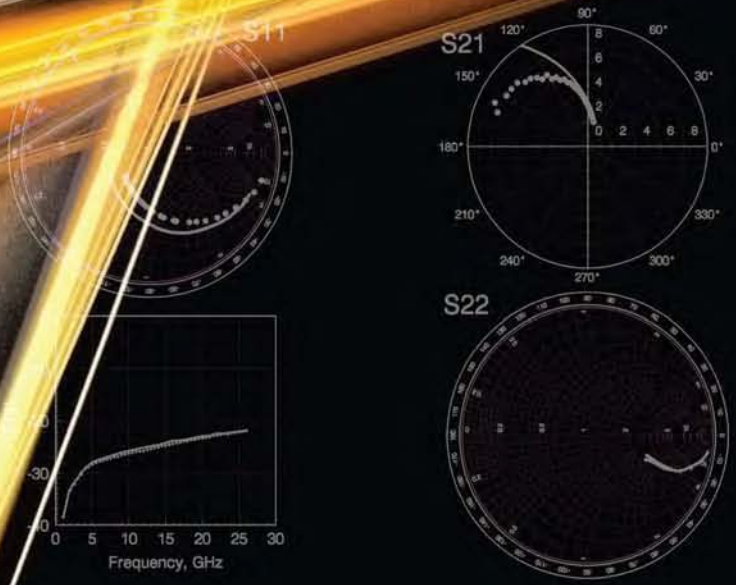
RF and Microwave Handbook

Second Edition

RF AND MICROWAVE CIRCUITS, MEASUREMENTS, AND MODELING



Editor-in-Chief

Mike Golio

Managing Editor

Janet Golio

CRC Press
Taylor & Francis Group

The RF and Microwave Handbook

Second Edition

Editor-in-Chief

Mike Golio

HVVi Semiconductors, Inc.

Phoenix, Arizona, U.S.A.

Managing Editor

Janet Golio

RF and Microwave Applications and Systems

RF and Microwave Circuits, Measurements, and Modeling

RF and Microwave Passive and Active Technologies

The Electrical Engineering Handbook Series

Series Editor

Richard C. Dorf
University of California, Davis

Titles Included in the Series

The Handbook of Ad Hoc Wireless Networks, Mohammad Ilyas
The Avionics Handbook, Second Edition, Cary R. Spitzer
The Biomedical Engineering Handbook, Third Edition, Joseph D. Bronzino
The Circuits and Filters Handbook, Second Edition, Wai-Kai Chen
The Communications Handbook, Second Edition, Jerry Gibson
The Computer Engineering Handbook, Vojin G. Oklobdzija
The Control Handbook, William S. Levine
The CRC Handbook of Engineering Tables, Richard C. Dorf
The Digital Avionics Handbook, Second Edition Cary R. Spitzer
The Digital Signal Processing Handbook, Vijay K. Madisetti and Douglas Williams
The Electrical Engineering Handbook, Second Edition, Richard C. Dorf
The Electric Power Engineering Handbook, Second Edition, Leonard L. Grigsby
The Electronics Handbook, Second Edition, Jerry C. Whitaker
The Engineering Handbook, Third Edition, Richard C. Dorf
The Handbook of Formulas and Tables for Signal Processing, Alexander D. Poularikas
The Handbook of Nanoscience, Engineering, and Technology, Second Edition
William A. Goddard, III, Donald W. Brenner, Sergey E. Lyshevski, and Gerald J. Iafrate
The Handbook of Optical Communication Networks, Mohammad Ilyas and
Hussein T. Mouftah
The Industrial Electronics Handbook, J. David Irwin
The Measurement, Instrumentation, and Sensors Handbook, John G. Webster
The Mechanical Systems Design Handbook, Osita D.I. Nwokah and Yidirim Hurmuzlu
The Mechatronics Handbook, Second Edition, Robert H. Bishop
The Mobile Communications Handbook, Second Edition, Jerry D. Gibson
The Ocean Engineering Handbook, Ferial El-Hawary
The RF and Microwave Handbook, Second Edition, Mike Golio
The Technology Management Handbook, Richard C. Dorf
The Transforms and Applications Handbook, Second Edition, Alexander D. Poularikas
The VLSI Handbook, Second Edition, Wai-Kai Chen

RF AND MICROWAVE CIRCUITS, MEASUREMENTS, AND MODELING

Editor-in-Chief

Mike Golio

HVVi Semiconductors, Inc.

