinear subcircuits, so, for example,

$$\frac{\partial f_2({}^{i}[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_1} = \frac{\partial I_2({}^{i}[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_1} + \frac{\partial \bar{I}_2({}^{i}V_2)}{\partial V_1}. \tag{4.10}$$

Since the linear current $\bar{I}_2$ is dependent only on V_2 (see Equation (4.5),

$$\frac{\partial \bar{I}_2({}^iV_2)}{\partial V_1} = 0, \tag{4.11}$$

and following trigonometric expansion of Equation (4.1),

$$\frac{\partial I_2({}^i[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_1} = \frac{\partial}{\partial V_1}\left(\tfrac{1}{2}{}^iV_1^2\right) = {}^iV_1. \tag{4.12}$$

Similarly

$$\left.\begin{array}{ll}
\dfrac{\partial f_0({}^i[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_0} = 1 + 2{}^iV_0 + 1 & \dfrac{\partial f_0({}^i[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_1} = {}^iV_1 \\
\dfrac{\partial f_0({}^i[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_2} = {}^iV_2 & \dfrac{\partial f_1({}^i[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_0} = 2{}^iV_1 \\
\dfrac{\partial f_1({}^i[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_1} = 1 + 2{}^iV_0 + {}^iV_2 + 1 & \dfrac{\partial f_1({}^i[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_2} = {}^iV_1 \\
\dfrac{\partial f_2({}^i[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_0} = 2{}^iV_2 & \dfrac{\partial f_2({}^i[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_1} = {}^iV_1 \\
\dfrac{\partial f_2({}^i[V_0, V_1, V_2]^{\mathrm{T}})}{\partial V_2} = 2{}^iV_0 + 2 &
\end{array}\right\}. \tag{4.13}$$

Thus the equations to be solved by the harmonic balance simulator are

$$\left.\begin{array}{lll}
{}^iI_0 = {}^iV_0 + {}^iV_0^2 + V_1^2/2 + V_2^2/2 & {}^i\bar{I}_0 = {}^iV_0 & f_0 = {}^iI_0 + {}^i\bar{I}_0 \\
{}^iI_1 = {}^iV_1 + 2{}^iV_0\,{}^iV_1 + {}^iV_1\,{}^iV_2 & {}^i\bar{I}_1 = {}^iV_1 - E_1 & f_1 = {}^iI_1 + {}^i\bar{I}_1 \\
{}^iI_2 = {}^iV_2 + 2{}^iV_0\,{}^iV_2 + {}^iV_1^2/2 & {}^i\bar{I}_2 = {}^iV_2 & f_2 = {}^iI_2 + {}^i\bar{I}_2
\end{array}\right\}. \tag{4.14}$$

This is solved iteratively using the Newton–Raphson algorithm described by Equation (4.8) to provide an updated estimate of the voltages (${}^{(i+1)}V_0$, ${}^{(i+1)}V_1$, and ${}^{(i+1)}V_2$). The output of a program implementing this algorithm is shown in Table 4-1. Note how quickly the iterations arrive at the steady-state solution. This quick convergence to a steady-state solution is typical of harmonic balance simulation of nonlinear RF circuits.

The harmonic balance approach to nonlinear circuit analysis can be extended to consider excitation by sums of nonharmonically related sinusoids, which, for example, enables distortion to be determined using a two-tone test.

Many nonlinear RF circuits can be solved with just a few harmonics considered and only a few iterations are required to obtain convergence. The harmonic balance analysis is many orders of magnitude faster than simulation using transient circuit simulation and the dynamic range of the simulation can exceed 150 dB while a transient simulator is often limited to dynamic ranges of 80 dB and often much less is achieved. Consider a cell phone that can have transmit signals of up to 30 dBm, yet receive signals as small as -100 dBm. Circuit simulation of such a system requires that the simulator have a dynamic range 20 dB greater than the 130 dB dynamic range of the cellphone signals. This is easily met by harmonic balance simulators. However, there are limitations to the use of the harmonic balance method.

ITERATION 0					
$\bar{I}_0$	= 0	$\bar{I}_1$	= 0	$\bar{I}_2$	= 0
V_0	= 0	V_1	= 0.5	V_2	= 0
I_0	= 0.5	I_1	= 1	I_2	= 0.5
ITERATION 1					
$\bar{I}_0$	= 0	$\bar{I}_1$	= −0.5	$\bar{I}_2$	= 0
V_0	= −0.0769231	V_1	= 0.557692	V_2	= −0.0769231
I_0	= 0.125	I_1	= 0.5	I_2	= 0.125
ITERATION 2					
$\bar{I}_0$	= −0.0769231	$\bar{I}_1$	= −0.442308	$\bar{I}_2$	= −0.0769231
V_0	= −0.0892028	V_1	= 0.57747	V_2	= −0.0912325
I_0	= 0.087463	I_1	= 0.428994	I_2	= 0.0904216
ITERATION 3					
$\bar{I}_0$	= −0.0892028	$\bar{I}_1$	= −0.42253	$\bar{I}_2$	= −0.0912325
V_0	= −0.089835	V_1	= 0.578574	V_2	= −0.0919462
I_0	= 0.0896515	I_1	= 0.421762	I_2	= 0.0917795
ITERATION 4					
$\bar{I}_0$	= −0.089835	$\bar{I}_1$	= −0.421426	$\bar{I}_2$	= −0.0919462
V_0	= −0.0898368	V_1	= 0.578577	V_2	= −0.0919482
I_0	= 0.0898363	I_1	= 0.421424	I_2	= 0.0919477
ITERATION 5					
$\bar{I}_0$	= −0.0898368	$\bar{I}_1$	= −0.421423	$\bar{I}_2$	= −0.0919482
V_0	= −0.0898368	V_1	= 0.578577	V_2	= −0.0919482
I_0	= 0.0898368	I_1	= 0.421423	I_2	= 0.0919482
ITERATION 6					
$\bar{I}_0$	= −0.0898368	$\bar{I}_1$	= −0.421423	$\bar{I}_2$	= −0.0919482
V_0	= −0.0898368	V_1	= 0.578577	V_2	= −0.0919482
I_0	= 0.0898368	I_1	= 0.421423	I_2	= 0.0919482
ITERATION 7					
$\bar{I}_0$	= −0.0898368	$\bar{I}_1$	= −0.421423	$\bar{I}_2$	= −0.0919482
V_0	= −0.0898368	V_1	= 0.578577	V_2	= −0.0919482
I_0	= 0.0898368	I_1	= 0.421423	I_2	= 0.0919482

stop

Table 4-1: Circuit quantities at each iteration in the harmonic balance analysis of the circuit in Figure 4-4(a).

A major drawback is that harmonic balance does not work well with circuits containing large numbers of transistors, say a few tens of transistors. This is largely because the Jacobian becomes very large and analysis becomes unwieldy.

4.2.3 User's Guide to Using Harmonic Balance Analysis

Three major factors limit the accuracy of harmonic balance circuit simulation:

(i) The number of tones included in the analysis. If the number of tones is too small, there will be truncation error. Truncation error arises because, theoretically, an infinite number of harmonics can be generated by the interaction of large signals with even simple nonlinear circuits. The error can be reduced, of course, by specifying additional frequency components.

(ii) The aliasing errors due to a finite transform spectrum. This error can be reduced by considering many tones. The aliasing error is a numerically introduced error. This sets an upper limit on resolution.

(iii) The final value of the harmonic balance error. The major limiting factor here is how closely the Jacobian describes the actual error function.

Both the error function and the Jacobian have truncation error so ideally the Jacobian evaluation reflects the same truncation errors as the error function evaluation. In the end this comes down to the accuracy of the models. That is, whether the derivatives calculated in the model reflect the actual nonlinear relationship.

Naturally errors also arise due to the quality of the models of the active devices and of the linear circuit elements. Also, as the number of tones included in a harmonic balance analysis increases, the simulation time rapidly increases.

4.2.4 Periodic Steady-State Simulation of RF Circuits

Periodic steady-state (PSS) analysis is used with a Spice simulator to establish the response of a circuit to a periodic excitation signal [5]. Typically there would be one large sinusoid, such as a local oscillator or a carrier signal. The idea is that the dynamic state of the circuit is established by a transient simulation with a single sinusoidal excitation. Then a linear (or perhaps quadratic) model of the dynamic circuit is used with smaller signals. If the excitation signal is large, then the circuit is a time-varying linear circuit as far as smaller signals are concerned.

The PSS technique uses what is called the shooting method in which the simulator guesses the initial values of voltages at all terminals, charges on all capacitors, and currents through all inductors (i.e., the state-variables of the circuit) [7–13]. Then the circuit is simulated using transient analysis for one period of the exciting waveform. The state variables after one period are compared to the assumed state variables at the beginning of the period. If there is a difference, then the initial guess is changed and the process repeated. Convergence is usually achieved after a few iterations. Then the Fourier components of the state variables are found and a time-varying circuit calculated. From this, a model akin to a conversion matrix describing the dynamic circuit is established. There are many similarities to harmonic balance, as the frequencies of all signals in the circuit must be specified by the user. An advantage of the PSS technique is that a conventional Spice simulator, and the all important transistor models, can be used.

4.3 Switching Amplifiers, Classes D, E, and F

Switching amplifiers are the most efficient RF amplifiers, but are also the hardest to design. With linear amplifiers, such as Class A, B, AB, and C amplifiers, there is appreciable simultaneous voltage across the transistor and current flowing through it. Thus power is dissipated in the transistor in such an amplifier. The DC and AC **loadlines** at the transistor output of a linear amplifier essentially coincide, as is seen in the transistor output characteristic shown in Figure 4-5(a). The DC loadline is a straight line and the AC loadline closely follows the linear DC loadline; this is where the linear in linear amplifier comes from. Even when reactive effects cause the AC loadline to loop (and a small loop is seen in Figure 4-5(a)), the term linear amplifier is still used.

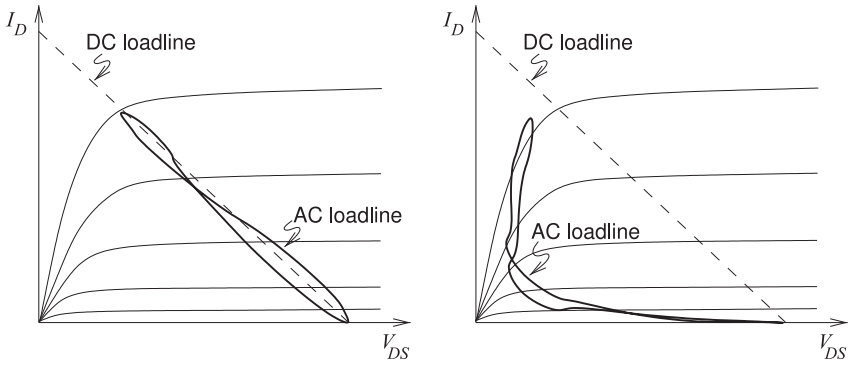
4.3.1 Dynamic Waveforms

In a switching amplifier there is little voltage across the output of the transistor when there is current flowing through it, and little voltage when there is current. This is seen in the **AC loadline**, also called **dynamic loadline**, of a switching amplifier, shown in Figure 4-5(b). This loadline is obtained through careful attention to the loading of the transistor at the harmonics.

The dynamic loadline of a switching amplifier is obtained by presenting the appropriate harmonic impedances to the transistor output. The particular scheme of harmonic termination (e.g., short or open circuits at the even and odd harmonics) leads to the designation of a switching amplifier as Classes D, E, F, etc. The key characteristic of all switching amplifiers is that when there is current through the transistor, there is negligible voltage across the output [14–20]. Also, when there is voltage across the transistor, there is little current through it (see Figure 4-5(b)).

Switching amplifiers are a conceptual departure from Class A, AB, B, and C linear amplifiers. The transistor output waveforms of a switching amplifier and those of Class A, AB, B, and C amplifiers are shown in Figure 4-6. Again it is seen, in Figure 4-6, that for a switching amplifier the voltages and current waveforms are shifted and voltage across the transistor and current through it do not occur simultaneously. The power dissipated by the transistor is the average of the product of the current through it and the voltage across the output. Thus the ideal switching amplifier consumes very little DC power, transferring nearly all of the DC power to the output RF signal. Bandpass filtering of the output of the amplifier results in a final RF output with little distortion. Switching amplifiers are the preferred amplifier in both handsets and basestations of cellular systems.

The theoretical maximum power-added efficiencies achieved by the various amplifier classes with a sinusoidal input signal are given in Table 4-2. With modulated signals, the maximum efficiencies cannot be achieved, as typically the average input power of the amplifier must be backed off by the peak-to-mean envelope power ratio (PMEPR) of the signal so that the peak carrier portion of the signal has limited distortion. Generally the acceptable distortion of the peak signal occurs at the 1 dB compression point of the amplifier. This is only an approximate guide, but useful. The PMEPRs of several digitally modulated signals are given in Table 4-3, together with their impact on efficiency. If there are two carriers, then the PMEPR of the



(a) Class A, AB, B, C linear amplifiers (b) Switching amplifier

Figure 4-5: DC and RF loadlines. An AC loadline is also called a dynamic loadline.

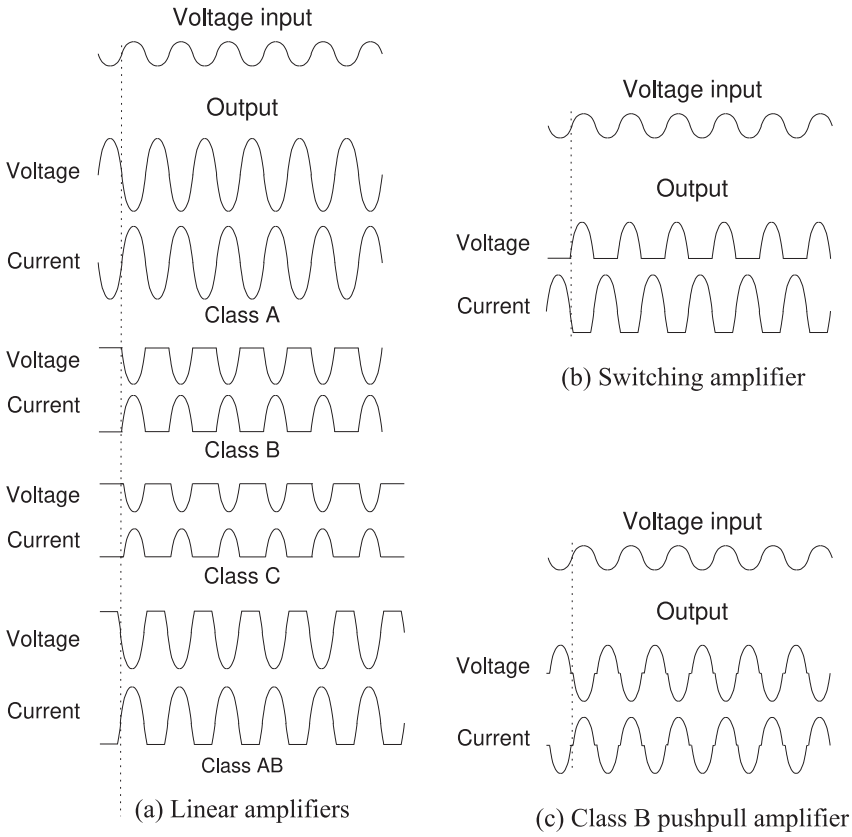


Figure 4-6: Input and output waveforms of a transistor for various classes of amplifier. The output voltage is measured across the transistor(s) and the current is through the transistor(s).

Table 4-2: Theoretical maximum efficiencies of amplifier classes with a sinusoidal excitation.

Amplifier class	Maximum efficiency
Class A (resistive bias)	25%
Class A (inductive bias)	50%
Class B	78.53%
Class C	100%
Class D	100%, but typically 75%
Class E	96%
Class F	88.36%

combined signal will be higher, requiring even greater amplifier backoff [21]. In practice, the efficiencies achieved will differ from these theoretical values. This is because the PMEPR does not fully capture the statistical nature of signals and because of coding and other technologies that can be used to reduce the PMEPR of a digital modulation scheme.

4.3.2 Conduction Angle

The conduction angle indicates the proportion of time an amplifying device, typically a transistor, is conducting current with a conduction angle of $360°$ indicating that the amplifying device is always on. The view is that a sinusoidal signal is being amplified which is appropriate since the majority of communication schemes produce an RF signal that looks like a sinewave with a very slowly-varying amplitude and phase. Class A amplifiers conduct current throughout the RF cycle and so have a conduction angle of $360°$. A Class B amplifier is biased so that only half of a sinewave is produced at the

Table 4-3: Efficiency reductions due to modulation type for a single modulated carrier. The Class A amplifier uses inductive drain biasing. The PMEPR increase is for multiple carriers. E.g. for GMSK PMEPR = 3.01 dB, 6.02 dB, 9.01 dB, 11.40 dB, 14.26 dB, and 17.39 dB for 1, 2, 4, 8, 16, and 32 carriers respectively. The PMEPR does not increase by 3 dB every time the number of carriers is doubled because statistically the envelope peaks of the individual carriers are less likely to align for more carriers.

Signal	PMEPR (dB)	Efficiency reduction factor	Class A (L bias) PAE	Class E PAE
FSK (ideal)	0	1.0	50%	96%
GMSK	3.0	0.501	25.1%	48.1%
QPSK	3.6	0.437	21.9%	42%
$\pi/4$DQPSK	3.0	0.501	25.1%	48.1%
OQPSK	3.3	0.467	23.4%	44.8%
8-PSK	3.3	0.467	23.4%	44.8%
64-QAM	7.8	0.166	8.3%	15.9%

output, so a Class B amplifier has a conduction angle is $180°$. A Class AB amplifier has a conduction angle between $180°$ and $360°$ with the bias, or one could say conduction angle, adjusted so that the distortion produced is acceptable. A Class C amplifier produces an output for less than half of the input sinewave and so it has a conduction angle of less than $180°$.

Switching amplifiers have high efficiency by ensuring that there is little current flowing through a transistor when there is voltage across it. So a switching amplifier is ideally either fully on or fully off. Ideally current flows for half the time and so the conduction angle is $180°$. However, the transistor must transition between these regions and the conduction angle indicates the degree of overlap. To minimize overlap, the conduction angle of an actual switching amplifier is less than $180°$.

4.3.3 Class D

The Class D amplifier was the first type of switching amplifier developed. The main concept of the Class D amplifier is using the transistor as a switch so that there is negligible current flowing through the transistor when there is voltage across it. The audio form of the Class D amplifier is shown in Figure 4-7(a) [14]. The two transistor inputs have the same RF signal applied but are level-shifted (but the circuit to do this has been omitted as well as the appropriate bias circuitry). Each transistor approximates Class C operation so only one transistor is switched on at a time. The current and voltage waveforms are shown in Figure 4-7(c). The transistors drive a resonant circuit with L_1 and C_1 acting as a bandpass filter. The filter reduces the distortion of the voltage and current waveforms presented to the output load, but results in the amplifier being narrowband. This Class D amplifier works very well at frequencies up to a few megahertz. Above that the transistors, having opposite polarity, are not well matched. At RF a more appropriate Class D amplifier is as shown in Figure 4-7(b) [14, 15, 19, 22] where, in this case, two nMOS transistors are used as they have higher mobility than pMOS transistors. Parasitic reactances result in substantial overlap of the current and voltage transition regions and hence there is loss of RF power in the

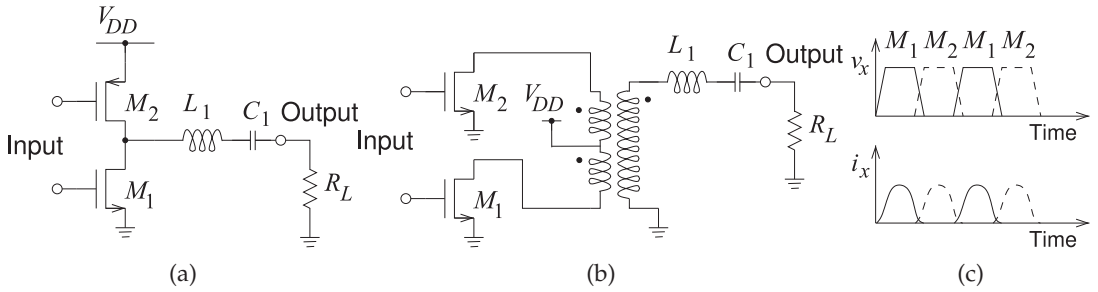


Figure 4-7: Class D amplifier: (a) low-frequency form; (b) microwave form; and (c) current and voltage waveforms (where v_x is the drain-source voltage and i_x is the drain current) indicating which transistor is turned on.

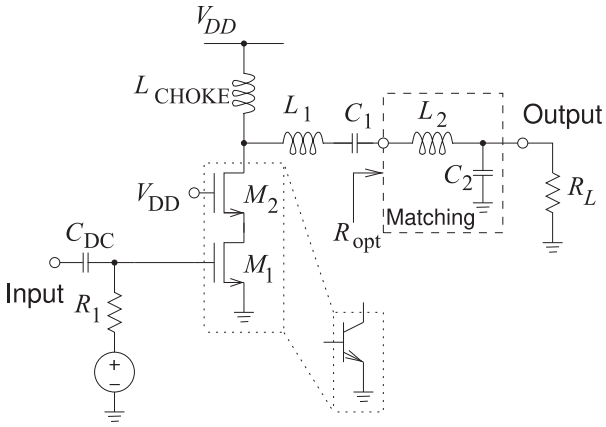


Figure 4-8: Class E amplifier.

transistors. This loss can be reduced using an alternative form of the Class D amplifier, called a current-mode Class D amplifier, which switches current rather than voltage [23]. Efficiencies of around 75% are achieved.

4.3.4 Class E

The Class E [15, 17] amplifier builds on the Class D amplifier concept of using a transistor as a switch rather than as a linear amplifier. An RF Class E amplifier is shown in Figure 4-8. The circuit shown uses two MOSFETS with M_1 being the switching transistor and M_2 acting in part as an active load. The main function of M_2 is to translate the drain current of M_1 into a voltage at the output. Bias is provided through L_{CHOKE}, which presents a high RF impedance. L_1 and C_1 provide a bandpass filtering function, while L_2 and C_2 provide matching to the load so that the impedance looking into the L_2, C_2 and R_L network is an optimum resistance, R_{opt}. The circuit is designed so that L_{CHOKE}, the parasitic output capacitance of the transistors, L_1, C_1, and R_{opt} form a damped oscillating circuit. The two MOSFETs can also be replaced by a single transistor, typically an HBT transistor, as shown by the insert in Figure 4-8, and sometimes a capacitor is added from the top of the transistor to ground if the parasitic transistor capacitance is insufficient. So at the output of the transistors there is a parallel LC circuit to ground

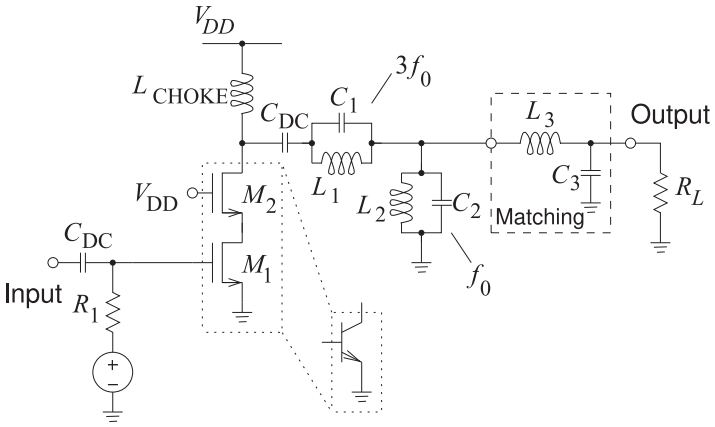


Figure 4-9: Class F amplifier.

(L_{CHOKE} and the parasitic capacitance of M_2) and a series LC circuit (L_1 and C_1). When the transistors are switched off (and have a high voltage across them), the series LC circuit provides current to the parallel LC circuit as well as to R_{opt}, instead of current being drawn through the transistors. When the transistors are switched on (and have little voltage across them), the oscillation proceeds in the opposite direction and current is delivered to R_{opt} through the transistors. This is the oscillation mechanism with the resistance R_{opt} providing damping. Design matches the natural oscillation frequency with the frequency of the RF signal.

4.3.5 Class F

In the Class E amplifier the voltage at the drain/collector of the transistor is approximately a square wave and the transistor current approximates a half sinusoid. The Class F amplifier takes this one step further and realizes an approximate square current wave through the transistors as well as an out-of-phase square voltage wave [14, 15, 19]. This is achieved using harmonic resonance.

A Class F amplifier is shown in Figure 4-9, where the narrowband RF signal has a center frequency f_0. Again M_1 operates as a switch producing a square voltage wave at the output of the transistors. The parallel L_1C_1 circuit is tuned to the third harmonic of the RF signal (i.e., $3f_0$), and this parallel circuit provides third-harmonic current when the transistors are turned off. The L_1C_1 circuit is an approximate short circuit at f_0, and the parallel L_2C_2 circuit, tuned to be resonant at f_0, presents an open circuit at f_0 and short circuits at the harmonics. The net result is that the current waveform passing through the transistors is a reasonable square wave with first and third harmonics and no second harmonic (note that a square wave consists of only odd frequency components). Also note that third-harmonic current does not pass through the load. This concept could be continued to provide similar behavior at the fifth harmonic as well, but it then becomes increasingly difficult to adjust the design to work as intended. With both the voltage and current waves being square, the lower the overlap, and the lower the power dissipated in the transistors.

4.3.6 Inverted Amplifiers

The Class D and F switching amplifiers described above are designed to switch the voltage between two states. The dual of these are the **inverted Class D amplifier** [20, 23–25] and the **inverted Class F amplifier** [26–29]. The design intent is that the transistors in these amplifiers switch the current rather than the voltage. They are also called **current-mode amplifiers**.

4.3.7 Summary

The efficiency advantages of switching amplifiers are significant and often justifies the higher design cost. They are used in power amplifiers in most basestations and are starting to be used in cellphones. At high microwave and millimeter-wave frequencies stability concerns and the high cost of design means that many amplifiers will continue to be linear amplifiers for some time.

4.4 Distortion and Digitally Modulated Signals

Digitally-modulated signals do not have constant envelopes so that the short-term peak power of the RF signal is higher than the average RF power. Efficient amplifiers must introduce minimal distortion and be efficient for peak signal levels as well as for the average signal level. There is a trade-off between distortion and efficiency and various amplifier architectures have been developed to implement good trade-offs. These architectures will be introduced in a later section. This section introduces approaches for describing distortion.

4.4.1 PMEPR and Probability Density Function

Amplitude variations occur with most digital modulation schemes. For example, a QPSK signal consists of two digital data streams, equal in amplitude, modulated in quadrature onto a carrier signal. If the data streams are not filtered prior to modulation, then the resulting modulated carrier signal has a constant envelope with instantaneous transitions from one constellation point to another. However, the occupied bandwidth of this signal is quite large, as the spectrum is that of a pulse train, $\sin(x)/x$, the sinc function. The first **sidelobe** of the sinc spectrum is only 13 dB down from the carrier level and typically is in the middle of the adjacent channel. To reduce the bandwidth of the modulated signal, a low-pass filter is applied to each digital data stream to minimize the out-of-band spectrum of the modulated signal. This comes with a drawback: the filters cause a finite memory effect resulting in amplitude variations as the ringing energy from a previous data pulse adds to the current filtered data pulse. As well, the transitions from one constellation point to another are slow and the amplitude of the modulated carrier varies significantly during the transition.

Amplitude variations of the modulated signal are characterized by waveform statistics such as the PMEPR. A signal with a high PMEPR requires that the RF system have high linearity to handle both the average power requirements and the peak amplitude excursions without generating excessive out-of-band distortion. A simple design technique is to reduce the average output power of an amplifier to a level that is below the 1 dB

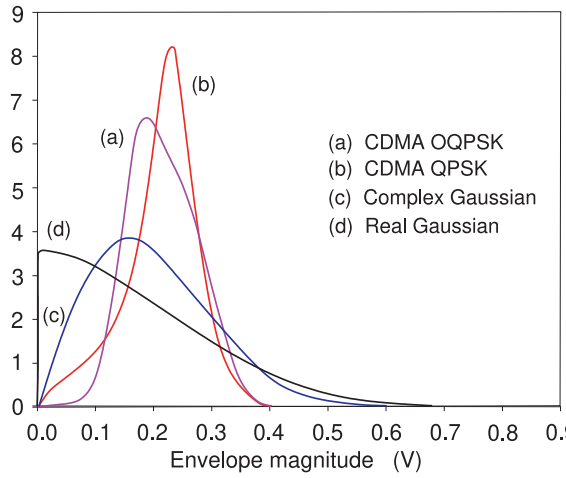


Figure 4-10: Amplitude PDF for CDMA and Gaussian modulated signals with the same power of 0 dBm. After [31].

Signal Modulation	PMEPR (dB)
CDMA OQPSK	5.4
CDMA QPSK	6.6
Real Gaussian	13.5
Complex Gaussian	11.8

Table 4-4: Peak-to-average ratios in decibels for OQPSK and QPSK used in CDMA compared to the PMEPRs of Gaussian signals.

gain compression power by the PMEPR, this is called **amplifier back-off**. However, it is possible for a signal with a high PMEPR to exhibit less nonlinear distortion than a signal with lower PMEPR [30]. The reason for this apparent inconsistency is because the signal peak may have a low probability of occurrence. Thus PMEPR by itself is an incomplete statistic for determining the linearity requirements of an amplifier.

The **amplitude probability density function** (APDF) is a more complete statistical description of the amplitude variations of a modulated signal. The APDF defines the maximum and minimum variation along with the relative probability of occurrence of amplitudes within the variation. The APDF is typically estimated from a histogram of envelope amplitudes, with a uniform bin size, by

$$f(A) = \frac{N}{\Delta A \times N_c}, \tag{4.15}$$

where N is the number of counts in the bin having amplitude A, ΔA is the bin width (i.e., the range of amplitudes in the bin), and N_c is the total number of samples. The shape of the amplitude density determines the sensitivity of a particular signal to spectral regrowth due to nonlinear gain compression or expansion. This is a more precise determinant of distortion than the single PMEPR metric.

As an example, consider Figure 4-10 which shows the APDF of a CDMA signal using OQPSK modulation, of the same signal but now using QPSK modulation of a real **Gaussian signal** and of a complex Gaussian QPSK signal (with I and Q each having Gaussian distributions) where the average power of each signal is set to 0 dBm. Gaussian signals are of particular interest because their simple statistics lend them, and their interaction with nonlinear circuits, to quasi-analytic treatment [32]. The PMEPR of each signal is given in Table 4-4. From the shape of the APDF it is possible to estimate which signal will be more sensitive to nonlinear gain compression.

For example, even though QPSK has a higher PMEPR than OQPSK, the probability that the OQPSK envelope is near the peak is higher than for the QPSK signal. This is seen in Figure 4-10. It is not surprising then that the measured spectral regrowth of an OQPSK signal is higher than that for a QPSK signal. So PMEPR is only a rough guide to the distortion that is produced.

4.4.2 Design Guidelines

Generally an amplifier is not operated in saturation nor near the **third-order intercept** (**TOI** or IP3) point. An exception is with constant or near-constant envelope modulation schemes such as FM, GMSK, SOQPSK, FOQPSK, and SBPSK. With little variation in the amplitude of the RF signal, saturating amplifiers can be used with the near-constant envelope schemes. Also some advanced amplifier technologies can operate with nonlinear transistor loadlines and still have low distortion. For example, with some switching amplifiers produce little intermodulation distortion but a lot of harmonic distortion which is easily filtered out. With nonconstant envelope modulation schemes (i.e., the signal PMEPR is more than 0 dB), the variation in the amplitude of the modulated RF carrier results in signal distortion and spectral regrowth (i.e., power will be transferred into neighboring communication channels). One general rule of design is to ensure that the peak of the RF pseudo-carrier is at or below the 1 dB gain compression point. The amplifier is backed off by an amount approximately PMEPR below the 1 dB gain compression point. Generally, however, third-order intermodulation is a much greater concern, as this relates more closely to the amount of power dumped into neighboring channels.

Modern communication schemes can require that the adjacent channel spectral regrowth be as much as 80 dB below the power of the main channel. As a result, the signal must be backed off considerably from the IP3 point. The input or output power at IP3 (IIP3 or OIP3) is obtained from two-tone characterization and thus is a weakly accurate indicator of distortion with a digitally modulated signal. However, since it is a simple measurement to make and understand, it is widely used. Experience provides a rule of thumb for the backoff needed to ensure a maximum level of spectral regrowth for a particular modulation format and transistor technology [21].

Since the complexity of designing high-power and highly efficient amplifiers is considerable, it is necessary to use measurements following design to optimize the system for high efficiency while ensuring acceptable distortion. The most common measurement approach is to use loadpull, described in the next section. Design of the baseband circuit can also affect intermodulation distortion and spectral regrowth [33–35]. It is very important that the transistor ground be the system ground and that there be good heat-sinking so that the thermal circuit does not contribute to spectral regrowth through the variation of thermally dependent transistor parameters as the signal envelope varies (and thus the heat generated varies).

While IP3 and the 1 dB compression point only refer to amplitude distortion, there will also be phase distortion. While amplitude distortion has no effect on constant amplitude signals, phase distortion does. Phase distortion affects the performance of constant envelope modulation schemes

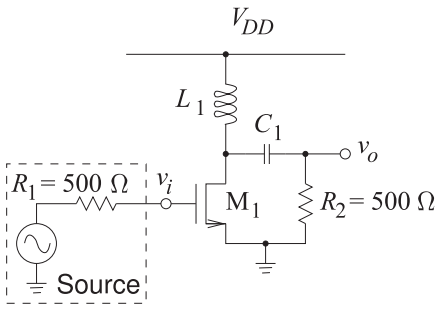


Figure 4-11: Class A inductively-biased MOSFET amplifier.

and the 1 dB compression point and IP3 provide some indication of possible phase distortion. Although EVM and BER are much better metrics of distortion for all digital modulation schemes, these are difficult to use in guiding initial design of RF hardware.

EXAMPLE 4.1 Intercept Point

The dynamic range derivations in Section 4.6 of [2] should be reviewed before working through this example.

The amplifier shown in Figure 4-11 is modeled, after expansion around the operating point, by a nonlinear transconductance with $i_{DS} = a_1 v_{GS} + a_3 v_{GS}^3$ with $a_1 = 0.01$ A/V and $a_3 = -0.1$ A/V^3. L_1 is an RF choke and can be assumed to present an open circuit at the operating frequency. C_1 is a biasing capacitor and can be treated as a short circuit at the operating frequency. R_2 is the load resistance.

(a) What is the small signal voltage gain of the amplifier?
(b) If the input signal is a two-tone signal consisting of two equal-amplitude sinewaves at 900 MHz and 901 MHz, what are the frequencies in the spectrum of v_o? Determine the output voltage across R_2?
(c) If a single sinusoid of amplitude 100 mV is applied to the gate of the MOSFET, what is the level of the fundamental tone at the output of the amplifier? What is the level of the third harmonic?
(d) What is the input-referred third-order intercept point, IIP3, of the amplifier?
(e) What is the output-referred third-order intercept point, OIP3, of the amplifier?

Solution:

(a) When the input signal v_i is small, $i_{DS} = a_1 v_i$, so the small signal output voltage is

$$v_o = -i_{DS} R_2 = -a_1 R_2 v_i$$

and the small signal voltage gain is

$$A = \frac{v_o}{v_i} = -a_1 R_2 = -0.01 \times 500 = -5.$$

(b) Eventually the symbolic amplitudes of the mixing terms will be required, so a trigonometric expansion is undertaken now. To minimize complexity, consider two cosinsoids, $A\cos(x)$ and $B\cos(y)$. So the input to the amplifier is

$$v_{GS} = A\cos(x) + B\cos(y).$$

The output of the amplifier is

$$
\begin{aligned}
v_o &= -i_{DS}R_2 = -R_2\left(a_1 v_{GS} + a_3 v_{GS}^3\right) \\
&= -R_2\left\{a_1\left[A\cos(x) + B\cos(y)\right] + a_3\left[A\cos(x) + B\cos(y)\right]^3\right\} \\
&= -R_2\left\{a_1\left[A\cos(x) + a_1 B\cos(y)\right] + a_3\left[A\cos(x) + B\cos(y)\right]\left[A\cos(x) + B\cos(y)\right]^2\right\} \\
&= -R_2\left\{a_1\left[A\cos(x) + a_1 B\cos(y)\right]\right. \\
&\qquad \left. + a_3\left[A\cos(x) + B\cos(y)\right]\left[A^2\cos^2(x) + 2AB\cos(x)\cos(y) + B^2\cos^2(y)\right]\right\} \\
&= -R_2\left\{a_1 A\cos(x) + a_1 B\cos(y)\right. \\
&\qquad + a_3\left[A\cos(x) + B\cos(y)\right] \\
&\qquad \left. \times \tfrac{1}{2}\left[A^2 + B^2 + A^2\cos(2x) + B^2\cos(2y) + 2AB\cos(x-y) + 2AB\cos(x+y)\right]\right\} \\
&= -R_2\left\{a_1\left[A\cos(x) + B\cos(y)\right]\right. \\
&\qquad + a_3\tfrac{1}{2}\left[\left(A^3 + AB^2\right)\cos(x) + A^3\cos(x)\cos(2x) + AB^2\cos(x)\cos(2y)\right. \\
&\qquad + 2A^2 B\cos(x)\cos(x-y) + 2AB^2\cos(x)\cos(x+y) + \left(A^2 B + B^3\right)\cos(y) \\
&\qquad + B^3\cos(y)\cos(2y) + A^2 B\cos(y)\cos(2x) + 2AB^2\cos(y)\cos(x-y) \\
&\qquad \left.\left. + \ 2AB^2\cos(y)\cos(x+y)\right]\right\} \\
&= -R_2\left(a_1\left[A\cos(x) + a_1 B\cos(y)\right]\right. \\
&\quad + a_3\tfrac{1}{2}\left\{\left(A^3 + AB^2\right)\cos(x) + \tfrac{1}{2}A^3\left[\cos(x) + \cos(3x)\right]\right. \\
&\qquad + \tfrac{1}{2}AB^2\left[\cos(2y-x) + \cos(x+2y)\right] + A^2 B\left[\cos(y) + \cos(2x-y)\right] \\
&\qquad + A^2 B\left[\cos(y) + \cos(2x+y)\right] + \left(A^2 B + B^3\right)\cos(y) \\
&\qquad + \tfrac{1}{2}B^3\left[\cos(y) + \cos(3y)\right] + \tfrac{1}{2}A^2 B\left[\cos(2x-y) + \cos(2x+y)\right] \\
&\qquad \left.\left. + AB^2\left[\cos(x) + \cos(2y-x)\right] + AB^2\left[\cos(x) + \cos(2y+x)\right]\right\}\right) \\
&= -R_2\left\{a_1 A\cos(x) + a_1 B\cos(y) \quad + a_3\tfrac{1}{4}\left[\left(3A^3 + 6AB^2\right)\cos(x)\right.\right. \\
&\qquad + \left(3B^3 + 6AB^2\right)\cos(y) + B^3\cos(3y) + 3A^2 B\cos(2x-y) \\
&\qquad \left.\left. + 3A^2 B\cos(2x+y) + A^3\cos(3x) + 3AB^2\cos(2y-x) + 3AB^2\cos(2y+x)\right]\right\}.
\end{aligned}
\tag{4.16}
$$

So with x representing 900 MHz and y representing 901 MHz, the frequencies in the output spectrum are 899, 900, 901, 902, 2700, 2701, 2702, and 2703 MHz.

(c) From Equation (4.16) and considering only one tone (i.e., $B = 0$), the output signal is

$$
v_o = -R_2\left[a_1 A\cos(x) + \frac{3a_3 A^3}{4}\cos(x) + \frac{a_3 A^3}{4}\cos(3x)\right]. \tag{4.17}
$$

So the coefficient of the fundamental at the output is

$$
v_o = -R_2\left[a_1 A\cos(x) + \frac{3a_3 A^3}{2}\cos(x))\right]. \tag{4.18}
$$

Here $A = 100$ mV, so the

$$
\begin{aligned}
\text{fundamental output} \quad &= -(500\ \Omega)\cdot\left[(0.01\ \text{A/V})\cdot(0.5\ \text{V}) + (-0.1\ \text{A/V}^3)\cdot(0.1\ \text{V})^3\right] \\
&= 0.05 - 0.0125\ \text{V} = 0.0375\ \text{V} = 37.5\ \text{mV}
\end{aligned}
\tag{4.19}
$$

$$
\text{third harmonic output} = \frac{a_3 R_2}{4}v_i^3 = (-500)\cdot(-0.1)\times(0.1)^3/4 = 0.0125\ \text{V}. \tag{4.20}
$$

(d) To answer this, determine the level of the fundamental and IM3 outputs for small v_{GS} and for two input tones having the same amplitude, $A = B = v_{GS}$. For a resistive nonlinearity, such as the transconductance here, the level of the lower and upper third-order intermods are the same. So, after examining Equation (4.16), consider $\cos(x)$ and $\cos(2x - y)$. The fundamental (at $\cos(x)$) is

$$v_o(900\text{ MHz}) = -R_2 a_1 v_i$$

and the level of the lower third-order intermod at $(2x - y)$ is

$$v_o(899\text{ MHz}) = -R_2 a_3 \tfrac{3}{4} (v_i)^3 .$$

IIP3 is the value of v_i when $v_o(900\text{ MHz}) = v_o(899\text{ MHz})$, that is, when

$$-R_2 a_1 v_i = -R_2 a_3 \tfrac{3}{4} (v_i)^3 .$$

That is, when $v_i = \sqrt{\left|\dfrac{4a_1}{3a_3}\right|} = \sqrt{\dfrac{4 \cdot 0.01}{3 \cdot 0.1}} = 0.3651\text{ V} = 365.1\text{ mV}.$

Thus $A_{\text{IIP3}} = 365.1\text{ mV}.$

Normally IIP3 is expressed in terms of power. Considering Figure 4-11, $E = 2v_i$, and so the available input power is

$$P_{\text{av}} = \frac{1}{2}\frac{v_i^2}{R_1} = \frac{v_i^2}{2R_1}.$$

Thus $\text{IIP3} = \frac{1}{2R_1} A_{\text{IIP3}}^2 = \dfrac{0.3651^2}{2 \cdot 500} = 0.0001333\text{ W} = 0.1333\text{ mW} = -0.875\text{ dBm}.$

(e) The voltage output-referred intercept point is

$$A_{\text{OIP3}} = |A|\, A_{\text{IIP3}} = 5 \cdot 0.3651\text{ V} = 1.8255\text{ V}$$

and $\text{OIP3} = (\text{Power gain}) \cdot \text{IIP3} = \dfrac{R_1}{R_2} A^2 \cdot 0.1333\text{ mW} = 3.333\text{ mW} = 5.23\text{ dBm}.$

4.5 Loadpull

Power amplifiers are designed beginning with an initial concept, detail designed using nonlinear circuit simulation tools, and finally optimized in the laboratory. However, computer-assisted design relies on models of components that can never capture all effects including, for example, thermal effects and parasitic coupling of components. A final experimental optimization for gain and efficiency while limiting distortion is nearly always required [36–38]. Once the design has been optimized in the laboratory, it is found that a design can be fixed for manufacturing. Minimal manual tuning of individual production units to peak performance is then usually all that is required. The most useful experimentally based design optimization approach uses the load-pull method. This technique uses computer-controlled variable input and output matching networks to search for the optimum input and output conditions. As well, performance with digitally modulated signals can be investigated, and this is very difficult to do in simulation.

A load-pull system is shown in Figure 4-12. The automated tuners realize variable input and output networks and are generally based on automated double-stub slabline tuners where the position of the stubs is also variable. The output impedance of the active device in a power amplifier is typically around a few ohms or less and it is often necessary to use an impedance transformer to scale the output impedance of the device up to the typically $50\ \Omega$ system impedance of the automated tuners.

The load-pull system shown in Figure 4-12 is configured to measure the reflected power at the input of the amplifier (through the input

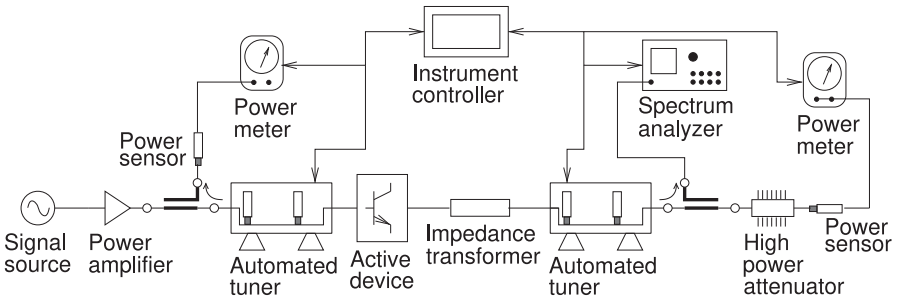


Figure 4-12: Load-pull system.

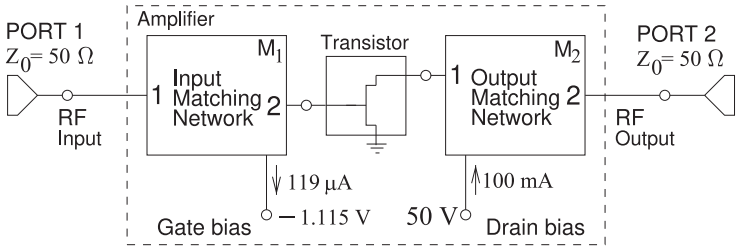


Figure 4-13: WiMAX amplifier block diagram showing the DC gate and drain currents.

directional coupler), the spectrum at the output of the amplifier (again through a directional coupler), and the output power (following a high-power attenuator). From the spectrum and power measurements, the gain, efficiency, output power, EVM, and spectral regrowth metrics are found. Many of these factors can be optimized [36] by systematically presenting impedances to the active device from a grid of possible input and output impedances. The tuners shown in Figure 4-12 provide controlled matching networks at the fundamental frequency. More elaborate systems can provide separate tuning at several harmonics. Microwave computer-aided design tools also support load-pull calculations for amplifiers.