Phoenix, Arizona, U.S.A.

Managing Editor

Janet Golio

CRC Press

Taylor & Francis Group

Boca Raton London New York

CRC Press is an imprint of the

Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

© 2008 by Taylor & Francis Group, LLC
CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works
Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number-13: 978-0-8493-7218-6 (Hardcover)

This book contains information obtained from authentic and highly regarded sources. Reprinted material is quoted with permission, and sources are indicated. A wide variety of references are listed. Reasonable efforts have been made to publish reliable data and information, but the author and the publisher cannot assume responsibility for the validity of all materials or for the consequences of their use.

No part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (http://www.copyright.com/) or contact the Copyright Clearance Center, Inc. (CCC) 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Library of Congress Cataloging-in-Publication Data

RF and microwave circuits, measurements, and modeling / [Ed.] Mike Golio.
p. cm.
Includes bibliographical references and index.
ISBN-13: 978-0-8493-7218-6 (alk. paper)
1. Microwave circuits. 2. Radio circuits. 3. Wireless communication systems. I. Golio, John Michael, 1954-

TK7876.R487 2008
621.381'32--dc22 2007016316

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Contents

SECTION II Circuits

SECTION III CAD, Simulation and Modeling

Preface

The first edition of the *RF and Microwave Handbook* was published in 2000. The project got off to an inauspicious start when 24 inches of snow fell in Denver the evening before the advisory board planned to hold their kick-off meeting. Two members of the board were trapped for days in the Denver airport since planes were not arriving or leaving. Because of road closures, one member was stranded only miles away from the meeting in Boulder. And the remainder of the board was stranded in a Denver hotel 10 miles from the airport. Despite this ominous beginning, a plan was formed, expert authors recruited, and the book was developed and published. The planning and development of this second edition have been very smooth and uneventful in comparison to our first efforts. Since publication in 2000, the value of the *RF and Microwave Handbook* has been recognized by thousands of engineers throughout the world. Three derivative handbooks have also been published and embraced by the microwave industry. The advisory board believes that this edition will be found to be of even greater value than the first edition.

Prior to the 1990s, microwave engineering was employed almost exclusively to address military, satellite, and avionics applications. In 1985, there were a few limited applications of RF and microwave systems that laymen might be familiar with such as satellite TV and the use of satellite communications for overseas phone calls. Pagers were also available but not common. In contrast, by 1990 the wireless revolution had begun. Cell phones were becoming common and new applications of wireless technology were emerging every day. Companies involved in wireless markets seemed to have a license to print money. At the time of the introduction of the first edition of the *RF and Microwave Handbook*, wireless electronic products were pervasive, but relatively simple, early generations of the advanced wireless products available today. At present, the number of people using wireless voice and data systems continues to grow. New systems such as 3G phones, 4G phones, and WiMAX represent emerging new wireless markets with significant growth potential. All of these wireless products are dependent on the RF and microwave component and system engineering, which is the subject of this book. During this time the military, satellite, and avionics systems have also become increasingly complex. The research and development that drives these applications continues to serve as the foundation for most of the commercial wireless products available to consumers.

This edition of the handbook covers issues of interest to engineers involved in RF/microwave system and component development. The second edition includes significantly expanded topic coverage as well as updated or new articles for most of the topics included in the first edition. The expansion of material has prompted the division of the handbook into three independent volumes of material. The chapters are aimed at working engineers, managers, and academics who have a need to understand microwave topics outside their area of expertise. Although the book is not written as a textbook, researchers and students will find it useful. Most of the chapters provide extensive references so that they will not only explain fundamentals of each field, but also serve as a starting point for further in-depth research.

This book, *RF and Microwave Circuits, Measurements, and Modeling*, examines three areas of critical importance to the RF and microwave circuit designer.

Characterization and measurement of components, circuits, and systems at high frequencies are unique and challenging tasks. Standard, low frequency equipment fails to provide meaningful information for the RF and microwave engineer. Small-signal, large-signal, phase, pulsed, waveform, and noise measurements are discussed in detail. Calibration procedures are extremely important for these measurements and are also described.

RF and microwave circuit designs are explored in terms of performance and critical design specifications. Transmitters and receivers are first discussed in terms of functional circuit blocks. The blocks are then examined individually. Separate chapters consider fundamental amplifier issues, low noise amplifiers, power amplifiers for handset applications, and high power amplifiers. Other circuit functions including oscillators, mixers, modulators, phase locked loops, filters, and multiplexers are each considered in individual chapters.

The unique behavior and requirements associated with RF and microwave systems establish a need for unique and complex models and simulation tools. The required toolset for a microwave circuit designer includes unique device models, both 2D and 3D electromagnetic simulators, as well as frequency domain based small-signal and large-signal circuit and system simulators. This unique suite of tools requires a design procedure that is also distinctive. Individual chapters examine not only the distinct design tools of the microwave circuit designer, but also the design procedures that must be followed to use them effectively.

Acknowledgments

Although the topics and authors for this book were identified by the editor-in-chief and the advisory board, they do not represent the bulk of the work for a project like this. A great deal of the work involves tracking down those hundreds of technical experts, gaining their commitment, keeping track of their progress, collecting their manuscripts, getting appropriate reviews/revisions, and finally transferring the documents to be published. While juggling this massive job, author inquiries ranging from, "What is the required page length?", to, "What are the acceptable formats for text and figures?", have to be answered and re-answered. Schedules are very fluid. This is the work of the Managing Editor, Janet Golio. Without her efforts there would be no second edition of this handbook.

The advisory board has facilitated the book's completion in many ways. Board members contributed to the outline of topics, identified expert authors, reviewed manuscripts, and authored several of the chapters for the book.

Hundreds of RF and microwave technology experts have produced the chapters that comprise this second edition. They have all devoted many hours of their time sharing their expertise on a wide range of topics.

I would like to sincerely thank all of those listed above. Also, it has been a great pleasure to work with Jessica Vakili, Helena Redshaw, Nora Konopka, and the publishing professionals at CRC Press.

Editors

Editor-in-Chief

Mike Golio, since receiving his PhD from North Carolina State University in 1983, has held a variety of positions in both the microwave and semiconductor industries, and within academia. As Corporate Director of Engineering at Rockwell Collins, Dr. Golio managed and directed a large research and development organization, coordinated corporate IP policy, and led committees to achieve successful corporate spin-offs.

As an individual contributor at Motorola, he was responsible for pioneering work in the area of large signal microwave device characterization and modeling. This work resulted in over 50 publications including one book and a commercially available software package. The IEEE recognized this work by making Dr. Golio a Fellow of the Institute in 1996.

He is currently RF System Technologist with HVVi Semiconductor, a start-up semiconductor company. He has contributed to all aspects of the company's funding, strategies, and technical execution.

Dr. Golio has served in a variety of professional volunteer roles for the IEEE MTT Society including: Chair of Membership Services Committee, founding Co-editor of *IEEE Microwave Magazine*, and MTT-Society distinguished lecturer. He currently serves as Editor-in-chief of *IEEE Microwave Magazine*. In 2002 he was awarded the N. Walter Cox Award for exemplary service in a spirit of selfless dedication and cooperation.

He is author of hundreds of papers, book chapters, presentations and editor of six technical books. He is inventor or co-inventor on 15 patents. In addition to his technical contributions, Dr. Golio recently published a book on retirement planning for engineers and technology professionals.

Managing Editor

Janet R. Golio is Administrative Editor of *IEEE Microwave Magazine* and webmaster of www.golio.net. Prior to that she did government, GPS, and aviation engineering at Motorola in Arizona, Rockwell Collins in Iowa, and General Dynamics in Arizona. She also helped with the early development of the personal computer at IBM in North Carolina. Golio holds one patent and has written six technical papers. She received a BS in Electrical Engineering Summa Cum Laude from North Carolina State University in 1984.