The input automated tuner, the active device, the impedance transformer, and the output automated tuner become the amplifier. Once the optimum settings of the automated tuners are found for a range of frequencies, the design task then becomes realizing matching network with S parameters that correspond to the tuner settings.

4.6 Case Study: Design of a WiMAX Power Amplifier

In this section the design and simulation of a 10 W power amplifier module using a GaN HEMT transistor, here a Eudyna EGN010MK device, is examined.[1] A GaN (gallium nitride) transistor can have a very large drain-source voltage and can produce substantial output power. The amplifier schematic is shown in Figure 4-13. The amplifier is targeted for use in a WiMAX data communication system operating from 3.4 to 3.8 GHz.

4.6.1 Input and Output Matching Networks

The harmonic balance simulation employed in design uses a large-signal transistor model supplied by the manufacturer of the transistor and, as is usual, it is proprietary and cannot be viewed in the simulator. A

[1] AWR Design Environment Project File: WiMAX_amp.emp.

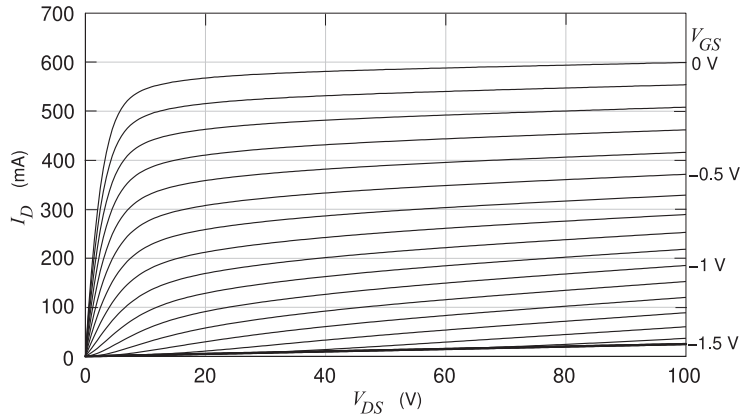
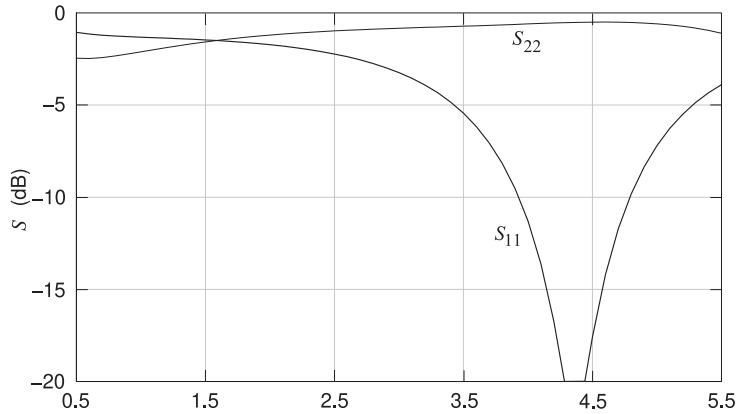
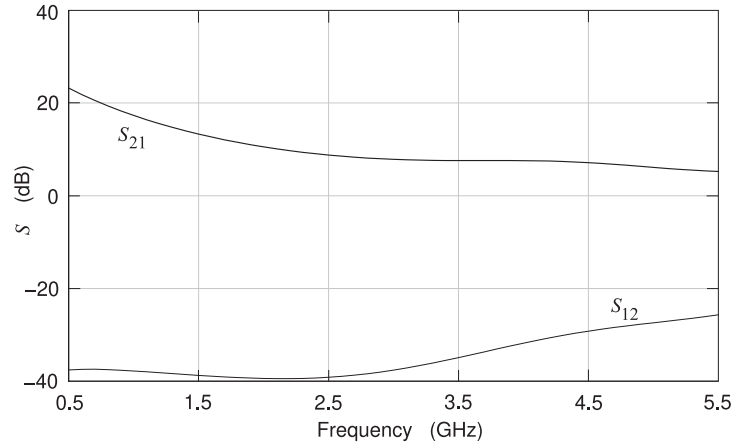


Figure 4-14: Output current-voltage characteristics of the active device in the WiMAX amplifier.



(a) S_{11} and S_{22}



(b) S_{21} and S_{12}

Figure 4-15: Small signal scattering parameters of the active device of the WiMAX amplifier.

designer must initially work with the transistor's output current-voltage characteristic, as shown in Figure 4-14. The transistor is biased with a gate-source voltage of -1.115 V, a drain-source voltage of 50 V, and a drain current of 100 mA. At this bias point the input and output 50 Ω S parameters of the transistor are as shown in Figure 4-15. Examination of S_{11} in Figure 4-15(a) indicates that the input impedance of the GaN transistor is approximately

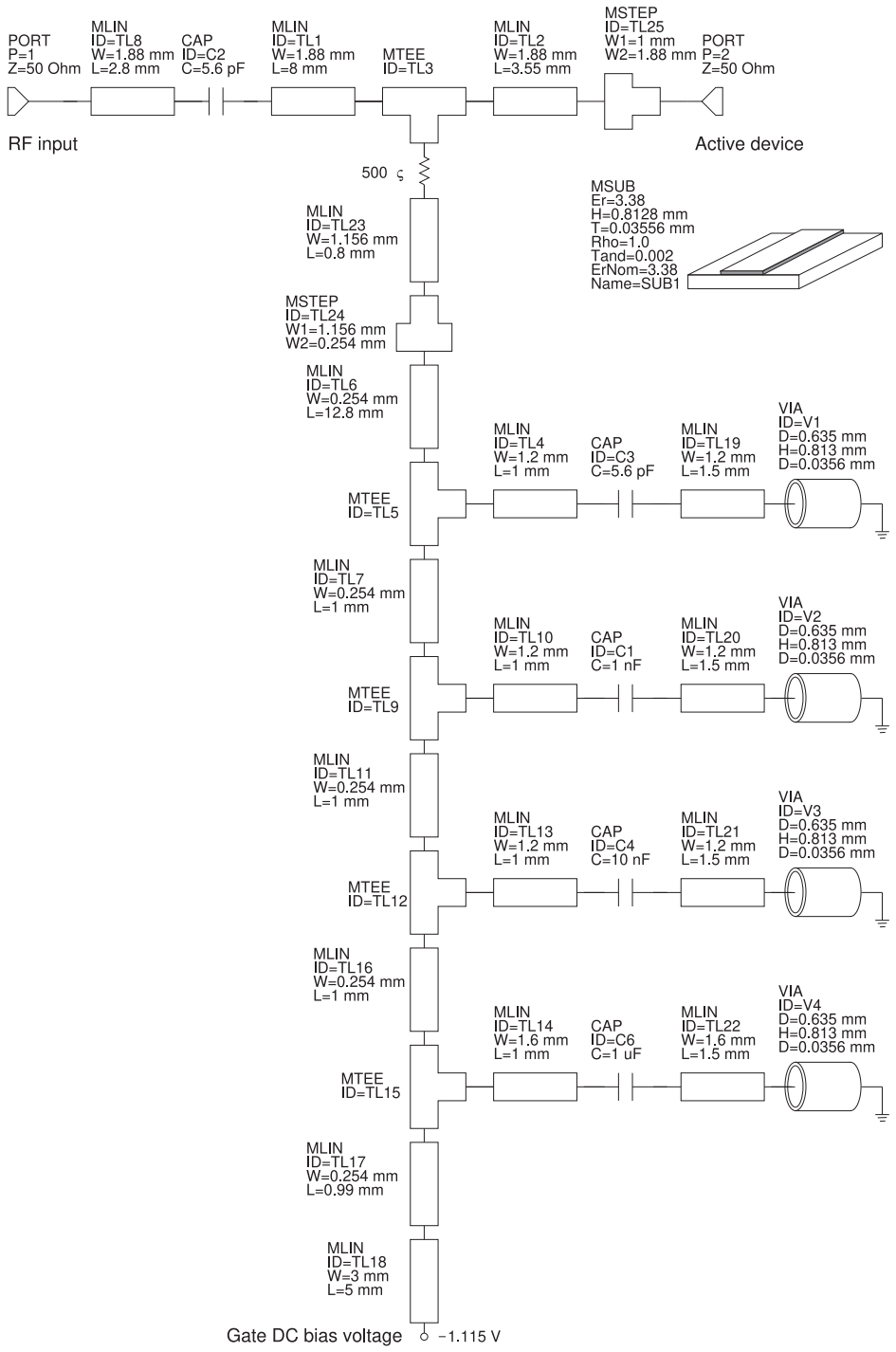


Figure 4-16: Input matching network of the WiMAX amplifier.

50 Ω over the WiMAX frequency range of 3.4–3.8 GHz. As a result, very little input matching is required. The circuit model of the input network is shown in Figure 4-16. The input network is shown with 50 Ω ports at Ports 1 and 2. These are used in evaluating the S parameters of the network but are removed when the network is incorporated in the complete amplifier shown

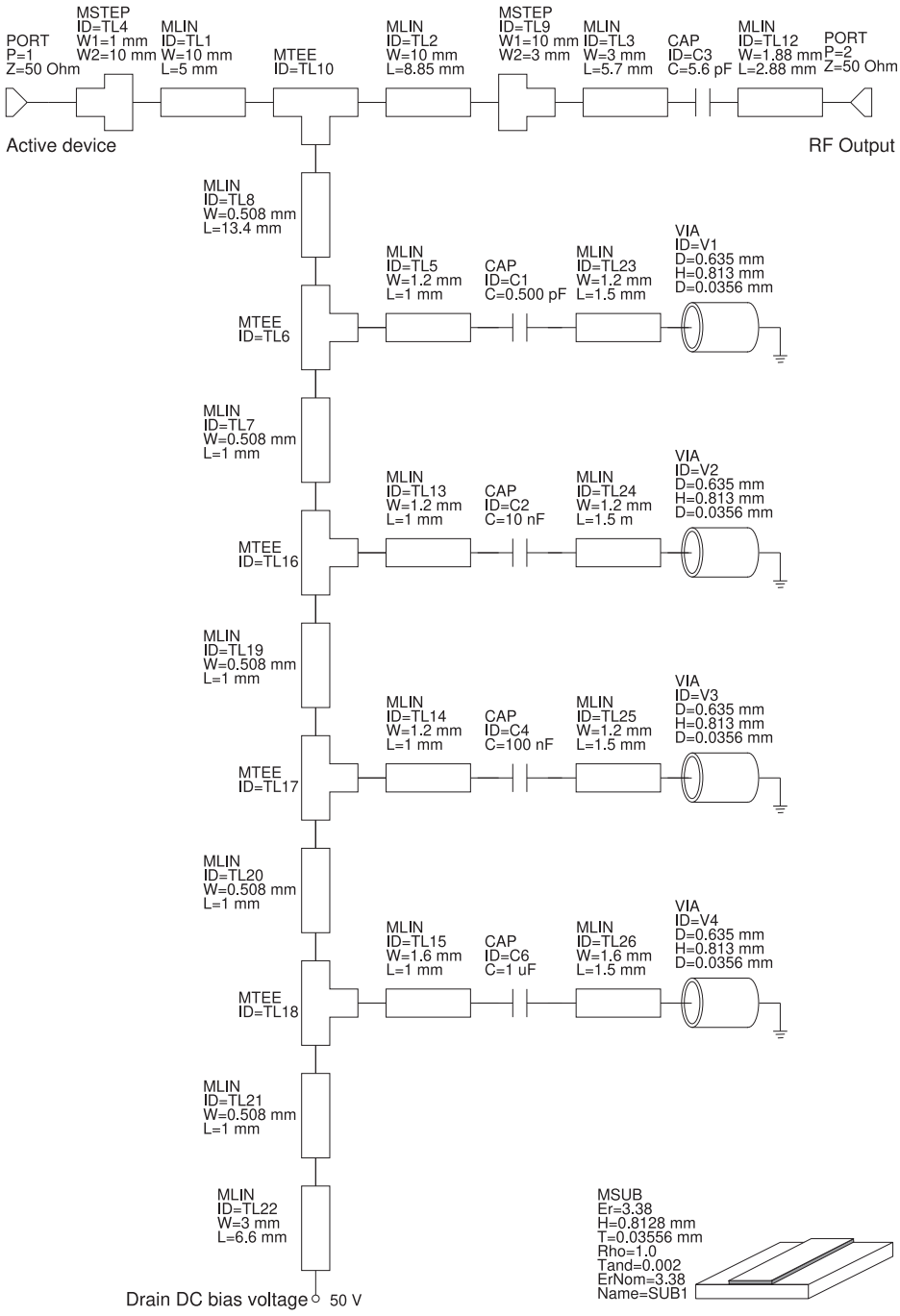


Figure 4-17: Output matching network of the WiMAX amplifier.

in Figure 4-13.

Examination of the output reflection coefficient of the transistor indicates that it can be approximately modeled by a current source in parallel with a 33 Ω resistor and a 4.4 pF capacitor (an impedance of $2.5 - j11\ \Omega$ at 3.6 GHz) and so output matching is required to transition to 50 Ω. The circuit model of the output network is shown in Figure 4-17. The main component of the output matching network is an approximate one-quarter wavelength long

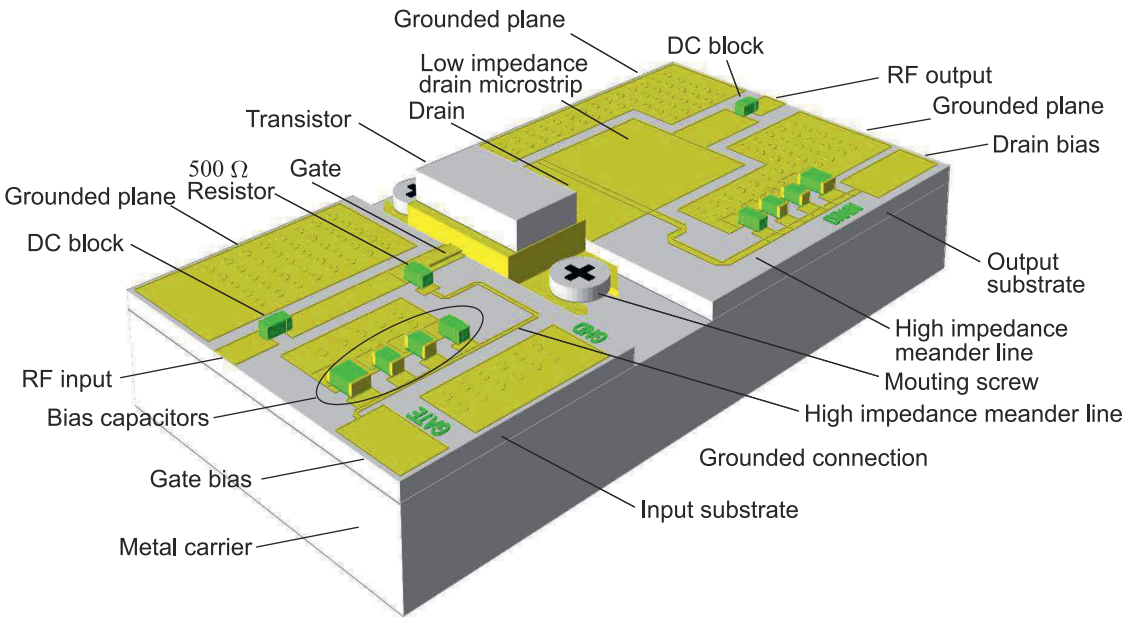


Figure 4-18: WiMAX amplifier operating from 3.4 to 3.8 GHz.

impedance transformer implemented by the microstrip transmission lines (MLIN) TL1 and TL2. This is the basis of the output network design and other components complete the network.

The complete power amplifier module is shown in Figure 4-18. A major consideration in the mechanical assembly is management of the heat produced by the active device. The semiconductor is valence-bonded by the transistor foundry to a metal plate which is secured to the metal carrier using mounting screws and thermal paste between the metal carrier and the metal backing of the transistor. The metal carrier is typically mounted on a heat sink.

The input and output networks are on separate substrates, as this enables the transistor to be mounted directly on the metal carrier. The input substrate supports a microstrip circuit with a 50 Ω RF input connected through a DC blocking capacitor to a microstrip line that drives the gate. Near the transistor, the gate microstrip line is connected through a 500 Ω resistor to a DC bias network. The 500 Ω resistor provides part of the RF isolation between the gate line and the gate bias network. The gate current drawn by the transistor is 11 μA so that there is a 0.06 V DC voltage drop across the 500 Ω resistor. From the 500 Ω isolation resistor, a high-impedance (i.e., narrow) microstrip meander line continues to a sequence of RF shorting capacitors of increasing capacitance ranging from 5.6 pF next to the meander line to 1 μF at the gate bias pad from which wirebonds lead to an external source. This arrangement minimizes the impact of the increasing parasitic series inductances of the capacitors as their capacitance values increase. Together, the 500 Ω resistor and the high-impedance line provide significant RF isolation from the external bias network. Note that

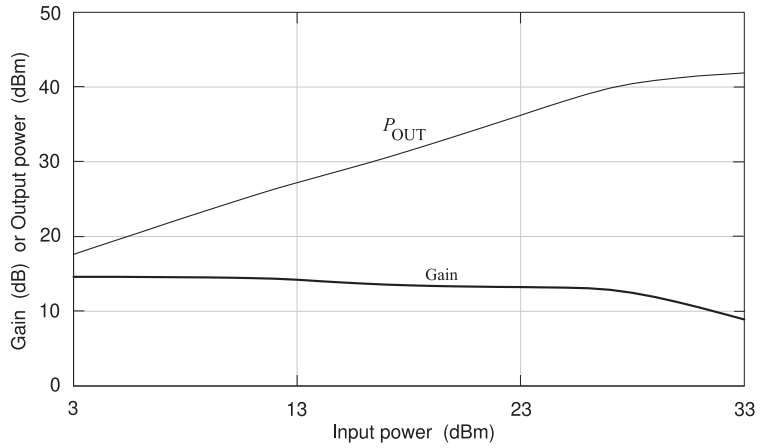


Figure 4-19: WiMAX amplifier simulated output power, P_{OUT}, and gain at 3.5 GHz.

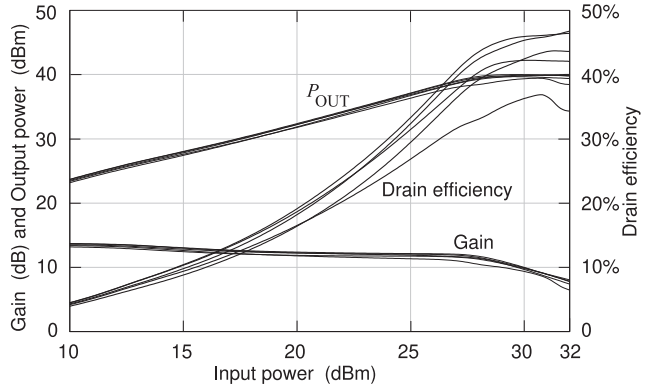


Figure 4-20: WiMAX amplifier simulated output power, gain, and drain efficiency at 3.4, 3.5, 3.6, 3.7, and 3.8 GHz.

areas of the input network without microstrip elements are grounded planes. These are regularly shorted by vias to the metal backing of the microstrip substrate, which is in turn connected to the metal carrier. This arrangement suppresses substrate modes.

The output network is similarly arranged, but now the bias resistance is omitted, as the voltage drop across it would be too great since the DC drain bias current is 100 mA. The output impedance of the active device two-port is very low, so the first microstrip line is wide and so has low characteristic impedance. Now RF isolation from the drain bias network is provided by the high-impedance meander line and there is a high-impedance contrast to the low-impedance drain line.

The nonlinear performance of the amplifier is characterized using one-tone and two-tone tests. The one-tone test results are shown in Figure 4-19, where the input power of a single input tone, a sinewave, at 3.5 GHz is swept from 3 dBm to 33 dBm. The output power at first increases linearly but then begins to saturate and the gain of the device drops. The small-signal gain is 13.3 dB and the output power at 1 dB gain compression is 28.2 dBm while the gain of the amplifier is 12.3 dB. While the gain and output power vary slightly over the WiMAX frequency range of 3.4 to 3.8 GHz (see Figure 4-20), the drain efficiency is more sensitive.

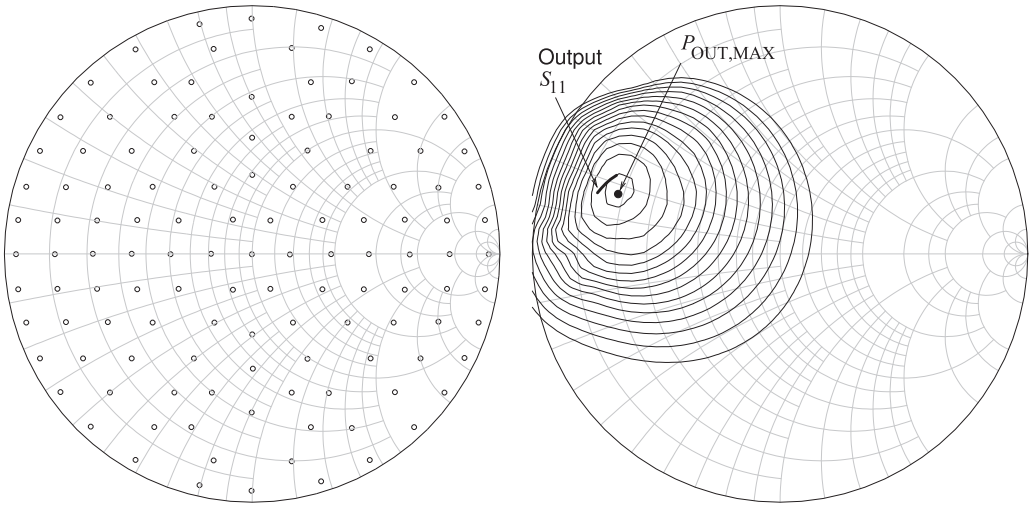


Figure 4-21: Simulated output load-pull characterization of the WiMAX amplifier: (a) load-pull points used; and (b) load-pull 0.1 dB contours of the amplifier at 3.5 GHz showing the locus of the reflection coefficient looking from the transistor into the output matching network as contours of constant output power. The family of contours begins with the point of maximum output power, $P_{\mathrm{OUT,MAX}}$ = 42.6 dBm. The output power for the first contour surrounding $P_{\mathrm{OUT,MAX}}$ is 42.5 dBm and the powers of the contours then reduce in 0.5 dB steps. Also shown is the S_{11} of the final output matching network design from 3.4 to 3.8 GHz. The 'Output S_{11}' is the reflection coefficient from 3.4 to 3.8 GHz of the output matching network looking from the active device indicating the simulated output power is between 42.4 dB and 42.5 dB (approximately). (A complete determination would require load-pull contours at several frequencies.)

4.6.2 Load-Pull

Optimum design of a power amplifier requires load-pull characterization in which, often, the input matching network is already designed, and in simulation a special load-pull element is used for the output matching network. This element can present a range of matching transformations between the output of the active device and the output load. An array of transformations is specified in terms of reflection coefficients such as that shown in Figure 4-21(a). At each of the points, the output power, P_{OUT}, is calculated and contours of constant output power are interpolated from these data, yielding the load-pull contours shown in Figure 4-21(b). The frequency locus (from 3.4 to 3.8 GHz) of S_{11} of the output matching network (looking from the active device) is also shown in Figure 4-21(b). The design choice made for the output matching network is for its S_{11} to be tangential to one of the load-pull contours so that the output power is constant across the frequency band. So rather than designing for maximum power transfer, which practically could only be achieved at one frequency, the design choice is to achieve flat gain across the band. Other quantities can also be plotted in the load-pull test. These include efficiency, the level of harmonics, and distortion terms.

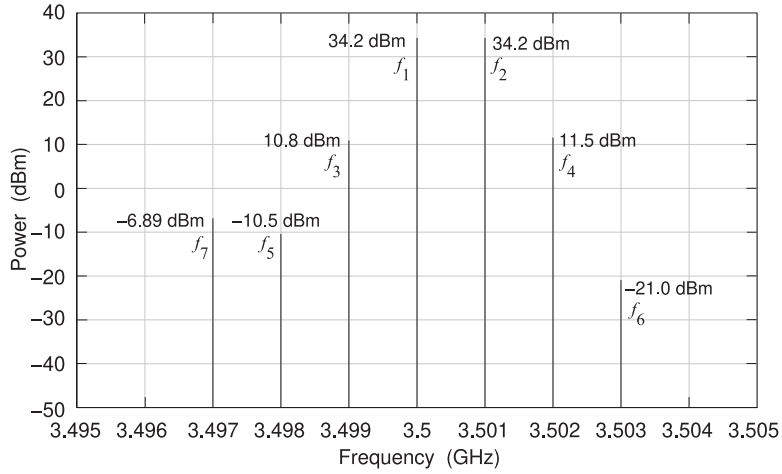


Figure 4-22: Simulated output spectrum with a two-tone input signal comprising tones at 3.500 GHz and 3.501 GHz each being 23 dBm.

4.6.3 Two-Tone Characterization

The second test used to quantify the nonlinear performance of a power amplifier is the two-tone test in which two closely spaced tones of equal power are applied at the input of the amplifier. The output spectrum is then calculated and tabulated for a range of input powers. The output spectrum under one condition is shown in Figure 4-22. The output spectrum consists of amplified versions of the two input tones at $f_1 = 3.500$ GHz and $f_2 = 3.501$ GHz as well as intermodulation products separated by multiples of the input tone spacing. The tones at $f_3 = 3.499$ GHz and $f_4 = 3.502$ GHz are called third-order intermodulation components, abbreviated as IM3, where $f_3 = 2f_1 - f_2$ and $f_4 = 2f_2 - f_1$. More specifically, the tone at f_3 is called the lower IM3 tone ($\mathrm{IM3_L}$) and the tone at f_4 is called the upper IM3 tone ($\mathrm{IM3_U}$). The next tones spaced one further interval out, the f_5 and f_6 tones, are called the fifth-order intermodulation distortion tones, the IM5 tones. Only one of the IM7 tones, f_7, has sufficient amplitude to be seen on the scale in the figure. The amplitudes of the output tones are not symmetrical around f_1 and f_2 and this effect is attributed to the impedance of the baseband circuit, as some of the input signal power is converted to baseband before being converted back up again [39–43]. The thermal modulation of the transistor characteristics, especially mobility, due to temperature variations at or slower than the rate at which the envelope varies, can also produce this effect. Minimizing temperature variations is an additional reason for very good heat-sinking of power devices.

In Figure 4-23 the output power at both the fundamental and $\mathrm{IM3_U}$ are plotted versus the input power in a two-tone test. The output power at the fundamental increases linearly with the input power until saturation is reached. The $\mathrm{IM3_U}$ signal initially increases with the third power of the input signal level, which corresponds to the model of the active device around the bias point at low signal levels being a cubic power series. As the power increases, high-order nonlinearities become significant and the IM3 behavior is less predictable.

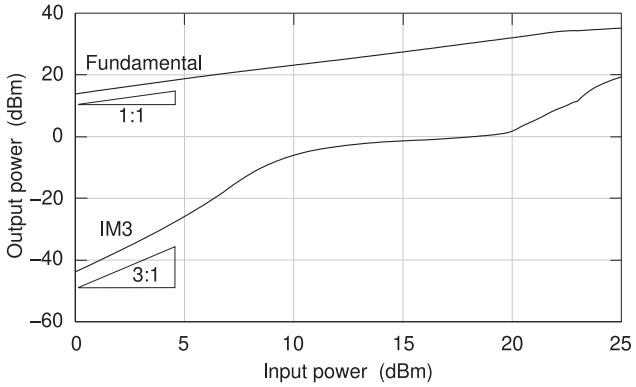
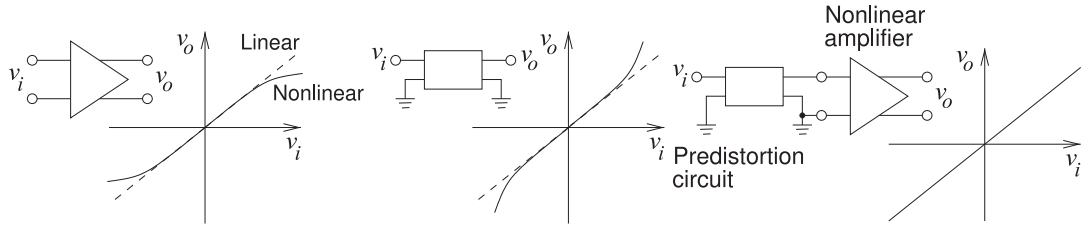


Figure 4-23: Response to a two-tone input signal: output power at the fundamental and at $\mathrm{IM3_U}$ versus input power. The two input tones are at 3.500 GHz and 3.501 GHz and the fundamental plotted is at 3.501 GHz.



(a) Nonlinear amplifier (b) Nonlinear compensator (c) Linearized amplifier

Figure 4-24: Predistortion linearizer concept.

4.6.4 Summary

Design of power amplifiers considers discrete tones and in particular uses two tones to characterize distortion. Discrete tones are required in simulation and two-tone characterization is convenient in a laboratory. Of course, digitally modulated signals are not discrete tones, but designers have found that the nonlinear responses to one tone and to two tones are good indications of the behavior with digitally modulated signals which are not convenient to use in simulation [6, 44].

4.7 Linearization

Linearization refers to strategies that compensate for the inherent phase and amplitude nonlinearities of RF amplifiers. There are three basic strategies that introduce a correction mechanism [45–50]:

(a) predistortion linearization,
(b) feed-forward linearization, and
(c) linearization by design.

Reducing distortion using the above techniques enables an amplifier to operate at higher efficiency while producing the maximum permissible amount of distortion.

4.7.1 Predistortion

The predistortion concept is illustrated in Figure 4-24. The essential nonlinearity of an RF amplifier is a tanh-like response, as shown in

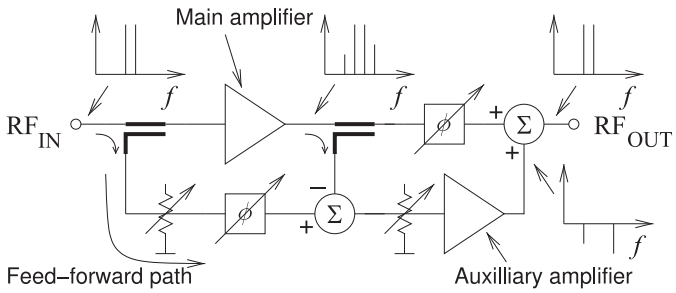


Figure 4-25: Feed-forward linearizer concept with the plots showing the spectra in different parts of the circuit.

Figure 4-24(a). With predistortion, a compensation network that has a complementary response, such as that shown in Figure 4-24(b), is developed [49, 50]. When the compensation network precedes the amplifier, ideally the output is linearized with little final distortion (see Figure 4-24(c)). Predistortion can be analog predistortion, as shown here, or the compensating function can be performed at baseband in a DSP unit. Predistortion requires a good behavioral analytic model of the amplifier from which distortion can be calculated. Then this model is reverted to synthesize the compensation network. Variations of the predistortion strategy include adaptive predistortion in which the distortion at the output of the amplifier is regularly sampled during operation and the coefficients of the compensation network updated. The linearization achievable is limited by the accuracy of the reversion synthesis, the stability of the amplifier circuit, and the stability of the adaptive process.

4.7.2 Feed-Forward Linearization

A linearization strategy that does not require characterization of the nonlinear process is feed-forward linearization, the concept of which is shown in Figure 4-25 [48]. In Figure 4-25 a two-tone input signal is considered for illustration. The two-tone signal is sampled prior to amplification by the main amplifier, which produces intermodulation distortion where third-order intermodulation distortion is shown in the spectrum at the main amplifier output. The original undistorted input signal is sampled by the directional coupler at the input and, following amplitude and phase adjustment, is compared to the sampled distorted signal at the output of the main amplifier. The difference is then the distorted waveform, which is amplified by an auxiliary amplifier and then added into the main signal path (after appropriate phase adjustment). This addition ideally cancels distortion in the high-power signal. Feed-forward linearization requires a very linear auxiliary amplifier, but this is easier to achieve than for the main amplifier since the auxiliary amplifier operates at much lower power levels.

4.7.3 Summary

The third category of linearization is linearization by design, achieved by choosing amplifier topologies that inherently compensate for distortion. Many of these topologies will be considered in the next sections.

The various linearization strategies can be combined, but linear by design topologies are very important, as predistortion and feed-forward linearization can consume significant DC power, especially for wideband signals.

4.8 Advanced Power Amplifier Architectures

A central challenge of power amplifier design is trading off power efficiency and distortion. Up to now discussion has focused on efficient design of single transistor amplifiers. In this section several architectures are described that enable the efficiency-distortion trade-off to go to a higher level. The three architectures discussed are the Doherty amplifier, the envelope tracking amplifier, the LITMUS amplifier, and the LINC amplifier.

4.8.1 Doherty Amplifier

A Doherty amplifier improves linearity and achieves high efficiency by using two amplifiers, one called a carrier amplifier and the other a peaking amplifier [45, 46, 51–55]. The basic Doherty amplifier configuration is shown in Figure 4-26(a). Here the input signal is split and half is applied to the carrier amplifier, typically a low-distortion Class AB amplifier, which amplifies small signals. However, as the input signal increases this amplifier becomes nonlinear and saturates, as seen in Figure 4-26(b). The other amplifier is typically a Class C amplifier and does not amplify small signals, but as the input becomes larger, the amplifier turns on. The Doherty amplifier has high linearity, as at high input powers the turn-on characteristic of the peaking amplifier complements the saturation of the carrier amplifier. The essential operation is that the input signal is split and half is shifted by $90°$. At the output this phase shift is matched by the $90°$ electrical length of the transmission line at the output of the carrier amplifier. This transmission line efficiently routes the output of the carrier amplifier to the load whether or not the peaking amplifier is turned on. When it is not turned on, the peaking amplifier presents an open circuit, but when it is on, it presents a Thevenin impedance Z_0. Then, the output power of both the carrier and peaking amplifiers are efficiently combined.

A more complete circuit design of an HBT Doherty amplifier is shown in Figure 4-27. The $90°$ hybrid at the input splits the signals and introduces the required phase shift. Base biasing voltages of the HBT transistors (i.e. V_{bc} and V_{bp}) are adjusted so that the carrier amplifier will amplify small signals, but the peaking amplifier will have a Class C-like characteristic and only turn on for large signals.

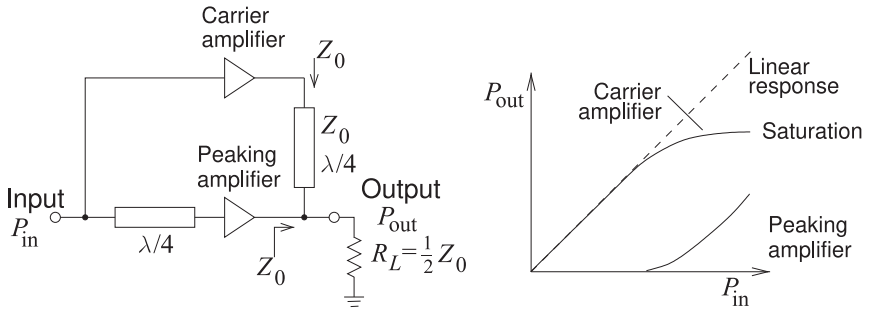


Figure 4-26: Doherty amplifier. (a) Doherty amplifier concept (b) Amplifier output characteristic

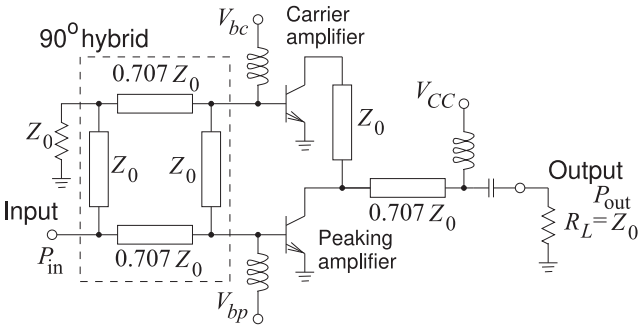


Figure 4-27: HBT Doherty amplifier with a 90° hybrid between P_{in} and the input of the transistors. All lines are $\lambda/4$ (i.e. 90°) long.

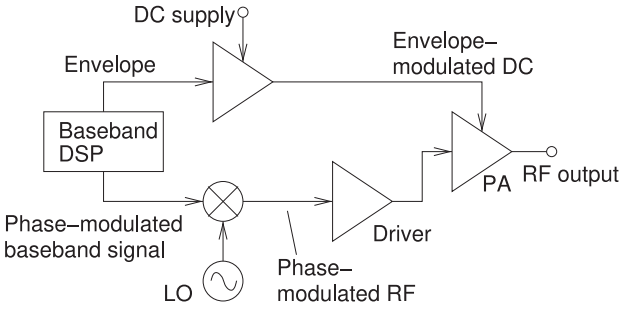


Figure 4-28: Khan transmitter using envelope elimination and restoration.

4.8.2 Envelope Tracking Amplifier

The concept of the envelope tracking amplifier was introduced with the **Khan transmitter** [56–59]. This technique is also called the **envelope-elimination and restoration** (**EER**) technique. The original amplifier concept [56], but updated to use modern DSP capabilities, is shown in Figure 4-28. The DSP operates at baseband and produces the digitally modulated signal to be transmitted. The DSP separately produces an amplitude-varying envelope signal and a constant-envelope phase-modulated baseband signal. These are amplified separately. This is in contrast to the usual approach of producing a baseband signal, or even a low-power RF signal, which has the complete modulation with both phase and amplitude variations. The phase-modulated signal is amplified by a driver before it is input to an efficient power amplifier that always operates in a highly linear mode since the DC source to the PA is adjusted to match the envelope of the final signal. Thus the PA is always operating at peak efficiency. These amplifiers have excellent linearity over a broad frequency range.

4.8.3 LINC Amplifier

The linear amplification with nonlinear component (LINC) amplifier separates a digitally modulated signal into two constant envelope components [18, 60–62]. This amplifier is also known as an **outphasing amplifier**. Each of the components is efficiently amplified by individual amplifiers and then recombined. The amplifier concept is shown in Figure 4-29.

The LINC amplifier operates as follows. The digitally modulated RF input signal can be expressed as

$$s(t) = a(t)\cos[\omega t + \theta(t)]. \tag{4.21}$$

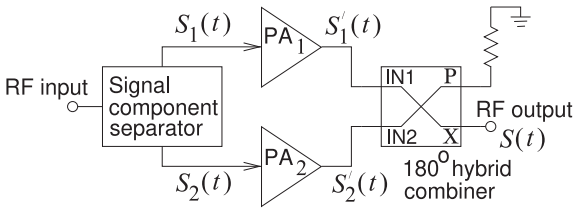


Figure 4-29:
LINC amplifier.

In complex exponential form this becomes

$$s(t) = a(t)\mathrm{e}^{\jmath\theta(t)}, \tag{4.22}$$

where $a(t)$ describes the amplitude modulation of the signal and $\theta(t)$ describes the phase modulation of the signal. This signal is split by the signal component separator to create two components, $S_1'(t)$ and $S_2'(t)$, with the same constant amplitudes and with complementary time-varying phase modulation:

$$S_1(t) = \tfrac{1}{2}\left[s(t) - e(t)\right]_{\mathrm{I}} = \tfrac{1}{2}V_m \cos\left[\omega t + \theta(t) - \psi(t)\right] \tag{4.23}$$

$$S_2(t) = \tfrac{1}{2}\left[s(t) + e(t)\right]_{\mathrm{I}} = \tfrac{1}{2}V_m \cos\left[\omega t + \theta(t) + \psi(t)\right], \tag{4.24}$$

where $\psi(t) = \cos^{-1}\left[a(t)/V_m\right]$, the subscript I denotes the in-phase component and $\frac{1}{2}V_m$ is the magnitude of the voltage signal applied to each of the PAs. The $e(t)$ term is called the quadrature signal and

$$e(t) = \jmath s(t)\sqrt{\frac{V_m^2}{a^2(t)} - 1}. \tag{4.25}$$

$S_1(t)$ and $S_2(t)$ are efficiently amplified separately to produce the amplified signals $S_1'(t)$ and $S_2'(t)$ and these are combined at the output by a $180°$ hybrid. Over time the amplitudes of $S_1'(t)$ and $S_2'(t)$ do vary but only when the average power of the RF signal changes. However in one data packet their amplitudes stay the same.

Ideally $S_1'(t)$ and $S_2'(t)$ are linearly scaled versions of $S_1(t)$ and $S_2(t)$ so that

$$S_1'(t) = kS_1(t) \quad \text{and} \quad S_2'(t) = kS_2(t), \tag{4.26}$$

and k may be complex. The output of the combiner is

$$S(t) = S_1'(t) + S_2'(t) = \tfrac{1}{\sqrt{2}}ka(t)\cos\left[\omega t + \theta(t) + \phi\right], \tag{4.27}$$

where k is the gain of each PA, ϕ is the phase shift introduced in the amplifier path, and the factor of $\frac{1}{\sqrt{2}}$ is introduced by the combiner. Thus $S(t)$ is a linearly amplified version of $s(t)$.

The LINC amplifier relies on the two PAs being well balanced and this must be maintained for a wide range of conditions and this can sometimes be difficult to achieve.

4.8.4 LITMUS Amplifier

The linear amplification by time-multiplexed spectrum (LITMUS) amplifier [63, 64], shown in Figure 4-30, utilizes the ability of a bandpass filter to

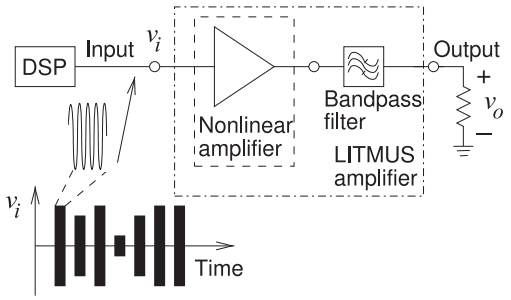


Figure 4-30: LITMUS amplifier. Each pulse of the input signal, v_i, has many sinusoids. For each pulse the sinusoids have different frequency, amplitude, and relative phase.

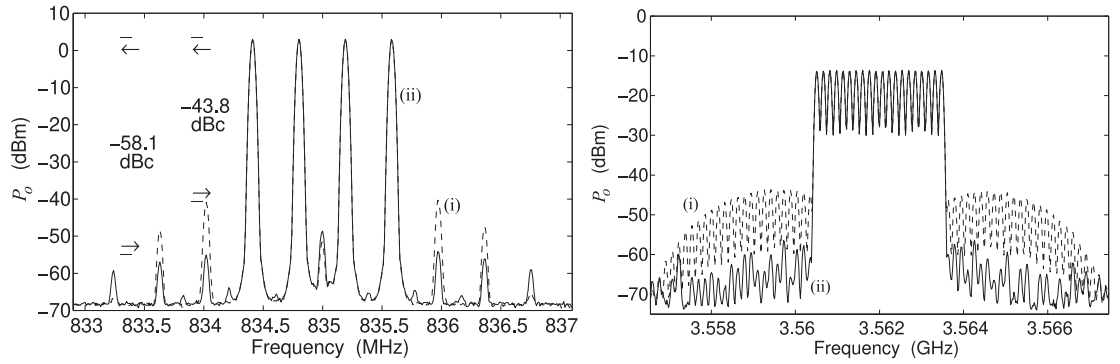
combine signals of different frequencies spaced in time. This is a result of the time-frequency behavior of filters, which is discussed in Section 2.18 of [2].

The most common way of reducing distortion of an amplifier is by setting the average output power of a modulated signal at a level that is below the 1 dB gain compression level by PMEPR. This is called backing off. So while an amplifier, such as a switching amplifier or a Class C amplifier could have close to 100% efficiency for a signal with a constant envelope, the back-off required for modulated signals with nonconstant envelopes significantly reduces the achievable efficiency.

A digitally modulated signal in a time interval Δt (coinciding with one or a few symbol interval times) can be decomposed into a sum of sinusoids that differ in amplitude, frequency, and relative phase [65]. (Generalizing this, the signal could be represented as a sum of constant envelope signals.) Now each sinusoid can be represented by a CW pulse of duration $\Delta\tau \ll \Delta t$ so that each pulse contains $\Delta\tau/T$ cycles of its sinusoidal constituent, where T is the period of the sinewave. These pulses can be separated in time, as shown for the input signal, v_i, in Figure 4-30. This input CW pulse train, generated in the DSP unit, is then applied to the input of the highly efficient nonlinear amplifier, which produces CW pulses at the output of the nonlinear amplifier. The output CW pulses are then combined in the bandpass filter. The bandpass filter spreads the pulses in time and reconstitutes the original, but now amplified, digitally modulated signal. During the time $\Delta\tau$ of each input pulse, the nonlinear amplifier is amplifying a constant envelope signal so the amplifier can operate in high-efficiency mode and saturation of the nonlinear circuit is not of concern. Also, since the individual frequency components are separated in time in the nonlinear part of the circuit, there can be no intermodulation distortion.

Theoretically a LITMUS amplifier completely suppresses intermodulation distortion, but the suppression achieved is limited by the bandwidth of the pulse train and also limited by reactive and memory effects in the nonlinear amplifier, especially baseband effects (including thermal effects), which provide long-term memory of the signals in the nonlinear circuit.