When not working, Golio actively pursues her interests in archaeology, trick roping, and country western dancing. She is the author of young adult books, *A Present from the Past* and *A Puzzle from the Past.*

Advisory Board

Peter A. Blakey

Peter A. Blakey obtained a BA in applied physics from the University of Oxford in 1972, a PhD in electronic engineering from the University of London in 1976, and an MBA from the University of Michigan in 1989. He has held several different positions in industry and academia and has worked on a wide range of RF, microwave, and Si VLSI applications. Between 1991 and 1995 he was the director of TCAD Engineering at Silvaco International. He joined the Department of Electrical Engineering at Northern Arizona University in 2002 and is presently an emeritus professor at that institution.

Nick Buris

Nick Buris received his Diploma in Electrical Engineering in 1982 from the National Technical University of Athens, Greece, and a PhD in electrical engineering from the North Carolina State University, Raleigh, North Carolina, in 1986. In 1986 he was a visiting professor at NCSU working on space reflector antennas for NASA. In 1987 he joined the faculty of the Department of Electrical and Computer Engineering at the University of Massachusetts at Amherst. His research work there spanned the areas of microwave magnetics, phased arrays printed on ferrite substrates, and broadband antennas. In the summer of 1990 he was a faculty fellow at the NASA Langley Research Center working on calibration techniques for dielectric measurements (space shuttle tiles at very high temperatures) and an ionization (plasma) sensor for an experimental reentry spacecraft.

In 1992 he joined the Applied Technology organization of Motorola's Paging Product Group and in 1995 he moved to Corporate Research to start an advanced modeling effort. While at Motorola he has worked on several projects from product design to measurement systems and the development of proprietary software tools for electromagnetic design. He currently manages the Microwave Technologies Research Lab within Motorola Labs in Schaumburg, Illinois. Recent and current activities of the group include V-band communications systems design, modeling and measurements of complex electromagnetic problems, RF propagation, Smart Antennas/MIMO, RFID systems, communications systems simulation and modeling, spectrum engineering, as well as TIA standards work on RF propagation and RF exposure.

Nick is a senior member of the IEEE, and serves on an MTT Technical Program Committee. He recently served as chair of a TIA committee on RF exposure and is currently a member of its Research Division Committee.

Lawrence P. Dunleavy

Dr. Larry Dunleavy, along with four faculty colleagues established University of South Florida's innovative Center for Wireless and Microwave Information Systems (WAMI Center—http://ee.eng.usf.edu/WAMI).

In 2001, Dr. Dunleavy co-founded Modelithics, Inc., a USF spin-off company, to provide a practical commercial outlet for developed modeling solutions and microwave measurement services (www.modelithics.com), where he is currently serving as its president.

Dr. Dunleavy received his BSEE degree from Michigan Technological University in 1982 and his MSEE and PhD in 1984 and 1988, respectively, from the University of Michigan. He has worked in industry for E-Systems (1982–1983) and Hughes Aircraft Company (1984–1990) and was a Howard Hughes Doctoral Fellow (1984–1988). In 1990 he joined the Electrical Engineering Department at the University of South Florida. He maintains a position as professor in the Department of Electrical Engineering. His research interests are related to microwave and millimeter-wave device, circuit, and system design, characterization, and modeling. In 1997–1998, Dr. Dunleavy spent a sabbatical year in the noise metrology laboratory at the National Institute of Standards and Technology (NIST) in Boulder, Colorado. Dr. Dunleavy is a senior member of IEEE and is very active in the IEEE MTT Society and the Automatic RF Techniques Group (ARFTG). He has authored or co-authored over 80 technical articles.

Jack East

Jack East received his BSE, MS, and PhD from the University of Michigan. He is presently with the Solid State Electronics Laboratory at the University of Michigan conducting research in the areas of high-speed microwave device design, fabrication, and experimental characterization of solid-state microwave devices, nonlinear and circuit modeling for communications circuits and low-energy electronics, and THz technology.

Patrick Fay

Patrick Fay is an associate professor in the Department of Electrical Engineering at the University of Notre Dame, Notre Dame, Indiana. He received his PhD in Electrical Engineering from the University of Illinois at Urbana-Champaign in 1996 after receiving a BS in Electrical Engineering from Notre Dame in 1991. Dr. Fay served as a visiting assistant professor in the Department of Electrical and Computer Engineering at the University of Illinois at Urbana-Champaign in 1996 and 1997, and joined the faculty at the University of Notre Dame in 1997.