The output spectrum following amplification of a signal decomposed into four tones is shown in Figure 4-31(a). Curve (i) is the spectrum of the output of a conventional amplifier and Curve (ii) is the output of a LITMUS amplifier. Both amplifiers produce the same average RF power at the output. The output power of each of the four tones is 0 dBm. In Figure 4-31(a) six intermodulation tones are seen, three below the four fundamental tones and three above. These intermodulation tones are IM3 tones, as their frequencies can be numerically calculated as twice the frequency of one of the fundamental tones minus the frequency of one of the other tones. For



(a) Four tones (b) Twenty tones

Figure 4-31: Measured spectra at the output of the LITMUS amplifier for a different number of input tones. Curve (i) in both figures is the spectrum of the output of a conventional amplifier with a multitone input signal. Curve (ii) is the spectrum of the output of a LITMUS amplifier after the bandpass filter. The spike seen at 835 MHz is an artifact of the arbitrary waveform generator used to produce the signal. After [64].

the conventional amplifier, the largest intermodulation tone has an output power of -43.8 dBm. For the LITMUS amplifier the largest intermodulation tone has a power of -58.1 dBm. The intermodulation distortion is reduced by 14.3 dB.

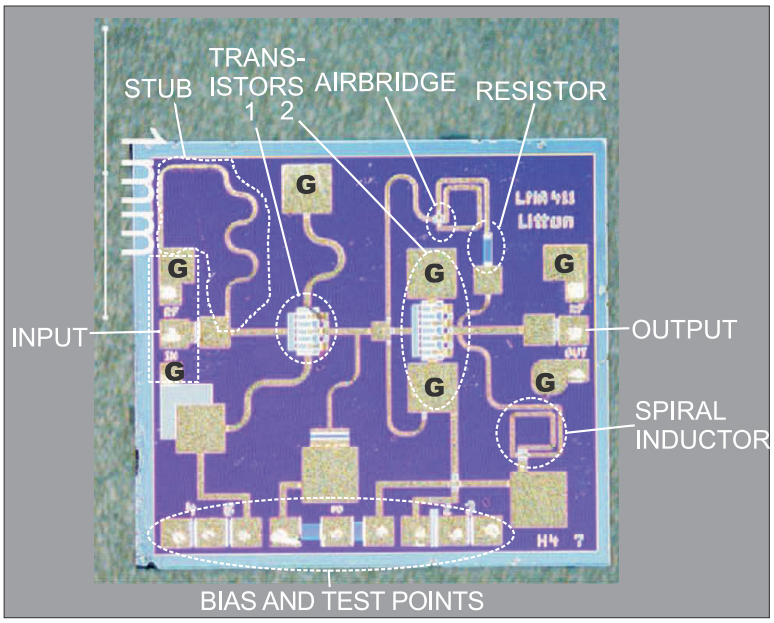
The amplification of a signal decomposed into 20 tones is shown in Figure 4-31(b). Again Curve (i) is the spectrum of the output of a conventional amplifier and Curve (ii) is the output of a LITMUS amplifier for the same output power. Now spectral regrowth is reduced by 13.9 dB.

4.8.5 Summary

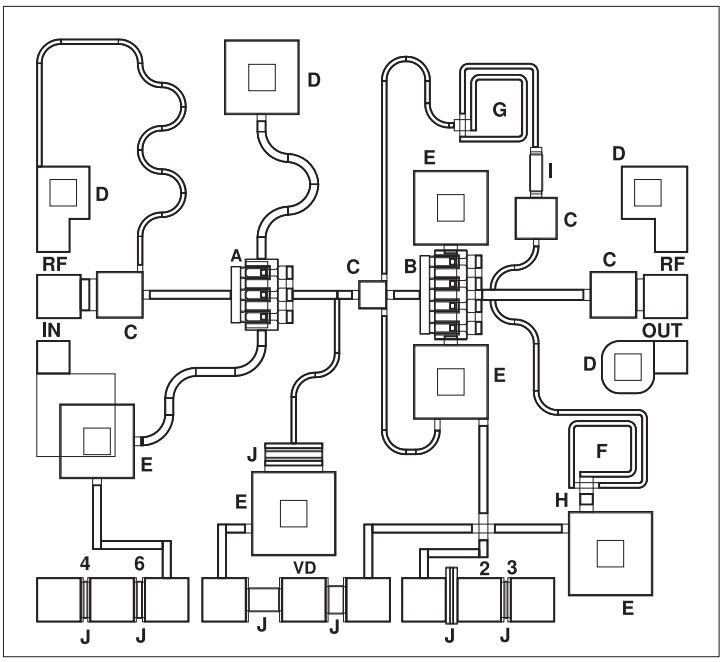
There are many variations on the advanced power amplifier architectures discussed in this section. These architectural concepts can be combined with each other and can also be combined with switching amplifier concepts. So the number of different power amplifier conceptual combinations is enormous and many papers have been written about them. Some amplifier topologies require high-performance DSP and high-performance transistors, while others require extensive manual tuning of each amplifier. These affect the cost of each amplifier. Also, the design cost of some amplifiers can be high and require the most experienced designers. These factors are considered in choosing the best amplifier architecture for a particular application.

4.9 MMIC Power Amplifiers

A linear Class A X-band amplifier implemented in GaAs MESFET **microwave monolithic integrated circuit** (**MMIC**) technology is shown in Figure 4-32. This amplifier operates from 8 to 12 GHz producing 100 mW of total power. This is a two-stage amplifier, as this results in the required gain, power, and bandwidth. The input and output pads are in a ground-signal-ground (GSG) configuration and can be used for both microwave



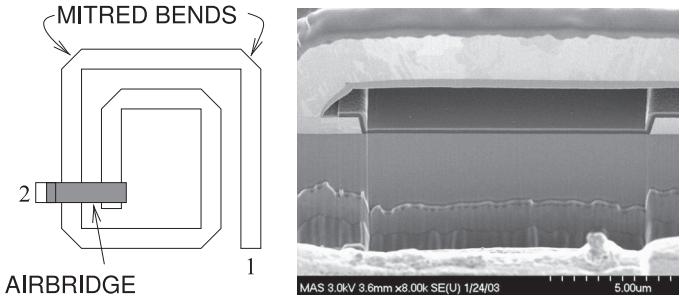
(a)



(b)

Figure 4-32: A two-stage X-band (8–12 GHz) MMIC amplifier producing 100 mW of power: (a) photomicrograph with key networks identified, **G** indicates ground; (b) annotated layout:
A, pHEMT in first stage.
B, pHEMT in second stage.
C, coupling capacitors.
D, via to backside metal.
E, via capacitor.
F, RF choke (bias) inductor.
G, RF feedback inductor.
H, airbridge.
I, feedback resistor.
J, bias (TaN) resistor.

probe testing and for wire-bonding. The first transistor has two source connections (at the top and bottom): a gate connection on the left and a drain connection on the right. The layout of the second transistor is the same. The larger second transistor has higher drain current. The source connections of the two transistors are grounded (indicated by the G connection). The matching network is implemented using stubs and capacitors. The second



(a) (b)

Figure 4-33: Details of the top spiral inductor in Figure 4-32(a): (a) layout of the top spiral inductor; and (b) scanning electron microscope image of the cross section of the **airbridge**.

transistor has a feedback network with a spiral inductor and series resistor. Drain bias to the second transistor is through a spiral inductor. Each of the spiral inductors use an airbridge (shown in Figure 4-33) to reduce parasitic capacitance by eliminating dielectric between the bridge metal and that of the spiral.

A second, higher-power X-band MMIC using pHEMT transistors is shown in Figure 4-34. This amplifier produces 1 W over the 8–12 GHz band. There are two stages, but the most striking contrast with the previous MMIC is the use of multiple transistors in each of the stages. The first stage has four transistors in parallel. The input matching network is integrated with a four-way power divider that drives the input of each transistor. An interstage matching network with four two-way power dividers drives the gates of 16 pHEMTS. A 16-way power combiner in the output matching network combines the signals at drains of each second-stage transistor, bringing them to a single output. The output combining network has four levels of two-way combiners and effectively enables transistors to be put in parallel. Putting transistors in parallel requires close matching of the transistors. Practically, the limit to the number of transistors that can be combined this way is 16 as there is loss with each level of combining.

4.10 RFIC Power Amplifiers

One of the distinguishing features of RFIC design is the synthesis of a circuit that intrinsically has the desired attributes. For an RFIC power amplifier the synthesized circuit must produce low levels of distortion while achieving high efficiency. The fundamental distortion mechanism in an RFIC is the near quadratic i-v characteristic of a MOS transistor and the tanh-like transfer characteristic of a MOS amplifier. This section has three examples of analytic and circuit techniques for calculating and managing distortion in MOS circuits.

4.10.1 Distortion in a MOSFET Enhancement-Depletion Amplifier Stage

In this section, power series analysis is applied to the MOS enhancement-depletion amplifier stage shown in Figure 4-35(a). This circuit is a Class A amplifier with a common-source enhancement gain stage and a depletion transistor as an active load. Recall that with a MOSFET, the voltage of the substrate has an important effect on operation of the transistor. With enhancement-mode transistors the substrate or body is typically connected

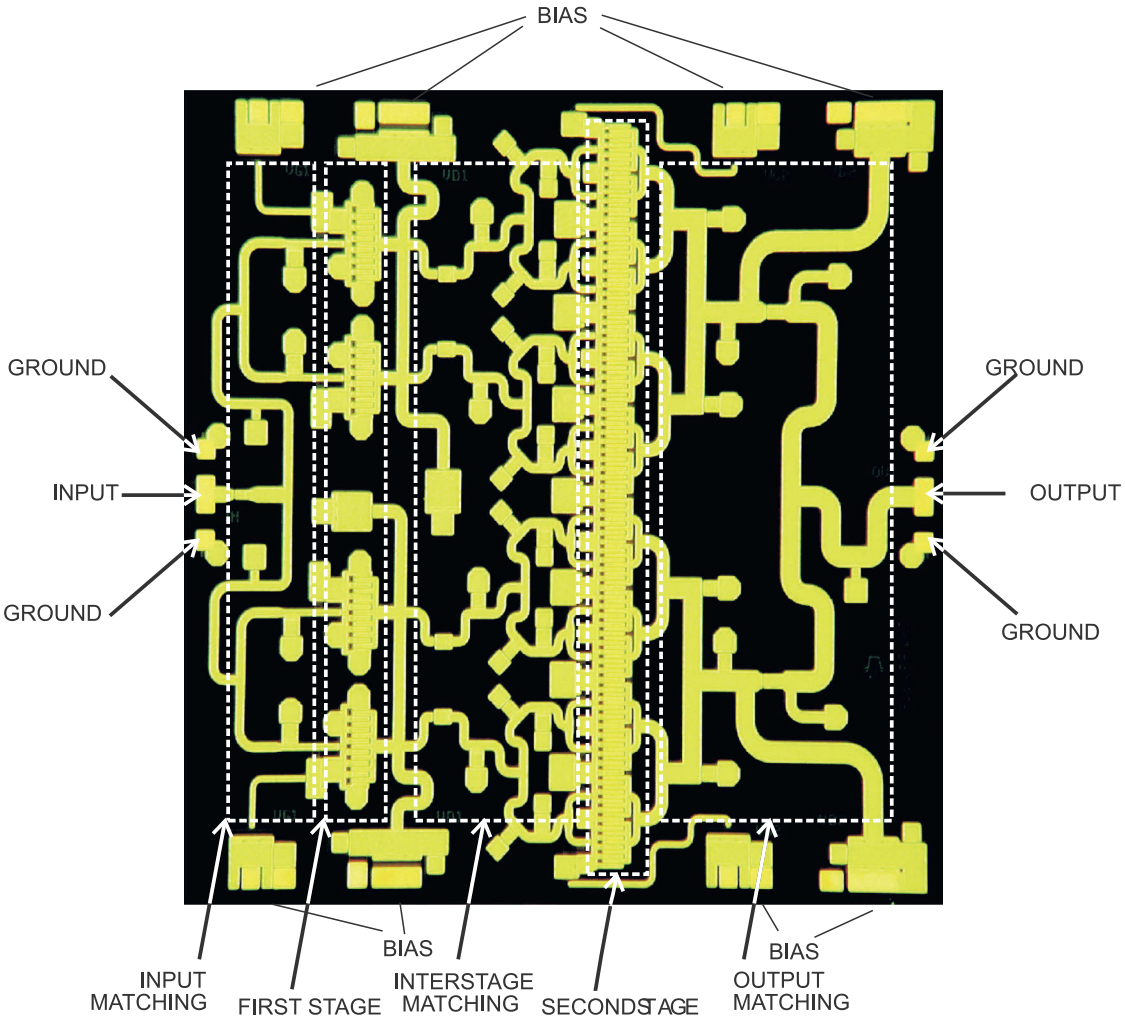


Figure 4-34: An 8–12 GHz MMIC amplifier producing approximately 1 W of output power with key networks identified. (Courtesy Filtronic, PLC, used with permission.)

to the most negative voltage in the circuit, in this case ground. So, since M_1 has the body connected to the source, there is no body effect for M_1. However, for M_2 the body effect must be considered. Using a simple MOSFET quadratic input-output relationship,

$$I_{D1} = \frac{K_1}{2}\left(V_A + V_X - V_{T0E}\right)^2 = I_{D2} = \frac{K_2}{2}\left(-V_{TD}^2\right), \tag{4.28}$$

where the threshold voltage of M_2, with the body effect, is

$$\begin{aligned} V_{TD} &= V_{T0D} + \gamma\left(\sqrt{2\phi_F + V_B + V_Y} - \sqrt{2\phi_Y}\right) \\ &= V_1 + \gamma\left(\sqrt{2\phi_F + V_B + V_Y}\right). \end{aligned} \tag{4.29}$$

(These equations are given in slightly different form in Section 1.A.1. The quantities ϕ_F and ϕ_Y are built-in inversion potentials.) Also K is

proportional to W/L, where W is the width of the transistor's channel and L is its length. The variable $V_1 = V_{T0D} - \gamma\sqrt{2\phi_Y}$ has been introduced, but as will be seen it will be canceled during the derivation.

The aim here is to develop a relationship between the input signal V_X and the output signal V_Y. The first step in developing a simple relationship that can be used in initial design is to relate the operating point voltage levels. Rearranging Equation (4.28),

$$\sqrt{\frac{K_1}{K_2}}\left(V_A + V_X - V_{TOE}\right) = -V_{TD}, \tag{4.30}$$

with the appropriate sign choice made. Combining Equations (4.28) and (4.29) yields

$$\sqrt{\frac{K_1}{K_2}}\left(V_A + V_X - V_{TOE}\right) = -\left(V_1 + \gamma\sqrt{2\phi_F + V_B + V_Y}\right) \tag{4.31}$$

$$\text{and } \sqrt{\frac{K_1}{K_2}}\left(V_X\right) + \sqrt{\frac{K_1}{K_2}}\left(V_A - V_{TOE}\right) = -\left(V_1 + \gamma\sqrt{2\phi_F + V_B + V_Y}\right). \tag{4.32}$$

Now when $V_X = 0 = V_Y$, that is, when there is no AC signal, Equation (4.32) becomes

$$\sqrt{\frac{K_1}{K_2}}\left(V_A - V_{TOE}\right) = -\left(V_1 + \gamma\sqrt{2\phi_F + V_B}\right). \tag{4.33}$$

Substituting Equation (4.33) into Equation (4.32) yields

$$\sqrt{\frac{K_1}{K_2}}V_X = \gamma\sqrt{2\phi_F + V_B} - \gamma\sqrt{2\phi_F + V_B + V_Y}, \tag{4.34}$$

$$\left(\sqrt{\frac{K_1}{K_2}}V_X - \gamma\sqrt{2\phi_F + V_B}\right)^2 = \gamma^2\left(2\phi_F + V_B + V_Y\right), \tag{4.35}$$

$$\frac{K_1}{K_2}V_X^2 - \left(2\gamma\sqrt{\frac{K_1}{K_2}}\sqrt{2\phi_F + V_B}\right)V_X + \gamma^2(2\phi_F + V_B) = \gamma^2 V_Y + \gamma^2(2\phi_F + V_B), \tag{4.36}$$

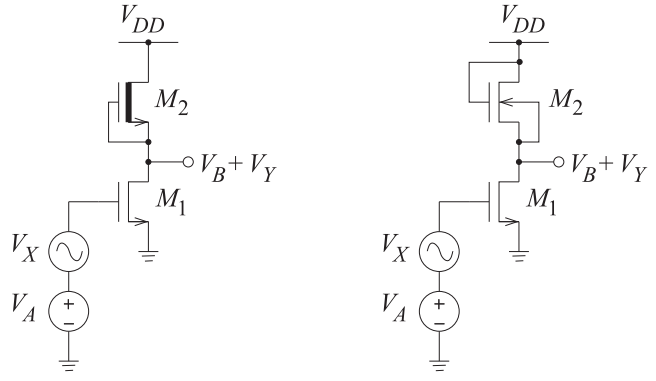


Figure 4-35: MOS amplifier stages.

(a) Enhancement-depletion stage (b) Ultralinear stage

and finally $$\left(\frac{K_1}{K_2}\right) V_X^2 - \left(2\gamma\sqrt{\frac{K_1}{K_2}}\sqrt{2\phi_F + V_B}\right) V_X = \gamma^2 V_Y. \tag{4.37}$$

Therefore the component of the output that differs from the quiescent point is

$$V_Y = \frac{-2}{\gamma}\sqrt{\frac{K_1}{K_2}}\sqrt{2\phi_F + V_B}V_X + \frac{1}{\gamma^2}\frac{K_1}{K_2}V_X^2. \tag{4.38}$$

For a sinusoidal input $V_X = |V_X|\cos(\omega t)$, $V_X^2 = \frac{1}{2}|V_X|^2 + \frac{1}{2}|V_X|^2\cos(2\omega t)$, which contains the second harmonic of the input signal.

So the second-harmonic distortion level, HD_2, the ratio of the second-harmonic component of V_Y to the fundamental component, is

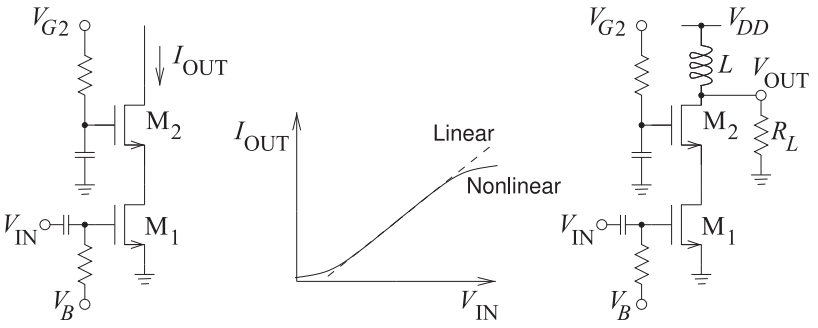
$$\begin{aligned}\mathrm{HD}_2 &= \left(\left|\tfrac{1}{2}\frac{K_1}{K_2}\frac{1}{\gamma^2}V_X^2\right|\right)\left(\left|\frac{-2\sqrt{2\phi_F + V_B}}{\gamma}\sqrt{\frac{K_1}{K_2}}V_X\right|\right)^{-1} \\ &= \frac{1}{4\gamma\sqrt{2\phi_F + V_B}}\sqrt{\frac{K_1}{K_2}}|V_X|. \end{aligned}\tag{4.39}$$

Examination of Equations (4.38) and (4.39) leads to the following conclusions. If there is a lot of voltage gain ($\propto\sqrt{K_1/K_2}$), then the input signal level must be small to keep the harmonic distortion down. Other design considerations that have the same effect are to use a wider device, which increases the transconductance (since $g_m \approx K \propto W/L$) of the transistor so that the drain current will be maintained for the lower input voltage levels. Of course, making the device wider increases capacitive parasitics, which will reduce the maximum operating frequency. Making the channel of the device shorter also increases the transconductance while not affecting the frequency performance. The important point is that there are trade-offs in modifying the performance of the circuit and these are only made apparent using the type of analysis presented here. The use of analysis and the optimization of design through synthesis is one of the cornerstones of RFIC design. In large part this is possible because of the (soft) quadratic-like current-voltage characteristics of MOSFETs.

4.10.2 Distortion in the Ultralinear MOS Connection

The circuit in Figure 4-35(b) is an enhancement-mode gain stage with an enhancement-mode load. Transistor M_2 is the n-channel enhancement load with its well tied to its source. That is, in the fabrication of M_2, a well is formed and the transistor is constructed in the well. The well now serves as the substrate for M_2. Since the well and the source of M_2 are connected together, M_2 is not subject to the body effect. One important property of this circuit is that the characteristics of M_1 and M_2 are matched. Using a similar approach to that used in the previous section, the input/output relationship of the circuit can be developed. Equating drain currents,

$$I_{D1} = \frac{K_1}{2}\left(V_A + V_X - V_{T0E}\right)^2 = I_{D2} = \frac{K_2}{2}\left(V_{DD} - V_B - V_Y - V_{T0E}\right)^2. \tag{4.40}$$



(a) Cascode transconductance cell

(b) Transconductance characteristic

(c) Cascode amplifier

Figure 4-36: A MOS cascode amplifier. L is a choke inductor and is an RF open circuit. V_B and V_{G2} are bias voltages. The cascode transconductance cell is used in this section as a tanh cascode cell (TCC).

With $V_X = V_Y = 0$, $$\sqrt{\frac{K_1}{K_2}}(V_A - V_{T0E}) = V_{DD} - V_B - V_{T0E}. \tag{4.41}$$

Therefore $$\sqrt{\frac{K_1}{K_2}}V_X + (V_{DD} - V_B - V_{T0E}) = (V_{DD} - V_B - V_Y - V_{T0E}), \tag{4.42}$$

and so $$V_Y = -\sqrt{\frac{K_1}{K_2}}V_X. \tag{4.43}$$

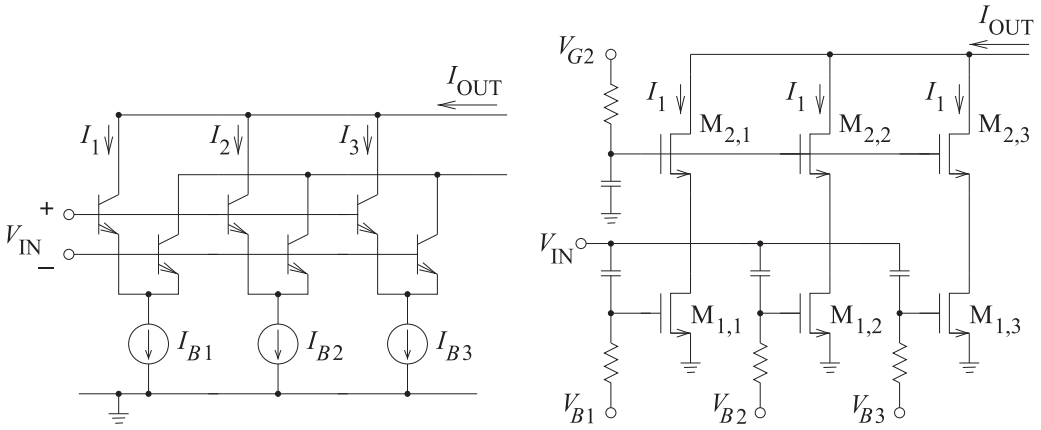
The result is that, provided the transistors are matched, the amplifier is inherently linear. So within the approximations of the drain current expressions there is no distortion, and this includes no third-order intermodulation distortion or spectral regrowth.

4.10.3 RFIC Power Amplifiers with Minimal Distortion

The most widely used linear-by-design strategy for linear RFIC amplifier design is to use what is called the multi-tanh method [66]. The idea is to scale several parallel amplifier stages to obtain an overall linear characteristic. Consider the MOS cascode amplifier shown in Figure 4-36(c), which is based on the transconductance cell shown in Figure 4-36(a). Applying an input voltage, $V_{\rm IN}$, produces an output current, $I_{\rm OUT}$, with the tanh-like characteristic shown in Figure 4-36(b). This characteristic is the dominant cause of amplitude distortion in a FET amplifier. The design strategy for linearizing the output characteristic of a FET amplifier is to put multiple amplifier stages in parallel so that the overall current-voltage characteristic optimally combines the tanh-like characteristic of each stage.

Figure 4-37 shows two circuits that exploit the multi-tanh strategy. Figure 4-37(a) is a three-stage multi-tanh differential amplifier transconductance cell in which the bias currents I_{B1}, I_{B2}, and I_{B3} are adjusted so that the overall input-output characteristic, $I_{\rm OUT}$ versus $V_{\rm IN}$, has greater linearity than that of the individual stages [66]. A similar concept is used to combine the output of multiple cascode stages (see Figure 4-37(b)) and this is a topology more suited to the development of RFIC power amplifiers [67].

As an example of the use of the multi-tanh design approach, consider the multiple tanh cascode amplifier in Figure 4-37(b) with three tanh cascode cells (TCCs). This amplifier has both amplitude and phase distortion. The amplitude distortion largely results from the tanh-like current-voltage



(a) Multi-tanh differential amplifier (b) Three tanh cascode cells (TCC)

Figure 4-37: Combining the output from multiple stages to obtain a highly linear overall transconductance characteristic.

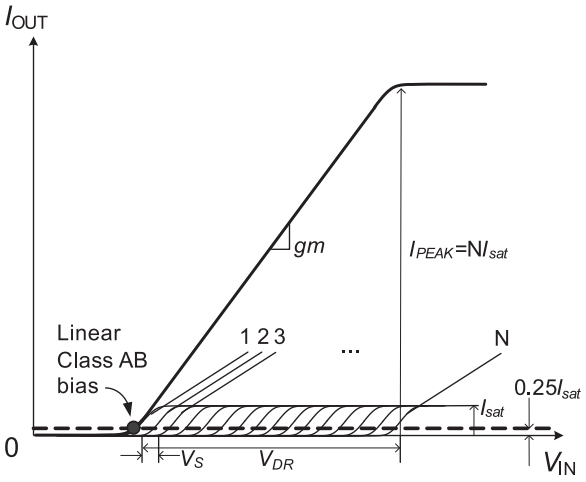


Figure 4-38: Combining the output from N Class AB cascode stages to obtain a highly linear overall transconductance characteristic. Shown are the individual $I_{\rm OUT}$ versus $V_{\rm IN}$ characteristics of each of N transistors each biased differently. The combined $I_{\rm OUT}$ versus $V_{\rm IN}$ characteristic is the sum of the individual transistor's $I_{\rm OUT}$ After [67].

characteristic of the individual cascode stages. Phase distortion largely results from the nonlinearity of the gate capacitance of the FETs. The tanh characteristic of each tanh cascode cell in Figure 4-37(b) is adjusted by scaling the transistors $\mathrm{M}_{1,1}$, $\mathrm{M}_{1,2}$, and $\mathrm{M}_{1,3}$, and changing their bias voltages V_{B1}, V_{B2}, and V_{B3}. The design approach is illustrated in Figure 4-38 and a synthesis approach is presented in references [68] and [67]. Each tanh cascode stage operates in high Class AB mode and with appropriate biasing the tanh characteristics are staggered. Note that for small input signals only one or a few stages are active. The currents from each stage are summed to yield an overall linearized transconductance response.

Phase distortion is reduced by using an analog predistortion circuit that realizes a nonlinear capacitance network that cancels the combined nonlinear gate capacitance of the cascode stages. The complete CMOS RF power amplifier topology is shown in Figure 4-39, where NCN is the nonlinear capacitor network. The die micrograph of this amplifier is shown in Figure

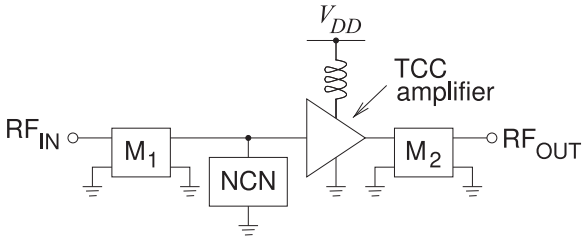
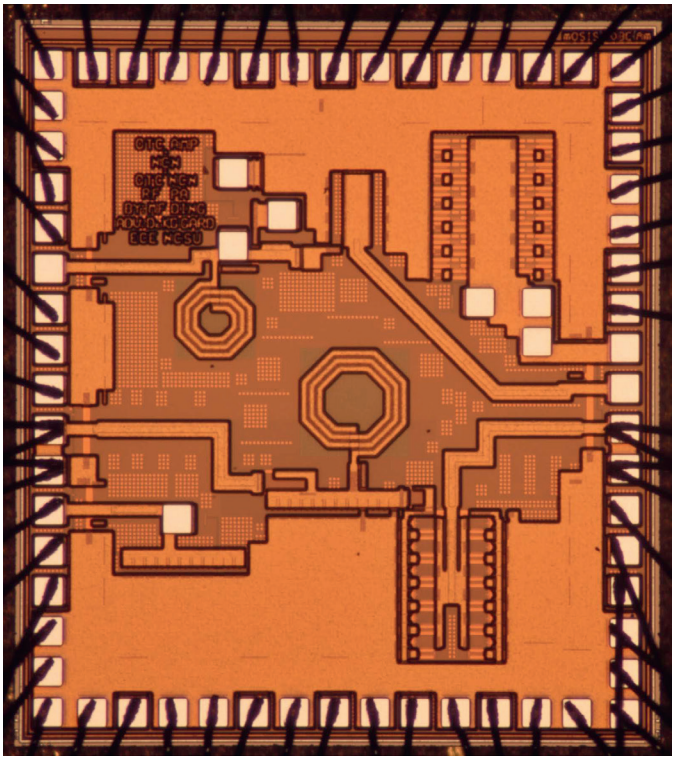
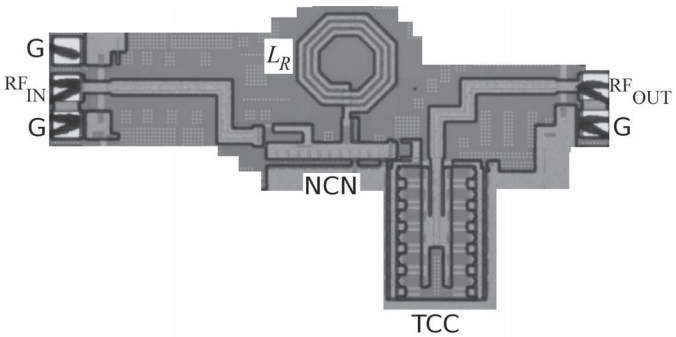


Figure 4-39: Complete RFIC power amplifier with a TCC amplifier with 12 Class AB TCC cells and a nonlinear capacitor network (NCN) and input and output matching networks (M_1 and M_2). After [67].



(a) Die



(b) 2D synthesized amplifier (NCN+TCC)

Figure 4-40: Die micrograph of the 2D synthesized amplifier with tanh cascode cell (TCC) amplifier stage and nonlinear capacitor network (NCN) stage. (There are two operating at different frequencies, one at the top of the die and one at the bottom. The bottom one is being referred to here.) The die also contains a TCC amplifier, an NCN network, and a conventional RF CMOS Class AB amplifier. The supply is 3.6 V and the ground connection is identified by G. The die is shown in (a) and the 2D synthesized amplifier in (b). The inductor L_R resonates out the linear component of the NCN capacitance and the linear component of the gate capacitance of the TCC amplifier. The matching networks and the coke inductor from V_{DD} are off-chip.

4-40(a) together with test circuits. The TCC amplifier with the nonlinear capacitor network is broken out in Figure 4-40(b). The performance of the RFIC power amplifier has excellent performance, as shown in Figure 4-41, and achieves an output power of 25 dBm with an efficiency of 42% and an ACPR of -22 dBc.

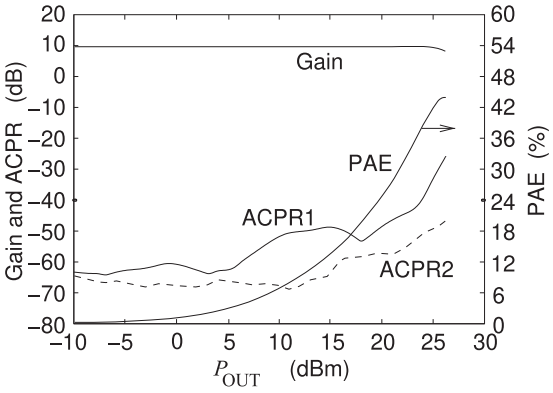


Figure 4-41: Performance of the TCC amplifier with the nonlinear capacitor predistortion circuit at 960 MHz with a WCDMA test signal. A gain of 9.4 dB with power-added efficiency of 41.6% is achieved with an output power of 24.9 dBm and meeting the 3G ACPR1 specifications of -33 dBc and an ACPR2 of -43 dBc. After [67].

4.11 Summary

The realization of efficient power amplifiers is one of the more competitive aspects of RF and microwave design. Power amplifiers must operate at high efficiency and hence operation is strongly nonlinear. Linear amplifier design is only an approximate guide to power amplifier design. Power amplifiers significantly affect the cost and reliability of RF front ends. This is true of base station amplifiers producing hundreds of watts, and equally true of handset power amplifiers producing hundreds of milliwatts. The cost of a power transistor is significant, and so it must operate close to its power output capability. It is important that the power amplifier designer extract every bit of performance from the power transistors, as every tenth of a decibel is significant.

In a basestation the cost of electricity for the RF front end, largely determined by the power amplifier, and by the required air-conditioning can be a significant part of the operating expense of the basestation. With handsets, the efficiency of the RF power amplifier impacts battery life. While it is known how to build amplifiers with high efficiency by combining the concepts presented in this chapter, combining too many concepts can result in a lengthy and costly design effort. The design cost and the possible need to manually tune individual amplifiers can appreciably raise the unit cost of each amplifier. The cost of design must be managed and designers should consider the requirement for individual amplifier tuning. For example, it is not reasonable to tune individual handset power amplifiers but it is for basestation power amplifiers. The trade-off for basestation amplifiers will be different from the trade-off for handset amplifiers, which have unit volumes that can be many orders of magnitude larger. Handset amplifiers are powered directly from the battery supply without voltage regulation. Thus handset power amplifier design must contend with limited supply voltages that can drop as the battery discharges. The limited supply voltage also restricts the choice of transistor technologies.

As with most aspects of RF design, intuition and experience is important in guiding initial topology selection. This is typically followed by computer-aided design and then optimization at the bench. All power amplifiers require manual, at-the-bench, optimization during design, as there are many effects that cannot be fully accounted for in the models used in microwave circuit simulators. Power amplifiers must contend with signals whose average power can vary significantly from packet to packet, and which can have a very large ratio of peak envelope power to mean envelope

power (i.e., large PMEPR). The standard design procedure is to arrive at a successful topology and initial layout using computer-aided design. Then load-pull is used in the laboratory to optimize the loading conditions for actual digitally modulated signals. It is possible that the topology may need to be changed to accommodate necessary changes identified during load-pull.

One of the significant costs and sources of reliability degradation in a handset is using multiple technologies for different parts of the RF front end. This requires heterogeneous integration rather than the more reliable monolithic integration. For example, in a handset it would be desirable to implement a power amplifier in silicon so that it can be integrated with the silicon driver and other circuitry. However, it is currently more feasible to obtain high efficiencies with compound semiconductor devices such as HBT and pHEMT transistors.

Ideally base station amplifiers would use silicon transistors, be able to efficiently amplifier many carriers (with PMEPRs exceeding 20 dB) simultaneously, be able to operate from 500 MHz to 5 GHz, and be reconfigurable for future unforeseen applications. Similarly, cellular handsets must support multiple bands and an important consideration is whether separate power amplifiers are used for each band. Ideally a handset would use a single silicon power amplifier as part of an RFIC that contains the rest of the RF front end, achieve efficiencies of 60% or greater, and operate from supply voltages that vary between 2 and 3 volts.

The RF spectrum is now being exploited beyond 200 GHz and power amplifiers are required from 100s of megahertz to more than 200 gigahertz. Currently the amplifiers with highest output power in this range merge semiconductor-based power amplifiers with vacuum-tube-based amplifier technologies.

Power amplifier design is a trade-off of available technologies that enable design with manageable design and operating costs while achieving requisite powers and efficiencies. There is a large trade-off space and there is a significant opportunity to provide competitive solutions.

4.12 References

[1] P. Aaen, J. Pla, and C. Balanis, "Modeling techniques suitable for CAD-based design of internal matching networks of high-power RF/microwave transistors," *IEEE Trans. on Microwave Theory and Techniques*, vol. 54, no. 7, pp. 3052–3059, Jul. 2006.

[2] M. Steer, *Microwave and RF Design, Modules*, 3rd ed. North Carolina State University, 2019.

[3] R. Gilmore and M. Steer, "Nonlinear circuit analysis using the method of harmonic balance—a review of the art: part i, introductory concepts," *Int. J. on Microwave and Millimeter Wave Computer Aided Engineering*, vol. 1, pp. 22–37, Jan. 1991.

[4] ——, "Nonlinear circuit analysis using the method of harmonic balance—a review of the art: part ii, advanced concepts," *Int. Journal on Microwave and Millimeter Wave Computer Aided Engineering*, vol. 1, pp. 159–180, Apr. 1991.

[5] K. Kundert, "Introduction to RF simulation and its application," *IEEE J. of Solid-State Circuits*, vol. 34, no. 9, pp. 1298–1319, Sep. 1999.

[6] J. Sevic, M. Steer, and A. Pavio, "Nonlinear analysis methods for the simulation of digital wireless communication system," *Int. Journal on Microwave and Millimeter Wave Computer Aided Engineering*, pp. 197–216, May 1996.

[7] F. Colon and T. Trick, "Fast periodic steady-state analysis for large-signal electronic circuits," *IEEE J. of Solid-State Circuits*, vol. 8, no. 4, pp. 260–269, Aug. 1973.

[8] T. Aprille and T. Trick, "A computer algorithm to determine the steady-state response of nonlinear oscillators," *IEEE Trans. on Cir-*

cuit Theory, vol. 19, no. 4, pp. 354–360, Jul. 1972.

[9] S. Director and K. Current, "Optimization of forced nonlinear periodic circuits," *IEEE Trans. on Circuits and Systems*, vol. 23, no. 6, pp. 329–335, Jun. 1976.

[10] M. Nakhla and F. Branin, "Determining the periodic response of nonlinear systems by a gradient method," *Circuit Theory Appl.*, vol. 4, no. 3, pp. 255–273, Jul. 1977.

[11] I. Norenkov, Y. Yevstifeyev, and V. Manichev, "A steady-state analysis method for multiperiod electronic circuits," *Radiotekhnika*, no. 11, pp. 86–89, 1987.

[12] J. Aprille, T.J. and T. Trick, "Steady-state analysis of nonlinear circuits with periodic inputs," *Proc. of the IEEE*, vol. 60, no. 1, pp. 108–114, Jan. 1972.

[13] S. Director, "A method for quick determination of the periodic steady-state in nonlinear networks," in *Allerton Conf. Circuit System Theory*, Oct. 1971, pp. 131–139.

[14] H. Chireix, "High power outphasing modulation," *Proc. of the Institute of Radio Engineers*, vol. 23, no. 11, pp. 1370–1392, Nov. 1935.

[15] F. Raab, P. Asbeck, S. Cripps, P. Kenington, Z. Popovic, N. Pothecary, J. Sevic, and N. Sokal, "Power amplifiers and transmitters for RF and microwave," *IEEE Trans. on Microwave Theory and Techniques*, vol. 50, no. 3, pp. 814–826, Mar. 2002.

[16] E. Hegazi and A. Abidi, "A 17-mw transmitter and frequency synthesizer for 900-mhz gsm fully integrated in 0.35-μm CMOS," *IEEE J. of Solid-State Circuits*, vol. 38, no. 5, pp. 782–792, May 2003.

[17] N. Sokal and A. Sokal, "Class e-a new class of high-efficiency tuned single-ended switching power amplifiers," *IEEE J. of Solid-State Circuits*, vol. 10, no. 3, pp. 168–176, Jun. 1975.

[18] D. Cox, "Linear amplification with nonlinear components," *IEEE Trans. on Communications*, vol. 22, no. 12, pp. 1942–1945, Dec. 1974.

[19] F. Raab, "Class-F power amplifiers with maximally flat waveforms," *IEEE Trans. on Microwave Theory and Techniques*, vol. 45, no. 11, pp. 2007–2012, Nov. 1997.

[20] M. Eron, B. Kim, F. Raab, R. Caverly, and J. Staudinger, "The head of the class," *IEEE Microwave Magazine*, vol. 12, no. 7, pp. S16–S33, Dec. 2011.

[21] J. Sevic and M. Steer, "On the significance of envelope peak-to-average ratio for estimating the spectral regrowth of an RF/microwave power amplifier," *IEEE Trans. on Microwave Theory and Techniques*, vol. 48, no. 6, pp. 1068–1071, Jun. 2000.

[22] S.-A. El-Hamamsy, "Design of high-efficiency RF class-D power amplifier," *IEEE Trans. on Power Electronics*, vol. 9, no. 3, pp. 297–308, May 1994.

[23] H. Kobayashi, J. Hinrichs, and P. Asbeck, "Current-mode class-d power amplifiers for high-efficiency RF applications," *IEEE Trans. on Microwave Theory and Techniques*, vol. 49, no. 12, pp. 2480–2485, Dec. 2001.

[24] J.-Y. Kim, D.-H. Han, J.-H. Kim, and S. Stapleton, "A 50 w ldmos current mode 1800 mhz class-d power amplifier," in *2005 IEEE MTT-S Int. Microwave Symp. Dig.*, Jun. 2005, p. 4 pp.

[25] H. Nemati, C. Fager, and H. Zirath, "High efficiency LDMOS current mode class-D power amplifier at 1 ghz," in *36th European Microwave Conf.*, Sep. 2006, pp. 176–179.

[26] D. Schmelzer and S. Long, "A GaN HEMT class F amplifier at 2 GHz with $> 80\%$ PAE," *IEEE J. of Solid-State Circuits*, vol. 42, no. 10, pp. 2130–2136, Oct. 2007.

[27] P. Saad, H. Nemati, M. Thorsell, K. Andersson, and C. Fager, "An inverse class-F GaN HEMT power amplifier with 78% PAE at 3.5 ghz," in *39th European Microwave Conf.*,, 2009, pp. 496–499.

[28] J. H. Kim, G. D. Jo, J. H. Oh, Y. H. Kim, K. C. Lee, and J. H. Jung, "Modeling and design methodology of high-efficiency class-F and class-F^{-1} power amplifiers," *IEEE Trans. on Microwave Theory and Techniques*, vol. 59, no. 1, pp. 153–165, Jan. 2011.

[29] E. Cipriani, P. Colantonio, F. Giannini, and R. Giofre, "Theoretical and experimental comparison of class F vs. class F^{-1} PAs," in *40th European Microwave Conf.*, Sep. 2010, pp. 1670–1673.

[30] J. Sevic and M. Steer, "Analysis of GaAs MESFET spectrum regeneration driven by a $\pi/4$-DQPSK modulated source," in *1995 IEEE MTT-S Int. Microwave Symp. Dig.*, May 1995, pp. 1375–1378.

[31] K. Gard, "Autocorrelation analysis of spectral regrowth generated by nonlinear circuits in wireless communication systems," Ph.D. dissertation, University of California at San Diego, 2003.

[32] H. Gutierrez, K. Gard, and M. Steer, "Nonlinear gain compression in microwave amplifiers using generalized power-series analysis and transformation of input statistics," *IEEE Trans. on Microwave Theory and Techniques*, vol. 48, no. 10, pp. 1774–1777, Oct. 2000.

[33] J. Brinkhoff and A. Parker, "Effect of baseband impedance on FET intermodulation," *IEEE Trans. on Microwave Theory and Techniques*, vol. 51, no. 3, pp. 1045–1051, Mar. 2003.

[34] F. Giannini, *Nonlinear Microwave Circuit Design*. Wiley, 2004.

[35] K. Lu, P. McIntosh, C. Snowden, and R. Pollard, "Low-frequency dispersion and its influence on the intermodulation performance of AlGaAs/GaAs HBTs," in *1996 IEEE MTT-S International Microwave Symposium Digest*, 1996, pp. 1373–1376.

[36] J. Sevic, K. Burger, and M. Steer, "A novel envelope-termination load-pull method for ACPR optimization of RF/microwave power amplifiers," in *1998 IEEE MTT-S Int. Microwave Symp. Dig.*, 1998, pp. 723–726.

[37] J. Sevic, M. Steer, and A. Pavio, "Large-signal automated load-pull of adjacent-channel power for digital wireless communication systems," in *1996 IEEE MTT-S Int. Microwave Symp. Dig.*, vol. 2, 1996, pp. 763–766.

[38] J. Sevic, R. Baeten, G. Simpson, and M. Steer, "Automated large-signal load-pull characterization of adjacent-channel power ratio for digital wireless communication systems," in *ARFTG Conf. Dig.-Fall, 46th*, vol. 28, 1995, pp. 64–70.

[39] W. Jang, A. Walker, K. Gard, and M. Steer, "Capturing asymmetrical spectral regrowth in RF systems using a multislice behavioral model and enhanced envelop transient analysis," *Int. Journal of RF and Microwave Computer-Aided Engineering*, vol. 16, no. 4, pp. 400–407, 2006.

[40] A. Walker, M. Steer, and K. Gard, "Capturing asymmetry in distortion of an RF system using a multislice behavioral model," *IEEE Microwave and Wireless Components Letters*, vol. 16, no. 4, pp. 212–214, 2006.

[41] J. Hu, K. Gard, N. Carvalho, and M. Steer, "Dynamic time-frequency waveforms for VSA characterization of PA long-term memory effects," in *69th ARFTG Conf., 2007*, 2007, pp. 1–5.

[42] N. Borges de Carvalho and J. Pedro, "A comprehensive explanation of distortion sideband asymmetries," *IEEE Trans. on Microwave Theory and Techniques*, vol. 50, no. 9, pp. 2090–2101, Sep. 2002.

[43] J. Pedro and N. Carvalho, *Intermodulation Distortion in Microwave and Wireless Circuits*. Artech House, 2003.