His educational initiatives include the development of an advanced undergraduate laboratory course in microwave circuit design and characterization, and graduate courses in optoelectronic devices and electronic device characterization. He was awarded the Department of Electrical Engineering's IEEE Outstanding Teacher Award in 1998–1999. His research interests include the design, fabrication, and characterization of microwave and millimeter-wave electronic devices and circuits, as well as high-speed optoelectronic devices and optoelectronic integrated circuits for fiber optic telecommunications. His research also includes the development and use of micromachining techniques for the fabrication of microwave components and packaging. Professor Fay is a senior member of the IEEE, and has published 7 book chapters and more than 60 articles in refereed scientific journals.

David Halchin

David Halchin has worked in RF/microwaves and GaAs for over 20 years. During this period, he has worn many hats including engineering and engineering management, and he has worked in both academia and private industry. Along the way, he received his PhD in Electrical Engineering from North Carolina State University. Dave currently works for RFMD, as he has done since 1998. After a stint as a PA designer, he was moved into his current position managing a modeling organization within RFMD's Corporate Research and Development organization. His group's responsibilities include providing compact models for circuit simulation for both GaAs active devices and passives on GaAs. The group also provides compact models for a handful of Si devices, behavioral models for power amplifier assemblies, and physics-based simulation for GaAs device development. Before joining RFMD, Dave spent time at Motorola and Rockwell working

in the GaAs device development and modeling areas. He is a member of the IEEE-MTT and EDS. Dave is currently a member of the executive committee for the Compound IC Symposium.

Alfy Riddle

Alfy Riddle is vice president of Engineering at Finesse. Before Finesse, Dr. Riddle was the principal at Macallan Consulting, a company he founded in 1989. He has contributed to the design and development of a wide range of products using high-speed, low-noise, and RF techniques. Dr. Riddle developed and marketed the Nodal circuit design software package that featured symbolic analysis and object-oriented techniques. He has co-authored two books and contributed chapters to several more. He is a member of the IEEE MTT Society, the Audio Engineering Society, and the Acoustical Society of America. Dr. Riddle received his PhD in Electrical Engineering in 1986 from North Carolina State University. When he is not working, he can be found on the tennis courts, hiking in the Sierras, taking pictures with an old Leica M3, or making and playing Irish flutes.

Robert J. Trew

Robert J. Trew received his PhD from the University of Michigan in 1975. He is currently the Alton and Mildred Lancaster Distinguished Professor of Electrical and Computer Engineering and Head of the ECE Department at North Carolina State University, Raleigh. He previously served as the Worcester Professor of Electrical and Computer Engineering and Head of the ECE Department of Virginia Tech, Blacksburg, Virginia, and the Dively Distinguished Professor of Engineering and Chair of the Department of Electrical Engineering and Applied Physics at Case Western Reserve University, Cleveland, Ohio. From 1997 to 2001 Dr. Trew was director of research for the U.S. Department of Defense with management responsibility for the $1.3 billion annual basic research program of the DOD. Dr. Trew was vice-chair of the U.S. government interagency group that planned and implemented the U.S. National Nanotechnology Initiative. Dr. Trew is a fellow of the IEEE, and was the 2004 president of the Microwave Theory and Techniques Society. He was editor-in-chief of the *IEEE Transactions on Microwave Theory and Techniques* from 1995 to 1997, and from 1999 to 2002 was founding co-editor-in-chief of the *IEEE Microwave Magazine.* He is currently the editor-in-chief of the *IEEE Proceedings.* Dr. Trew has twice been named an IEEE MTT Society Distinguished Microwave Lecturer. He has earned numerous honors, including a 2003 Engineering Alumni Society Merit Award in Electrical Engineering from the University of Michigan, the 2001 IEEE-USA Harry Diamond Memorial Award, the 1998 IEEE MTT Society Distinguished Educator Award, a 1992 Distinguished Scholarly Achievement Award from NCSU, and an IEEE Third Millennium Medal. Dr. Trew has authored or co-authored over 160 publications, 19 book chapters, 9 patents, and has given over 360 presentations