[44] J. Sevic and M. Steer, "On the significance of envelope peak-to-average ratio for estimating the spectral regrowth of an RF/microwave power amplifier," *IEEE Trans. on Microwave Theory and Techniques*, vol. 48, no. 6, pp. 1068–1071, 2000.

[45] S. Cripps, *RF Power Amplifiers for Wireless Communications*. Artech House, 1999.

[46] P. Kenington, *High-Linearity RF Amplifier Design*. Artech House, 2000.

[47] J. Pothecary, *High-power GaAs FET Amplifiers*. Artech House, 1999.

[48] N. Pothecary, *Feedforward Linear Power Amplifiers*. Artech House, 1999.

[49] J. Cavers, "Amplifier linearization using a digital predistorter with fast adaptation and low memory requirements," *IEEE Trans. on Vehicular Technology*, vol. 39, no. 4, pp. 374–382, Nov. 1990.

[50] A. D'Andrea, V. Lottici, and R. Reggiannini, "RF power amplifier linearization through amplitude and phase predistortion," *IEEE Trans. on Communications*, vol. 44, no. 11, pp. 1477–1484, Nov. 1996.

[51] C. Campbell, "A fully integrated ku-band doherty amplifier mmic," *IEEE Microwave and Guided Wave Letters*, vol. 9, no. 3, pp. 114–116, Mar. 1999.

[52] M. Iwamoto, A. Williams, P.-F. Chen, A. Metzger, L. Larson, and P. Asbeck, "An extended doherty amplifier with high efficiency over a wide power range," *IEEE Trans. on Microwave Theory and Techniques*, vol. 49, no. 12, pp. 2472–2479, Dec. 2001.

[53] W. Doherty, "A new high efficiency power amplifier for modulated waves," *Proc. of the Institute of Radio Engineers*, vol. 24, no. 9, pp. 1163–1182, Sep. 1936.

[54] B. Kim, J. Kim, I. Kim, and J. Cha, "The doherty power amplifier," *IEEE Microwave Magazine*, vol. 7, no. 5, pp. 42–50, Oct. 2006.

[55] F. Raab, "Efficiency of doherty RF power amplifier system," *IEEE Trans. Broadcast.*, vol. 33, no. 3, pp. 77–83, Sep. 1987.

[56] L. Kahn, "Single-sideband transmission by envelope elimination and restoration," *Proc. of the IRE*, vol. 40, no. 7, pp. 803–806, Jul. 1952.

[57] F. Raab and M. Poppe, "Kahn-technique transmitter for L-band communication/radar," in *2010 IEEE Radio and Wireless Symp.*, Jan. 2010, pp. 100–103.

[58] F. Raab, "Intermodulation distortion in Kahn-technique transmitters," *IEEE Trans. on Microwave Theory and Techniques*, vol. 44, no. 12, pp. 2273–2278, Dec. 1996.

[59] ——, "Drive modulation in Kahn-technique transmitters," in *1999 IEEE MTT-S Int. Microwave Symp. Dig.*,, Jun. 1999, pp. 811–814.

[60] B. Stengel and W. Eisenstadt, "LINC power amplifier combiner method efficiency optimization," *IEEE Trans. on Vehicular Technology*, vol. 49, no. 1, pp. 229–234, Jan. 2000.

[61] A. Birafane and A. Kouki, "An analytical approach to LINC power combining efficiency estimation and optimization," in *33rd European Microwave Conf.*, Oct. 2003, pp. 1227–1229.

[62] X. Zhang, L. Larson, and P. Asbeck, *Design of Linear RF Outphasing Power Amplifiers*. Artech House, 2003.
[63] G. Mazzaro, K. Gard, and M. Steer, "Low distortion amplification of multisine signals using a time-frequency technique," in *2009 IEEE MTT-S Int. Microwave Symp. Dig.*, Jun. 2009, pp. 901–904.
[64] ——, "Linear amplification by time-multiplexed spectrum," *IET Circuits, Devices Systems*, vol. 4, no. 5, pp. 392–402, Sep. 2010.
[65] N. Carvalho, K. Remley, D. Schreurs, and K. Gard, "Multisine signals for wireless system test and design [application notes]," *IEEE Microwave Magazine*, vol. 9, no. 3, pp. 122–138, Jun. 2008.
[66] B. Gilbert, "The multi-tanh principle: a tutorial overview," *IEEE J. of Solid-State Circuits*, vol. 33, no. 1, pp. 2–17, Jan. 1998.
[67] M. Ding, K. Gard, and M. Steer, "A highly linear and efficient CMOS RF power amplifier with a new circuit synthesis technique," *IEEE Trans. on Microwave Theory and Techniques*, vol. 60, no. 8, pp. 1–2, Nov. 2012.
[68] D. M. and G. K.G., "Tanh cascode cell amplifier: an arbitrary transfer characteristics amplifier," *Electronics Letters*, vol. 46, pp. 1495–1497, Oct. 2010.

4.13 Exercises

1. Section 4.2.1 presented the harmonic balance equations for the analysis of a nonlinear resistor. DC, the fundamental, and the second harmonic were retained. Derive the harmonic balance equations if only the DC and fundamental tones are retained in the analysis. That is, derive the analysis in Section 4.2.1. You do not need to implement the computer program to perform the analysis.
2. A new digital modulation scheme produces a signal with a PMEPR of 5.6 dB. What is the maximum power-added efficiency of an inductively biased Class A amplifier if the amplifier gain is 40 dB? Explain your reasoning. [Hint: Consider Table 4-3.]
3. A digitally modulated signal has a PMEPR of 10 dB. What is the maximum power-added efficiency of a resistively biased Class A amplifier if the amplifier gain is 20 dB? Explain your reasoning.
4. An amplifier is driven by a digitally modulated signal with a PMEPR of 20 dB. What is the efficiency reduction factor?
5. A Class C amplifier is used to amplify a GSM signal.
   (a) What is the PMEPR of the GSM signal?
   (b) The 1 dB gain compression metric is not relevant when a Class C amplifier amplifies a GSM signal. Explain.
6. An amplifier with a single-tone output 1 dB gain compression power of 26 dBm is used to amplifier a digitally modulated signal with a PMEPR of 6 dB. What is the average power in dBm of the output signal if the 1 dB gain compression level is taken as the maximum acceptable distortion?
7. The RF input to a 10 dB power amplifier is an FM signal.
   (a) What is the PMEPR of the input RF signal?
   (b) What is the maximum power-added efficiency of an inductively biased Class A amplifier if the amplifier gain is 20 dB? Explain your reasoning.
   (c) What is the RF power at the output of the amplifier?
   (d) What is the DC power consumed by the amplifier if the 1 dB gain compression point sets the peak RF output power? That is, the maximum RF output power is when the amplifier is operating at the gain compression point.
8. The RF input to a 20 dB power amplifier is the combination of ten 10 W FM signals.
   (a) What is the PMEPR of one FM signal?
   (b) What is the PMEPR of the combined RF signal? (Consider that the FM signals are narrowband and that they are uncorrelated.)
   (c) What is the gain of the amplifier when it is outputting the largest distorted signal with acceptable distortion? (Assume that the maximum distorted signal is defined by the output power at the 1-dB gain compression.)
   (d) What is the total RF power at the output of the amplifier?
   (e) What is the DC power consumed by the amplifier if the 1 dB gain compression point sets the peak RF output power? That is, the maximum RF output power is when the amplifier is operating at the gain compression point.
   (f) What is the maximum undistorted power-added efficiency of an inductively biased Class A amplifier if the amplifier gain is 20 dB? Explain your reasoning.
9. The input of a 10 dB power amplifier consists of 10 GSM signals.

(a) What is the PMEPR of one GSM signal
(b) What is the PMEPR of the 10-GSM signal? (Consider that the GSM signals are narrowband and that they are uncorrelated.)
(c) What is the maximum power-added efficiency of an inductively biased Class A amplifier? Explain your reasoning.

10. What is the conduction angle of an ideal inductively biased Class A amplifier?

11. What is the conduction angle of an ideal Class E amplifier?

12. What is the conduction angle of an ideal Class F amplifier?

13. A matched amplifier with a $50\ \Omega$ source and a $50\ \Omega$ load has an RF input with peak voltage $V_{\rm IN}$ and input power $P_{\rm IN}$, and an output with peak voltage $V_{\rm OUT}$ and output power $P_{\rm OUT}$. The transfer function of the amplifier (including the input and output matching networks and the active device) is described by $V_{\rm OUT} = 5 \tanh V_{\rm IN}$.
(a) Sketch the transfer characteristic as $V_{\rm OUT}$ versus $V_{\rm IN}$.
(b) Derive an expression for $P_{\rm OUT}$ versus $P_{\rm IN}$.
(c) What is the linear voltage gain of the amplifier?
(d) What is the linear power gain of the amplifier in decibels?
(e) What is the saturated output power of the amplifier in dBm if the RF signal is a single sinusoid?
(f) What is the RF output power in dBm at the 1 dB gain compression point if the RF signal is a single sinusoid?

14. A single-stage amplifier has a linear gain of 16 dB, an output 1 dB gain compression point of 10 dBm, and an OIP3 of 30 dBm.
(a) What is the maximum sinusoidal input signal when the output of the amplifier is compressed by 1 dB?
(b) What is the input-referred third-order intercept point (IIP3)?

15. A single-stage amplifier has a linear gain of 16 dB, an output 1 dB gain compression point of 10 dBm, and an OIP3 of 30 dBm. A communication signal with a PAR of 6 dB is used. What is the maximum average power of the input signal before the output suffers significant compression? This is defined at the point at which the peak signal is compressed by 1 dB.

16. A matched amplifier with a $50\ \Omega$ source and a $25\ \Omega$ load has a sinusoidal RF input with peak voltage $V_{\rm IN}$ and input power $P_{\rm IN}$, and an output with peak voltage $V_{\rm OUT}$ and output power $P_{\rm OUT}$. The transfer function of the amplifier is described by $v_{\rm out}(t) = 5v_{\rm in}(t) - 0.5v_{\rm in}^3(t)$, where $v_{\rm in}(t)$ and $v_{\rm out}(t)$ are the instantaneous values of the input and output voltage. The maximum value of $v_{\rm in}(t)$ is 2 V.
(a) Sketch the instantaneous voltage transfer function of the amplifier. That is, plot $v_{\rm out}$ versus $v_{\rm in}$.
(b) Derive an expression for $V_{\rm OUT}$ versus $V_{\rm IN}$. That is, plot $V_{\rm OUT}$ versus $V_{\rm IN}$.
(c) On your previous sketch, overlay the magnitude voltage transfer function of the amplifier.
(d) Derive an expression for $P_{\rm OUT}$ versus $P_{\rm IN}$.
(e) What is the linear voltage gain of the amplifier?
(f) What is the linear power gain in dB of the amplifier?

17. The RF output of a cell phone has a driver amplifier followed by a power amplifier. The driver amplifier has a linear gain of 30 dB and an output 1 dB compression point of 20 dBm. The power amplifier has a linear gain of 12 dB and an output 1 dB gain compression power of 39 dBm.
(a) What is the linear gain of the driver-power amplifier cascade?
(b) What is the output 1 dB gain compression power in dBm of the cascade?

18. A matched single-stage amplifier in a $50\ \Omega$ system has a linear gain of 16 dB and an output 1 dB gain compression power of 10 dBm. What is the amplitude of the maximum sinusoidal input signal when the gain of the amplifier is compressed by 1 dB?

19. An inductively biased Class A HBT amplifier is biased with a collector-emitter quiescent voltage of 5 V and a quiescent collector-emitter current of 100 mA. When operated at the 1 dB compression point, the input RF power is 20 mW and the output power is 300 mW. Consider that the RF signal is a sinewave, and note that the quiescent collector-emitter voltage will be the supply rail voltage.
(a) What is the quiescent DC power consumed by the transistor? Express your answer in milliwatts?
(b) What is the output power in dBm?
(c) What is the efficiency of the amplifier? Note that the efficiency of an inductively biased Class A amplifier can be more than 50% if distortion can be tolerated.
(d) What is the power-added efficiency of the amplifier?
(e) If the input power is reduced by 10 dB so

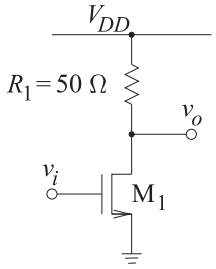
that the amplifier is no longer in compression, will the DC quiescent point change? Explain your answer.

(f) With 1 mW input power, what is the power-added efficiency of the amplifier if the quiescent point does not change? (The small-signal gain of the amplifier is 13 dB.)

20. An amplifier has a gain of 10 dB and an output power of 1 W. The amplifier has a power-added efficiency of 25%.
    (a) What is the total efficiency of the amplifier?
    (b) What is the efficiency of the amplifier?

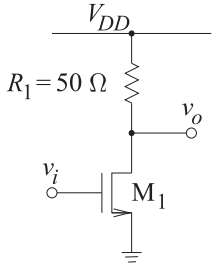
21. An amplifier with a gain of 20 dB and with a single-tone input-referred 1 dB gain compression power of 0 dBm is used to amplify a digitally modulated signal with a PMEPR of 8 dB. What is the average power in dBm of the output signal if the peak RF power is set equal to to the 1 dB gain compression level?

22. The distortion properties of the MOSFET circuit below are captured by the nonlinear transconductance equation $i_{DS1} = a_1 V_{GS1} + a_3 V_{GS1}^3$, where $a_1 = 0.01\text{A/V}$ and $a_3 = -0.1\text{A/V}^3$. [You can use interim results from Example 4.1.]



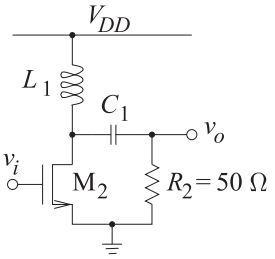
    (a) What is IIP3 in terms of voltage?
    (b) What is OIP3 in terms of voltage?

23. The distortion properties of the MOSFET circuit below are captured by the nonlinear transconductance equation $i_{DS1} = 0.02 V_{GS1} - 0.5 V_{GS1}^3$. [You can use interim results from Example 4.1.]



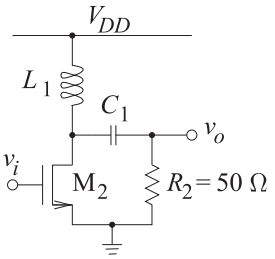
    (a) What is IIP3 in terms of voltage?
    (b) What is OIP3 in terms of voltage?

24. The distortion properties of the MOSFET circuit below are captured by the nonlinear transconductance equation $i_{DS2} = b_1 V_{GS1} + b_3 V_{GS2}^3$, where $b_1 = 0.05\text{A/V}$ and $b_3 = -0.2\text{A/V}^3$. [You can use interim results from Example 4.1.]



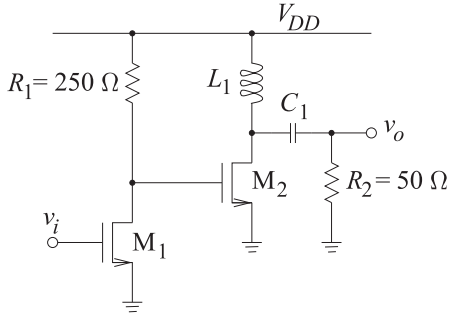
    (a) What is the IIP3 in terms of voltage?
    (b) What is the OIP3 in terms of voltage?

25. The distortion properties of the MOSFET circuit below are captured by the nonlinear transconductance equation $i_{DS2} = 0.1 V_{GS1} - 0.4 V_{GS2}^3$. [You can use interim results from Example 4.1.]



    (a) What is the IIP3 in terms of voltage?
    (b) What is the OIP3 in terms of voltage?

26. The distortion properties of the MOSFET circuit below are captured by the nonlinear transconductance equations $i_{DS1} = a_1 V_{GS1} + a_3 V_{GS1}^3$ and $i_{DS2} = b_1 V_{GS1} + b_3 V_{GS2}^3$, where $a_1 = 0.01\text{A/V}$, $a_3 = -0.1\text{A/V}^3$, $b_1 = 0.05\text{A/V}$, and $b_3 = -0.2\text{A/V}^3$. [You can use interim results from Example 4.1.]



    (a) What is the IIP3 in terms of voltage?
    (b) What is the OIP3 in terms of voltage?

4.13.1 Exercises By Section

†challenging, ‡very challenging

§4.2 1†
§4.3 2, 3, 4, 5, 6, 7†, 8†, 9†, 10, 11, 12, 13‡
§4.4 14†, 15†, 16†, 17†, 18†, 19†, 20, 21
§4.5 22†, 23†, 24†, 25†, 26†

4.13.2 Answers to Selected Exercises

2 13.75%
5 0 dB
7(d) 796 mW
8(d) 10 kW
11 180°
13(d) 13.98 dB
16(b) $5V_{\mathrm{IN}} - 0.375V_{\mathrm{IN}}^3$
16(f) 17.0 dB
19(f) 3.8%
24 4.33 V

CHAPTER 5

Oscillators

5.1 Introduction 159
5.2 Two-Port Oscillators 160
5.3 Reflection Oscillator 165
5.4 Case Study: Reflection Oscillator 167
5.5 Voltage-Controlled Oscillator 172
5.6 Case Study: Design of a C-Band VCO 176
5.7 Negative Transconductance Differential Oscillator 190
5.8 Advanced Discussion of Oscillator Noise 194
5.9 Case Study: Oscillator Phase Noise Analysis 206
5.10 Summary 209
5.11 References 210
5.12 Exercises 212

5.1 Introduction

Most RF oscillators generate sinusoidal signals that are either used to drive mixers or, if modulated, to produce frequency modulated signals directly. In some designs the microwave oscillators drive flip-flop circuits that produce periodic square-wave signals with multiple phases as required for the LO drives of I/Q modulators. RF and microwave oscillators can be designed using either a two-port or a one-port approach. The classic treatment of oscillators is based on a two-port gain device with a feedback loop, but the oscillator can nearly always be viewed, and thus more conveniently designed at microwave frequencies, as a one-port in which a resonant circuit, called the **tank** circuit, is connected to an active circuit that presents a negative resistance. However, the stability, noise, and start-up analyses of an oscillator are based on a two-port with feedback.

All microwave oscillator designs are based on one of three basic oscillator configurations. Suitable configurations must have as few reactive elements as possible while enabling stable single-frequency oscillation. These configurations have three or four reactive element. Mapping a microwave oscillator design on to one of the standard oscillator designs is not simple mainly because active devices at microwave frequencies have substantial parasitic capacitances and also adjustments must be made to accommodate biasing. Stability of a microwave oscillator is of great concern

and the essence is that there should be as few significant energy storage elements as possible. If there are parasitic energy storage elements these should be quite small or absorbed into the capacitors of one of the basic oscillator configurations. With more energy storage elements than necessary the chances of unwanted resonances is much higher and hence instability of an oscillator is more likely.

At microwave frequencies the Q of lumped elements is limited. A lumped inductor in the tank circuit has a particularly low Q and if there is room it is replaced by a transmission line. Oscillators that are fixed in frequency can use a resonant circuit with high Q circuit elements. This contrasts with voltage-controlled oscillators (VCOs) that have a lossy variable element, nearly always a varactor, in the tank circuit in what is now a low Q resonant circuit. The voltage-controlled variation of (invariably) the capacitance of this element changes the frequency of the oscillator.

A microwave oscillator could be realized on-chip or realized as a hybrid circuit with a packaged active device and packaged lumped-elements and possibly transmission lines for the resonant circuit. A hybrid design has much more flexibility than an on-chip design, and if designed correctly has better performance than a monolithically integrated design. Hybrid design techniques are much more mature than chip-based designs but over time some of the techniques used with hybrid designs will migrate to on-chip designs. Also, on-chip designs are preferably differential with transistors in a push-pull, i.e. differential, configuration. Better performance of an on-chip oscillator can be obtained by using an off-chip resonator.

Oscillator theory derives from the analysis of a two-port with gain and feedback. This theory is described in Section 5.2. The following sections explore practical oscillator configurations. Section 5.3 presents the design technique for designing a fixed-frequency microwave oscillator and this is followed up with a design case study in Section 5.4. The distinguishing feature here is that the resonant circuit consists of high Q elements. Section 5.5 presents a design approach for a voltage-controlled microwave oscillator. A state-of-the-art case study of a 5 GHz VCO design is presented in Section 5.6. Section 5.7 describes the design of an on-chip differential oscillator. The last two sections, Sections 5.8 and 5.9 in this chapter describe oscillator phase noise, a characteristic that is the fundamental performance limiting parameter of an oscillator as, for example, it affects the sensitivity and performance of communication systems and the range of radar.

5.2 Oscillator Theory

Microwave oscillators are usually implemented as reflection oscillators with two connected one-port circuits with one being an active device configured as a one-port and presenting a negative conductance, and a second one-port network being the tank or resonator network which must be designed to have specific admittance versus frequency characteristics that ensure stability. However the great body of network theory available derives from control theory and general circuit theory, and is based on a two-port oscillator with linear frequency-selective feedback.

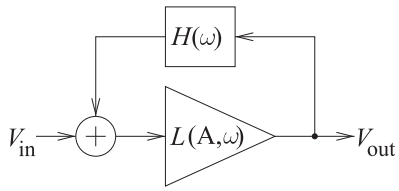


Figure 5-1: Representation of a feedback oscillator based on a two-port active device.

5.2.1 Theory of Oscillation

The two-port view of oscillators is based on the amplifier plus feedback loop shown in Figure 5-1. Here $L(A,\omega)$ describes the characteristics of an amplifier whose transfer function is dependent on both amplitude, A, and radian frequency, ω, and has a magnitude greater than 1. $H(\omega)$ describes the transfer characteristics of a linear feedback network that is dependent only on frequency. The output of the feedback system is described by

$$V_{\text{out}} = L(A,\omega)V_{\text{in}} + L(A,\omega)H(\omega)V_{\text{out}}, \tag{5.1}$$

and so

$$V_{\text{out}} = \frac{L(A,\omega)V_{\text{in}}}{1 - L(A,\omega)H(\omega)}. \tag{5.2}$$

The aim of most oscillator design is to use an active device whose characteristics are independent of frequency but whose transfer response depends on the output amplitude. Then Equation (5.2) becomes

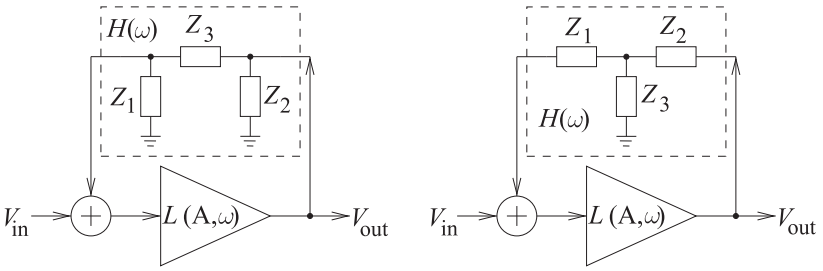
$$V_{\text{out}} = \frac{L(A)V_{\text{in}}}{1 - L(A)H(\omega)}. \tag{5.3}$$

Oscillation begins with input noise when the oscillator is powered on. If the denominator of Equation (5.3) is close to zero, oscillations build up at a frequency determined by the feedback network. As the amplitude of the oscillations builds, $L(A)$ compresses until the denominator is finite but close to zero[1], and there are stable oscillations. Stable oscillation is no accident and is the result of careful design, and much depends on the nature of the feedback network. When an amplifier becomes unstable, for example, the signal produced is often chaotic, with rapid variations in amplitude and frequency, and is not the single frequency sinusoid required of an oscillator.

5.2.2 Basic Oscillator Configurations

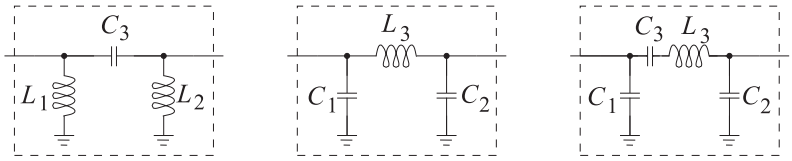
There are two dominant types of feedback networks, the Pi- and T-type networks, shown in Figure 5-2. Three Pi-type networks have proven to be particularly suited to the amplitude saturation characteristics of FET and BJT active devices and, ignoring parasitic capacitances, they result in stable oscillation. These are the Hartley, Colpitts, and Clapp circuits shown in Figure 5-3. The grounds shown can be considered to be common terminals instead.

[1] This description of oscillation, based on Equation (5.3), is behind the erroneous **Barkhausen stability criterion**, which is also known as the **Barkhausen oscillation criterion**. Barkhausen himself used the criterion, known as the **Barkhausen criterion**, to establish the frequency of oscillation as $L(A,\omega)H(\omega) = 1$. This was misinterpreted as a stability or oscillation criterion. It is a necessary criterion for two-port feedback oscillation, but not sufficient. It does not indicate whether a system is unstable. Instead, the Nyquist criterion is the necessary and sufficient criterion for oscillation in feedback oscillators [1–4]. The Barkhausen criterion should not be used in determining whether oscillation occurs.



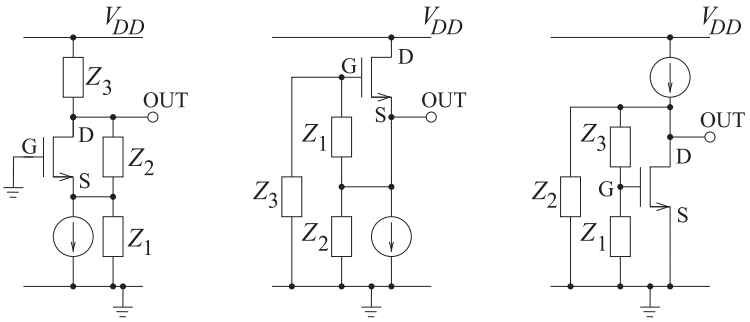
(a) Pi feedback network (b) T feedback network

Figure 5-2: Feedback oscillator with Pi- and T-type feedback networks.



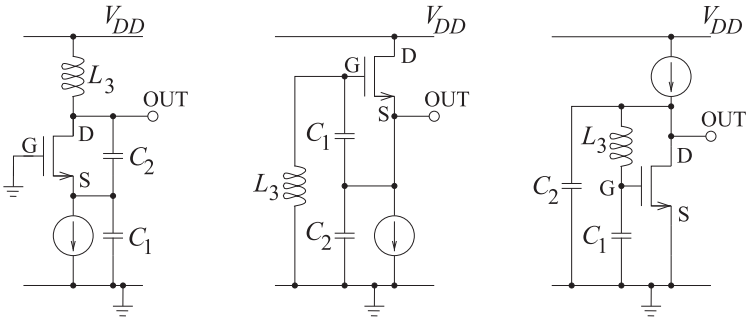
(a) Hartley (b) Colpitts (c) Clapp

Figure 5-3: Basic oscillator feedback networks.



(a) Common gate (b) Common drain (c) Common source

Figure 5-4: Circuit schematics for FET feedback oscillators using Pi-type feedback networks. The current sources provide bias.



(a) Common gate (b) Common drain (c) Common source

Figure 5-5: Circuit schematics for Colpitts FET oscillators. The current sources provide bias. (In (a) C_1 is across the inputof the gate-source transistor with connection through the ground.)

The Hartley and Colpitt's basic oscillator configurations have three large energy storage elements. The Clapp oscillator has four significant energy storage elements but it is important because sometimes it is desirable to use a very high Q resonator such as a piezoelectric crystal which is electrically modeled as a series LC circuit.

Single transistor oscillators using the feedback networks arranged in common-gate, common-source or common-drain configurations are shown in Figure 5-4. FET Colpitts oscillators are shown in Figure 5-5. The Colpitt's configuration is the most common configuration for a microwave oscillator

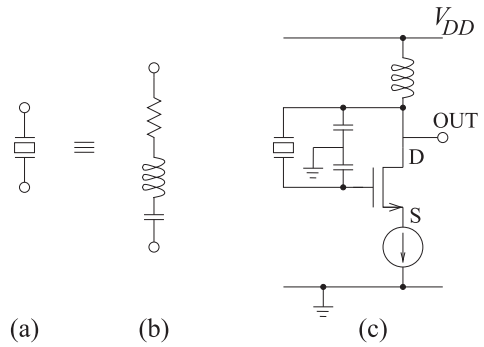


Figure 5-6: Crystal oscillator: (a) schematic symbol for a crystal; (b) its equivalent circuit; and (c) a FET crystal oscillator. The most commomly used form of a crystal oscillator is a Clapp oscillator because of the series LC electrical model of a piezolectric crystal.

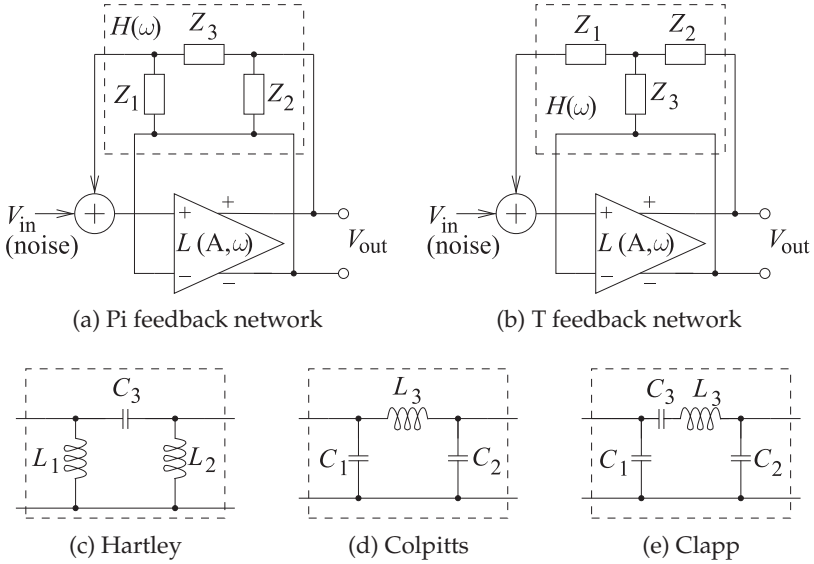


Figure 5-7: Differential feedback oscillator with Pi- and T-type feedback networks.

as it can absorb the parasitic capacitances of a transistor into the C_1 and C_2 capacitors, as in Figures 5-5(a and c). The circuits shown in Figures 5-2–5-5 are the bases of nearly every microwave oscillator. Many modifications lead to better stability, compensate for transistor parasitics, and accommodate differential signaling.

Crystal references, for example, are commonly used for fixed-frequency oscillators, as the piezoelectric (usually quartz) crystals used have a very high Q and the oscillation frequency is very stable, typically to a few parts in a billion, and better if the crystal is temperature stabilized. A circuit with a crystal-based feedback loop creating a Clapp FET oscillator is shown in Figure 5-6. Common quartz reference crystals available as off-the-shelf components are 10 MHz, 20 MHz, and 40 MHz but are available up to 300 MHz. So they are not microwave oscillators but they are used as reference signals in a phase-locked loop to precisely set the frequency of microwave VCOs. Fixed-frequency microwave oscillators do not have this method available for stabilizing the oscillation frequency.

The Hartley, Colpitts, and Clapp networks are also widely used as the bases of oscillators in CMOS RFICs, where the feedback network is

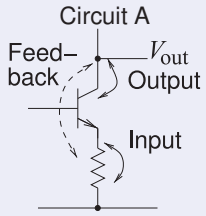
closely approximated using differential circuits [5–10]. The basic differential configurations are shown in Figure 5-7.

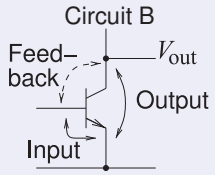
EXAMPLE 5.1 Common Emitter BJT Clapp Oscillator

Draw the schematic of a common-emitter BJT Clapp oscillator without biasing and then draw it with biasing.

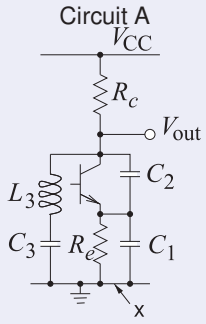
Solution:

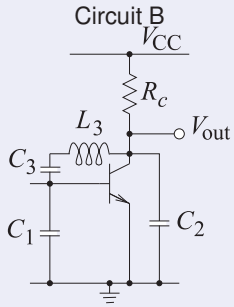
The first step is defining the input and the output connections to be used in the oscillator configuration. There are several ways to do this and the solution here considers two configurations, Circuit A and Circuit B.



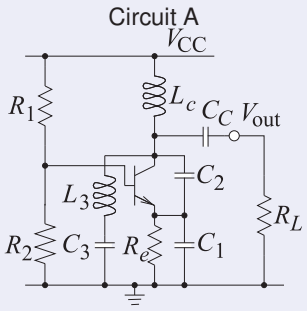


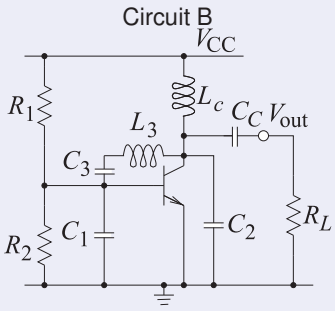
Thus the possible common-emitter Clapp oscillators are (as in Figure 5-4 but with a BJT)





In Circuit A the point x is the input terminal of the oscillator and the emitter terminal is the common point of the basic Clapp oscillator configuration. Appropriate bias must be applied to the base of the transistor. Also it would be good not to deliver power to R_c and instead deliver it to the load. In the circuits below a choke inductor L_c, an open circuit at RF, is used to apply bias. L_c also enables a large output voltage swing. C_c is a large coupling capacitor and has very low impedance at RF and blocks DC.





Of these Circuit B is the more attractive as the parasitic base-emitter capacitance will be absorbed into C_1, and the parasitic collector-emitter capacitance will be absorbed into C_2. Neither configuration can absorb the collector-base capacitance, C_{cb}, so extra care must be made that C_{cb} does not result in instability.

5.3 Reflection Oscillators

Design of stable microwave oscillators traditionally uses the one-port oscillator stability criterion outlined by Kurokawa [11, 12]. Most two-port oscillators can be designed by casting them in the form of a one-port oscillator. In applying the condition, each of the networks—the active device, the resonator load, and the device termination—are characterized as one-ports.

5.3.1 Kurokawa Oscillation Condition

The oscillation condition for a stable reflection oscillator can be expressed in terms of the admittances of the resonator network, $Y_r = G_r + \jmath B_r$, and of the active network, $Y_d = G_d + \jmath B_d$, see Figure 5-8(a). The Kurokawa oscillator condition establishes that for stable (single-frequency) oscillation [11]

$$\left(\frac{\partial G_d}{\partial V}\frac{\partial B_r}{\partial \omega} - \frac{\partial B_d}{\partial V}\frac{\partial G_r}{\partial \omega}\right)\Bigg|_{V=V_0,\omega=\omega_0} > 0, \tag{5.4}$$

where the subscript 0 refers to the operating point, r refers to the resonator, and d refers to the active device. The active device and resonator characteristics shown in Figures 5-8(b and c) satisfy the Kurokawa condition. In Equation (5.4) V_0 is the amplitude of the oscillation at the interface of the active and resonator networks. If the condition in Equation (5.4) is not met, then the oscillator may simultaneously oscillate at multiple frequencies. The Kurokawa oscillation condition must be met at all times which includes during start-up, including when bias is applied, of the oscillator.

With a fixed-frequency oscillator the resonator network is linear so G_r and B_r are independent of amplitude and G_r (by design) is independent of frequency. B_r varies with frequency. Ideally the active device has a frequency independent G_d and amplitude independent B_d Achieving this is a major design task. With these conditions $\partial B_d/\partial V \approx 0$ and $\partial G_r/\partial \omega_0 \approx 0$ and the

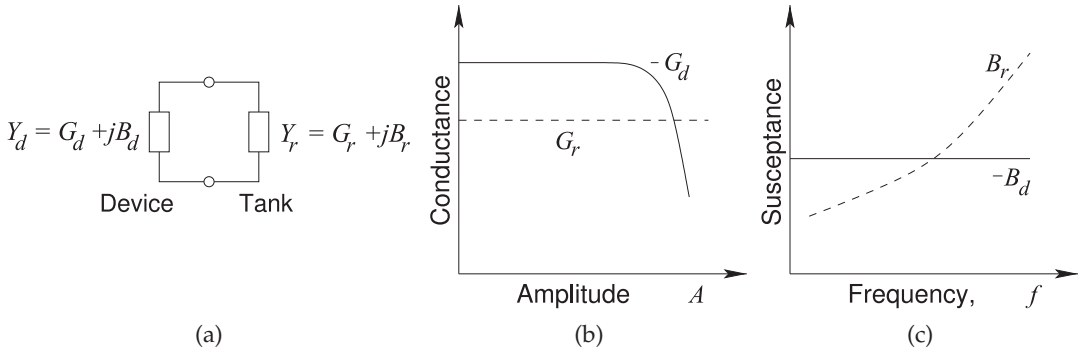


Figure 5-8: Reflection oscillator operation: (a) one-port oscillator; (b) as the amplitude of the oscillation increases, the magnitude of the device conductance, $|G_d|$, decreases while the conductance of the tank circuit, G_r, is constant; and (c) as the frequency of the oscillation increases, the susceptance of the tank circuit, B_r, changes while, B_d (ideally) does not change.

Kurokawa condition in Equation 5.4 simplifies to

$$\left(\frac{\partial G_d}{\partial V}\frac{\partial B_r}{\partial \omega_0}\right)\Bigg|_{V=V_0,\omega=\omega_0} > 0. \tag{5.5}$$

Thus design is greatly simplified for a fixed-frequency oscillator with a high-Q resonator.

5.3.2 Reflection Oscillator Design Approach

When a device with admittance $Y_d = G_d + \jmath B_d$ is connected in shunt to a resonator of admittance $Y_r = G_r + \jmath B_r$ (see Figure 5-8(a)), the voltage amplitude A and radian frequency ω of the resulting equilibrium oscillation are determined when $-G_d(A) = G_r(\omega)$ and $-B_d(A) = B_r(\omega)$. Here the assumption is that the device conductance only is a strong function of voltage amplitude, while the resonator admittance is a function only of angular frequency. This condition can be represented graphically by first denoting the locus of the negative of the device's complex admittance as $-Y_d(A) = -[G_d(A)+\jmath B_d(A)]$ (also referred to as the inverse device reflection coefficient, or $1/\Gamma$ locus (sometimes referred to as the $1/S$ locus)) and the locus of the resonator admittance as $Y_r(\omega) = G_r(\omega)+\jmath B_r(\omega)$. Then, for stable single-frequency oscillation, the intersection of these loci occurs at a single point (i.e., at a single amplitude and frequency combination).

In most oscillator design, the aim is to make the device admittance independent of frequency, and, of course, the admittance of the linear tank circuit is independent of the amplitude of the oscillating signal. These oscillation conditions are depicted in Figure 5-8(b and c). In Figure 5-8(b), as the amplitude of the oscillation increases, the magnitude of the device conductance, $|G_d|$, decreases while the conductance of the tank circuit, G_r, is constant. In Figure 5-8(c), as the frequency of the oscillation increases, the susceptance of the tank circuit, B_r, changes while B_d (ideally) is slow to change. The intersections define the amplitude and frequency of the oscillation.

If the device admittance is dependent on frequency, then it is difficult to avoid multiple intersections of the $-Y_d$ $(1/\Gamma_d)$ and Y_r (Γ_r) loci as viewed in the complex plane. The angle of intersection of the Y_r and $-Y_d$ loci is an important indicator of stability relating to multiple oscillations, oscillator start-up problems, and excess noise [13]. Thus the appropriate angular intersection of these loci is critical. It is difficult to achieve all of the objectives in design unless Y_d is frequency independent.

Ideally, resonator design requires that Q be maximized so that $G_r \approx 0$. This can be achieved with a fixed-frequency microwave oscillator as the resonator can be implemented with a capacitor and a transmission line segment both of which have very low loss. However the tunable capacitors in a VCO design are lossy and so the Q is not high. The two types of oscillators need a different design approach. Furthermore, for a VCO, voltage tuning of the resonator must satisfy the stability criteria, including a single point of intersection and appropriate angle of intersection, over the tuning range. With emphasis on these characteristics and since there are device capacitive parasitics, achieving a proper stable resonator-device interface can be troublesome. An alternative and equally viable approach to stability analysis of a broad class of oscillators, particularly for those using

three-terminal devices, is application of the two-port criteria developed for amplifier stability assessment. However, the one-port design approach is preferred by microwave designers because the one-port connection is closer to the intended operation. The one-port assessment of oscillator stability is not unlike the Bode criteria applied to two-port feedback systems [14, 15]. However, unlike the two-port open-loop assessment of stability, the one-port characterization technique is conveniently aligned with the measurements that can be made by a VNA [16, 17]. As well, the nonlinear limiting effect of the active device is readily measured.

5.3.3 Summary

Microwave oscillator design invariably uses the reflection oscillator approach in which a one-port active device network is connected to a one-port resonator network. Design is complicated enough and it is necessary to simplify design and limit the design space. The procedure almost universally followed is to design for the characteristics shown in Figure 5-8. Design of a fixed-frequency oscillator is further simplified because the conductance of the resonator is almost zero.

5.4 Case Study: Reflection Oscillator

In this case study[2] the design of the reflection oscillator shown in Figure 5-9(a) is examined. This is an 18 GHz common-gate oscillator with series inductive feedback provided by the transmission line TL_1. This circuit is derived from the Colpitts oscillator configuration. The resonator is resonant considerably below the oscillation frequency, and so presents a capacitance to the source of the transistor at the oscillation frequency.

5.4.1 Design Procedure

The circuit in Figure 5-9(a) is a modified common-base Colpitts oscillator and the RF equivalent circuit required to understand oscillator operation, and how the circuit corresponds to the standard Colpitts oscillator, are shown in Figure 5-10. This is the most common fixed-frequency microwave oscillator topology.

The Colpitts feedback network is shown in Figure 5-10(a) where an inductor provides feedback from the output to the input of the active device. Returning to the modified Colpitts oscillator of Figure 5-9, L_1 is a large choke inductor presenting an RF open circuit and C_1 is a large DC blocking capacitor presenting an RF short circuit. The gate transmission line, TL_1, presents a small inductance and most importantly provides inductive feedback from the output to the input of the active device and closely corresponds to L_3 in the Colpitts feedback network of Figure 5-10(a). The gate-source and drain-source parasitic capacitances of the transistor, C_{GS} and C_{DS}, correspond approximately to capacitors C_1 and C_2 in the Colpitts feedback network. This results in the RF equivalent circuit shown in Figure 5-10(b). In Figure 5-10(b) X_R is the reactance of the resonator and compensates for the nonideal nature of the modified Colpitts oscillator. Stable operation

[2] AWR Design Environment Project File: Reflection_Oscillator.emp.

of this oscillator requires that X_R be capacitive but have a variation with frequency substantially less than that of a capacitor. This is accomplished by having the oscillation frequency above the resonant frequency of the resonator.

In Figure 5-9 the reflection oscillator is composed of two networks, the active device network and the resonator network. The third network shown, the OscProbe network, is a probe used to control oscillator analysis using the harmonic balance method. The design approach here is to develop a topology incorporating the active device that presents a negative resistance at the interface, X, between the active device network and the resonator network. Since the active device and the resonator are each best modeled as parallel circuits, it is best to refer to admittance and so the active device network presents a negative conductance to the resonator. In addition, the small-signal admittances of the active network, $Y_d = G_d + jB_d$, and of the resonator network, $Y_r = G_r + jB_r$, are shown in Figure 5-11. Here the active network has a small-signal conductance that is negative and a small-signal

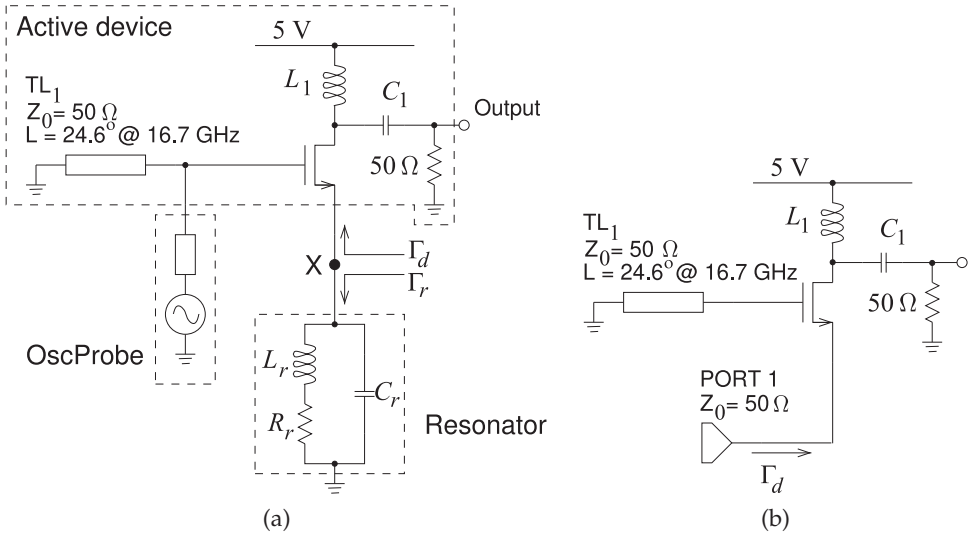


Figure 5-9: Reflection oscillator: (a) complete oscillator circuit used in simulation; and (b) configuration for measuring the large-signal reflection coefficient of the active device. $L_r = 5.6$ nH, $R_r = 10\ \Omega$, $C_r = 445$ fF, $L_1 = 15$ nH, and $C_1 = 10$ pF. The resonant frequency of the resonator is 3.19 GHz, but this is not the oscillation frequency. It presents the required slope of susceptance with respect to frequency at the oscillator frequency of 17.76 GHz.

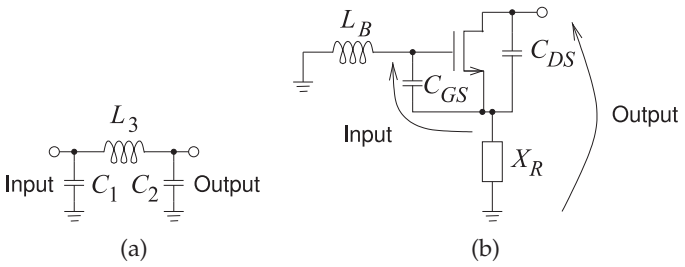


Figure 5-10: Operation of the modified Colpitts oscillator: (a) Colpitts feedback network; (b) the essential RF equivalent circuit of the Modified Colpitts oscillator of Figure 5-9.

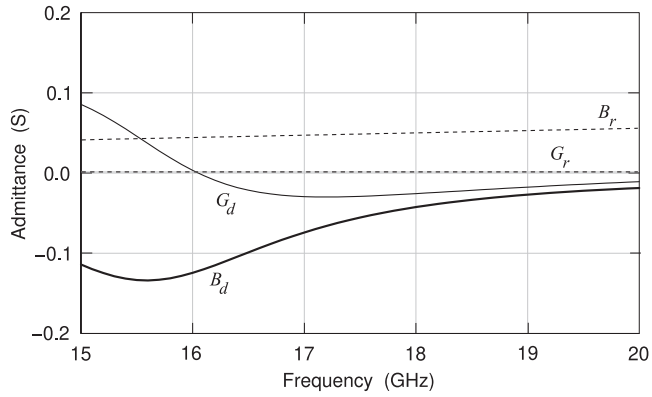


Figure 5-11: Small-signal admittance of the active device, $Y_d = G_d + \jmath B_d$, and admittance of the resonator, $Y_r = G_r + \jmath B_r$. B_d varies with frequency largely because of the frequency-dependent feedback provided by TL_1.

susceptance that is inductive. The resonator has negligible conductance and a capacitive susceptance.

The admittance looking into the source of the active network depends on the level of the signal at X in Figure 5-9. However, the admittance of the resonator is independent of the signal level since the resonator is a linear network. The design strategy is then to develop a feedback network, here the transmission line, TL_1, in the gate of the FET, that then presents a negative conductance to a frequency-selective structure, the resonator network.

The effect of the signal level is examined using the active network test circuit shown in Figure 5-9(b). Here a 50 Ω generator, at Port 1, drives the source of the active device. The power of the signal at the port is varied and it is found that the negative conductance varies from -0.0274 S at a small applied signal, to -0.0224 S at an applied signal level of -10 dBm, to -0.004 S at -7 dBm, and to 0.001 S at -6 dBm. Oscillation will occur when the conductance looking into the source terminal of the active device is approximately zero (since the conductance of the resonator is negligible) and the susceptances of the active network and the resonator network cancel.

Unfortunately the susceptance of the active network changes as the signal level varies since the active device capacitances are nonlinear. The effect of this will now be examined. It is instructive to view the effect of signal level on the oscillation condition by considering the reflection coefficients Γ_d for the active device network and Γ_r for the resonator network. This is shown in Figure 5-12, where a polar plot is used since $|\Gamma_d| > 1$ (except for a very large signal at X). Γ_r is approximately on the unit circle and is capacitive being in the lower half plane. Γ_d is shown for several signal levels, with a signal level of -20 dBm corresponding to the small-signal condition. As the signal level increases eventually to -6 dBm, both the conductance and susceptance of the active device change. The device conductance becomes positive as Γ_d crosses inside the unit circle. Resonance occurs when $\Gamma_d\Gamma_r = 1$. Since $G_r \approx 0$, this is when $\Gamma_d \approx 1/\Gamma_r$. That is, resonance will occur at the frequency where the locus of Γ_d intersects with $1/\Gamma_r$.

The results presented in Figure 5-12 for the reflection coefficient of the active network in Figure 5-9(b) are derived from a large-signal solution calculated using harmonic balance analysis. In most harmonic balance analyses it is only necessary to consider a few harmonics to obtain good results. In the simulation here, five harmonics are considered, with the fundamental frequency set by the frequency of the source at the driven port. The fundamental frequency was stepped from 15 GHz to 20 GHz. Oscillator

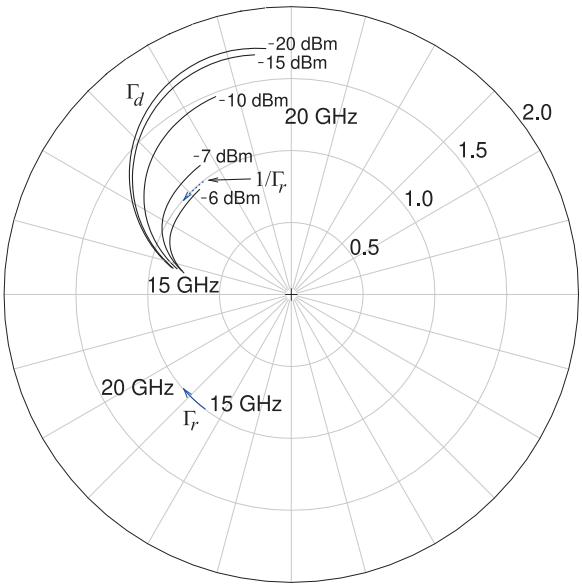


Figure 5-12: Reflection coefficient of the resonator, Γ_r, and of the active device, Γ_d, at various signal levels. Γ is plotted on a polar plot with the outer circle corresponding to $|\Gamma| = 2$. As desired, the locus of $1/\Gamma_r$ is parallel to Γ_d at a fixed signal level (here about -6 dB), and the direction of $1/\Gamma_r$ with increasing frequency is opposite that of Γ_d. The close parallel match is ensured by adjusting the susceptance-versus-frequency slope of the resonator. Considering the intersection of $1/\Gamma_r$ and Γ_d is equivalent to considering the intersection of $1/\Gamma_d$ and Γ_r.

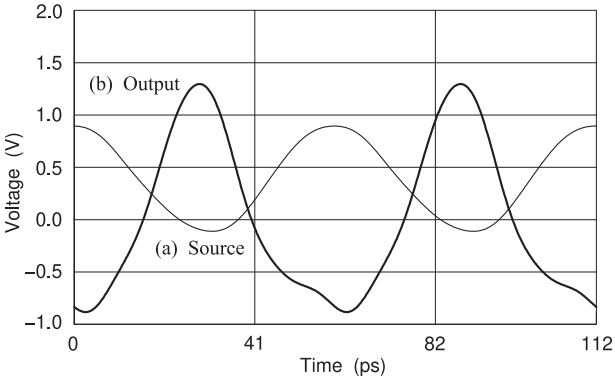


Figure 5-13: Voltage waveforms at the output of the oscillator (Curve (b)), and at the source terminal of the active device, Terminal x, which is the interface between the active device and the resonator (Curve (a)).

simulation also uses harmonic balance analysis, but now the frequency of oscillation is not known ahead of time. It is necessary to introduce another condition to enable the simulator to find the oscillation frequency. One of the techniques used in harmonic balance simulators is the introduction of an oscillator probe element such as the OscProbe element shown in Figure 5-9(a). The frequency of the source, F_{OSC}, in the OscProbe element is initially guessed by the simulator and the series impedance of the OscProbe element is a short circuit at the oscillation frequency and an open circuit at the harmonics. This extra condition is included in the harmonic balance equations and F_{OSC} is allowed to vary as well as the amplitude of the source. The oscillation solution is obtained when the current through the impedance in the OscProbe element is zero. The oscillation frequency is found to be 17.76 GHz. Note that this is not the resonant frequency of the resonator, as here the active device presents an inductance to the resonator. The resonator must present an effective capacitance that has the required susceptance-versus-frequency slope so that the frequency locus of $1/\Gamma_r(f)$ is parallel but oppositely directed to that of $\Gamma_d(f)$ (see Figure 5-12).

Figure 5-13 shows the waveforms at the output of the oscillator (Curve (b)) and at the interface (X) between the resonator and active device networks

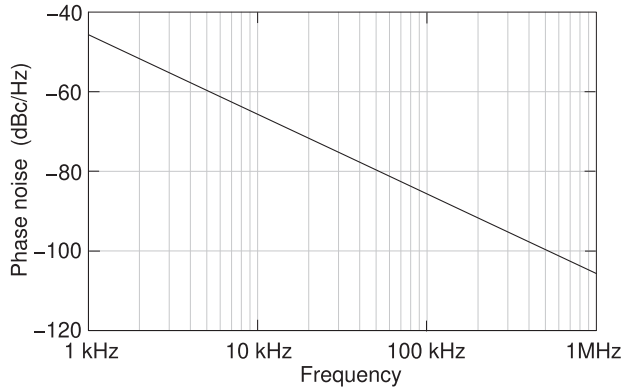


Figure 5-14: Phase noise of the reflection oscillator.

(Curve (a)). The output of the oscillator would be followed by a bandpass filter so that a single harmonic-free sinewave is presented to the external output terminal.

The phase noise of the oscillator is calculated from the oscillator steady-state conditions and is shown in Figure 5-14. The phase noise level is relatively high and this is principally due to the resistance R_r in the resonator, which models the loss in the lumped inductor. Performance would be improved if, instead, a distributed circuit (with transmission lines) was used to realize the required admittance and its derivative at the oscillation frequency.

5.4.2 Summary

The zero combined suspectance point in Figure 5-11 was at 17.76 GHz while the oscillation frequency shown in Figure 5-13 is at 17.29 GHz.The reason for this discrepancy is that the susceptance plot in Figure 5-11 is based on small-signal conditions whereas the simulation is for a large signal and includes harmonics. It can be expected that the oscillation frequency calculated using harmonic balance analysis will be different from that obtained from a small-signal analysis. Consider the Γ_d curves in Figure 5-12. (Generating this data took many simulation runs and did not consider harmonics.) The magnitude of Γ_d reduces as the applied signal level increases. This rotation is not desired and means that the susceptance of the active device is amplitude-dependent. This corresponds to the magnitude of the negative conductance of the active device getting smaller. Note that there is also a small rotation of Γ_d indicating that the susceptance of the active device also changes as the signal level increases. This means that the oscillation frequency will depend on the amplitude of the oscillation. In addition, the applied signal used to calculate the Γ_d at different power levels is not the actual oscillating signal.

5.5 Voltage-Controlled Oscillator (VCO)

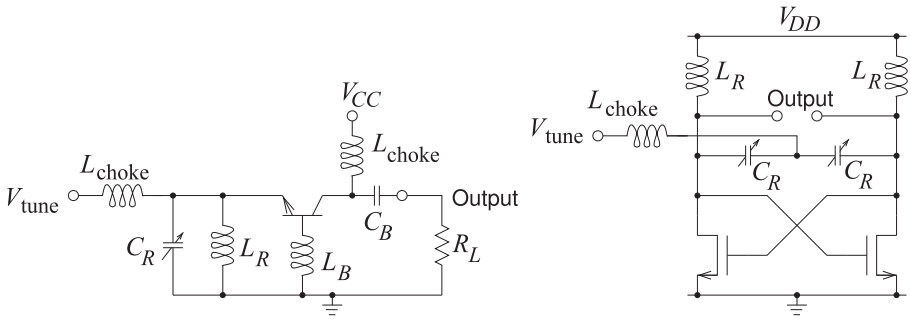
This section describes the special challenges in designing a microwave VCO. The main challenge derives from the resonator network now being lossy because of the use of a tuning element that unfortunately has low Q. In nearly all cases the tuning element is a varactor, a reverse-biased pn-junction or Schottky diode, which has substantial series resistance at microwave frequencies. This means that the full Kurokawa conditions for stable oscillation must be used and also these conditions must be met across the tuning range of the oscillator and during start-up.

5.5.1 Design Procedure

A broadband VCO is a key element in a communication system and is commonly used to drive a mixer to enable different frequency bands to be selected simply by changing the oscillation frequency. The design of a broadband VCO is relatively complex, with issues of simultaneous oscillation at multiple frequencies, phase noise, and power efficiency being of primary concern. In a battery-powered communication device the power drawn by the VCO is a substantial fraction of the total power consumed by the RF front end.

The design of a VCO builds on the standard oscillator design procedure which matches the inverse reflection coefficient of the active device, $1/\Gamma_d$, to the reflection coefficient of a tank or resonator circuit, Γ_r, so that $\Gamma_r = 1/\Gamma_d$ only at oscillation. This is complicated at microwave frequencies as parasitic capacitances are significant and special design strategies are required to mitigate their effect. The required stable oscillation conditions must be achieved over the entire tuning range while also maintaining constant output power and low phase noise.

Two VCO circuits are shown in Figure 5-15. The circuit in Figure 5-15(a) is a common-base oscillator with the feedback provided by the base inductor, L_B, presenting a negative conductance and a small shunt inductance at the emitter. The resonant circuit comprising C_R and L_R presents either a capacitance or an inductance at the oscillation frequency, and cancels the susceptance presented by the active device. The resonant circuit has a non-negligible conductance due to loss in the resonator. Ideally the active device's susceptance is independent of amplitude, however, the transistor's parasitics complicate the situation so that this susceptance also has amplitude dependence. The resonator is tunable because of the variable



(a) Common-base BJT VCO

(b) Differential VCO

Figure 5-15: VCO with tunable resonator comprising C_R and L_R. L_{choke} is an RF open circuit and C_B is an RF short circuit.

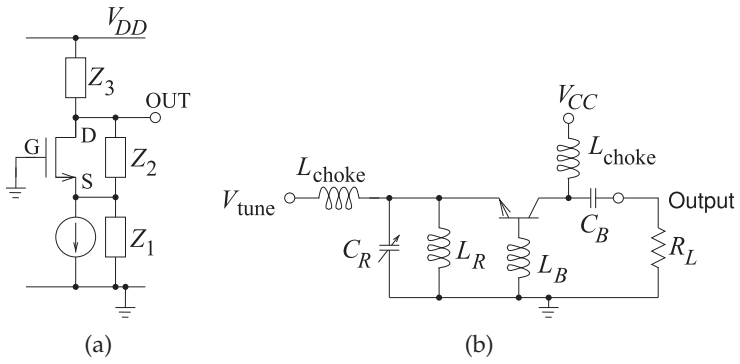


Figure 5-16: Common-gate/base feedback oscillator using a Pi-type feedback networks: (a) common-gate FET oscillator; and (b) common-base BJT oscillator.

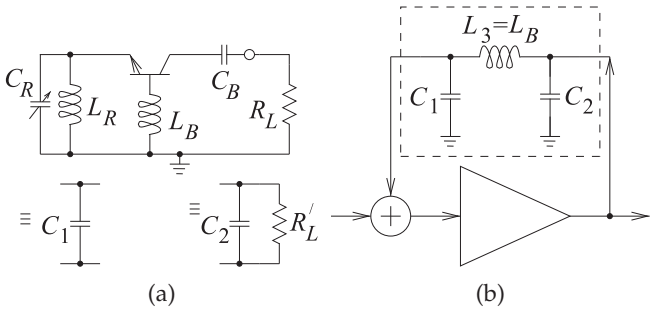


Figure 5-17: Transformation of the oscillator of Figure 5-16 into the form of a standard Colpitts feedback oscillator: (a) first stage of transformation replacing the L_{choke}s with RF open circuits, replacing C_B with an RF short circuit; and (b) final feedback form (compare with the circuits in Figures 5-2(a) and 5-3(b)). C_1 combines the inductance of the L_r-C_r resonator and the capacitive susceptance introduced by the device parasitics. C_2 derives mostly from the collector-emitter capacitance.

capacitance, usually a varactor.

The most common microwave single-transistor reflection oscillator topology is the common-base Colpitts BJT/HBT oscillator, or the similar common-gate FET oscillator, as shown in Figure 5-16. Consider the BJT oscillator in Figure 5-16(b). This is called a common-base configuration, although it is only approximate as will be seen when the RF signal flow path is traced. The parallel L_R-C_R combination is called a resonator, but it is resonant below (or above) the oscillation frequency, so at the oscillation frequency it is effectively a capacitor (or an inductor) but with an admittance-versus-frequency slope magnitude that is less than that of a lumped capacitor (or inductor).[3] Design of the susceptance slope is important to obtaining successful operation as a VCO. Thus one of the purposes of L_R is to enable $\partial B_r/\partial\omega$ to be selected in design. The base inductance L_B couples the output of the active device at the collector to the input at the emitter. So this circuit is a common-base Colpitt's oscillator (the equivalence is illustrated in Figure 5-17).

Parasitic and design elements (especially L_B and compensating capacitors) in the active network may require that the resonant network present an inductance. As long as the Kurokawa condition is met oscillation will be stable. VCO design requires considerable flexibility.

Figure 5-15(b) is the schematic of a differential oscillator realized using

[3] The base inductor L_B can be adjusted so that either a capacitive or inductive susceptance is presented to the resonator. As will be seen stable oscillation can be achieved either way.

MOSFETs, where the resonator comprises a pair of variable capacitances, the C_Rs, and a pair of inductors, the L_Rs. The chokes, the L_{choke}s, are short circuits at DC but block RF currents. On-chip design has many restrictions and it cannot be expected that the same performance can be obtained as with a hybrid design.

5.5.2 Managing Multioscillation and Phase Noise

The oscillation condition for a stable reflection oscillator needs to be revisited and discussed with respect to a VCO. Referring to the reflection oscillator in Figure 5-8, the Kurokawa oscillator condition for stable (single-frequency) oscillation is [11]

$$\left(\frac{\partial G_d}{\partial V}\frac{\partial B_r}{\partial \omega_0} - \frac{\partial B_d}{\partial V}\frac{\partial G_r}{\partial \omega_0}\right)\Bigg|_{V=V_0,\omega=\omega_0} > 0, \tag{5.6}$$

where the the subscript 0 refers to the operating point and V_0 is the amplitude of the oscillation at the interface of the active and resonator networks. The active network includes elements added to the active device mainly to compensate for parasitics. If the condition in Equation (5.6) is not met, then the oscillator may simultaneously oscillate at multiple frequencies or be chaotic.

Achieving single-frequency oscillation can be a challenge for fixed-frequency oscillator design but is especially challenging for a VCO as the condition must be met across the tuning range. Complicating design further is that with a VCO the resonator network is lossy so that the Kurokawa condition is not simplified as it is for a fixed-frequency oscillator, see Section 5.3.1. The condition is simplified then because of of the high Q resonator of a fixed-frequency oscillator where terms in Kurokawa's oscillation condition disappear because $G_r \approx 0$.

In VCO design the design procedure must be kept simple, and this opens up the design space to enable optimization of other characteristics such as minimizing phase noise and DC power consumption. One aspect of the preferred oscillator design procedure is to choose a topology that results in an effective device susceptance that as much as possible is independent of signal amplitude (i.e., $\partial B_d/\partial V|_{V=V_0} \approx 0$) and a loaded resonator conductance that is independent of frequency (i.e., $\partial G_r/\partial \omega_0|_{\omega=\omega_0} \approx 0$). Then the criterion for stable oscillation is the much simpler

$$\left(\frac{\partial G_d}{\partial V}\right)\left(\frac{\partial B_r}{\partial \omega_0}\right)\Bigg|_{V=V_0,\omega=\omega_0} > 0, \tag{5.7}$$

something that is much easier to satisfy in design. This equation is the same as the simplified criteria developed for the fixed-frequency oscillator but now there are two additional design requirements, i.e. $\partial B_d/\partial V|_{V=V_0} \approx 0$ and $\partial G_r/\partial \omega|_{\omega=\omega_0} \approx 0$.

Referring to the common-base circuit of Figure 5-17(a), the element in the base, L_B, induces a negative conductance at the interface with the L_R-C_R resonator [18]. It also induces an inductive susceptance and with the active device capacitance and other capacitors the total susceptance presented by the active network is either inductive or capacitive. A feature of the circuit in Figure 5-17(a) is that the L_R-C_R resonator is isolated from the load

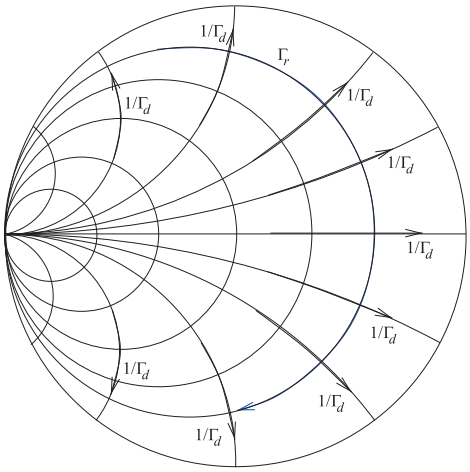


Figure 5-18: Requirements on Γ_r and Γ_d for stable oscillation as derived from Equation (5.7) with assumptions behind the derivation of the equation (i.e., $\partial B_d/\partial V|_{V=V_0} \approx 0$ and $\partial G_r/\partial \omega_0|_{\omega=\omega_0} \approx 0$). The locus of Γ_r is for increasing frequency. Several loci are shown for $1/\Gamma_d$ are shown at different frequencies. The arrowed loci of $1/\Gamma_d$ are for increasing frequency.

and so the load has little effect on the oscillation frequency. The output of the transistor is modeled as a shunt connection of a current source, a conductance, and a susceptance. Thus it is natural to treat the oscillator as a negative conductance oscillator. It is found that, as desired, this topology leads to a negative conductance that reduces in magnitude as the oscillation level increases. Another way of stating the same issue is saying that the L_r-C_r resonator network does not contain all of the reactance that should be assigned to C_1 in the simple Colpitts oscillator. Now $\partial G_d/\partial V|_{V=V_0}$ in Equation (5.7) is positive. (That is, G_d becomes less negative as the magnitude of the oscillating signal increases.) Thus for stable oscillation, $\partial B_r/\partial \omega|_{\omega=\omega_0}$ must be positive. The conditions for stable oscillation, using the design simplifications that led to Equation (5.7), are then as shown in Figure 5-18. What this figure indicates is that if Γ_d is capacitive ($1/\Gamma_d$ is in the top half plane) then Γ_r needs to be inductive, while if Γ_d is inductive ($1/\Gamma_d$ is in the bottom half plane) then Γ_r needs to be capacitive.

At microwave frequencies, parasitic reactances of a transistor are significant and so the active device does not have a signal-level independent susceptance (i.e., $\partial B_d/\partial V|_{V=V_0}$ is finite). In part this is because the parasitic capacitances of the transistor are voltage dependent. It is also due to the reactive feedback element used to produce the negative conductance. Parasitic elements further isolate the negative output conductance of a transistor from the port of the active network that interfaces with the resonator. As a result it may be necessary to introduce compensatory elements to translate the behavior of the active device into the correct form. This will be presented in the case study of the next section. The bottom line is that designing for $\partial B_d/\partial V|_{V=V_0} \approx 0$ is a major challenge. It is a less significant a problem to design the resonator to have the required conductance property (i.e., $\partial G_r/\partial \omega|_{\omega=\omega_0} \approx 0$).

5.5.3 Negative Resistance Oscillator

The above discussion concerns a negative conductance oscillator in which the magnitude of the device's negative conductance reduces with signal amplitude. Alternatively, a negative resistance oscillator could be realized using a capacitor in series with the emitter. This leads to an oscillator model in which the negative resistance is in series with a load resistance and

a series resonant circuit. Stable oscillation of this oscillator requires that the magnitude of the negative resistance reduce as the oscillation signal increases in size. However, it is difficult to realize a negative resistance that reduces in magnitude as the signal level increases and simultaneously achieve a reactance that is unchanged by the signal level. Thus it is better to use feedback to create a negative conductance shunting a susceptance. Design of a stable negative resistance oscillator is quite difficult and is nearly always avoided.

5.5.4 Summary

With the low Q resonator of a VCO the design procedure is quite different to that for a fixed-frequency oscillator which has a high Q resonator. The central reflection-based VCO design concept is the development of an active network presenting a near-voltage-independent susceptance (i.e., $\partial B_d/\partial V|_{V=V_o} \approx 0$) and a frequency-independent negative conductance ($\partial G_d/\partial \omega|_{\omega=\omega_0} \approx 0$). Once this is achieved, the conventional reflection oscillator design approach can be used and the resonator designed so that $\partial G_r/\partial \omega|_{\omega=\omega_0} \approx 0$. Design is both an art and a science. Sometimes the problem must be simplified for the designer to be able to conceptualize and synthesize the required circuit. This is usually the best approach to achieving microwave circuits that have near optimum performance. Simply optimizing a given topology is not enough. The correct topology must be synthesized first.

The requirements for oscillation start-up have not been considered yet, and these will be addressed in the following VCO design case study.

5.6 Case Study: Design of a C-Band VCO

This section presents the design of a high-performance microwave VCO operating from 4.5 to 5.3 GHz reported in [19]. The design objective is the generation of a frequency-independent negative conductance, $G_d(A)$, with a prescribed reflection coefficient shape, Γ_d, using a three-terminal active device in a common-base configuration with series-inductive feedback. The oscillator schematic is shown in Figure 5-19(a). This is a common-base oscillator in which the 2.2 nH base inductor L_3 provides negative feedback between the input and output of the transistor and the resonator presents an inductance at the oscillation frequency. So this circuit is a common-base Colpitts oscillator; the equivalence is illustrated in Figure 5-20.

5.6.1 Design Philosophy and Topology

Referring to Figure 5-19(a), the oscillator is partitioned into an active network to the right of the line x-x and a resonator network to the left of the dividing line. The negative conductance presented to the resonator is principally because of the feedback provided by the base inductor L_3. The choke, $L_{\text{CHOKE}} = 8.2$ nH, and associated elements provide DC bias to the transistor. The choke has an impedance magnitude of $258\ \Omega$ at 5 GHz and is effectively an RF open circuit. The output of the oscillator is taken from the collector of the transistor through a 2.2 pF capacitor that drives a $50\ \Omega$ load, Z_L. The emitter is connected to the resonator network through the parallel $47.5\ \Omega$ resistor and 10 pF capacitor. The 10 pF capacitors have an impedance

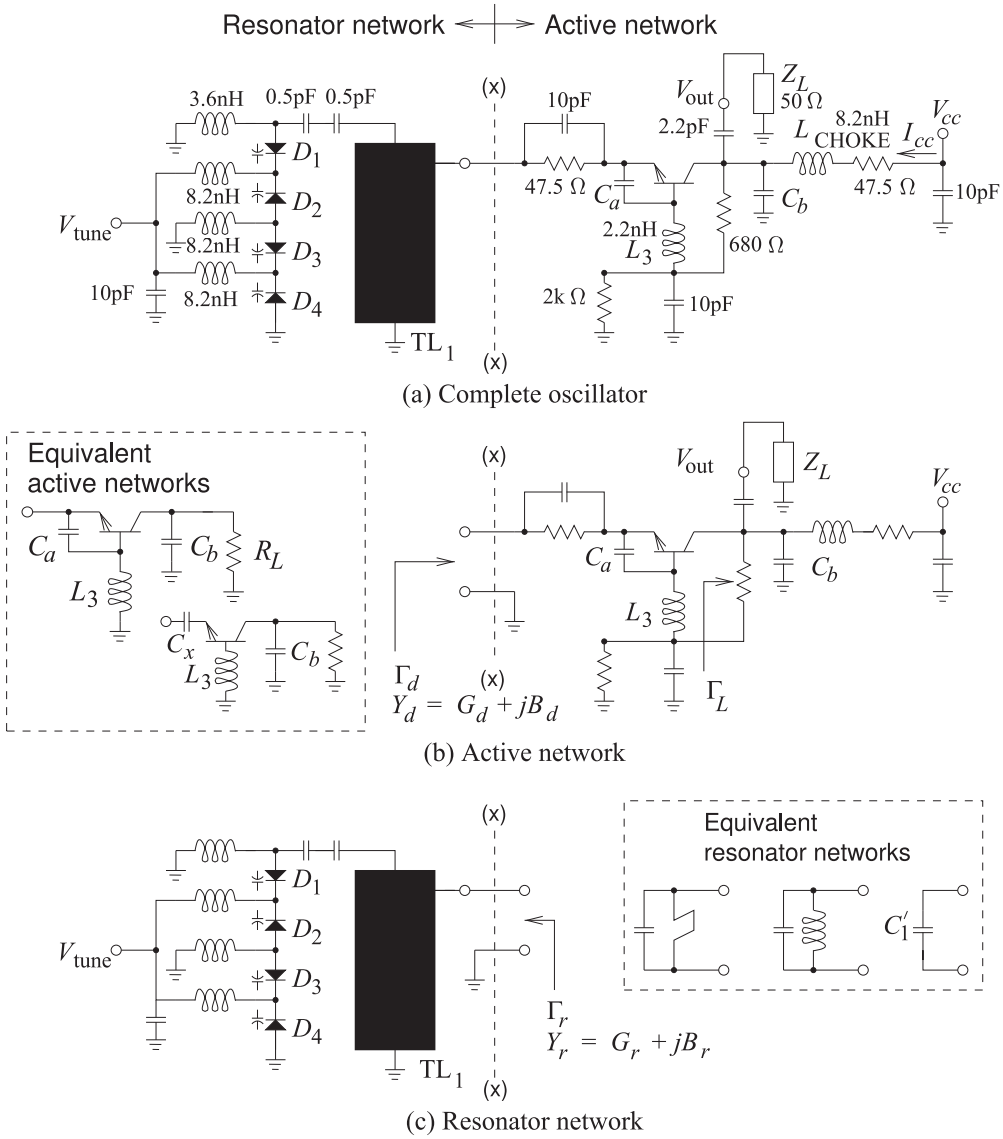


Figure 5-19: A 5 GHz common-base SiGe BJT oscillator: (a) oscillator showing the interface (x-x) between the resonator network (also called a tank circuit) and the device; (b) active network; and (c) resonator network. The element labeled TL_1 is a low impedance microstrip line. Capacitors C_a and C_b compensate for the frequency-dependent feedback provided by the base inductance. The choke inductor, $L_{CHOKE} = 8.2$ nH, presents an RF open circuit and is part of the bias circuit. $V_{CC} = 30$ V and $I_{CC} = 30$ mA. Each varactor diode (D_1–D_4) is model JDS2S71E. The transistor is a Si BJT model NE894M13, which is designed for oscillator applications above 3 GHz.

magnitude of 3 Ω at 5 GHz and are effectively RF short circuits.

The base inductor, L_3, provides feedback that results in negative conductance from the emitter to ground. Since the feedback is frequency dependent, this induced negative conductance is also frequency dependent.

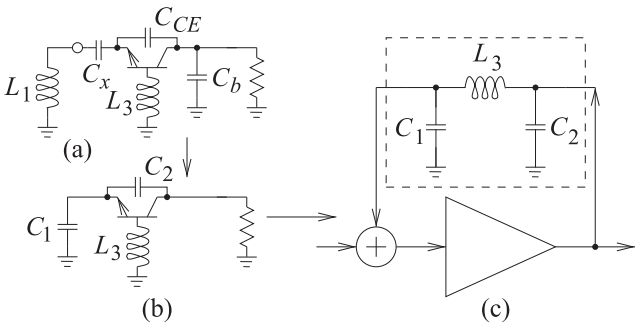


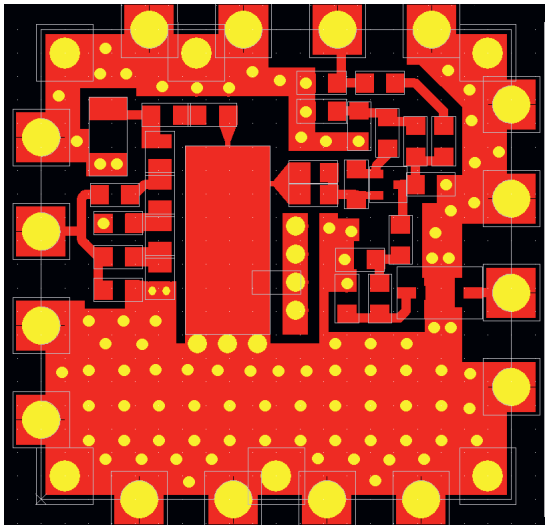
Figure 5-20: Transformation of the oscillator of Figure 5-19 into the form of a standard Colpitts feedback oscillator: (a) first stage of transformation combining the equivalent active and resonator networks in Figure 5-19; (b) combining L_1 and C_x (due to C_a and parasitics) to obtain an equivalent capacitance C_1; and (c) final feedback form (compare with the circuits in Figures 5-2(a) and 5-3(b)).

However, the reactive loading by C_a and C_b modifies the effective device conductance so that it becomes frequency independent. The design of C_a and C_b will be considered in depth later. The capacitors C_a and C_b are key to presenting an admittance to the resonator network that has the required characteristics for stable oscillation.

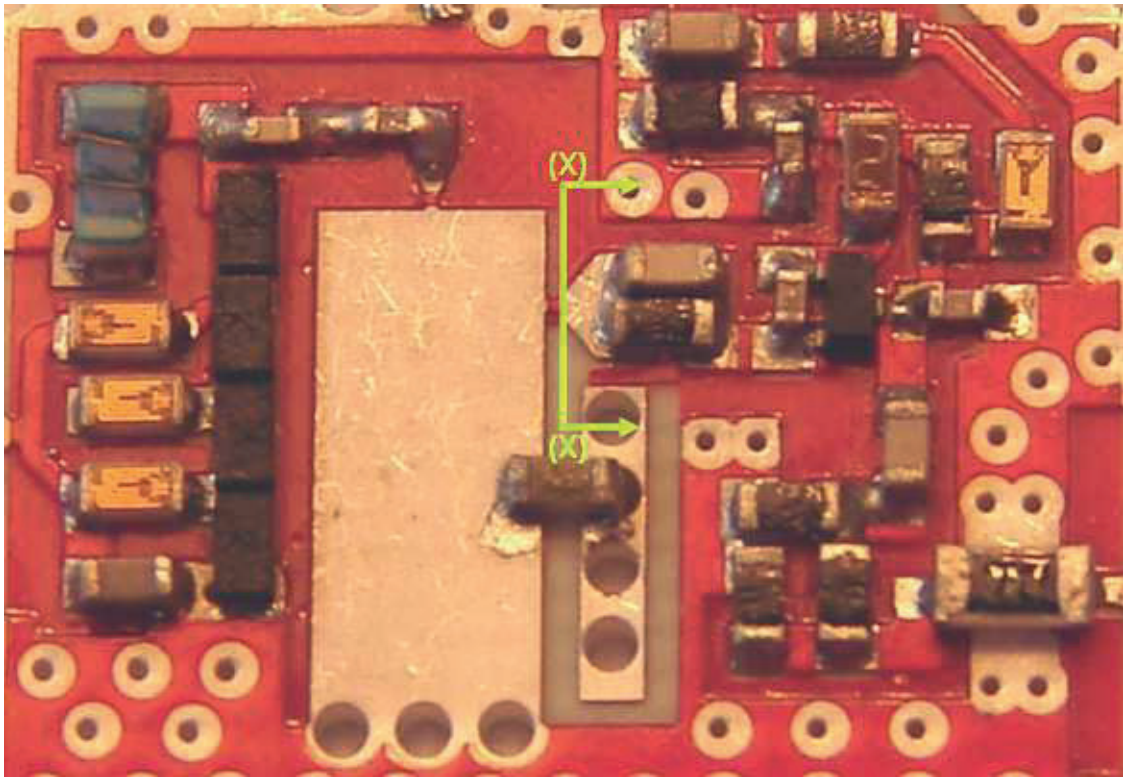
The resonator network, to the left of (x-x) in Figure 5-19(a), consists of a transmission line TL_1 that is coupled to a variable capacitance provided by the stack of four varactors. A single varactor would provide a voltage-tunable capacitance, but the stack of four varactors allows a four-times-larger RF voltage swing [20]. The 3.6 nH and 8.2 nH inductors provide DC shorts while presenting RF open circuits. The tapped transmission line, referring to the connection between the active and resonator networks not being at the top of the transmission line, improves the loaded Q of the resonator network. The resonator network is resonant at a frequency below the oscillation frequency, with the transmission line being inductive at the oscillation frequency. So at the oscillation frequency the resonator network presents an inductance with the required slope of admittance with respect to frequency[4]. The design of the varactor stack is discussed further in Section 5.3 of [21].

The layout and populated microstrip circuit board are shown in Figure 5-21(a and b). The resonator network is shown in Figure 5-22(a) and the active network is shown in Figure 5-22(b). Note the many vias to the backing ground plane. This is typical of microwave designs, as the vias eliminate substrate modes and the large grounded regions minimize parasitic coupling. The wide microstrip resonator, TL_1, is seen in Figure 5-22(a) and there is a shorting bar (available in chip form as a $0\ \Omega$ resistor) across it to a ground strip. The bar can be repositioned and used to tune the resonator network. While this is necessary during manual oscillator design optimization, it is retained in the final design (as is usual). The shorting bar effectively reduces the length of TL_1. At the output of the oscillator (see Figure 5-22(b)) a Pi attenuator (with $294\ \Omega$ resistors in the shunt legs and a $17.4\ \Omega$ series resistor) is between V_{out} and the $50\ \Omega$ bandpass filter. The filter blocks harmonics from the final output of the circuit. The impedance seen from the V_{out} terminal looking toward the filter is $50\ \Omega$.

[4] The resonator could be adjusted to present a capacitive susceptance.



(a)



(b)

Figure 5-21: C-band VCO circuit: (a) layout showing metalization and vias to ground planes (in yellow); and (b) populated circuit board with the resonator network to the left of the cutaway line (x-x) separated from the active circuit to the right.

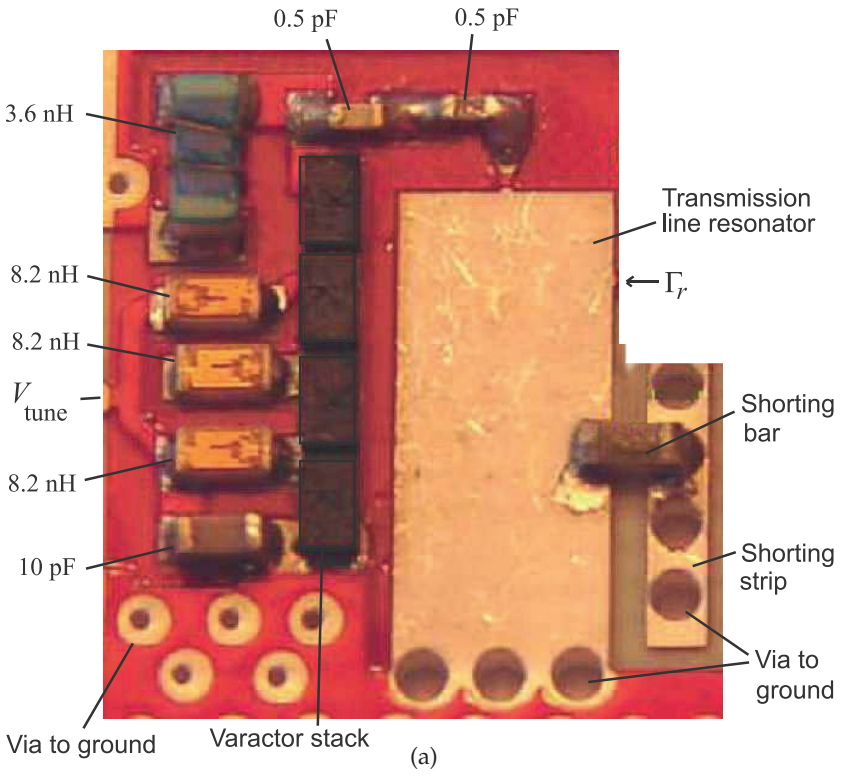


Figure 5-22: C-band VCO circuit: (a) annotated resonator network; and (b) annotated active network. The Pi attenuator (with 294 Ω resistors in the shunt legs and a 17.4 Ω series resistor) is between V_{out} and the 50 Ω bandpass filter.

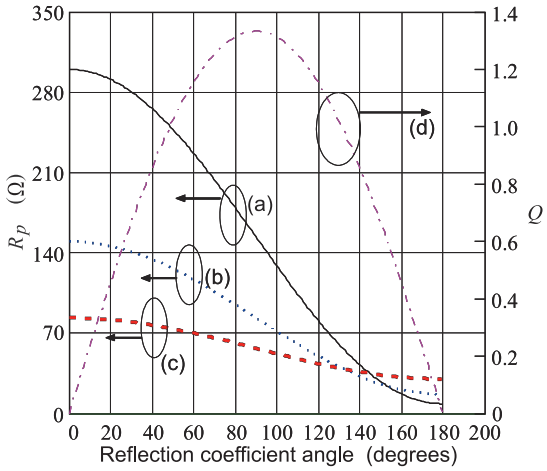


Figure 5-23: The resistance, R_P, of a parallel (or shunt-tuned) resonator required to satisfy the condition of oscillation for (a) $|\Gamma_d| = 1.4$, (b) $|\Gamma_d| = 2$, and (c) $|\Gamma_d| = 4$ versus the magnitude of the device reflection coefficient angle. Curve (d) is the active device $Q = Q_d = |B_d/G_d|$ for $|\Gamma_d| = 2$.

5.6.2 Design Strategy

Small-signal S parameters are generally good indicators of oscillator operation, particularly for the frequency of oscillation [22]. However, they do not provide sufficient information to determine if stable oscillation will occur. The design strategy here is to use simulations and measurements of the individual resonator and active networks to select the required modifications, particularly the selection of C_a and C_b and the position of the shorting bar. Measurements are needed as the characteristics are very sensitive to coupling parasitics, principally between the sides of the lumped elements, which cannot be captured in simulation.

5.6.3 Oscillator Start-Up Considerations

Another consideration in oscillator design is that the right conditions must be provided for oscillation start-up, which initially begins with the amplification of noise. Simply stated, the condition for oscillator start-up of a shunt-tuned oscillator is that for small signals (small A) $|G_d(A,\omega)| > G_r(\omega)$. That is, for small signals, the active network must present a negative conductance that is greater in magnitude than the positive conductance of the resonator network. Since $B_d(\omega) \approx -B_r(\omega)$, the condition for start-up of oscillation can also be expressed in terms of the Qs of the active and resonator networks. That is, for oscillation start-up, $Q_r > Q_d$ for small signals where the Q of the active network is $Q_d = |B_d/G_d|$ and the Q of the resonator network is $Q_r = |B_r/G_r|$. Since G_d is negative, $|\Gamma_d| > 1$, however, it is not sufficient to simply have a large value of $|\Gamma_d|$.

There is a specific angular range of Γ_d that provides the right condition for oscillator start-up. Now $1/\Gamma_d \approx \Gamma_r$ at steady-state oscillation and so designing for a particular Γ_d also determines Γ_r. It is found that the angle of Γ_d, $\angle\Gamma_d$, must be constrained so that losses in the resonator network are accommodated. The appropriate range is selected using Figure 5-23, which plots the required minimum parallel resistance, R_p $(= 1/G_r)$, of the resonator network as a function of $|\angle\Gamma_d|$ for several values of $|\Gamma_d|$. If $|\angle\Gamma_d|$ is less than $50°$, then the resonator network would need to have a higher Q, Q_r, in order to satisfy the oscillator start-up requirement. Now if a large tuning range of the VCO is required, then a large range of $\angle\Gamma_d$ must be considered. A

reasonable range derived from Figure 5-23 is that $\angle\Gamma_d$ between $30°$ and $70°$ should be supported. Thus, in the case of a shunt-tuned oscillator, the design of the interface of the resonator and active networks is a methodical process to provide an appropriate admittance to enable oscillator start-up over the tuning bandwidth of the VCO.

What the above means is that if the active network is designed to present a negative conductance and little susceptance (so that $\angle\Gamma_d \approx 180°$) to the resonator network, then the equivalent parallel resistance, R_p, of the resonator network needs to be very high to ensure start-up. A high R_p means that the Q of the resonator network, Q_r, must be high. However such a high Q_r would be difficult to achieve with a tunable resonator. To provide for greater likelihood of oscillator start-up as well as achievable Q_r, it is better for the active network to present $\angle\Gamma_d \approx 50°$. The wide tuning range requirement extends this range to something like $30° \leq \angle\Gamma_d \leq 70°$. It is thus not reasonable to simply embed reactances in the resonator network to compensate for parasitics in the active network and then to expect that the required start-up conditions be achieved. It may be possible to do this, but this approach would limit the ability to make other trade-offs that would optimize oscillator performance. So the point is that better oscillator performance can be achieved by going beyond the simple form of the Kurokawa oscillation condition embodied in Equation (5.7).

With a fixed-frequency common-base Colpitts oscillator the oscillation operating point (the intersection of Γ_r and $1/\Gamma_d$) can be in the left-half plane of the Smith chart as the loss of the resonator is negligible. However with a microwave VCO the resonator loss, due to the varactors, is appreciable and the location of the intersection of Γ_r and $1/\Gamma_d$ is important.

In summary, designing for a slightly reactive Γ_d is a subtle point that leads to a VCO design of superior performance. It is not necessary to understand this issue to follow the design procedure that follows. It is sufficient to follow the VCO design procedure by noting that one of the requirements is that $30° \leq |\angle\Gamma_d| \leq 70°$, that is, the input admittance of the active network at the resonator-active network interface should be slightly capacitive (or slightly inductive). In matching $1/\Gamma_d$ to Γ_r, this corresponds to a slightly inductive (or slightly capacitive) resonator network with $70° \geq \angle\Gamma_r \geq 30°$ (or $-70° \leq \angle\Gamma_r \leq -30°$).

5.6.4 Initial Design

The initial design of the oscillator with $C_a = 0.5$ pF and with $C_b = 0$ resulted in the simulated resonator and active network characteristics shown in Figure 5-24, where the locus of Γ_r at two bias voltages, 1 V and 3 V, are shown. Also shown are the small-signal characteristics of the active network, plotted as $1/\Gamma_d$, and determined (in harmonic balance simulation) using a $50\ \Omega$ port. The port impedance was high enough to suppress oscillation. The $1/\Gamma_d$ curve intersects with each of the resonator curves at multiple places so that multiple oscillations are possible.