Contributors

Peter A. Blakey
Northern Arizona University
Flagstaff, Arizona

John C. Cowles
Analog Devices–Northwest Labs
Beaverton, Oregon

Walter R. Curtice
W. R. Curtice Consulting
Washington Crossing,
Pennsylvania

Lawrence P. Dunleavy
Modelithics, Inc.
Tampa, Florida

Patrick Fay
University of Notre Dame
Notre Dame, Indiana

Joseph M. Gering
RF Micro Devices
Greensboro, North Carolina

Mike Golio
HVVi Semiconductor
Phoenix, Arizona

Paul D. Hale
National Institute of Standards
and Technology
Boulder, Colorado

Ronald E. Ham
MW and RF Consulting
Austin, Texas and
Kitzbuhel, Austria

H. Mike Harris
Georgia Tech Research Institute
Atlanta, Georgia

Todd Heckleman
MKS Instruments, Inc.
Rochester, New York

Brent Irvine
MKS Instruments, Inc.
Rochester, New York

Christopher Jones
M/A-COM Tyco Electronics
Lowell, Massachusetts

J. Stevenson Kenney
Georgia Institute of Technology
Atlanta, Georgia

Ron Kielmeyer
RF Micro Devices
Chandler, Arizona

Jakub Kucera
AnaPico AG
Zürich, Switzerland

Jean-Pierre Lanteri
M/A-COM Tyco Electronics
Lowell, Massachusetts

Urs Lott
AnaPico AG
Zürich, Switzerland

John R. Mahon
M/A-COM Tyco Electronics
Lowell, Massachusetts

Charles Nelson
California State University
Sacramento, California

Robert Newgard
Rockwell Collins
Cedar Rapids, Iowa

Troels S. Nielsen
RF Micro Devices
Greensboro, North Carolina

Anthony E. Parker
Macquarie University
Sydney, Australia

Anthony M. Pavio
Microwave Specialties
Paradise Valley, Arizona

Aaron Radomski
Harris RF Communications
Rochester, New York

James G. Rathmell
University of Sydney
Sydney, Australia

Kate A. Remley
National Institute of Standards
and Technology
Boulder, Colorado

Alfy Riddle
Finesse, LLC
Santa Clara, California

Jonathan B. Scott
University of Waikato
Hamilton, New Zealand

Warren L. Seely
Ubidyne, Inc.
Tempe, Arizona

John F. Sevic
Maury Microwave Corporation
Ontario, California

Richard V. Snyder
RS Microwave
Butler, New Jersey

Edward T. Spears
RF Micro Devices
Chandler, Arizona

Joseph Staudinger
Freescale Semiconductor, Inc.
Tempe, Arizona

Michael B. Steer
North Carolina State University
Raleigh, North Carolina

Daniel G. Swanson, Jr.
Tyco Electronics
Lowell, Massachusetts

Douglas A. Teeter
RF Micro Devices
Billerica, Massachusetts

Manos M. Tentzeris
Georgia Institute of Technology
Atlanta, Georgia

Robert J. Trew
North Carolina State University
Raleigh, North Carolina

John P. Wendler
Tyco Electronics
Wireless Network Solutions
Lowell, Massachusetts

Dylan F. Williams
National Institute of Standards
and Technology
Boulder, Colorado

Introduction to Microwaves and RF

Patrick Fay
University of Notre Dame

I.1 Introduction to Microwave and RF Engineering

Modern microwave and radio frequency (RF) engineering is an exciting and dynamic field, due in large part to the symbiosis between recent advances in modern electronic device technology and the explosion in demand for voice, data, and video communication capacity that started in the 1990s and continues through the present. Prior to this revolution in communications, microwave technology was the nearly exclusive domain of the defense industry; the recent and dramatic increase in demand for communication systems for such applications as wireless paging, mobile telephony, broadcast video, and tethered as well as untethered computer networks has revolutionized the industry. These communication systems are employed across a broad range of environments, including corporate offices, industrial and manufacturing facilities, infrastructure for municipalities, as well as private homes. The diversity of applications and operational environments has led, through the accompanying high production volumes, to tremendous advances in cost-efficient manufacturing capabilities of microwave and RF products. This in turn has lowered the implementation cost of a host of new and cost-effective wireless as well as wired RF and microwave services. Inexpensive handheld GPS navigational aids, automotive collision-avoidance radar, and widely available broadband digital service access are among these. Microwave technology is naturally suited for these emerging applications in communications and sensing, since the high operational frequencies permit both large numbers of independent channels for the wide variety of uses envisioned as well as significant available bandwidth per channel for high-speed communication.

Loosely speaking, the fields of microwave and RF engineering together encompass the design and implementation of electronic systems utilizing frequencies in the electromagnetic spectrum from approximately 300 kHz to over 100 GHz. The term "RF" engineering is typically used to refer to circuits and systems having frequencies in the range from approximately 300 kHz at the low end to between 300 MHz and 1 GHz at the upper end. The term "microwave engineering," meanwhile, is used rather loosely to refer to design and implementation of electronic systems with operating frequencies in the range from 300 MHz to 1 GHz on the low end to upwards of 100 GHz. Figure I.1 illustrates schematically the electromagnetic spectrum from audio frequencies through cosmic rays. The RF frequency spectrum covers the medium frequency (MF), high frequency (HF), and very high frequency (VHF) bands, while the microwave portion of the

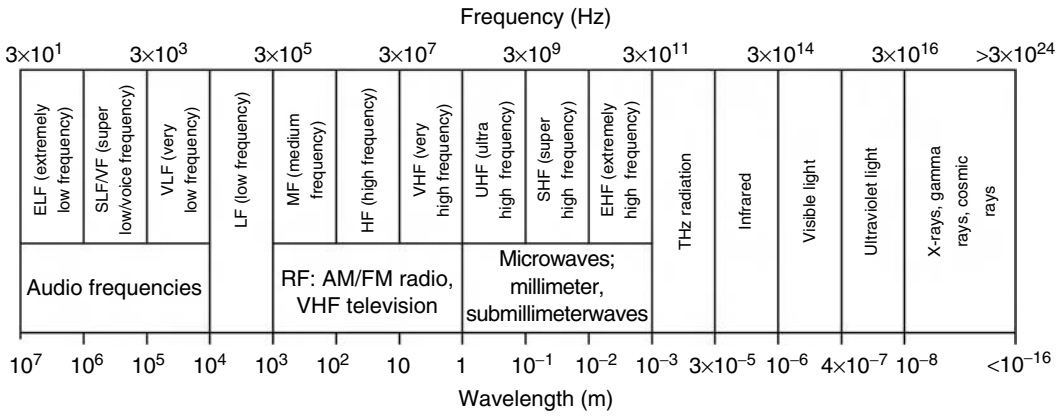


FIGURE I.1 Electromagnetic frequency spectrum and associated wavelengths.

electromagnetic spectrum extends from the upper edge of the VHF frequency range to just below the THz radiation and far-infrared optical frequencies (approximately 0.3 THz and above). The wavelength of free-space radiation for frequencies in the RF frequency range is from approximately 1 m (at 300 MHz) to 1 km (at 300 kHz), while those of the microwave range extend from 1 m to the vicinity of 1 mm (corresponding to 300 GHz) and below.

The boundary between "RF" and "microwave" design is both somewhat indistinct as well as one that is continually shifting as device technologies and design methodologies advance. This is due to implicit connotations that have come to be associated with the terms "RF" and "microwave" as the field has developed. In addition to the distinction based on the frequency ranges discussed previously, the fields of RF and microwave engineering are also often distinguished by other system features as well. For example, the particular active and passive devices used, the system applications pursued, and the design techniques and overall mindset employed all play a role in defining the fields of microwave and RF engineering. These connotations within the popular meaning of microwave and RF engineering arise fundamentally from the frequencies employed, but often not in a direct or absolute sense. For example, because advances in technology often considerably improve the high frequency performance of electronic devices, the correlation between particular types of electronic devices and particular frequency ranges is a fluid one. Similarly, new system concepts and designs are reshaping the applications landscape, with mass market designs utilizing ever higher frequencies rapidly breaking down conventional notions of microwave-frequency systems as serving "niche" markets.