The similar measured characteristics of the actual oscillator are shown in Figure 5-25. The resonator network is shown to the left of the (x-x) line in the oscillator schematic of Figure 5-19(a) and again in Figure 5-19(c). It is also shown in Figure 5-22(a). Measurement of the tank circuit yielded the set of resonator curves (Curves a–g) in Figure 5-25. Curve a is the locus of the

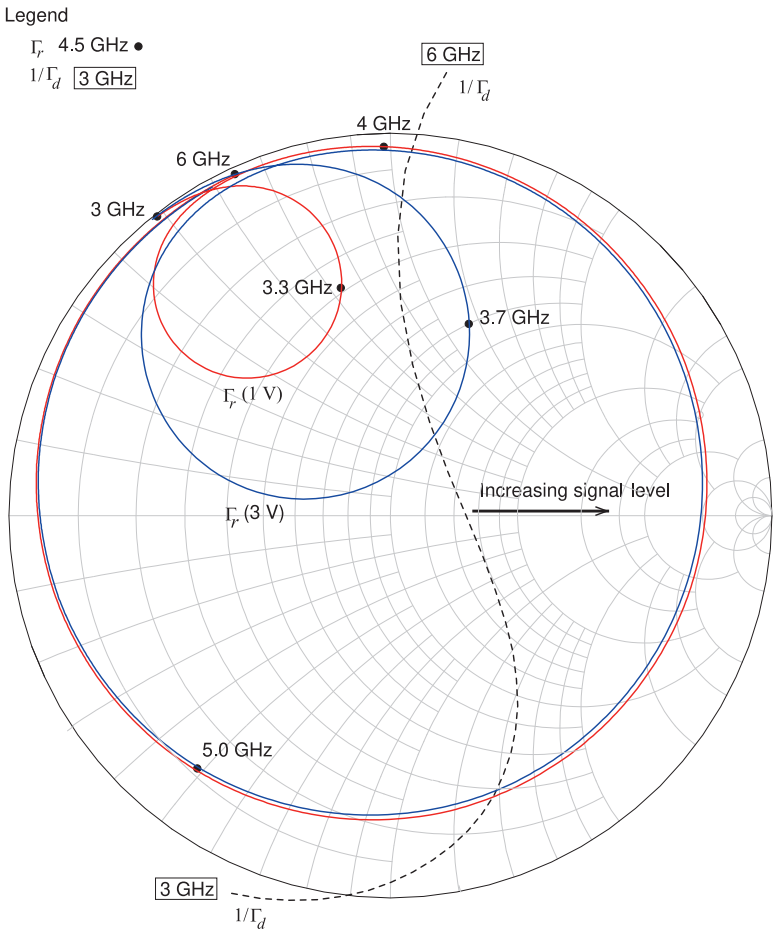


Figure 5-24: Simulated 50 Ω reflection coefficient of the resonator, Γ_r, and the inverse of the small-signal reflection coefficient of the active network, $1/\Gamma_d$, of the C-band VCO with $C_a = 0.5$ pF and without C_b. This is a 50 Ω Smith chart.

reflection coefficient of the resonator network when the tuning voltage is 0 V, that is, Curve a is the locus with respect to frequency of Γ_r(0 V). The seven resonator responses shown in Figure 5-25, Curves a–g, are the resonator reflection coefficients for equally spaced tuning voltages from 0 V through 9 V. Comparing Figures 5-24 and 5-25 it is seen that there is reasonable agreement between the simulated and measured results. The difference can be attributed to the difficulty of performing the measurements at the correct location, as well as coupling between the components themselves not being captured in simulation.

The possibility of multiple oscillations is seen in Figures 5-24 and 5-25, as there can be multiple intersections of a Γ_r curve (for a particular bias) and the $1/\Gamma_d$ curve. (Note that the $1/\Gamma_d$ locus will shift to the right as the level of the oscillation signal increases.) Multiple oscillations are observed as seen in the spectrum at the output, V_{out}, of the oscillator (see Figure 5-26). An oscillator that oscillates at multiple frequencies is not desirable, of course, so the design must be altered to avoid the multiple intersections of the Γ_r (at a particular tuning voltage) and $1/\Gamma_d$ curves.

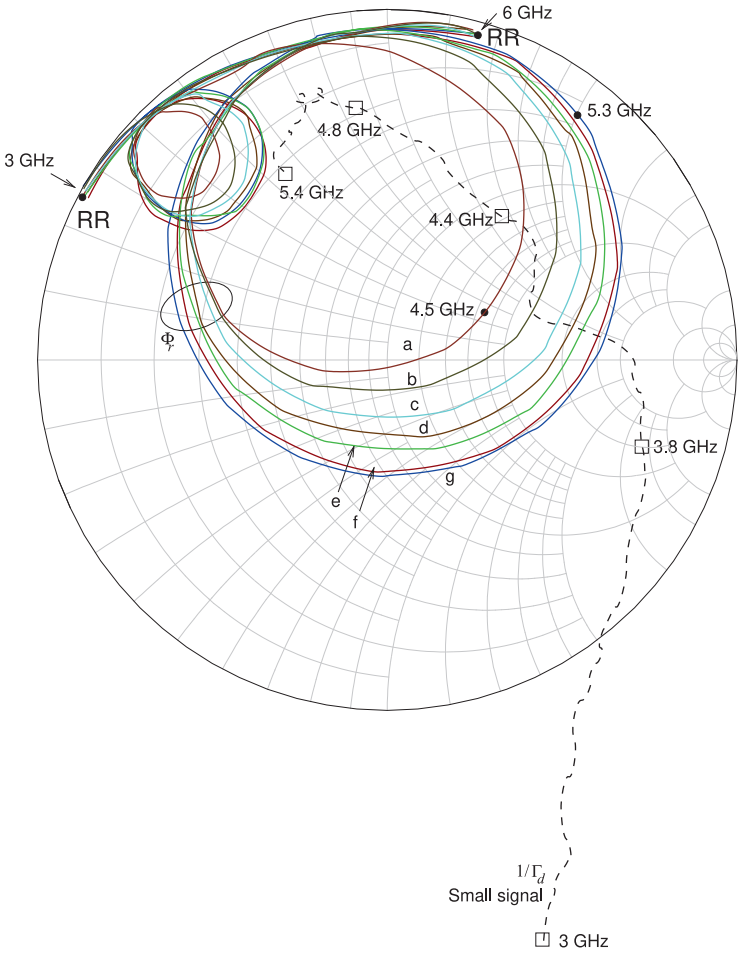


Figure 5-25: Measured reflection coefficient of the resonator network, Γ_r, and the inverse of the small-signal reflection coefficient of the active device network, $1/\Gamma_d$, of the C-band VCO with $C_a = 0.5$ pF and without C_b. Γ_r, Curve a, is the characteristic of the resonator network for a tuning voltage, V_{tune}, of 0 V. Curve g is for $V_{\text{tune}} = 9$ V. V_{tune} is equally spaced for Curves a–g. The locus of $1/\Gamma_d$ moves to the right as the signal level increases and the intersection of $1/\Gamma_d$ and Γ_r determines the oscillation frequency and the signal level.

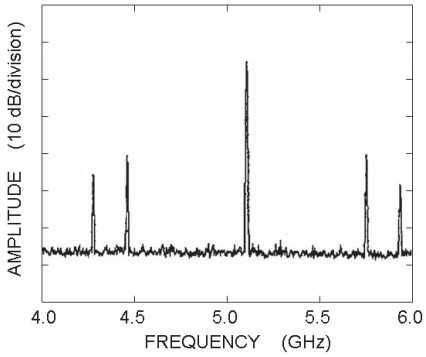


Figure 5-26: Multiple oscillations observed prior to reflection coefficient shaping by C_a and C_b. The spectrum was measured using a resolution bandwidth of 3 MHz and a video bandwidth of 1 MHz.

5.6.5 Avoiding Multiple Oscillations Through Reflection Coefficient Shaping

At an early stage in design (with $C_a = 0.5$ pF and $C_b = 0$) multiple oscillations were observed in Figure 5-26. The cause of these multiple oscillations is multiple intersections of the $1/\Gamma_d$ locus of the active circuit and Γ_r at a particular bias voltage. This is seen in both the simulated results in Figure 5-24 and the measured results in Figure 5-25. The next step in design is to shape $1/\Gamma_d$ of the active network so that there is a unique intersection

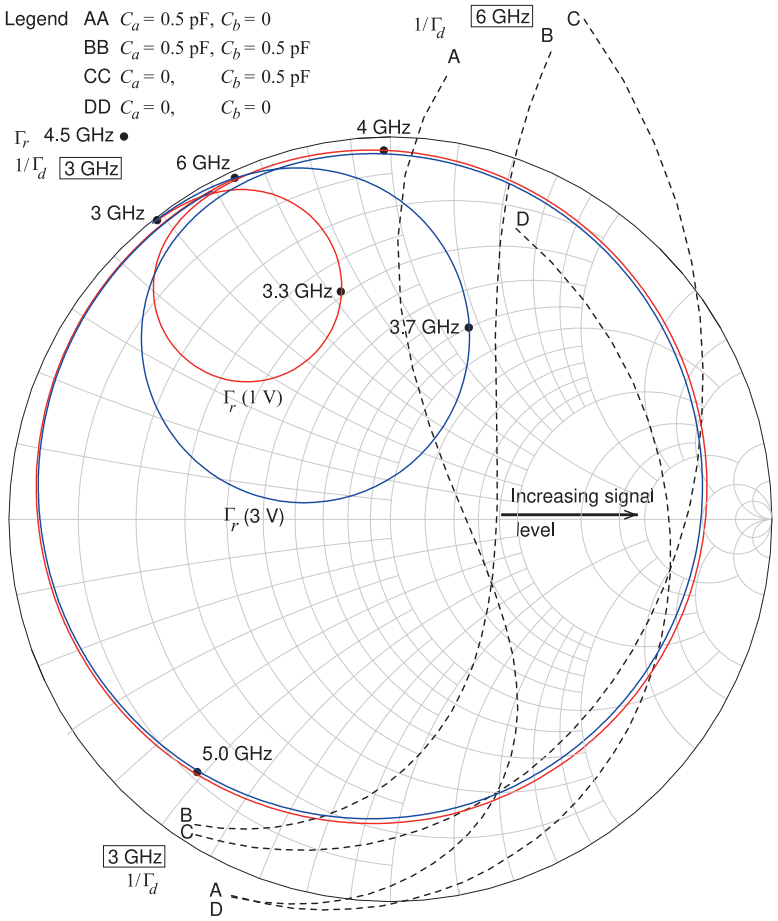


Figure 5-27: Simulated reflection coefficient of the resonator, Γ_r, and the inverse of the small-signal reflection coefficient of the active network, $1/\Gamma_d$, of the C-band VCO for various values of the compensation capacitors C_a and C_b.

of $1/\Gamma_d$ and Γ_r at each bias voltage. The elements used to shape $1/\Gamma_d$ are C_a and C_b. At the same time that these are adjusted the design must achieve $(\partial B_d/\partial V|_{V=V_0}) \approx 0$. Also, the discussion in Section 5.6.3 indicated that the preferred angle of Γ_d to ensure start-up of oscillations is around $50°$. However, for a wide tuning range of the VCO it is only possible to achieve a specific angle of Γ_d at a single oscillation frequency. The trade-off is to choose $-70° \leq \angle\Gamma_d \leq -30°$, indicating that the active network should be slightly capacitive, which corresponds to a slightly inductive resonator network with $70° \geq \angle\Gamma_r \geq 30°$. Thus the intersection of Γ_r and $1/\Gamma_d$ should be in the top right quadrant of the Smith chart. (An alternative design choice which still would have resulted in a successful start-up of the oscillator is $70° \geq \angle\Gamma_d \geq 30°$.)

The simulated characteristics of the resonator and active networks are shown in Figure 5-27. The loci of Γ_r for two tuning voltages are shown, and the small-signal $1/\Gamma_d$ is shown for various values of C_a and C_b. Curve AA (of $1/\Gamma_d$) is for $C_a = 0.5$ pF and $C_b = 0$, as considered before, and will resulted in multiple oscillations. Curve DD is for $C_a = 0$ and $C_b = 0$ and is very close to Γ_r and it may be difficult for oscillation to begin. Another way of describing this is that Q_r is very close to the small-signal Q_d. So even if oscillation did start, it would reach steady state at a low signal level. The active networks represented by Curves CC and DD do not provide sufficient

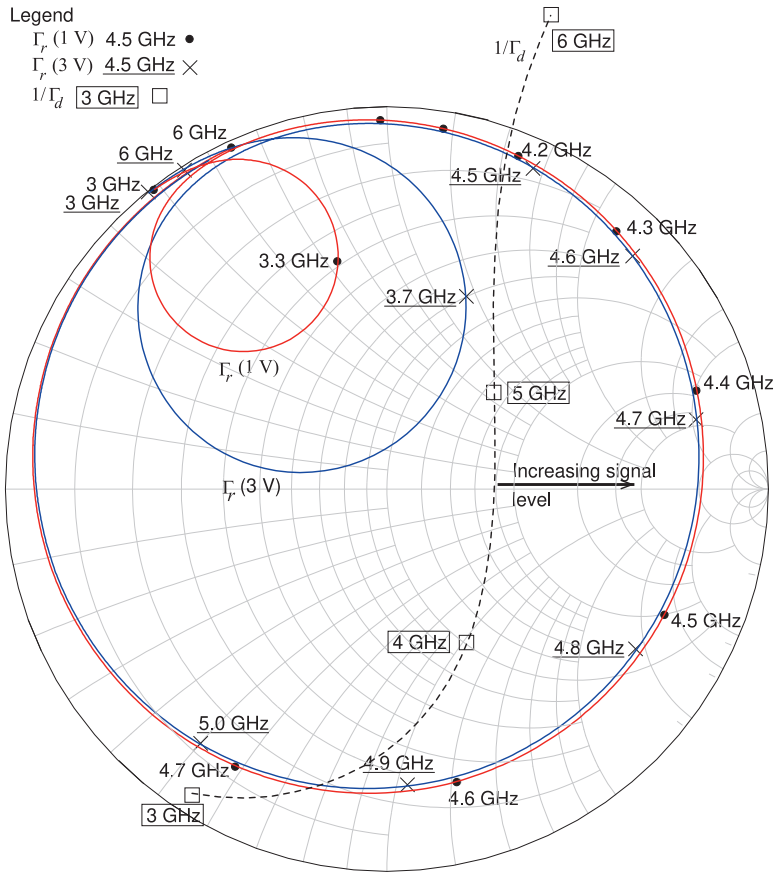


Figure 5-28: Simulated reflection coefficient of the resonator, Γ_r, and the inverse of the small-signal reflection coefficient of the active network, $1/\Gamma_d$, of the C-band VCO.

negative conductance and only enable oscillation over a narrow frequency range. The best characteristic here is Curve BB for which $C_a = 0.5$ pF and $C_b = 0.5$ pF. This response provides a single intersection of the resonator curve for each tuning voltage and the active network curve. Also it indicates a fairly large magnitude of negative conductance so that the output oscillator power will be high. That is, the magnitude of the negative conductance reduces as the signal level increases and the large magnitude of the negative conductance at small signals means the signal level can grow significantly before the conductances of the resonator and active networks match.

The simulated characteristics of the resonator and active networks are repeated in Figure 5-28 for $C_a = 0.5$ pF and $C_b = 0.5$ pF (i.e. Curve BB in Figure 5-27). The response of the circuit now has the desired properties. First consider the locus of the reflection coefficient of the resonator network with 1 V across the varactor diodes (this is the Γ_r(1V) curve). There are two resonances between 3 GHz and 6 GHz, but it is the resonance between 4.0 GHz and 5.3 GHz that is close to the $1/\Gamma_d$ locus and so (for this varactor bias) is the only resonance that will affect oscillation. The locus of Γ_r(1V) approximately follows a constant conductance curve so that $(\partial G_r/\partial\omega|_{\omega=\omega_0}) \approx 0$ as required. As the signal level across the active network increases, the $1/\Gamma_d$ locus shifts to the right in the direction of constant susceptance so that $(\partial B_d/\partial V|_{V=V_0}) \approx 0$ as desired. Provided that the frequencies match, the point at which the Γ_r(1V) locus intersects the $1/\Gamma_d$ locus determines both the oscillation frequency and the oscillation level as

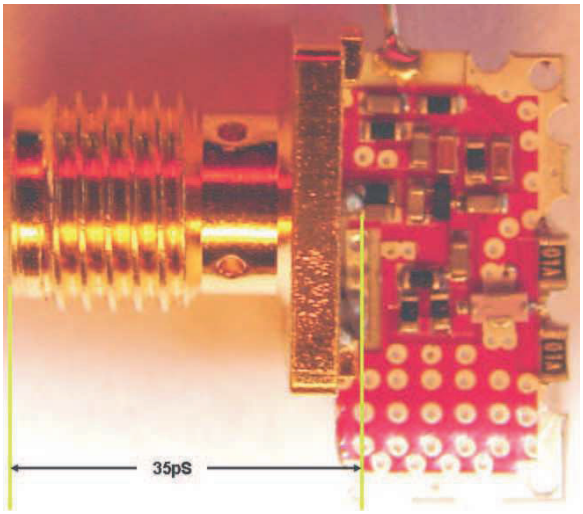


Figure 5-29: Measurement of the active circuit with a 50 Ω test fixture at the interface of the resonator and active networks. The card was cut at the interface to make the connection. A 35 ps delay due to the length of the SMA connector must be subtracted from measurements to reference measurements to the circuit card edge.

the locus of $1/\Gamma_d$ shifts to the right.

The simulated results discussed in the previous paragraph indicate that the oscillator will work as desired. However, there are many parasitics and coupling effects that are not fully captured in simulation. Final design optimization requires experimental investigation of the reflection coefficients looking into the resonator circuit and into the active circuit.

The next step in design of the VCO is using a VNA to measure the reflection coefficient of the active network in Figures 5-19(b) and 5-22(b) (shown again in its measurement configuration in Figure 5-29). The large signal locus in Figure 5-30 was measured with a 10 dBm signal applied to the active network at the 50 Ω measurement port. This curve is an indication of what the active network presents to the resonator under large signal conditions and it is used as a guide since it does not capture the full loading complexity, for example, the harmonic terminations are not correct.

Oscillation occurs when the characteristics of the active network (see the $1/\Gamma_d$ curve in Figure 5-30) match the characteristic of the resonator network, Curves a–g in Figure 5-30. First, for small signals the active network should provide

$$|1/\Gamma_d(A,\omega)| < |\Gamma(\omega)| \tag{5.8}$$

at all desired frequencies of operation. This is the requirement for oscillation start-up. Second, the rotation of $\Gamma_r(\omega)$ with respect to ω near the oscillation point (i.e., ω_0) should be positive (i.e., $(\partial B_r/\partial\omega|_{\omega=\omega_0}) > 0$ as developed in Section 5.5.2) and in the opposite direction to the $1/\Gamma_d(A,\omega)$ locus with respect to ω (i.e., in the same direction as the rotation of $\Gamma_d(A,\omega)$). This is indeed what happens and can be seen by closer examination of Curves a–g. In addition, the locus of $\Gamma_r(\omega)$ should approximately follow a line of constant conductance so that $(\partial G_r/\partial\omega|_{\omega=\omega_0}) \approx 0$. Now, device self-limiting stabilizes oscillation when the angles of $\Gamma_d(A,\omega)$ and $\Gamma_r(\omega)$ sum to zero. For single-frequency oscillation this must be obtained at each tuning voltage. Finally, the trajectory of the limiting $1/\Gamma_d(A,\omega)$ locus (i.e as the amplitude of oscillation, A or V, increases) should intersect the $\Gamma_r(\omega)$ locus, ideally at right angles to minimize phase noise [11, 17]. These requirements are referred to as a complementary relationship between the active and resonator networks.

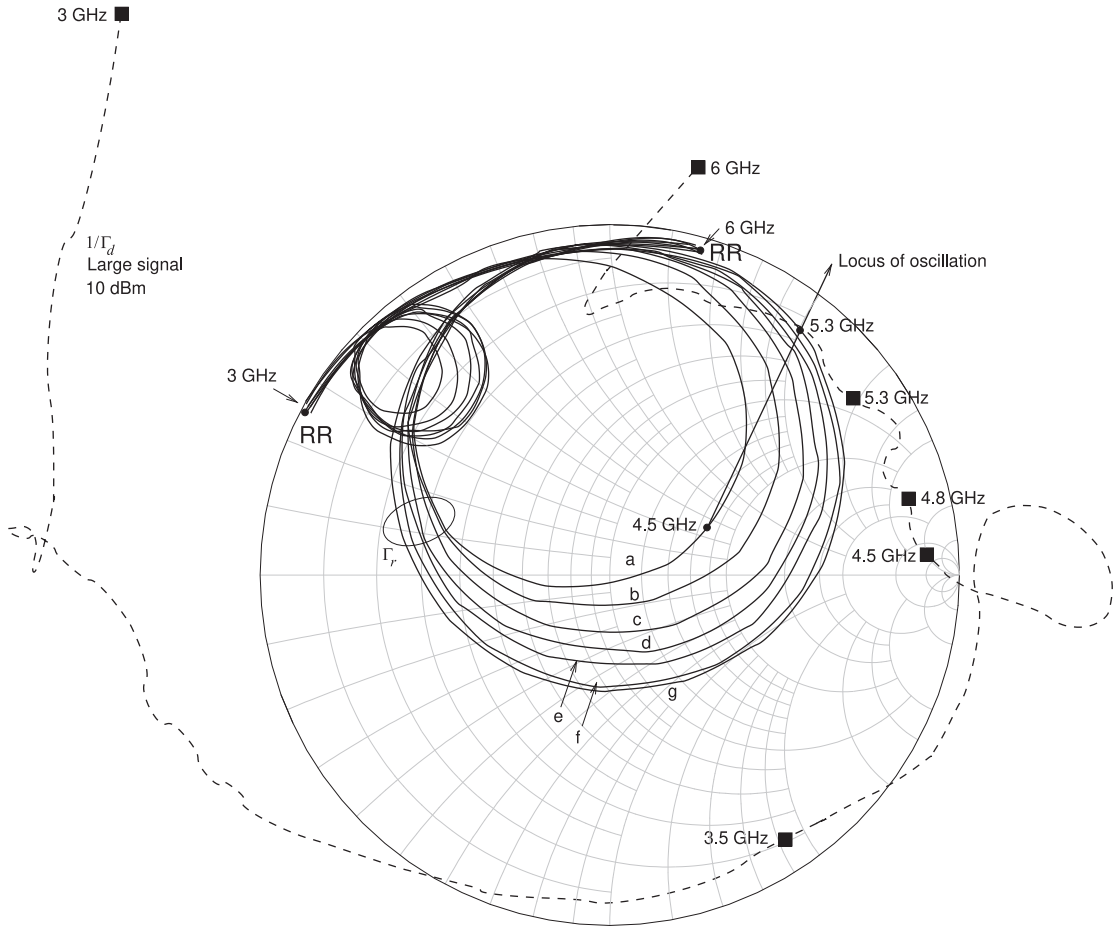


Figure 5-30: Measured reflection coefficient of the resonator network, Γ_r, and the inverse of the large signal reflection coefficient of the active network, $1/\Gamma_d$, of the C-band VCO. For the large signal Γ_d measurement, the active network is driven at 10 dBm from a $50\ \Omega$ port. The Γ_r curves are identical to those in Figure 5-25. The curves with end points labelled RR identify the resonator curves.

While under small-signal conditions the loci may not coincide, the important point is that they do when limiting occurs, as well as providing for the start-up of oscillation. A slight counterclockwise rotation of the modified active device $1/\Gamma_d$ locus as the signal level increases (as well as a general right shift) ensures stable, single-frequency oscillation. Put another way, the trajectory of the negative conductance as device limiting occurs must be such that $1/\Gamma_d$ just intersects the Γ_r locus, and $\angle\Gamma_d$ must complement $\angle\Gamma_r$. This is the situation shown in Figure 5-30, where the modified device network characteristic is achieved by adding capacitive terminations to the collector and the emitter base terminals. Here, unlike the conventional common-base series feedback oscillator situation (as considered in Section 5.4), the input of the active network is now capacitive above 4.5 GHz (see Figure 5-30). Consequently the inductance of the resonator is successfully absorbed. Thus

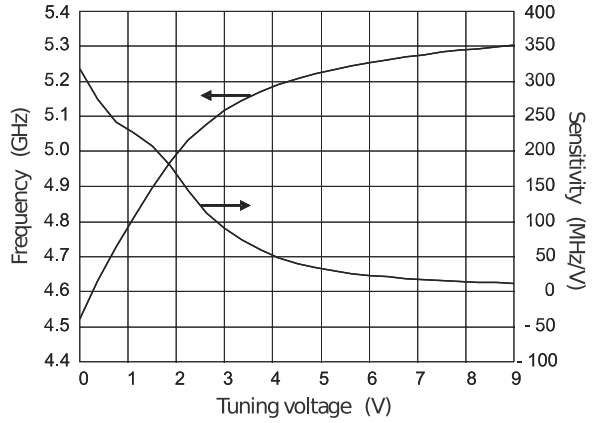


Figure 5-31: Measured tuning characteristic showing oscillation frequency and VCO sensitivity as a function of tuning voltage.

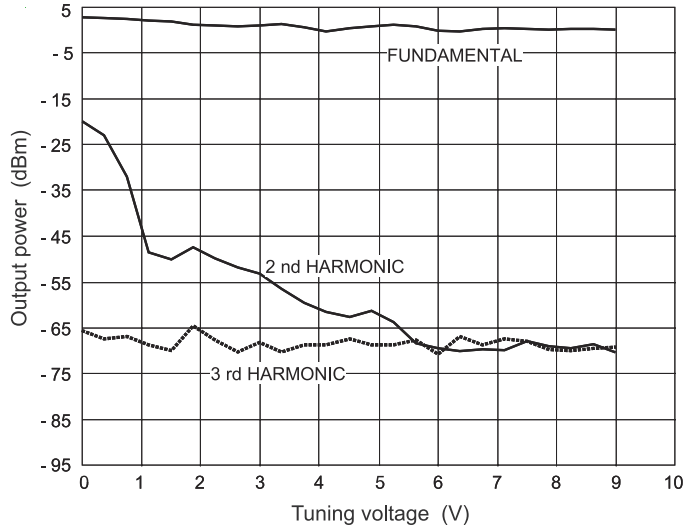


Figure 5-32: Measured output power and harmonics at V_{out} (before the bandpass filter) indicating low-level harmonic content.

the small-signal one-port reflection coefficient of the resonator is inductive initially. In effect the resonator is operated as a tunable shunt inductance rather than a tunable capacitive reactance.

5.6.6 VCO Performance

The most important metrics that describe the performance of the VCO are the phase noise, tuning bandwidth, output power, tuning gain or sensitivity, the output harmonic content, and the DC power consumption. The VCO characterized here, shown in Figure 5-19, includes the compensating capacitors C_a and C_b, both 0.5 pF. Figures 5-31 and 5-32 document the bandwidth and output powers of the VCO. As the varactor tuning voltage, V_{tune}, goes from 0 V to 9 V the filter tunes from 4.5 GHz to 5.3 GHz, producing a minimum output power of 0 dBm and varies by no more than 2 dB over the range. The DC power consumed is 150 mW. The tuning bandwidth is adjusted by varying the coupling (provided by the two 0.5 pF capacitors) between the varactor stack and the microstrip line, TL_1. Figure 5-32 also shows the power at the harmonics. At the final output these are further reduced by the bandpass filter.

The measured phase noise is shown in Figure 5-33 at 4.5 GHz

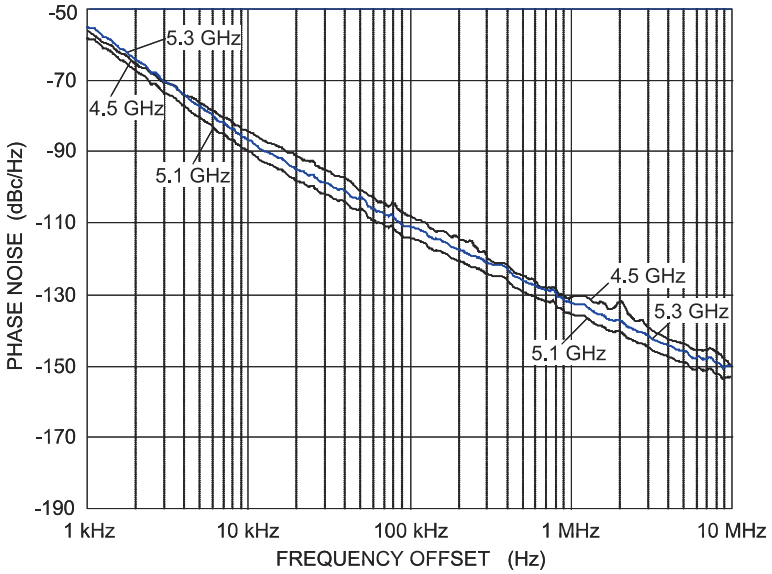


Figure 5-33: Phase noise measured at the top and bottom of the tuning range as well as at 5.1 GHz, where phase noise is optimum. Minimum phase noise floor -116 dBc/Hz at 1 kHz offset, -160 dBc/Hz at 10 MHz offset.

(corresponding to a tuning voltage of 0 V), and at 5.3 GHz (a tuning voltage of 9 V), as well as at 5.1 GHz where the best phase noise is obtained. The phase noise is approximately the same across the tuning range with a $1/f^2$ noise corner frequency, $f_{c,-2}$ (the transition from an f^{-1} dependency to an f^{-2} dependency) of 30 kHz. The phase noise at 10 kHz offset, $\mathcal{L}(10\text{ kHz})$, is better than -85 dBc/Hz, while $\mathcal{L}(1\text{ MHz})$ is better than -130 dBc/Hz. The best measured phase noise, $\mathcal{L}(1\text{ MHz})$, near band center (5.1 GHz) is -135 dBc/Hz.

The performance of a VCO should be quoted as the worst performance over the tuning bandwidth. For a tuning bandwidth of 770 MHz and center frequency of 4.92 GHz, the maximum phase noise of this VCO is -128 dBc/Hz. This improves to a maximum phase noise of -130 dBc/Hz for a bandwidth of 500 MHz centered at 5.05 GHz.

5.6.7 Summary

The VCO design here used a standard one-port reflection oscillator design approach with a technique that introduced compensation capacitors to manage the otherwise frequency-dependent susceptance of the active network. These compensation elements also resulted in the reflection coefficient of the augmented active device having the characteristics necessary to ensure stable oscillation and oscillator start-up.

5.7 Negative Transconductance Differential Oscillator

In RFICs it is common to use an oscillator with a tank circuit across a pair of matched transistors in a differential configuration. Such an oscillator is shown in Figure 5-34(a). So while this circuit is in a differential configuration, it is analyzed and designed as a reflection oscillator at RF.

The cross-connected differential common source pair creates a negative resistance while the fixed inductors (the Ls) and the voltage-tunable capacitors, the Cs, form the variable LC tank circuit. The tunable capacitors are typically implemented using semiconductor varactor diodes whose

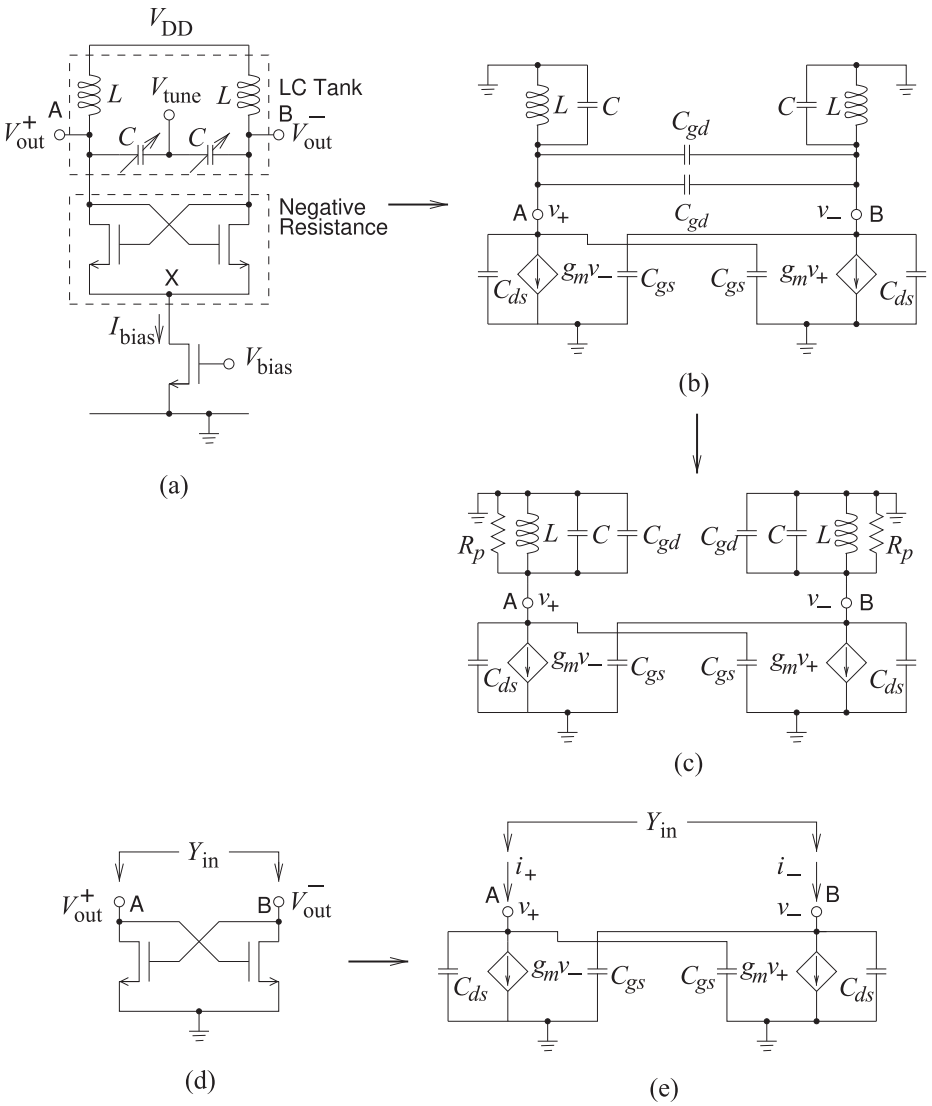


Figure 5-34: Negative-gm differential FET VCO: (a) schematic; and (b) small-signal model used in analyzing the oscillator; (c) small-signal model with C_{gd} incorporated in the tank circuit; (d) negative resistance network of the VCO; and (e) small-signal model used in deriving the input impedance of a negative resistance network. This is a modified form of a Colpitts oscillator. The L-C-C_{gd} resonant circuit operates below resonance and presents an effective inductance (a positive reactance) but with an admittance derivative with respect to frequency that is less than that of an actual inductor. This is essential for stability. The effective inductor, L_3 in Figure 5-7(b), connects the output of each of the transistors to its respective input. For each transistor C_{gs} is C_1, and C_{ds} is C_2, in 5-7(b)

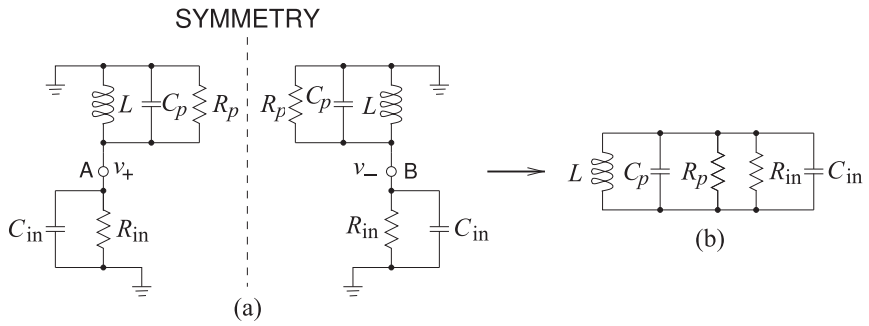


Figure 5-35: Reduced model of the differential FET VCO of Figure 5-34:(a) small-signal model with negative resistance FET network replaced by the equivalent resistance and capacitance; and (b) simplest parallel small-signal model combining the tank and the negative resistance network model.

capacitance can be adjusted by the tuning voltage V_{tune}. The bias transistor, the bottom FET, sets the current in the differential transistors and this current directly impacts the power consumption of the oscillator and the phase noise. The circuit is symmetrical so that the node between the two variable capacitors, the V_{tune} terminal, looks like an RF short as does the common source node of the differential pair, the node labeled X. This is key to developing the small-signal model shown in Figure 5-34(b), where the dominant parasitic capacitances of the transistors, the drain-source capacitance (C_{ds}), the gate-source capacitance (C_{gs}), and the gate-drain capacitance (C_{gd}) are seen. C_{gd} becomes part of the tank circuit. This leads to the simpler small-signal model shown in Figure 5-34(c). Removing the tank circuit leads to the small-signal active device models shown in Figure 5-34(d and e) which present a negative resistance to the tank circuit and load.

The input admittance of the negative resistance network (Figure 5-34(e)) can now be determined. Analysis begins by summing currents at the A and B nodes, respectively:

$$i_+ = (\jmath\omega C_{gs})v_+ + (\jmath\omega C_{ds})v_+ + g_m v_- \tag{5.9}$$

$$i_- = (\jmath\omega C_{gs})v_- + (\jmath\omega C_{ds})v_- + g_m v_+. \tag{5.10}$$

The differential input admittance is then

$$Y_{\text{in}} = \frac{i_+ - i_-}{v_+ - v_-} = -g_m + \jmath\omega(C_{gs} + C_{ds}). \tag{5.11}$$

Thus the negative resistance network is modeled as a negative resistance of value $R_{\text{in}} = (-1/g_m)$ in parallel with a capacitance $C_{\text{in}} = (C_{gs} + C_{ds})$. The dependence of R_{in} on g_m gives this oscillator its name "negative transconductance oscillator" or "negative-gm oscillator." The gate-drain capacitance C_{gd} is in parallel with the tank capacitance C and so a new equivalent capacitance $C_p = C + C_{gd}$ can be defined. Loss in the resonator circuit is modeled by a resistor R_p in parallel with C_p. The small-signal model of the oscillator is now as shown in Figure 5-35(a). This further reduces to the model shown in Figure 5-35(b). Oscillations will initiate if

$|1/R_{\text{in}}| = |g_m| > 1/R_p$. Also the oscillation frequency, f_0, is the frequency at which the shunt reactance is zero, that is,

$$f_0 = \frac{1}{2\pi}\frac{1}{\sqrt{L(C_p + C_{\text{in}})}} = \frac{1}{2\pi}\frac{1}{\sqrt{L(C + C_{gd} + C_{gs} + C_{ds})}}. \tag{5.12}$$

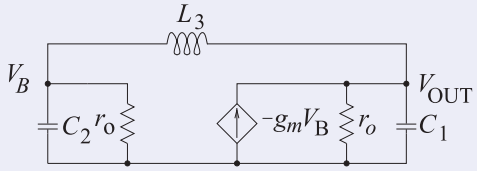
As oscillation builds up, $|g_m|$ reduces to the value of $1/R_p$ and stable oscillation is obtained. The negative-gm oscillator has the ideal characteristic if the negative conductance is the only element dependent on amplitude. Unfortunately the values of C_{gd}, C_{gs}, and C_{ds} also vary as the amplitude of the signal increases. This complicates design at microwave frequencies as these variations could lead to multiple simultaneous oscillations.

EXAMPLE 5.2 Oscillator Analysis

Determine the frequency of oscillation of a Colpitts common emitter BJT oscillator.

Solution:

Figure 5-36 shows two different implementations of a common emitter Colpitts BJT oscillator. The form in Figure 5-36(a) is the most direct implementation, with a clearly defined insertion of the Colpitts network in the collector-to-base feedback path. In Figure 5-36(a), the resistors R_1 and R_2 provide base biasing, and L_C is an RF choke. The oscillation frequency of this oscillator can be derived from the small-signal model of the oscillator. Since R_1 and R_2 will be relatively large resistances, and since L_C is an RF choke (it will look like an RF open circuit), the small-signal model of the oscillator is as shown below.



In this small-signal model, r_π is the base input resistance and r_o is the output resistance—both of these will be relatively large. The transconductance of the transistor is g_m. The network equations are obtained by summing the currents leaving the base node, with Y_1, Y_2, and Y_3 being the admittances of C_1, C_2, and L_3 respectively:

$$Y_2 V_B + G_\pi V_B + Y_3(V_B - V_{\text{OUT}}) = 0, \tag{5.13}$$

$$Y_1 V_{\text{OUT}} + g_m V_B + Y_3(V_{\text{OUT}} - V_B) + G_o V_{\text{OUT}} = 0, \tag{5.14}$$

and $G_\pi = 1/r_\pi$, $G_o = 1/r_o$. In matrix form

$$\begin{bmatrix} (Y_2 + Y_3 + G_\pi & -Y_3) \\ (g_m - Y_3) & (Y_1 + Y_3 + G_o) \end{bmatrix} \begin{bmatrix} V_B \\ V_{\text{OUT}} \end{bmatrix} = 0. \tag{5.15}$$

This can be simplified by noting that r_π and r_o will have admittances smaller than Y_1, Y_2, and Y_3. Thus Equation (5.15) becomes

$$\begin{bmatrix} (Y_2 + Y_3) & -Y_3) \\ (-Y_3) & (Y_1 + Y_3) \end{bmatrix} \begin{bmatrix} V_B \\ V_{\text{OUT}} \end{bmatrix} = 0. \tag{5.16}$$

Equation (5.16) has a solution only if the determinant of the matrix is zero. That is,

$$\begin{aligned}(Y_2 + Y_3)(Y_1 + Y_3) - Y_3 Y_3 &= Y_1 Y_2 + Y_2 Y_3 + Y_1 Y_3 + Y_3^2 - Y_3^2 \\ &= Y_1 Y_2 + Y_2 Y_3 + Y_1 Y_3 = 0.\end{aligned} \tag{5.17}$$

Now $Y_1 = j\omega C_1$, etc., where $\omega = 2\pi f$ is the radian oscillation frequency. Thus Equation (5.17) becomes

$$-\omega^2 C_1 C_2 + \frac{C_1}{L_3} + \frac{C_2}{L_3} = -\omega^2 C_1 C_2 + \frac{C_1 + C_2}{L_3} = 0. \tag{5.18}$$

Rearranging, the frequency of oscillation is

$$f = \frac{1}{2\pi}\sqrt{\frac{1}{L_3}\frac{(C_1 + C_2)}{C_1 C_2}}. \tag{5.19}$$

The same result is obtained for the alternative form of the Colpitts oscillator shown in Figure 5-36(b).

5.8 Advanced Discussion of Oscillator Noise

This section presents a discussion of oscillator noise, and particularly the rapid increase of phase noise close to the carrier. Noise can be partitioned into amplitude and phase noise components. The nonlinear saturation of an oscillator suppresses amplitude noise so only phase noise is of concern. While usually associated with oscillators, phase noise is also added to a signal by an amplifier.

There is not a consensus as to the origins of close-to-carrier phase noise. This section begins with observations of oscillator noise in the frequency domain and in the time domain. Then three theories of excess oscillator noise, Leeson's theory, the linear time-invariant model, and the chaotic map model, are presented.

Not having a complete model of the physical origin of phase noise means that a simulator cannot reliably predict the phase noise of an oscillator. Also designing an oscillator with good phase noise performance currently relies heavily on experience and projections based on what has been achieved by a designer previously.

5.8.1 Observations of Oscillator Noise in the Frequency Domain

The most puzzling noise observed with oscillators is the noise observed at a small frequency offset from the carrier (i.e., the average oscillation signal). To develop an appreciation for the breadth of observations, the signals produced by several different oscillators will be considered. First, Figure 5-37

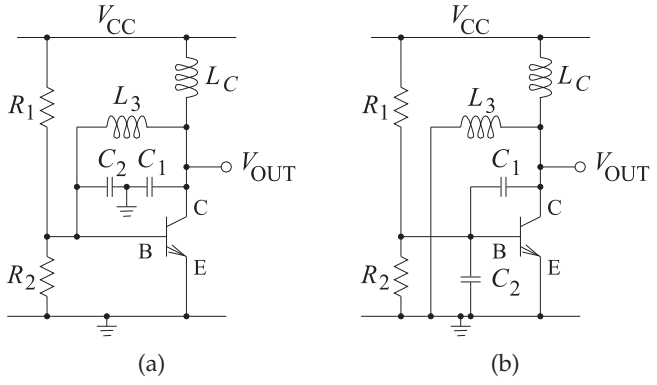


Figure 5-36: Common emitter BJT Colpitts oscillators: (a) configuration with a feedback network between the collector and base of the transistor; and (b) alternative configuration.

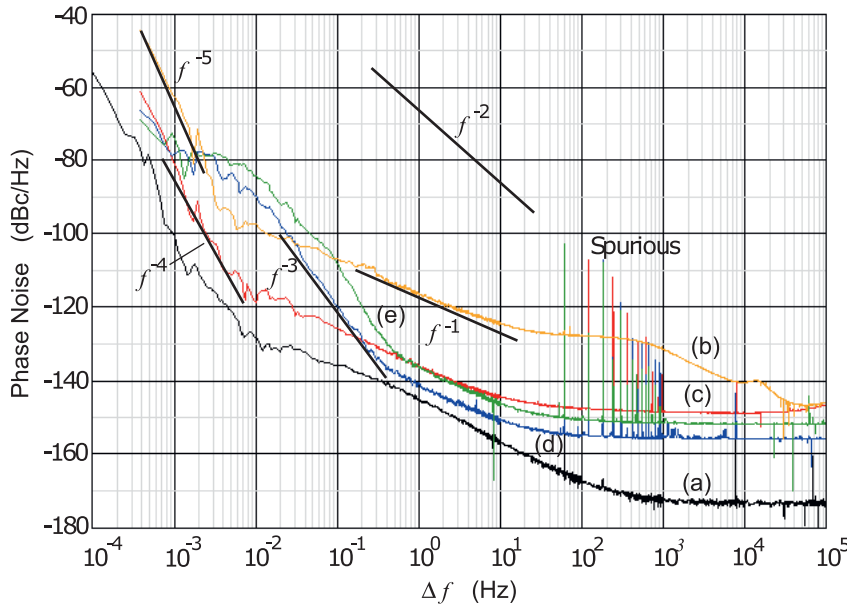


Figure 5-37: Measured phase noise of low-frequency oscillators: (a) instrument noise floor; (b) HP 5087A frequency distribution amplifier at 5 MHz (used to drive the external reference input of several test instruments using a single high-quality oscillator); (c) TADD-1 frequency distribution amplifier at 10 MHz; (d) TADD-1 frequency distribution amplifier at 5 MHz; (e) Spectracom 8140T frequency distribution amplifier at 10 MHz. Five phase noise regions are identified as f^{-5}, f^{-4}, f^{-3}, f^{-1}, and white noise. The spurious signals are related to injected harmonics of the 60 Hz power mains. Used with permission of John Ackermann [23].

is a plot of the phase noise observed at the output of several oscillators and amplifiers operating at 5 MHz and 10 MHz. Curve (a) is the noise floor of the noise measurement instrument and spurious tones are observed at multiples of 60 Hz, the power mains frequency. Curves (b), (c), (d), and (e) show phase noise varying in straight-line segments. Being a log-log plot, these curves show phase noise varying as f^{-5}, f^{-4}, f^{-3}, f^{-1}, and f^{0}. None of the phase noise plots here show a region with an f^{-2} dependence, although this is observed with other oscillators.

Another oscillator to consider is the VCO circuit shown in Figure 5-38 [24]. This is a 50 MHz VCO with a semiconductor varactor being the variable element with a zero-bias capacitance of 100 pF. The capacitance of the varactor is controlled by the voltage, V_b. With $V_b = 0$ V, the phase noise shown in Figure 5-39 was observed. The distinct phase noise regions have frequency dependencies of f^{0}, f^{-1}, f^{-2}, and f^{-3}. The phase noise of this oscillator is plotted again in Figure 5-40 for three different varactor bias voltages. The phase noise characteristics of the oscillator change even though the underlying physical sources of noise do not change (of course). Curve (a), with $V_b = 6$ V, and Curve (b), with $V_b = 0$ V, have an f^{-1} region around 20 kHz (see Figure 5-39 for more details), but the f^{-1} region is not observed in Curve (c) where $V_b = 18$ V. One interpretation is that the crossover frequencies have shifted. So what is particularly interesting here is

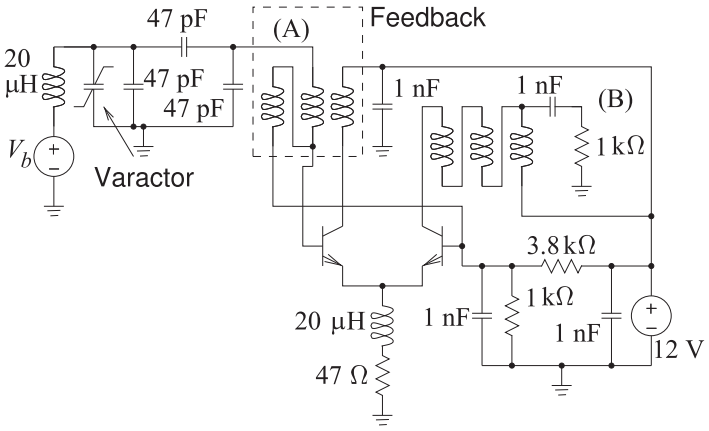


Figure 5-38: Varactor-tuned VCO schematic, from [24].

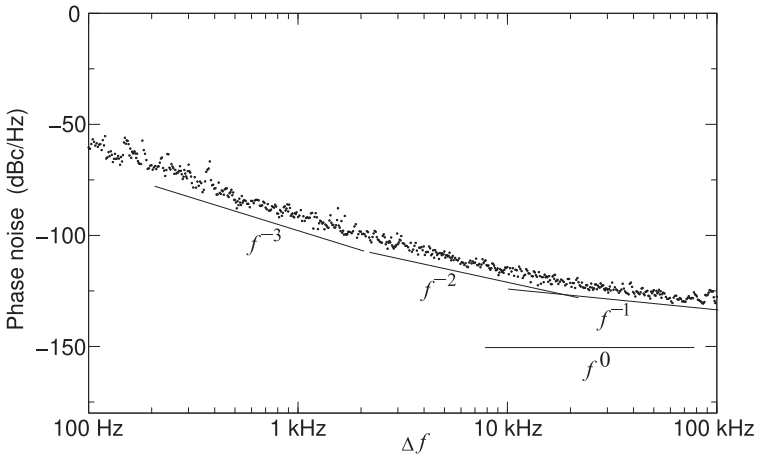


Figure 5-39: Measured phase noise of a 50 MHz BJT varactor-based VCO with the varactor biased at 0 V [25, 26]. Three phase noise regions are identified as f^{-3} (having a slope of -9 dB/octave), f^{-2} (having a slope of -6 dB/octave), and f^{-1} (having a slope of -3 dB/octave).

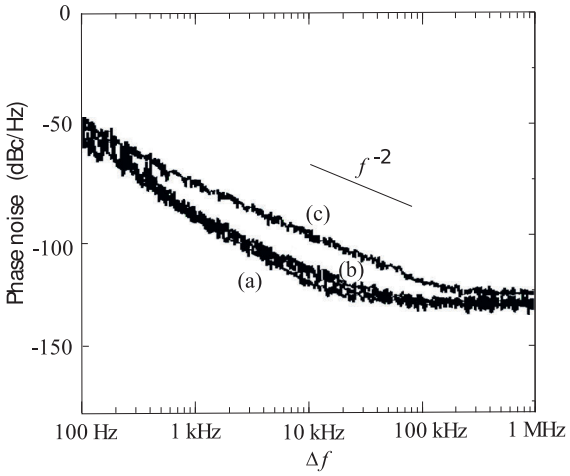


Figure 5-40: Measured phase noise of a 50 MHz BJT varactor-based VCO at three varactor bias voltages: (a) 6 V; (b) 0 V; and (c) 18 V [25]. The varactor breakdown voltage is 30 V Curve (b) was also plotted in Figure 5-39.

that the same physical source of noise can be manifested quite differently at the output of an oscillator when the circuit bias is changed.

The third phase noise example is for a 2.4 GHz power oscillator that has the output spectrum shown in Figure 5-41 with regions having dependencies of f^{-3} and f^{-0}, but nothing in between. (The slight increase in noise power spectral density at 40 kHz offset is due to dynamics of the oscillator's

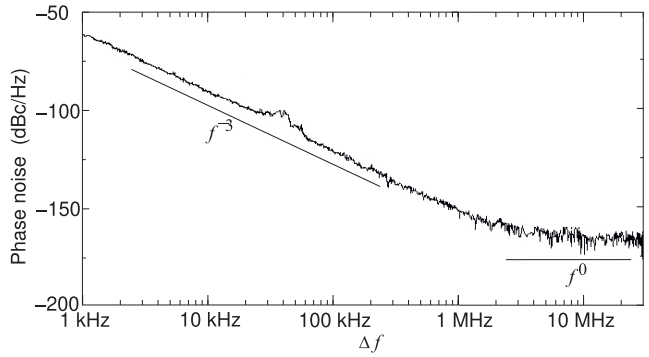


Figure 5-41: Phase noise of a 2.4 GHz power oscillator with an output power of 34.5 dBm [27, p. 323]. Two phase noise regions are identified as f^{-3} (having a slope of -9 dB/octave) and white noise (with an f^0 dependency).

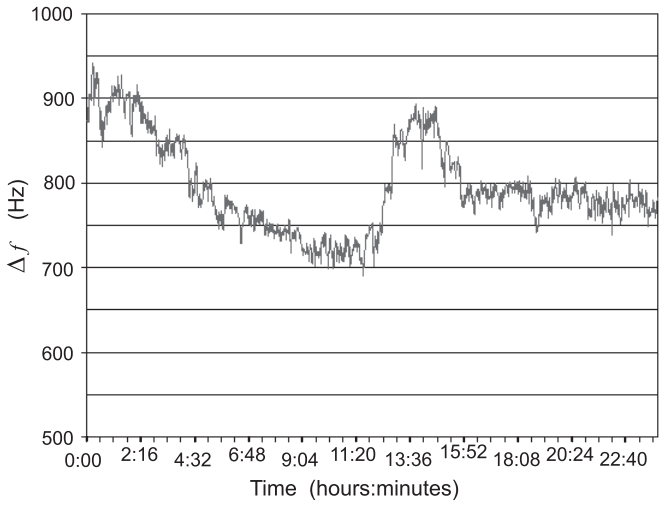


Figure 5-42: Long-term stability of a 10 GHz oscillator measured over a 24 hour interval after being on for 3 weeks. Used with permission of John Ackermann [23].

feedback loop.) Finally, the 5 GHz oscillator considered in Section 5.6 has f^{-3} and f^{-2} phase noise regions (see Figure 5-33).

So the whole range of phase noise dependencies on frequency offset are observed, but the universal observation is that the dependency of the noise power spectral density is to a non-positive integer power of frequency (i.e., $f^{-n}, n = 0, 1, \ldots$).

5.8.2 Observations of Oscillator Noise in the Time Domain

An important time-domain characterization of noise is referred to as random walk noise. An example of this is the variation of the oscillation frequency over a long period of time. The long-term stability of a 10 GHz oscillator is shown in Figure 5-42. This noise cannot be characterized in the frequency domain and instead is described by its **Allan variance**, $\sigma_y^2(\tau)$, or **Allan deviation**, $\sigma_y(\tau) = \sqrt{\sigma_y^2(\tau)}$, defined as follows.

If the frequency measured at time t is $f(t)$ and the nominal oscillation frequency is f_n, then the fractional frequency at time t is defined as

$$y(t) = \frac{f(t) - f_n}{f_n}. \tag{5.20}$$

Then the average fractional frequency over an observation time interval τ is

defined as

$$\bar{y}(t,\tau) = \frac{1}{\tau}\int_0^{\tau} y(t+t_v)dt_v. \tag{5.21}$$

This leads to the definition of the Allan variance as

$$\sigma_y^2(\tau) = \tfrac{1}{2}\left\langle \bar{y}_{n+1} - \bar{y}_n \right\rangle, \tag{5.22}$$

where τ is the observation period and $\bar{y}_n$ is the nth fractional frequency average over the time interval τ. Note that there is no dead-time between the nth and $(n+1)$th measurement time intervals.

The random walk shown in Figure 5-42 is an important clue to unraveling flicker noise. Figure 5-42 shows long-term memory and here it is shown that there is memory over several hours. Even on smaller time scales, random walk is apparent and there is a self-similar property—the hallmark of chaotic behavior. Could this random walk effect and $1/f$ noise arise from the same physical process? Most likely, but there is no accepted theory.

5.8.3 Excess Oscillator Noise: The Leeson Effect and Flicker Noise

As seen in Figures 5-37 to 5-41, oscillators have noise that increases as the offset Δf, from the mean oscillation frequency decreases. This noise has separable regions where noise varies as Δf^n, where n is an integer ranging from 0 to -5. There are transition regions between these discrete states, but there is not a region where n is a fractional number. Not all of the discrete states are observed because, presumably, either the crossover frequencies have changed order, or the frequency offset, Δf, was not low enough.

In 1966 Leeson [28] examined the effect of feedback on noise in oscillators (see Figure 5-1). The phase noise mechanism treated by this analysis is now called the Leeson effect. Leeson showed that white phase noise and white flicker noise (white here meaning independent of frequency) of the amplifier in the feedback loop translate to noise on the oscillation signal with power law dependencies of f^{-2}, called white frequency noise, and f^{-3}, called flicker frequency noise, respectively. These were the dominant "nonwhite" forms of noise observed in his time. However, his analysis did not predict the level of the noise accurately and sometimes was off by an order of magnitude.

The Leeson effect is briefly summarized here. First, it was observed that nearly every physical system has fluctuations that vary as $1/f$ at low frequencies. This includes electrical devices such as the amplifier in an oscillator feedback loop. This leads to equal amplitude phase and amplitude noise superimposed on the oscillation. Since noise is small, the amplitude fluctuations are suppressed by the saturation of the active device so that the only noise observed in good designs is phase noise. Leeson determined that the oscillator phase noise has a region with Δf^{-3} dependence that is due to low-frequency f^{-1} noise (i.e. around DC), a Δf^{-2} region due to white noise in the bandwidth of the oscillator's tank circuit, and also a white noise region outside the bandwidth of the tank circuit. The basis for the development of Leeson's oscillator phase noise model is shown in Figure 5-43. Mathematically [28],

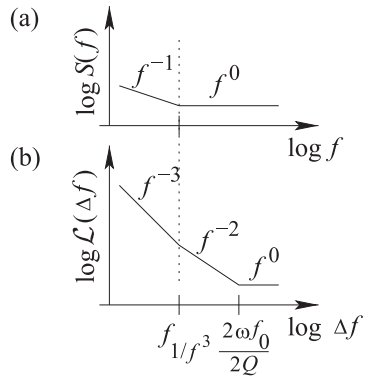


Figure 5-43: Derivation of oscillator noise spectra: (a) the noise spectra of an electronic material with noise increasing as the frequency reduces; and (b) the noise spectra close to the oscillation frequency of an oscillator.

$$\mathcal{L}(\Delta f) = \mathcal{L}(\Delta\omega) = \frac{2FkT}{P_0}\left[1 + \left(\frac{f_0}{2Q\Delta f}\right)^2\right], \tag{5.23}$$

where Q is the loaded Q factor of the oscillator's tank circuit and F is an empirical factor. $\mathcal{L}$ has the units of $\text{radians}^2/\text{Hz}$, or in decibels,

$$\mathcal{L}|_{\text{dB}}(\Delta f) = 10\log\left\{\frac{2FkT}{P_0}\left[1 + \left(\frac{f_0}{2Q\Delta f}\right)^2\right]\right\}, \tag{5.24}$$

which has the units of dB/Hz or more usually expressed as "decibels below the carrier" of power P_0, or dBc/Hz. This is the power at a specified offset such as the phase noise of a 5.05 GHz oscillator at an offset of 1 MHz and with an output power of 0 dBm being -130 dBc/Hz [19].

The derivation of the oscillator noise characteristics from first principles, which led to Equations (5.23) and (5.24), predicts noise levels that are much lower than those observed in practice [29, 30]. As well, the prediction inherent in Equation (5.23) is that by increasing the Q of the tank circuit the noise level will be reduced. However, this is not always obtained in practice. A further complication is that Equation (5.23) provides no mechanism for the generation of $1/(\Delta f)$ and $1/(\Delta f)^3$ noise in the oscillator phase noise spectrum. An adhoc modification of Equation (5.23) accounts for this:

$$\mathcal{L}(\Delta f) = \frac{2FkT}{P_0}\left[1 + \left(\frac{f_0}{2Q\Delta f}\right)^2\right]\left(1 + \frac{f_{c,-3}}{|\Delta f|}\right). \tag{5.25}$$

Given the inadequacy of this model it is just as well to use

$$\mathcal{L}(\Delta f) = \sum_{i=0}^{-5} b_i f_i^n \tag{5.26}$$

where the b_i coefficients are extracted from measurements.

The Leeson effect can be stated as oscillator phase noise being up-converted white noise around DC.

5.8.4 Excess Oscillator Noise: Linear Time-Variant Model

The Leeson effect model described in the previous subsection uses a linear time-invariant model of the oscillator and does not consider down-conversion of noise from frequencies near harmonics. The linear time-variant

model, also called the Hajimiri and Lee model, incorporates these higher-order conversion mechanisms [31].

Noise injected into an oscillator has a different impact depending on whether it is injected at the peak of the oscillating signal or at the zero-crossings. The noise injected at the peaks of the oscillating signal are quenched by the saturating effect of the active device in the oscillator. However, noise at or near the zero-crossings of the waveform introduces jitter and phase noise. This effect on phase noise can be described by an impulse sensitivity function [31]. Consider an impulse injected at phase $x = \omega_0 t$, then the time-domain impulse response is

$$h_\phi(t,\tau) = \frac{\Gamma(\omega_0 t)}{q_{\max}} u(t-\tau), \tag{5.27}$$

where $\Gamma(\ )$ is the impulse sensitivity function, $q_{\max}$ is the maximum charge displacement on the capacitor forming the tank circuit, t is the observation time, and τ is the time of the excitation. The excess phase of the oscillator (the additional phase induced onto the phase of the carrier) is

$$\phi(t) = \frac{1}{q_{\max}} \int_{-\infty}^{t} \Gamma(\omega_0 t) i(\tau) d\tau, \tag{5.28}$$

where $i(\tau)$ is the noise current injected in the oscillator.

$\Gamma(\ )$ can be derived approximately for some oscillators such as the CMOS LC oscillator in [32], where it was shown that the impulse sensitivity function can be expressed as a Fourier series with a fundamental component corresponding to the frequency of oscillation. The excess phase at the zero-crossings of the oscillator is

$$\phi(t) = \frac{1}{q_{\max}} \left[c_0 \int_{-\infty}^{t} i(\tau) d\tau + \sum_{m=1}^{\infty} c_m \int_{-\infty}^{t} i(\tau) \cos(n\omega_0 \tau) d\tau \right], \tag{5.29}$$

where c_m are coefficients of the Fourier series. The first term with the c_0 coefficient indicates noise that is up-converted from baseband, while the term in the summation is the contribution to the oscillator phase noise due to down-conversion of noise near the harmonic frequencies. With the assumption that the noise at the harmonics is white noise with mean-square current $\bar{i_n^2}$, then the noise spectral density is [32, 33]

$$\mathcal{L}(\Delta\omega) = 10 \log \left(\frac{\bar{i_n^2}}{\Delta f} \frac{\sum_{m=0}^{\infty} c_m^2}{4 q_{\max}^2 \Delta\omega^2} \right). \tag{5.30}$$

Here $\Delta f = 1$ Hz for noise in a 1 Hz bandwidth and $\Delta\omega$ is the radian offset frequency. Thus white noise at the harmonics is down-converted to f^{-2} noise at the oscillation frequency, and f^{-1} noise at baseband is up-converted to f^{-3} noise at the oscillation frequency.

The time-variant model provides a richer description of phase noise on an oscillating signal than does the Leeson model, but it does not describe f^{-1}, or $f^{-n}, n > 3$, noise that is observed with oscillators.

Thus the Hajimiri and Lee model relates to down-conversion of white noise at harmonics of the oscillation frequency to the near carrier noise which is in addition to Leeson's model of up-converted near-DC white noise. The Hajimiri and Lee, and Leeson models of phase noise led to designers developing microwave oscillators with significantly lower phase design yet there remains appreciable near-carrier phase noise.

5.8.5 Excess Oscillator Noise: Chaotic Maps and Flicker Noise

While not firmly established, it is possible that flicker noise originates from nonlinear dynamics and chaos [25, 26, 34–41]. In this model flicker noise derives from a nonlinear process with delayed feedback. The mathematical foundation is well established [42], describing a phenomenon called intermittency [43] that occurs when a physical process transitions between stable periodic states and chaotic states. Inherent to some forms of intermittency is long-term memory with a $1/f$ spectrum [44]. This has been established for many physical and biological systems.

Logistics Map

A classic example of **intermittency**, and the first widely accepted, is the following model of population dynamics. If t_n denotes discrete time and (the real number) x_n denotes the ratio of the existing population to the maximum possible population at t_n (so x_n is between 0 and 1), then what is called the logistics map provides the population ratio, $x_{(n+1)}$, at time t_{n+1}. The logistics map is [45]

$$F_\lambda(x) = \lambda x_n(1 - x_n), \tag{5.31}$$

and so

$$x_{n+1} = F_\lambda(x). \tag{5.32}$$

Here λ is a positive number representing the combined rate of reproduction and starvation. So environmental conditions determine λ, which is constrained so that $0 < \lambda \leq 4$. Depending on λ, the logistics map (i.e., Equation (5.31)) will produce a stable population or a random population depending on the value of λ, with $\lambda = 4$ producing white noise.

Thermal noise produces random fluctuations in the amplitude and phase of a sinusoidal signal that is being processed in a nonlinear electronic system such as an amplifier or an oscillator. Denoting the thermal amplitude fluctuations by $a_{t,I}(t)$ and the thermal phase fluctuations by $\phi_{t,I}(t)$, a sinusoidal signal with mean amplitude A and an initial phase of zero is

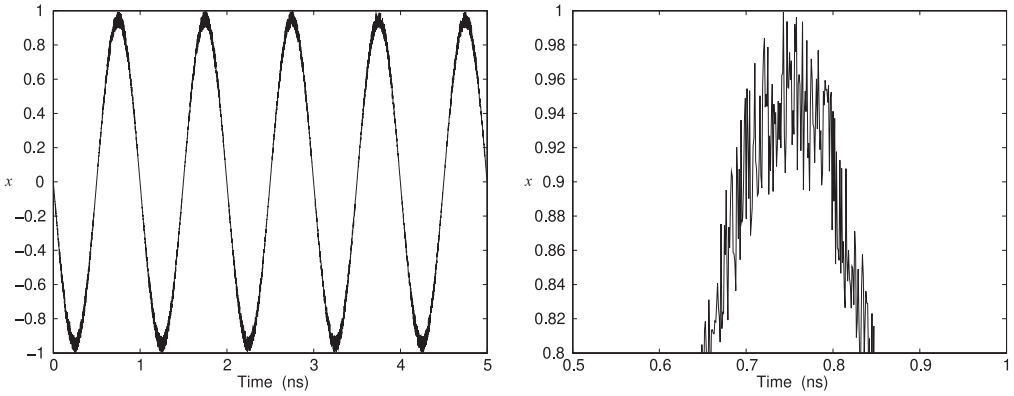
$$x(t) = A[1 + a_{t,I}(t)]\cos[\omega t + \phi_{t,I}(t)]. \tag{5.33}$$

Using the logistics map with $\lambda = 4$ (which produces white noise) to determine $a_{t,I}(t)$ and $\phi_{t,I}(t)$, a sinusoidal signal with thermal (white) noise is as shown in Figure 5-44. Using $a_{t,I}(t)$ and $\phi_{t,I}(t)$ determined from a Gaussian distribution would yield the same qualitative result. Of course most of this noise would be easy to remove by bandpass filtering but thermal noise will still appear within the finite bandwidth of the signal. It is just easier to visualize the effect of noise by plotting it on this scale.

Equation (5.31) is a simple nonlinear equation with delayed feedback that mixes x over time. What is called the rate of mixing describes the extent of correlation to past events and can be thought of as an exponential decay rate. However, with the logistics map, $F_\lambda(x)$ in Equation (5.31), the rate of this mixing is not controllable.

Logarithmic Map

There are many maps that will lead to $1/(\Delta f)$ effects and one of the most convenient to use in modeling flicker noise in electronics is called the loga-



(a) Waveform (b) Detail

Figure 5-44: Sinusoidal signal with superimposed white noise calculated using the logistics map rather than calculating noise as a Gaussian process.

rithmic map [46, 47]:

$$F_\beta(x) = \begin{cases} x(1 + Y(\beta)x\,|\log(x)|^{1+\beta}) & \text{if} \quad 0 \le x \le 1/2 \\ 2x - 1 & \text{if} \quad 1/2 < x \le 1 \end{cases} \tag{5.34}$$

and so $$x_{n+1} = F_\beta(x). \tag{5.35}$$

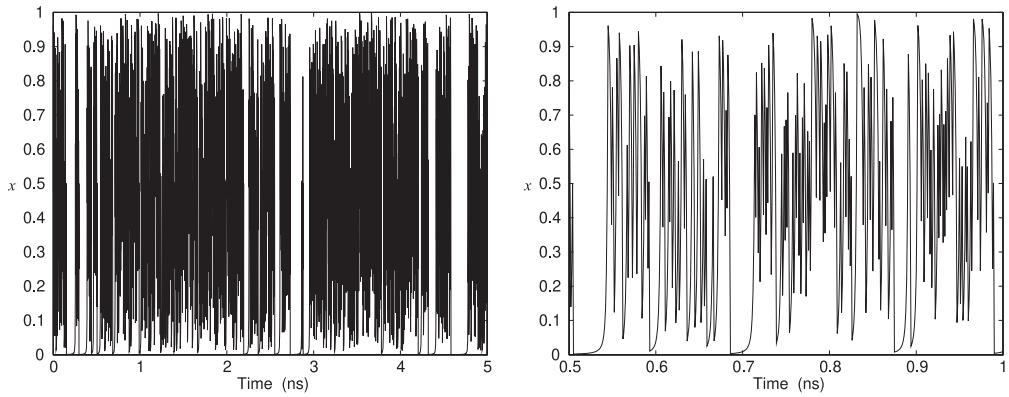
$F_\beta(x)$ is defined on the interval $0 < x \le 1$ and $Y(\beta) = 2(\log 2)^{-(1+\beta)}$ is chosen to ensure that $\lim_{x\to 1/2^-} f_\beta(x) = 1$. (Note that the map is discontinuous at $x = 1/2$.) If Δt is a fixed time interval, then $x(t + \Delta t) = F_\beta(x(t))$.

The logarithmic map uses only a single parameter, β, that controls the rate of mixing and thus the long-term memory property. The solution of the logarithmic map, Equation (5.34), with $\beta = 0.000005$ is shown in Figure 5-45.

The Fourier transform of the sequence plotted in Figure 5-45 is shown in Figure 5-46 and its autocorrelation is shown in Figure 5-47. The spectrum in Figure 5-46 (above 1 Hz) has an $f^{-0.5}$ dependence, and since the sequence corresponds to voltage, squaring this yields a $1/f$ power characteristic. The autocorrelation plot in Figure 5-47 shows the slow long-term decay in the correlation with respect to the discrete interval, i, between the sequence points. That is, the logarithmic map describes a process with slowly decaying correlations. The extended correlation is a measure of the rate of mixing. This interpretation corresponds very well to the understanding of physical systems, and in particular to electronic systems. It can be shown [47] that the rate of decay of correlation of this map is bounded as

$$R(n) \le B(\log n)^{-\beta}, \tag{5.36}$$

where n is the nth time interval. Thus the map is said to describe a logarithmic mixing rate that can be made as slow as desired by varying the value of β. It is this long-range dependence that produces f^{-1} (and f^{-2}, f^{-3}, etc.) noise. In an oscillator these result in phase noise with $1/(\Delta f)$, $1/(\Delta f)^2$, $1/(\Delta f)^3$, etc. characteristics. In semiconductors, for example, the



(a) Solution (b) Detail

Figure 5-45: Solution of the logarithmic map of Equation (5.34) with (a random initial seed) $x_0 = 0.477347$, $\beta = 0.000005$,and $t_{n+1} - t_n$ = 1 ps.

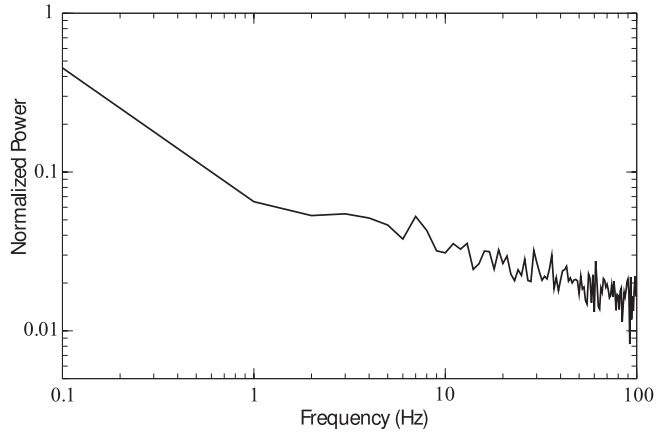


Figure 5-46: Spectrum of the logarithmic map with $\beta = 0.000005$.

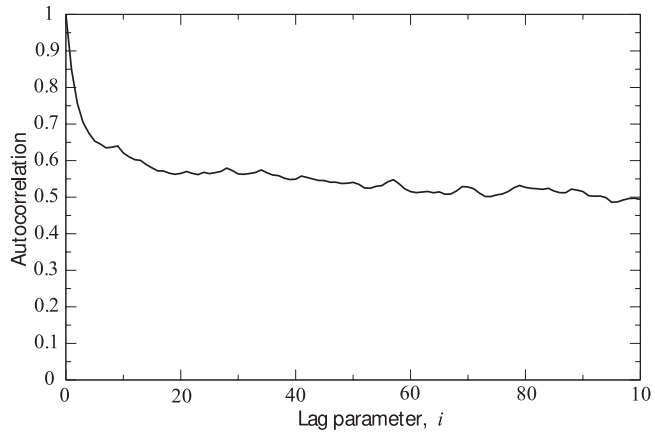


Figure 5-47: Correlation plot of the logarithmic map with $\beta = 0.000005$ for a million-point sequence.

rate of mixing is influenced by the density of traps [48] and lattice scattering [49] with the lower the density of traps (i.e. better quality semiconductor material) and the lower the amount of scattering, the lower the rate of mixing and hence the lower the level of flicker noise. The delayed feedback in the chaotic-map model is consistent with trapping and the observation that reducing traps improves phase noise performance.

The long-range mixing indicated by the slow decay of the correlation function (see Figure 5-47) is key to the f^{-1} response. Another function that yields a long-term correlation is the Ornstein–Uhlenbeck process [50–52] developed to describe Brownian motion. The autocorrelation function of this process decays exponentially and predicts f^{-2} noise but not f^{-1} noise [52]. The Ornstein–Uhlenbeck process decays too rapidly to predict the f^{-1} response. This is discussed further in [52].

So the solution of Equation (5.35) (the logarithmic map), shown in Figure 5-45, has complicated dynamics with long periods of stability with rapid transitions between stable and rapidly varying levels. The sequence of x_ns depends on the starting condition (i.e., x_0), but no matter how it starts, the power spectrum of the solution has an inverse frequency dependence (i.e., it is exactly f^{-1}). The logarithmic map, as with all chaotic maps, describes a nonlinear process with delayed feedback. This matches the situation in physical, biological, chemical, and financial systems. Since nearly every physical process can be described as a (perhaps weak) nonlinear process with delayed feedback, the widespread observation of $1/f$ fluctuations is not surprising. So the basis of $1/f$ noise is the most basic of physical processes.

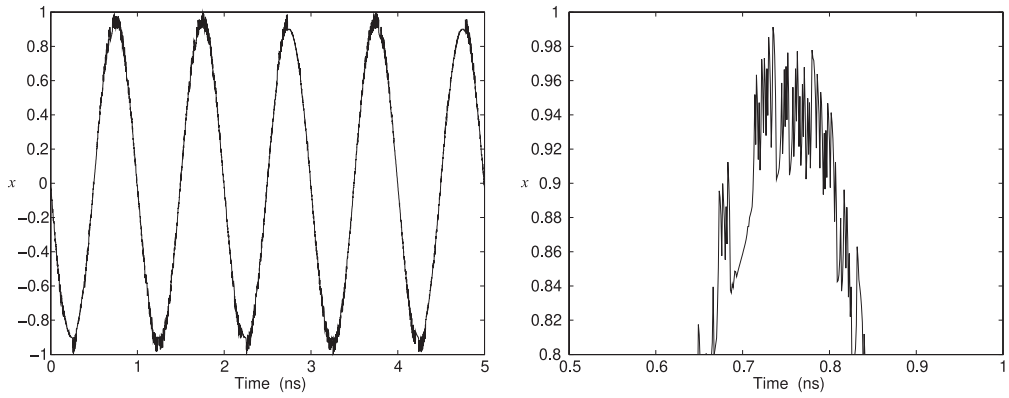
Intermittency (described by chaotic maps) results in random fluctuations in the amplitude and phase of a sinusoidal signal processed by a nonlinear electronic system such as an amplifier or oscillator. Denoting the amplitude intermittency by $a_I(t)$ and the phase intermittency by $\phi_I(t)$, a sinusoidal signal with mean amplitude A and an initial phase of zero is

$$x(t) = A[1 + a_I(t)]\cos[\omega t + \phi_I(t)]. \tag{5.37}$$

This signal is shown in Figure 5-48 with the logarithmic intermittency fluctuations $a_I(t)$ and $\phi_I(t)$ as shown in Figure 5-45, calculated using different seeds. The effect of intermittency fluctuations is greatly exaggerated here for visualization purposes. In practice, the fluctuations at the scale shown in Figure 5-48 could be eliminated using a bandpass filter. However, the fluctuations are self-similar (another property of chaotic processes) and is repeated at all scales. In a bandpass electronic system the in-band amplitude fluctuations are suppressed by device nonlinearity, but the phase fluctuations appear as phase noise on an oscillator.

5.8.6 Summary

Three oscillator phase noise models were presented. The Leeson model is based on up-conversion of white noise from baseband producing noise around the oscillator carrier with an f^{-2} dependency. The Hajimiri and Lee model is based on a linear time-varying model of the oscillator with f^{-1} noise at baseband resulting in oscillator phase noise with a $1/(\Delta f)^3$ dependency, and white noise at the baseband and harmonics resulting in noise around the oscillating frequency with an $1/(\Delta f)^2$ dependency. Up-conversion of noise has been demonstrated as a mechanism that describes



(a) Waveform (b) Detail

Figure 5-48: Sinusoidal signal with superimposed intermittency noise. Using Equation (5.37) with $A = 0.9$, $a_I(t)$ scaled to the interval $[0, 0.09]$, $a_I(0) = 0.477347$ (before scaling), $\phi_I(t)$ scaled to the interval $[0, 5]$ radians, and $\phi_I(0) = 0.00915926$ (before scaling).

some of the observed oscillator noise. Neither the Leeson nor the linear time-varying models describe the full set of observations of phase noise with $1/(\Delta f)^5, \ldots 1/(\Delta f)$ dependencies.

The chaotic map model is physically appealing and describes the origin of flicker noise as the time-delayed feedback of the output of a nonlinear process. This can produce a chaotic response called intermittency that embodies long-term memory. It has been shown through simulation that this model predicts the $1/(\Delta f)^3$, $1/(\Delta f)^2$, $1/(\Delta f)^1$, and $1/(\Delta f)^0$ dependencies of phase noise. It is also consistent with the random walk seen in time-domain observations of oscillator noise. However, the chaotic map-based model has not yet led to a compact formula for phase noise similar to Leeson's formula. The development of a compact phase noise model (e.g., like Leeson's model) will not be simple as integer calculus and transfer function-based analyses cannot be directly used with a chaotic map. However, it is clear that the description of phase noise is getting close to a satisfying physical explanation.

Phase noise can also be induced by vibrations [53, 54], and spurious signals from the environment (such as those coupled from the power mains) can also appear as phase noise.

5.9 Case Study: Oscillator Phase Noise Analysis

In this section the oscillator shown in Figure 5-38 is modeled and simulated in the time-domain [25, 26, 36]. The circuit includes three nonlinear devices, two identical BJTs and one varactor, which must be modeled. The process of developing a device model involves fitting the coefficients of a physically based model to measured characteristics. The development of models of the noise sources also requires fitting some noise parameters to measurements.

Device Modeling

The physical model used with each of the BJTs is shown in Figure 5-49 along with thermal, shot, and flicker noise sources. The Gummel-Poon BJT model was used with parameters provided by the manufacturer. The thermal noise sources, $i_{t,rb}$, $i_{t,rc}$, and $i_{t,re}$, are associated with the parasitic resistors at the collector, base, and emitter, respectively. The thermal noise current source models are

$$i_{t,rc} = \sqrt{\frac{2kT}{R_c}}\xi_c, \quad i_{t,rb} = \sqrt{\frac{2kT}{R_{bb}}}\xi_b, \quad \text{and} \quad i_{t,re} = \sqrt{\frac{2kT}{R_e}}\xi_e, \tag{5.38}$$

where ξ_c, ξ_b, and ξ_e are sequences of white noise generated by the logistic map of Equation (5.31) with $\lambda = 4$. The ξ_cs have values between 0 and 1. In this case the model is based on physical mechanisms and fitting of the noise model to measurements is not required.

Based on the development in [55] and [56], when the transistor is in the forward active region, minority carriers diffuse and drift across the base into the base-collector region. Since there is an electric field, the charges undergo acceleration when they enter the collector-base depletion region and are swept to the collector. This is a random process and is the source of shot noise in the collector. Recombination effects in the base-emitter region and carrier injection from the base into the emitter are also random processes contributing to shot noise in the base and emitter, respectively. Thus there are three shot noise current sources, $i_{\mathrm{s,ce}}$, $i_{\mathrm{s,be}}$, and $i_{\mathrm{s,bc}}$, which are proportional to the instantaneous collector-emitter, base-emitter, and base-collector currents, respectively. They are modeled as follows:

$$i_{s,ce} = \sqrt{e|i_{ce}|}\xi_{ce}, \quad i_{s,be} = \sqrt{e|i_{be}|}\xi_{be}, \quad \text{and} \quad i_{s,bc} = \sqrt{e|i_{bc}|}\xi_{bc}, \tag{5.39}$$

where ξ_{ce}, ξ_{be}, and ξ_{bc} are white noise sequences between 0 and 1 and generated by the logistics map with $\lambda = 4$, and e is the elementary charge.

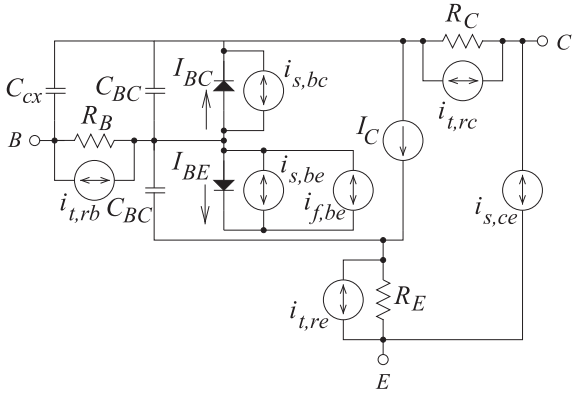


Figure 5-49: The Gummel-Poon BJT model, along with noise sources.

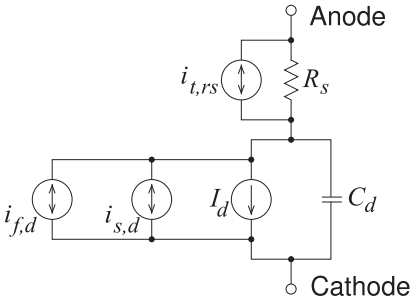


Figure 5-50: The p-n junction diode model with noise source.

Although there is more than one known source of flicker noise in a BJT [57], it has been shown that the dominant source of flicker noise can be modeled as a single noise current between the base and emitter terminals. It is a function of the instantaneous base-emitter recombination current [58] and so the flicker noise source is

$$i_{f,be} = k_f \sqrt{|i_{be}|^\alpha} \xi_f. \tag{5.40}$$

Here ξ_f is between 0 and 1 and is a sequence generated by the logarithmic map in Equation (5.34) and α controls the dependence of the flicker noise component on the non-ideal base current; here α is set to 2. The logarithmic map sequence, ξ_f, is parameterized by β, which was fit to noise measurements yielding $\beta = 0.000005$. The other parameter, k_f, sets the amplitude of the flicker noise and fitting to noise data yielded $k_f = 0.001$. The same flicker noise parameters were used with both BJTs. All of the noise sources were uncorrelated and this was achieved by randomly choosing initial seeds of each sequence, the ξs.

The diode model is shown in Figure 5-50. The noisy model of the diode includes a thermal current source for the parasitic resistance, a shot noise current source, and a flicker noise current source that is dependent on the current flowing through the diode [56]. The diode's thermal, shot, and flicker noise current sources are

$$i_{t,rs} = \sqrt{\frac{2kT}{R}} \xi_t, \quad i_{s,d} = \sqrt{e i_d} \xi_s, \quad \text{and} \quad i_{f,d} = k_{fd} \sqrt{i_d^\alpha} \xi_f, \tag{5.41}$$

respectively. Here the parameter k_{fd} is the scaling coefficient for the diode flicker noise and α controls the dependence of the flicker noise component on the current in the diode. As with the BJT, $\alpha = 2$ in the simulations reported here. The parameter β controls the slope of the autocorrelation characteristic. As before, the random variables ξ_t and ξ_s, describing the thermal and shot noise processes, were generated using the logistic map, and ξ_f, describing the flicker noise process, was generated using the logarithmic map. The amplitude of the flicker noise source was fit to measurements and the same $\beta = 0.00005$ parameter was used as was determined for the BJT model.

The noise model of the oscillator is completed by modeling the thermal noise current of each resistor as in Equation (5.38).

Oscillator Simulation

The VCO circuit of Figure 5-38 was simulated in the time domain using the transient simulator described in [25] and [26]. In the circuit model there are a total of three flicker noise parameters fit to measurements, k_f for the BJTs, k_{fd}

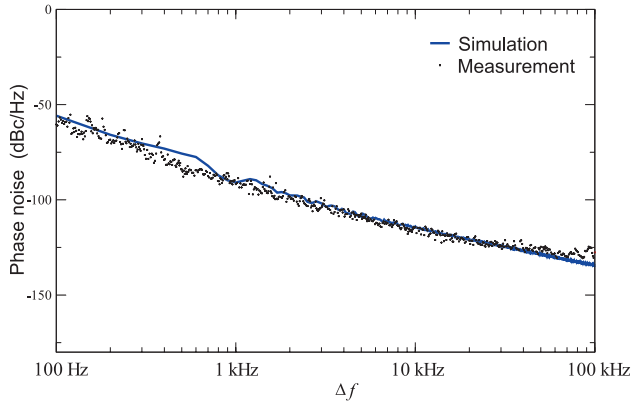


Figure 5-51: Phase noise comparison between data and experiment with bias voltage at 0 V.

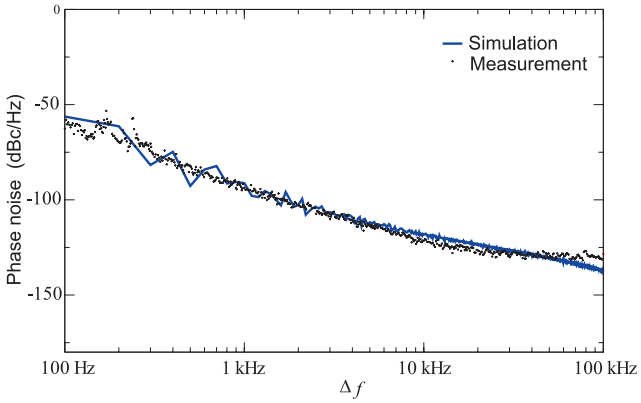


Figure 5-52: Phase noise comparison between data and experiment with bias voltage at 6 V.

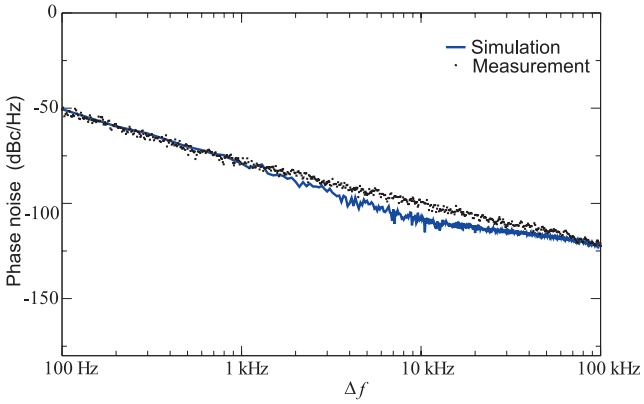


Figure 5-53: Phase noise comparison between data and experiment with bias voltage at 12 V.

for the diode, and the same value of β was used for the BJTs and the diode. Thus there are three noise parameters to be set in the VCO circuit. These were unchanged in simulations of the VCO with different varactor bias voltages. These are the only noise parameters that are not fixed by the device currents and parasitic resistance values. The phase noise output following simulation was Fourier analyzed yielding the phase noise results shown in Figures 5-51 to 5-53. These measured results were also presented in Figures 5-39 and 5-40, where the phase noise slopes were identified as f^0, f^{-1}, f^{-2}, and f^{-3}. So the simulated phase noise results in Figures 5-51 to 5-53 correctly calculate the various phase noise slopes and crossover frequencies for diverse varactor tuning conditions.

5.10 Summary

This chapter focused on the design of two categories of microwave oscillators: fixed-frequency oscillators and VCOs. All microwave oscillators can be considered to be an amplifier with a feedback network with the amplifier establishing the amplitude of oscillation and the feedback network setting the frequency of oscillation. As well, nearly all microwave oscillators can be considered as reflection oscillators in which the active device, with appropriate feedback, presents a negative resistance, or equivalently, a negative conductance to a linear frequency-selective circuit. Whether design based on negative conductance or negative resistance is used depends on which, resistance or conductance, reduces in magnitude as the signal level increases. Since transistors are essentially voltage-controlled current sources, and the current tends to saturate at high output levels, the natural view is to consider transistor-based oscillators as having a negative conductance that reduces in magnitude as the signal level increases. With negative-conductance reflection oscillators part or all of the feedback network appears as a linear two-terminal, that is one-port, circuit in parallel with the negative conductance from the active device. This two-terminal circuit is often called the resonator or tank circuit.

The chapter presented two case studies of oscillator design. This enabled design decisions to be illustrated that could not be presented in an algorithmic way. While the design case studies considered common-base Colpitts designs, the principles apply to other types of oscillators. Still the Colpitts common-base/common-drain oscillator core has proved to yield microwave VCOs with stellar performance. RFICs necessarily use differential circuits, but even then they are conveniently designed as negative conductance oscillators. Treating the design problem as that of interconnected one-port circuits considerably simplifies making design trade-offs, and so makes it easier to achieve an optimal VCO.

VCO design is the most complicated of microwave designs with significant trade-offs of low phase noise, stability, broad tuning range, rapid turn-on transient, high output power, and high efficiency. Several decades ago fixed-frequency oscillators were most common. With these the feedback network incorporates a high-Q resonant element that results in the phase noise on the oscillating signal being insignificant in nearly all microwave applications. The high-Q resonator of the fixed-frequency oscillators largely assures single frequency of operation, and the desired constant amplitude output is assured by ensuring the active device enters saturation. With a VCO, stability is not as easy to achieve. Stability here refers to the oscillator producing a sinewave of a single frequency and constant amplitude.

With VCOs the resonator is nearly always of relatively low Q, as it must be variable. Then a major concern in oscillator design is phase noise appearing on the output signal. Managing phase noise is complicated because the origins of phase noise are not well known and so physically based device models, that would enable the accurate determination of phase noise in computer-based RF circuit simulation, are not available. For noncompetitive VCOs there are many noise sources other than the intrinsic phase noise of an active device that dominate the phase noise of the output signal. For example, it is known that up-conversion of low-frequency noise, and down-conversion of harmonic noise, produce oscillator phase noise. However,

once these sources of phase noise are minimized in design, there is a remaining intrinsic phase noise component. It is this phase noise, intrinsic to active devices, that is the concern of competitive VCO design. While the origins of intrinsic phase noise in well-designed oscillators is not completely understood, there are best practices to follow that minimize the coupling of external noise to the oscillating signal. External noise at low frequencies will be up-converted unless care is taken. Good design practice is to eliminate low-frequency noise from the power supply and from the tuning voltage source for VCOs. Some more specific guidelines:

1. Good grounding is required with decoupling capacitors between the supply and ground.
2. Attention must be given to the signal return path for the supply and tuning voltage source to avoid common impedance coupling. Any noise on the tuning voltage in a VCO will result in phase noise on the VCO output. In a microstrip circuit, only minimum metal on the microstrip layer should be removed with floating metal connected to the ground through multiple vias. This suppresses substrate modes, minimizes parasitic coupling, and minimizes electro-thermal passive intermodulation distortion.
3. The oscillator output should drive a resistive load and it is common to use a resistive pad (i.e., attenuator) followed by a bandpass filter. This reduces the effect of the load on the oscillation frequency. Also it is common to isolate the tank circuit from the load as was seen in the case studies presented in this chapter.
4. Internal oscillator paths should be as small as possible to minimize the coupling of noise from the environment.

A VCO can be incorporated in a phase-locked loop which further reduces phase-noise and locks the oscillator to a low-frequency highly stable frequency reference. The oscillation frequency is then controlled by the fractional frequency division in the phase-locked loop.

5.11 References

[1] M. Odyniec, Ed., *RF and Microwave Oscillator Design*. Artech House, 2002.

[2] A. Rohde, U.L. Poddar and G. Bock, *The Design of Modern Microwave Oscillators for Wireless Applications*. Wiley, 2005.

[3] E. Faulkner, *Introduction to the Theory of Linear Systems*. Chapman & Hall, 1969.

[4] A. Pippard, *Response and Stability*. Cambridge University Press, 1985.

[5] L. Dai and R. Harjani, *Design of High Performance CMOS Voltage-Controlled Oscillators*. Kluwer Academic Publishers, 2003.

[6] M. Tiebout, *Low Power VCO Design in CMOS*. Springer, 2006.

[7] A. Aktas and M. Ismail, *CMOS PLLs and VCOs for 4G wireless*. Kluwer, 2004.

[8] D. Pederson and K. Mayaram, *Analog Integrated Circuits for Communication: Principles, Simulation and Design*. Springer, 2008.

[9] B. Razavi, *Design of Analog CMOS Integrated Circuits*. McGraw-Hill, 2001.

[10] T. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*. Cambridge University Press, 2004.

[11] K. Kurokawa, "Some basic characteristics of broadband negative resistance oscillator circuits," *The Bell System Technical J.*, vol. 48, no. 69, pp. 1937–1955, Jul.-Aug. 1969.

[12] G. Gonzalez, *Foundations of Oscillator Circuit Design*. Artech House, 2007.

[13] C. Rauscher, "Large-signal technique for designing single-frequency and voltage-controlled GaAs FET oscillators," *IEEE Trans. on Microwave Theory and Techniques*, vol. 29, no. 4, pp. 293–304, Apr. 1981.

[14] R. Jackson, "Criteria for the onset of oscillation in microwave circuits," *IEEE Trans. On Microwave Theory and Techniques*, vol. 40, no. 3, pp. 566–569, Mar. 1992.

[15] D. Maclean, *Evaluating Feedback in Amplifiers*

and Oscillators: Theory, Design and Analogue Applications. Research Studies Press LTD., 2004.

[16] J. Boyles, "The oscillator as a reflection amplifier: an intuitive approach to oscillator design," *Microwave Journal*, vol. 29, no. 6, pp. 83–98, Jun. 1986.

[17] D. Esdale and M. Howes, "A reflection coefficient approach to the design of one-port negative impedance oscillators," *IEEE Trans. on Microwave Theory and Techniques*, vol. 29, no. 8, pp. 770–776, Aug. 1981.

[18] G. Gonzalez and O. Sosa, "On the design of a series-feedback network in a transistor negative-resistance oscillator," *IEEE Trans. on Microwave Theory and Techniques*, vol. 47, no. 1, pp. 42–47, Jan. 1999.

[19] A. Victor and M. Steer, "Reflection coefficient shaping of a 5-GHz voltage-tuned oscillator for improved tuning," *IEEE Trans. on Microwave Theory and Techniques*, vol. 55, no. 12, pp. 2488–2494, Dec. 2007.

[20] A. Knights and M. Kelly, "Laterally stacked varactor formed by ion implantation," *Electronics Letters*, vol. 35, no. 10, pp. 846–847, May 1999.

[21] M. Steer, *Microwave and RF Design, Modules*, 3rd ed. North Carolina State University, 2019.

[22] P. J. Topham, A. Dearn, and G. Parkinson, "GaAs bipolar broadband oscillators," in *21st European Microwave Conf.*, Sep. 1991, pp. 178–183.

[23] http //www.febo.com.

[24] A. Victor, J. Nath, D. Ghosh, B. Boyette, J. Maria, M. Steer, A. Kingon, and G. Stauf, "Noise characteristics of an oscillator with a barium strontium titanate (BST) varactor," *IEE Proc., Part H, Microwaves, Antennas and Propagation*, vol. 153, no. 1, pp. 96–102, 2006.

[25] N. Kriplani, A. Victor, and M. Steer, "Time-domain modelling of phase noise in an oscillator," in *36th European Microwave Conf.*, 2006, pp. 514–517.

[26] N. Kriplani, S. Luniya, and M. Steer, "Integrated deterministic and stochastic simulation of electronic circuits: Application to large signal–noise analysis," *Int. Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, vol. 21, no. 6, pp. 381–394, 2008.

[27] A. Victor, "Microwave power oscillator utilizing thin-film ferroelectic varactors," Ph.D. dissertation, North Carolina State University, 2010.

[28] D. Leeson, "A simple model of feedback oscillator noise spectrum," *Proc. of the IEEE*, vol. 54, no. 2, pp. 329–330, Feb. 1966.

[29] T. Lee and A. Hajimiri, "Oscillator phase noise: a tutorial," *IEEE J. of Solid-State Circuits*, vol. 35, no. 3, pp. 326–336, Mar. 2000.

[30] S. Kogan, *Electronic Noise and Fluctuations in Solids*. Cambridge University Press, 1996.

[31] A. Hajimiri and T. Lee, "A general theory of phase noise in electrical oscillators," *IEEE J. of Solid-State Circuits*, vol. 33, no. 2, pp. 179–194, Feb. 1998.

[32] ——, "Design issues in cmos differential lc oscillators," *IEEE J. of Solid-State Circuits*, vol. 34, no. 5, pp. 717–724, May 1999.

[33] ——, "Corrections to "a general theory of phase noise in electrical oscillators"," *IEEE J. of Solid-State Circuits*, vol. 33, no. 6, p. 928, Jun. 1998.

[34] G. Wornell and A. Oppenheim, *Signal Processing with Fractals: A Wavelet-Based Approach*. Prentice Hall, 1996.

[35] G. Wornell, "Wavelet-based representations for the 1/f family of fractal processes," *Proc. of the IEEE*, vol. 81, no. 10, pp. 1428–1450, Oct. 1993.

[36] N. Kriplani, "Modelling colored noise under large-signal conditions," Ph.D. dissertation, North Carolina State University, 2005.

[37] E. Bullmore, C. Long, J. Suckling, J. Fadili, G. Calvert, F. Zelaya, T. Carpenter, and M. Brammer, "Colored noise and computational inference in neurophysiological (fmri) time series analysis: resampling methods in time and wavelet domains," *Human Brain Mapping*, vol. 12, no. 2, pp. 61–78, 2000.

[38] B. Mandelbrot, *The Fractal Geometry of Nature*. Times Books, 1982.

[39] B. Mandelbrot and J. Van Ness, "Fractional brownian motions, fractional noises and applications," *SIAM Review*, vol. 10, no. 4, pp. 422–437, 1968.

[40] M. Schroeder, *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise*. Dover Publications, 2009.

[41] R. Voss, *Fractals in Nature: From Characterization to Simulation*. Springer-Verlag, 1988.

[42] R. Devaney, *A First Course in Chaotic Dynamical Systems*. Perseus Book, 1992.

[43] Y. Pomeau and P. Manneville, "Intermittent transition to turbulence in dissipative dynamical systems," *Commun. Math. Phys*, vol. 74, no. 2, pp. 189–197, 1980.

[44] R. Bhansali, M. Holland, and P. Kokoszka, "Chaotic maps with slowly decaying correlations and intermittency," in *Fields Institute Communications: Asymptotic Methods in Stochastics*, L. Horváth and B. Szyszkowicz, Eds., 2004, p. ●.

[45] R. May, "Simple mathematical models with very complicated dynamics," *Nature*, vol.

261, no. 5560, pp. 459–467, June 1976.

[46] T. Kawabe and Y. Kondo, "Intermittent chaos generated by logarithmic map," *Progress of Theoretical Physics*, vol. 86, no. 3, pp. 581–586, 1991.

[47] M. Holland, "Slowly mixing systems and intermittency maps," *Ergodic Theory and Dynamical System*, vol. 25, no. 1, pp. 133–159, Feb. 2005.

[48] E. Simoen, A. Mercha, L. Pantisano, C. Claeys, and E. Young, "Low-frequency noise behavior of SiO2–HfO2 dual-layer gate dielectric nMOSFETs with different interfacial oxide thickness," *IEEE Trans. on Electron Devices*, vol. 51, no. 5, pp. 780–784, May 2004.

[49] F. Hooge, T. Kleinpenning, and L. Vandamme, "•experimental studies on 1/f noise," *Reports on Progress in Physics*, 1981.

[50] G. Uhlenbeck and L. Ornstein, "On the theory of brownian motion," *Physical Review*, vol. 36, pp. 823–841, Sep. 1930.

[51] E. Bibbona, G. Panfilo, and P. Tavella, "The ornstein-uhlenbeck process as a model of a low pass filtered white noise," *Metrologia*, vol. 45, no. 6, pp. S117–S126, Dec. 2008.

[52] A. Suárez, *Analysis and Design of Autonomous Microwave Circuits*. John Wiley & Sons, Inc., 2009.

[53] A. Hati, C. Nelson, and D. Howe, "Effect of vibration on P and AM noise of oscillatory and non-oscillatory components at 10 GHz," in *2009 IEEE Int. Frequency Control Symp., Joint with the 22nd European Frequency and Time Forum*, Apr. 2009, pp. 524–529.

[54] W. Robins, *Phase Noise in Signal Sources*. Peter Peregrinus Ltd., 1982.

[55] M. Buckingham, *Noise in Electronic Devices and Systems*. Ellis Horwood, 1983.

[56] P. Antognetti and G. Massobrio, *Semiconductor Device Modeling with SPICE*. McGraw-Hill, 1988.

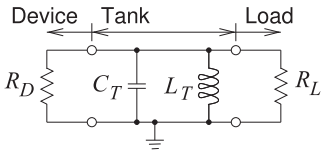
[57] A. van der Ziel, X. Zhang, and A. Pawlikiewicz, "Location of 1/f noise sources in BJT's and HBJT's—i. theory," *IEEE Trans. on Electron Devices*, vol. 33, no. 9, pp. 1371–1376, Sep. 1986.

[58] C. Green and B. Jones, "1/f noise in bipolar transistors," *J. Phys. D: Appl. Phys.*, vol. 18, pp. 77–91, 1985.

5.12 Exercises

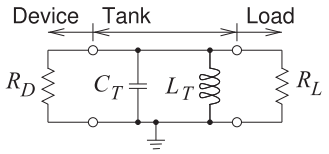
1. Draw the schematic of an opamp-based Hartley oscillator circuit.

2. Consider the circuit below. This is the equivalent circuit of an active device with a negative conductance $= 1/R_D$ connected to a tank circuit comprising capacitor C_T and inductor L_T, and then a load R_L. The oscillation amplitude will adjust so that $R_D = R_L$.

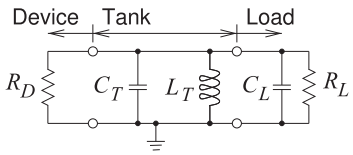


   (a) Write down a formula for the oscillation frequency f_0.
   (b) What is f_0 if $C_T = 1$ pF and $L_T = 1$ nH?

3. The circuit below is the equivalent circuit of a reflection oscillator with a negative conductance $= 1/R_D$ connected to a tank circuit $C_T = 0.1$ pF and $L_T = 0.5$ nH. What is the oscillation frequency assuming that there is sufficient negative conductance for oscillation to occur?



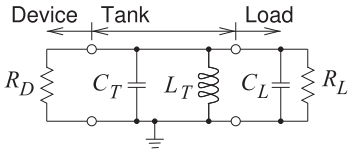
4. The circuit below is the equivalent circuit of a reflection oscillator with a negative conductance $= 1/R_D$ connected to a tank circuit with a capacitance $C_T = 1$ pF and inductance $L_T = 1$ nH. The load consists of a capacitor C_L in parallel with a resistor R_L. The oscillation amplitude will adjust so that $R_D = R_L$.



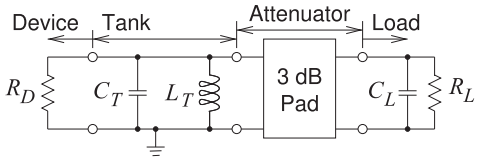
   (a) Write down a symbolic formula for the oscillation frequency f_0.
   (b) What is f_0 if $C_L = 0.2$ pF and $R_L = 50\ \Omega$?

5. The circuit below is the equivalent circuit of a reflection oscillator with a negative conductance $= 1/R_D$ connected to a tank circuit $C_T = 0.1$ pF and $L_T = 0.5$ nH. The capacitance of the load is $C_L = 0.05$ pF and the load resistance is $50\ \Omega$. What is the oscillation frequency assuming that

there is sufficient negative conductance for oscillation to occur?

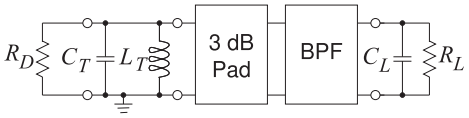


6. The circuit below is the equivalent circuit of a reflection oscillator with a negative conductance $= 1/R_D$ connected to a tank circuit with a capacitance $C_T = 1$ pF and inductance $L_T = 1$ nH. The load consists of a capacitor $C_L = 0.2$ pF in parallel with a resistor $R_L = 50\ \Omega$. Between the tank circuit is a 3 dB 50 Ω attenuator (i.e., the system is designed for system impedance of 50 Ω). The oscillation amplitude will adjust so that $R_D = R_L$.



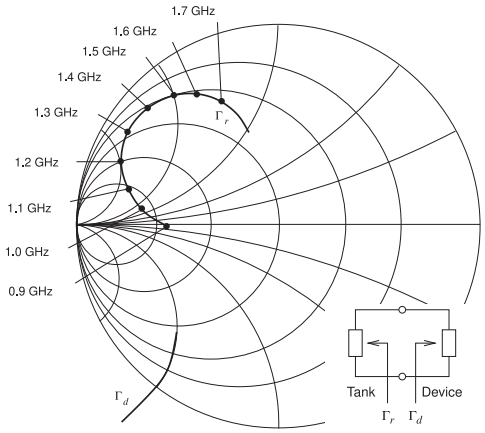
(a) Ignore the load capacitor C_L, what is the frequency of oscillation, f_0, of the oscillator?
(b) From now on consider the load capacitance. What is the oscillation frequency without the attenuator?
(c) At the frequency calculated in (a), what is the admittance looking into the attenuator from the tank circuit with the load comprising C_L and R_L? [Hint use the resistive PI network equivalent of the attenuator.]
(d) What is the equivalent shunt resistor and capacitor circuit looking into the attenuator from the oscillator? This is the effective load seen by the active device and tank circuit.
(e) What is the oscillation frequency with the attenuator and the RC load?
(f) Using your results above, discuss the effect of the attenuator on reducing the sensitivity of oscillation on the loading conditions.

7. The circuit below is the equivalent circuit of a reflection oscillator with a negative conductance $= 1/R_D$ connected to a tank circuit with a capacitance $C_T = 1$ pF and inductance $L_T = 1$ nH. The load consists of a capacitor $C_L = 0.2$ pF in parallel with a resistor $R_L = 50\ \Omega$. Between the tank circuit is a 3 dB 50 Ω attenuator (ti.e., the system is designed for system impedance of 50 Ω) and a bandpass filter with an insertion loss at the oscillation frequency of 2 dB. The oscillation amplitude will adjust so that $R_D = R_L$.



(a) Ignore the load capacitor C_L, what is the frequency of oscillation, f_0, of the oscillator?
(b) From now on consider the load capacitance. What is the oscillation frequency without the attenuator and bandpass filter?
(c) At the frequency calculated in (a), what is the admittance looking into the attenuator from the tank circuit with the bandpass filter and the load comprising C_L and R_L? [Hint: Consider that the effect of the insertion loss of the filter is an attenuation. Consider using a resistive PI network equivalent of the attenuator plus bandpass filter.]
(d) What is the equivalent shunt resistor and capacitor circuit looking into the attenuator for the oscillator? This is the effective load seen by the active device and tank circuit.
(e) What is the oscillation frequency with the attenuator and the RC load?

8. A reflection oscillator is shown below along with the Γ_r and Γ_d loci plotted on a Smith chart. Γ_r, the reflection coefficient looking into the resonator, rotates clockwise as frequency increases. Γ_d, the reflection coefficient of the active device, is amplitude dependent but is frequency independent. The arrowed Γ_d curve plots the loci of Γ_d as amplitude increases. [Hint: Consider Figure 5-18.]



(a) On the Γ_r plane show the region of the Smith Chart identifying oscillation.
(b) On the Γ_d plane show the region of the Smith Chart identifying oscillation.
(c) What is the frequency of oscillation?

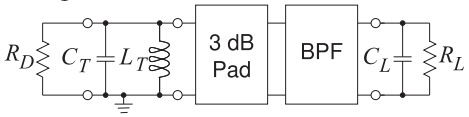
9. Consider an oscillator that can be modeled as shunt connections of a linear conductance, G_r,

a linear susceptance, B_r, a nonlinear or device conductance, G_d, and a nonlinear or device susceptance, B_d.

(a) Draw the circuit.
(b) Using G_r, B_r, G_d, and B_d, write down the Kurokawa oscillator condition that establishes stable single frequency oscillation.
(c) What are common design choices G_r, B_r, G_d, and B_d made that simplify the Kurokawa oscillator condition? Write down the resulting simplified Kurokawa oscillation condition.

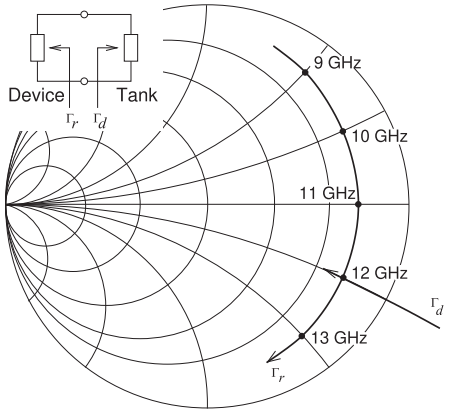
10. An oscillator has a linear conductance, G_r, a linear susceptance, B_r, a nonlinear or device conductance, G_d, and a nonlinear or device susceptance, B_d, all in shunt. The reflection coefficient looking into the linear network is Γ_r and the reflection coefficient looking into the device is Γ_d.
    (a) Use a Smith chart sketch to show the loci of the linear and nonlinear admittances (or equivalently their reflection coefficients Γ_r and Γ_d) for the case when loss in the linear network is low (i.e. $G_r \approx 0$). Do this when the oscillating signal is small and when it has reached steady state. Indicate the steady-state oscillation point. (You can use the negative or the inverse of either Γ_r or Γ_d as appropriate.)
    (b) Use a Smith chart sketch to show the loci of the linear and nonlinear admittances (or equivalently their reflection coefficients for the case when loss in the linear network is high (i.e. G_r cannot be ignored). Do this when the oscillating signal is small and when it has reached steady state. (You can use the negative of the inverse of either Γ_r or Γ_d as appropriate.)

11. The equivalent circuit of an oscillator is shown below with $R_L = 50\ \Omega$, $C_T = 1$ pF and inductance $L_T = 0.1$ nH.



    (a) What is the oscillation frequency assuming that there is sufficient negative conductance for oscillation to occur?
    (a) What is R_D when there is oscillation?

12. A reflection oscillator is shown below along with the Γ_r and Γ_d loci plotted on a Smith chart. Γ_r, the reflection coefficient looking into the resonator, rotates clockwise as frequency increases. Γ_d, the reflection coefficient of the active device, is amplitude dependent but is frequency independent. The arrowed curve plots the loci of Γ_d as amplitude increases. [Hint: Consider Figure 5-18.]



    (a) On the Γ_r plane show the region of the Smith Chart identifying oscillation.
    (b) On the Γ_d plane show the region of the Smith Chart identifying oscillation.
    (c) What is the frequency of oscillation?

13. The case study presented in Section 5.6 described the design of a 5 GHz VCO. The output of the oscillator was followed by a resistive Pi attenuator and then a bandpass filter. What is the attenuation (in decibels) and the system impedance of the attenuator. [Hint: You will need to carefully read Section 5.6.]

14. Consider a common emitter BJT Clapp oscillator.
    (a) Draw the schematic of the oscillator without the bias circuit.
    (b) Redraw the oscillator circuit including the bias circuit.

15. Consider a BJT Hartley oscillator.
    (a) Draw the schematic of the oscillator circuit in the common-base configuration. Do not show the bias circuit.
    (b) Redraw the schematic of the oscillator circuit in the common-base configuration, this time showing the bias circuit.

16. Consider a FET Clapp oscillator.
    (a) Draw the schematic of the oscillator circuit in the common drain configuration. Do not show the bias circuit.
    (b) Redraw the schematic of the oscillator circuit, this time showing the bias circuit.

17. Consider a FET Hartley oscillator.
    (a) Draw the schematic of the oscillator circuit in the common source configuration. Do not show the bias circuit.

(b) Redraw the schematic of the oscillator circuit, this time showing the bias circuit.

18 Consider a common source FET Clapp oscillator.
(a) Draw the schematic of the oscillator without biasing.
(b) Redraw the oscillator circuit including bias current sources.

19. Derive an expression for the oscillation frequency of the Colpitts BJT oscillator in the common emitter configuration shown in Figure 5-36(b).

20. A two-port feedback oscillator is shown in Figure 5-1.
(a) Draw the schematic of a feedback Colpitts oscillator.
(b) Considering that the amplifier in the feedback oscillator has a gain that is independent of frequency, what is the oscillation frequency if the components of the Colpitts feedback network are C_1 = C_2 = 2 pF and L_3 = 5 nH. Ignore any phase shift introduced by the amplifier.

21. A two-port feedback oscillator is shown in Figure 5-1.
(a) Draw the schematic of a feedback Colpitts oscillator.
(b) Considering that the amplifier in the feedback oscillator has a gain that is independent of frequency, what is the oscillation frequency if the components of the Colpitts feedback network are C_1 = 1 pF, C_2 = 3 pF, and L_3 = 1 nH. Ignore any phase shift introduced by the amplifier.

22. A two-port feedback oscillator is shown in Figure 5-1.
(a) Draw the schematic of a feedback Colpitts oscillator.
(b) Considering that the amplifier in the feedback oscillator has a gain that is independent of frequency, what is the oscillation frequency if the components of the Colpitts feedback network are C_1 = 5 pF, C_2 = 1 pF, and L_3 = 10 nH? Ignore any phase shift introduced by the amplifier.

23. A negative-gm differential FET VCO, as shown in Figure 5-34, has $C = 0.2$ pF and $L = 0.2$ nH. $V_{DD} = 5$ V and the circuit is biased so that for each transistor $g_m = 1$ mS. The output at the collector of the transistor drives a 1 kΩ load. Ignore the internal parasitics of the transistor and you must consider the possibility that the circuit does not oscillate. What is the oscillation frequency of the oscillator?

24. A negative-gm differential FET VCO, as shown in Figure 5-34, has $C = 0.2$ pF and $L = 0.2$ nH. $V_{DD} = 5$ V and the circuit is biased so that for each transistor $g_m = 1$ mS. The output at the collector of the transistor drives a $50\ \Omega$ differential load that is in parallel with a 0.5 pF capacitor. Ignore the internal parasitics of the transistor and you must consider the possibility that the circuit does not oscillate.
(a) Draw the schematic of the oscillator with the load.
(b) Draw the equivalent tank circuit of the oscillator. This will need to include the effect of the load.
(c) What is the oscillation frequency of oscillator?

25. A negative-gm differential FET VCO, as shown in Figure 5-34, has $C = 0.2$ pF and $L = 0.2$ nH. $V_{DD} = 5$ V and the circuit is biased so that for each transistor $g_m = 1$ mS. The output at the collector of the transistor drives a $50\ \Omega$ differential load that is in parallel with a 1 pF capacitor. Ignore the internal parasitics of the transistor and you must consider the possibility that the circuit does not oscillate. What is the oscillation frequency of the oscillator?

26. A common emitter Colpitts oscillator, as shown in Figure 5-36(a), has $C_1 = 1$ pF, $C_2 = 2$ pF, and $L_3 = 2$ nH. L_C is a choke inductor. $V_{CC} = 5$ V and the circuit is biased so that $g_m = 1$ mS. Ignore the internal parasitics of the transistor. What is the oscillation frequency of the oscillator? [Parallels Example 5.2]

27. A common emitter Colpitts oscillator, as shown in Figure 5-36(a), has $C_1 = 0.1$ pF, $C_2 = 0.2$ pF, and $L_3 = 0.5$ nH. L_C is a choke inductor. $V_{CC} = 5$ V and the circuit is biased so that $g_m = 1$ mS. Ignore the internal parasitics of the transistor. What is the oscillation frequency of the oscillator? [Parallels Example 5.2]

28. A digital communication system has a symbol rate of 1 MS/s.
(a) What is the highest frequency phase noise that will affect the bit error rate of the communication system? (It could be 1 MHz, 2 MHZ, no limit, 0.5 MHz, etc.)
(b) What changes would you make to the system (e.g., added components such as an attenuator, filter, amplifier, etc.) that will reduce the impact of high-offset phase noise?

29. A QPSK communication system has a transmitted bit rate of 100 kbit/s. Consider that the QPSK modulation scheme is ideal.
(a) What is the symbol rate?

(b) What would you do to the system (e.g., added components) to reduce the impact of high-offset phase noise?

30. A QPSK communication system has a transmitted bit rate of 1 Mbit/s. The QPSK modulation scheme is ideal (2 bit/s/Hz. DSP processing is such that phase noise slower than the duration of 5 symbols has no affect on the communication system throughput.
    (a) What is the symbol rate?
    (b) What is the lowest frequency phase noise that will affect the communication system?

31. An oscillator has phase noise that reduces away from the oscillation center frequency, f_{osc} at the rate of $1/\Delta f^2$ where Δf is the frequency offset from f_{osc}. If the phase noise at 100 MHz frequency offset from the oscillation frequency is -100 dBc/Hz, what is the phase noise at 10 MHz?

32. The phase noise of an oscillator measured at 1 MHz is -136 dBc/Hz. If the phase noise varies as $1/(\Delta f)$, where Δf is the offset from the center frequency of oscillation, what is the phase noise at 100 kHz offset?

33. A phase-locked microwave oscillator typically utilizes a low-Q oscillator. For such an oscillator the phase noise at the frequency that affects microwave systems has an inverse square relationship to frequency. The phase noise measured at 100 kHz is -106 dBc/Hz, what is the phase noise referred to 1 MHz?

34. A phase-locked microwave oscillator typically utilizes a low-Q oscillator. For such an oscillator the phase noise at the frequency that affects microwave systems often has an inverse square relationship to frequency. The phase noise measured at 1 MHz is -125 dBc/Hz, what is the phase noise at 100 kHz?

5.12.1 Exercises by Section

†challenging, ‡very challenging

§5.2 1
§5.3 $2^{\dagger}, 3^{\dagger}, 4^{\dagger}, 5^{\dagger}, 6^{\ddagger}, 7^{\ddagger}, 8, 9, 10, 11$
§5.5 12
§5.6 $13^{\dagger}$
§5.7 $14^{\dagger}, 15^{\dagger}, 16^{\dagger}, 17^{\dagger}, 18^{\dagger}, 19^{\dagger}, 20^{\dagger}, 21^{\dagger}, 22^{\dagger}, 23^{\dagger}, 24^{\dagger}, 25^{\dagger}, 26^{\dagger}, 27^{\dagger}$
§5.8 $28^{\dagger}, 29^{\dagger}, 30^{\dagger}, 31, 32^{\dagger}, 33^{\dagger}, 34^{\dagger}$

5.12.2 Answers to Selected Exercises

4 4.594 GHz
6(d) 49.8 Ω, 99.6 fF
7(e) 4.883 MHz
13 49.8 Ω, 2.97 dB
19 $\sqrt{\frac{C_1 + C_2}{L_3 C_1 C_2}}$
20 1.592 GHz
23 25.16 GHz
24(c) 10.27 GHz
30 100 kHz
34 -105 dBc/Hz

Index

η, 38
η_{overall}, 37
η_{total}, 37
η_D, 38
η_{PAE}, 37

AC loadline, 42, 117
active
  device
    models, 14
airbridge, 144
Allan
  deviation, 197
  variance, 197
amplifier, 25, 30, 38, 40
  k, 34
  back-off, 123
  balanced, 86
  cascode, 97, 148
    variable gain, 97
  case study
    distributed, 72
    distributed biasing, 96
    narrowband design, 57
    power amplifier, 128
    wideband design, 77
    WiMAX amplifier, 128
  Class A, 38, 40
    efficiency, 118
  class A, 116
  Class AB, 38, 39, 43
  class AB, 116
  Class B, 38, 39, 43
    efficiency, 118
    push-pull, 38
  class B, 116
  Class C, 38, 39
    efficiency, 118
    push-pull, 38
  class C, 116
  Class D, 116, 117, 119
    efficiency, 118
  Class E, 116, 117, 120
    efficiency, 118
  class E, 120
  Class F, 116–118
    efficiency, 118
  class F, 121
  classes, 38
    efficiencies, 118
  design
    narrowband, 57
    wideband, 77
  differential, 86, 94
    fully, 95
    pseudo, 95
  distributed, 71
  Doherty, 138
  efficiency, 37
  envelope tracking, 139
  FDA, 86
  figures of merit, 26
  FOM, 26
  fully differential, 86
  gain, 32
    definitions, 26
    maximum unilateral
      transducer, 32
    unilateral, 32
    unilateral transducer, 32
  inverted
    Class F, 122
  inverted class D, 122
  inverted class F, 122
  inverting
    Class D, 122
    Class F, 122
  Khan transmitter, 139
  LINC, 139
  linear, 26
  LITMUS, 140
  loadline, 40
  loadpull, 127
  MOSFET, 89
  multi-tanh, 148
  narrowband design, 57
  outphasing, 139
  PDA, 86, 88
  pseudo-differential, 86,
    88, 94, 96
  push-pull, 38
  RFIC, 97
  scattering parameters
    generalized, 30
  single-ended, 38, 41
  stability, 44
  switching, 116, 117
    Class D, 119
    Class E, 120
    class F, 121
    inverted class D, 122
    inverted class F, 122
  tanh cascode cell, 148
  TCC, 148
  wideband, 69
  wideband design, 77
amplitude
  probability density
    function, 123
APDF, 123
available
  gain, 29

balanced
  amplifier, 86
balun
  Marchand, 96
Barkhausen
  oscillation criterion, 161
  stability criterion, 161
Barkhausen criterion, 161
biasing, 96
  distributed, 96
bipolar junction transistor,
    4, 20
  HBT, 4
  heterostructure, 4
BJT, 4, 20
  device model, 20
  schematic symbol, 7

cascode amplifier, 97, 148
  variable gain, 97, 100
case study
  distributed amplifier, 72
  distributed biasing, 96
  narrowband amplifier, 57
  oscillator
    phase noise analysis, 206
    voltage-controlled, 176
  power amplifier, 128
  reflection oscillator, 167
  VCO, 176
  voltage-controlled
    oscillator, 176
  wideband amplifier, 77
  WiMAX amplifier, 128
CDMA, 123
chaos
  logarithmic map, 201
  logistics map, 201
  noise, 201
Clapp oscillator, 162
Class D amplifier, 119
Class E amplifier, 120
class F amplifier, 121
CMOS
  RFIC, 100
CMRR, 87, 96
collector efficiency, 38
Colpitts oscillator, 162
combiner, 94
common
  mode
    gain, 87
    rejection ratio, 87
compound semiconductor,
    11
conduction angle, 118
current
  mirror, 98
  source, 98

DC loadline, 40, 117
depletion mode
  FET, 7
  HEMT, 11
  JFET, 11
  MOSFET, 10
device model
  BJT, 20
  HEMT, 19
  MESFET, 19
  MOSFET, 14
differential
  amplifier, 86, 94, 95
    CMRR, 87
    common-mode rejection
      ratio, 87
    multi-tanh, 100
  circuit, 96
  mode
    gain, 87
digitally controlled
    circuits, 98
diode
  Schottky, 8
distortion, 122
  design guidelines, 124
distributed
  amplifier, 71
  biasing, 96
Doherty amplifier, 138
double
  stub tuner, 127
drain efficiency, 38
dynamic
  loadline, 42, 117
  waveform, 117
dynamic loadline, 117

Edwards–Sinsky stability criterion, 52
EER, 139
efficiency, 37, 38
  amplifier, 37, 118
  collector, 38
  drain, 38
  overall , 37
  power-added, 37
  power-added , 37
  total power-added, 37
  transmit chain, 37
enhancement mode
  FET, 7
  HEMT, 11
  JFET, 11
  MOSFET, 10
envelope
  elimination and restoration, 139
  tracking amplifier, 139

FDA, 86, 94, 95
feed-forward linearization, 136
FET, 4, 6
  depletion mode, 7
  enhancement mode, 7, 10
  HEMT, 10
  high electron mobility, 11
  IGFET, 4
  JFET, 10
  junction, 4, 10
  MESFET, 10, 11
  metal-epitaxy-semiconductor, 11
  metal-oxide-semiconductor, 4
  MODFET, 11
  modulation-doped, 11
  MOSFET, 4
  pHEMT, 11
  pseudomorphic HEMT, 11
field
  effect transistor, 4
figure of merit
  amplifier, 26
flicker noise, 198, 201
FOM
  amplifier, 26

GaAs, 3, 11, 142
gain
  actual power gain, 29
  available, 29
  circles, 35
  common mode, 87
  compression, 117
  definitions, 26
  differential mode, 87
  maximum unilateral transducer, 32
  power, 29
  system, 29
  transducer, 29, 32
  unilateral transducer, 32
gallium arsenide, 3
gallium nitride, 3
  MOSFET, 8
GaN, 3, 8
Gaussian
  signal, 123
generalized
  scattering parameters, 30
  amplifier, 30
Gummel–Poon model, 4

harmonic balance simulation, 111
Hartley
  oscillator, 162
HB simulation, 111
HBT, 4
  schematic symbol, 7
HEMT, 10, 11
  junction, 10
  model, 11, 19
  pseudomorphic, 11
high electron mobility transistor, 11
hybrid, 94, 140
  180° equivalents, 95
  combiner, 94
  RFIC, 94
  splitter, 94

IGFET, 4
indium phosphide, 3
inductor
  spiral, 144
InP, 3
instability
  oscillation, 45
integrated circuit
  RFIC, 86
intermittency, 201
inverted
  Class D amplifier, 122
  Class F amplifier, 122
inverted class D amplifier, 122
inverted class F amplifier, 122
inverting amplifier
  Class F, 122
IP3, 124

JFET, 10, 12
  model, 11, 19

k, 50
k factor, 50
Khan transmitter, 139
Kurokawa, 165
  simplified oscillation condition, 166
  stable oscillation condition, 165, 174

LINC amplifier, 139
linear
  amplifier, 26
linearization, 136
LITMUS amplifier, 140
loadline, 40, 42
  AC, 42, 117
  DC, 40
  dynamic, 42, 117
loadlines, 116
loadpull, 127
logarithmic map, 201
logistics map, 201

Marchand balun, 96
Materka–Kacprzak model, 19
maximum
  unilateral transducer gain, 32
MESFET, 10, 11, 142
  junction, 10
  model, 11, 19
metal
  -epitaxy-semiconductor FET, 11
  -oxide-semiconductor FET, 4
microwave
  monolithic integrated circuit, 142
Miller effect capacitance, 97
minority carrier, 4
mixer
  linear region, 9
mixing
  long-term correlation, 202
MMIC, 142, 143
mobility, 9
model
  BJT, 20
  HEMT, 19
  MESFET, 19
  MOSFET, 14
MODFET, 11
modulation
  doped FET, 11
MOSFET, 4, 6
  amplifier, 89
  charge model, 18
  current source, 98
  cutoff region, 8
  depletion mode, 10
  device model, 14
  gallium nitride, 8
  gate oxide capacitance, 9
  level 3 model, 14
  linear region, 8
  pentode region, 8
  saturation region, 8
  threshold voltage, 9
  triode region, 8
multi-tanh amplifier, 148
multi-tanh, differential amplifier, 100

noise
  Allan variance, 197
  chaotic map, 201
  correlation, 202
  excess oscillator noise, 201
  flicker, 198, 201
    model, 206
  intermittency, 201
  Leeson effect, 198
  linear time variant, 199
  logarithmic map, 201
  logistics map, 201
  model, 206
  observation
    frequency domain, 194
    time domain, 197
  phase, 165, 174
  random walk, 198
  shot
    model, 206
  source
    flicker noise, 206
    shot noise, 206
    thermal noise, 206
  substrate, 87
  thermal noise
    model, 206
nonlinear
  capacitor network, 149
  simulation, 110
    harmonic balance, 111
    HB, 111
    periodic steady-state, 116
    PSS, 116
Nyquist
  stability criterion, 53

OQPSK, 123, 124

oscillation, instability, 45
oscillator, 159
  case study
    phase noise analysis, 206
    reflection, 167
    VCO, 176
    voltage-controlled, 176
  Clapp, 162
  Colpitts, 162
  design guidelines, 210
  differential, 190
  feedback, 161
  Hartley, 162
  Kurokawa stable
    oscillation condition, 165, 174
  Leeson, 198
  multioscillation, 165, 174, 184
  negative gm, 190
  negative
    transconductance, 190
  noise, 194
    analysis, 206
    observation, frequency domain, 194
    observation, time domain, 197
  one port, 165
  phase noise, 165, 174, 194
    advanced, 194
    analysis, 206
    chaotic map, 201
    flicker noise, 201
    Hajimiri and Lee, 199
    Leeson effect, 198
    linear time variant, 199
    observation, frequency domain, 194
    observation, time domain, 197
  reflection, 165
    case study, 167
    coefficient shaping, 184
  RFIC, 190
  start-up, 181
  two-port, 161
  VCO, 172
    design, 176
  voltage-controlled, 172
    case study, 176
outphasing amplifier, 139
overall amplifier efficiency, 37

PAE, 37
PAR, 39
passive
  convention, 111
PDA, 86, 88, 95
pentode region, 8
periodic steady-state
  simulation, 116
phase
  noise, 165, 174
    chaotic map, 201
    excess oscillator noise, 198, 199, 201
    Leeson effect, 198
    linear time variant, 199
pHEMT, 11
PMEPR, 117, 122–124
  amplifier back-off, 123
polysilicon, 8
power
  added efficiency, 37
    total, 37
  amplifier
    loadpull, 127
  gain, 29
predistortion, 136
probability
  density function, 122
    amplitude, 123
pseudo
  differential
    amplifier, 94, 96
pseudomorphic HEMT, 11
PSS simulation, 116

QPSK, 122–124

radio
  frequency integrated circuit, 86
random walk, 198
rate of mixing, 202
RF
  integrated circuits, 97
RFIC, 86, 94, 97
  amplifier, 148
  cascode amplifier, 97
    variable gain, 97
  current mirror, 98
  digital control, 98
  power amplifier, 148
  transmitter, 100
  variable gain, 100
  VGA, 100
Rollet's stability criterion, 50
Rollet's stability factor, 34

S parameters
  amplifier, 30
  gain, 32
  generalized, 31
  transducer gain, 32
  unilateral transducer gain, 32
saturation region, 97
scattering parameters
  generalized, 30
Schottky
  diode, 8
shooting method, 116
sidelobe, 122
SiGe, 4
signal
  pseudo-differential, 94
silicon, 3
  RFIC, 97
simulation
  harmonic balance, 111
  HB, 111
  periodic steady-state, 116
  PSS, 116
slabline, 127
Smith chart
  gain circles, 35
Spice, 116
spiral inductor, 144
splitter, 94
stability, 44
  circles, 46
  criterion
    μ factor, 52
    k factor, 50
    Edwards–Sinsky, 52
    Nyquist, 53
    Rollet's, 50
  k factor, 50
  Kurokawa stable
    oscillation condition, 165, 174
  unconditional, 45
steady-state nonlinear
  simulation, 110
substrate
  noise, 87
switching
  amplifier, 116, 117
    Class D, 119
    Class E, 120
    class F, 121
    inverted class D, 122
    inverted class F, 122
system
  gain, 29

tanh cascode cell amplifier, 148
tank, 159
TCC amplifier, 148
Thevenin, 30
third
  -order intercept, 124
TOI, 124
total
  power-added efficiency, 37
transducer
  gain, 29
    maximum unilateral, 32
    unilateral, 32
  power gain, 32
transformer
  center-tapped, 94
transistor
  bipolar, 4
  bipolar junction, 41
  BJT, 4, 41
  FET, 4, 6, 10
    depletion mode, 7
    enhancement mode, 7
  field effect, 4
  HBT, 4, 41
  HEMT, 10, 11
  heterojunction bipolar, 41
  IGFET, 4
  insulated gate, 4
  JFET, 4, 10
  MESFET, 10, 11, 142
  MODFET, 11
  MOSFET, 4, 6
  pHEMT, 11
  technology, 3
  types, 3
transmit chain efficiency, 37
transmitter
  WCDMA, 100
triode
  region, 8
tuner
  automated, 127
  loadpull, 127

unconditional stability, 45
unilateral
  gain, 32
  transducer gain, 32
unstable amplifier, 44

VCO, 172
  case study, 176
  design, 176
VGA, 100
voltage
  controlled oscillator, 172
    case study, 176

waveform
  dynamic, 117
WCDMA
  transmitter, 100
WiMAX
  power amplifier, 128

MICROWAVE AND RF DESIGN: AMPLIFIERS AND OSCILLATORS Third Edition

Microwave and RF Design: Amplifiers and Oscillators presents the design of amplifiers and oscillators in a way that enables state-of-the-art designs to be realized. Detailed strategies and case studies are presented. Design of competitive microwave amplifiers and oscillators is particularly challenging as many trade-offs are required in design, and the design decisions cannot be reduced to a formulaic flow. The emphasis is on developing design skills. This book is suitable as both an undergraduate and graduate textbook, as well as a career-long reference book.

KEY FEATURES

- The fifth volume of a comprehensive series on microwave and RF design
- Open access ebook editions are hosted by NC State University Libraries at: https://repository.lib.ncsu.edu/handle/1840.20/36776
- 9 worked examples
- An average of 23 exercises per chapter
- Answers to selected exercises
- 6 extensive case studies following the design of competitive amplifiers and oscillators with world leading performance
- A companion book, *Fundamentals of Microwave and RF Design*, is suitable as a comprehensive undergraduate textbook on microwave engineering

ABOUT THE AUTHOR

Michael Steer is the Lampe Distinguished Professor of Electrical and Computer Engineering at North Carolina State University. He received his B.E. and Ph.D. degrees in Electrical Engineering from the University of Queensland. He is a Fellow of the IEEE and is a former editor-in-chief of *IEEE Transactions on Microwave Theory and Techniques*. He has authored more than 500 publications including twelve books. In 2009 he received a US Army Medal, "The Commander's Award for Public Service." He received the 2010 Microwave Prize and the 2011 Distinguished Educator Award, both from the IEEE Microwave Theory and Techniques Society.

OTHER VOLUMES

Microwave and RF Design
Radio Systems
Volume 1
ISBN 978-1-4696-5690-8

Microwave and RF Design
Transmission Lines
Volume 2
ISBN 978-1-4696-5692-2

Microwave and RF Design
Networks
Volume 3
ISBN 978-1-4696-5694-6

Microwave and RF Design
Modules
Volume 4
ISBN 978-1-4696-5696-0

ALSO BY THE AUTHOR

Fundamentals of Microwave and RF Design
ISBN 978-1-4696-5688-5

Published by NC State University

NC STATE UNIVERSITY

Distributed by UNC Press