The most fundamental characteristic that distinguishes RF engineering from microwave engineering is directly related to the frequency (and thus the wavelength, λ) of the electronic signals being processed. This distinction arises fundamentally from the finite speed of propagation of electromagnetic waves (and thus, by extension, currents and voltages). In free space, $\lambda = c/f$, where f is the frequency of the signal and c is the speed of light. For low-frequency and RF circuits (with a few special exceptions such as antennae), the signal wavelength is much larger than the size of the electronic system and circuit components. In contrast, for a microwave system the sizes of typical electronic components are often comparable to (i.e., within approximately 1 order of magnitude of) the signal wavelength. A schematic diagram illustrating this concept is shown in Figure I.2. As illustrated in Figure I.2, for components much smaller than the wavelength (i.e., $\ell < \lambda/10$), the finite velocity of the electromagnetic signal as it propagates through the component leads to a modest difference in phase at opposite ends of the component. For components comparable to or larger than the wavelength, however, this end-to-end phase difference becomes increasingly significant. This gives rise to a reasonable working definition of the two design areas based on the underlying approximations used in design. Since in conventional RF design, the circuit components and interconnections are generally small compared to a wavelength, they can be

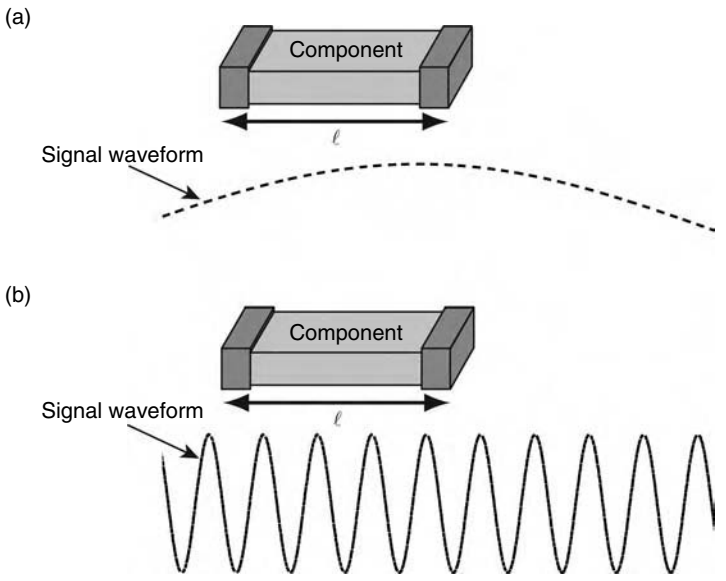


FIGURE I.2 Schematic representation of component dimensions relative to signal wavelengths. Conventional lumped-element analysis techniques are typically applicable for components for which $\ell < \lambda/10$ (a) since the phase change due to electromagnetic propagation across the component is small, while for components with $\ell > \lambda/10$ (b) the phase change is significant and a distributed circuit description is more appropriate.

modeled as lumped elements for which Kirchoff's voltage and current laws apply at every instant in time. Parasitic inductances and capacitances are incorporated to accurately model the frequency dependencies and the phase shifts, but these quantities can, to good approximation, be treated with an appropriate lumped-element equivalent circuit. In practice, a rule of thumb for the applicability of a lumped-element equivalent circuit is that the component size should be less than $\lambda/10$ at the frequency of operation. For microwave frequencies for which component size exceeds approximately $\lambda/10$, the finite propagation velocity of electromagnetic waves can no longer be as easily absorbed into simple lumped-element equivalent circuits. For these frequencies, the time delay associated with signal propagation from one end of a component to the other is an appreciable fraction of the signal period, and thus lumped-element descriptions are no longer adequate to describe the electrical behavior. A distributed-element model is required to accurately capture the electrical behavior. The time delay associated with finite wave propagation velocity that gives rise to the distributed circuit effects is a distinguishing feature of the mindset of microwave engineering.

An alternative viewpoint is based on the observation that microwave engineering lies in a "middle ground" between traditional low-frequency electronics and optics, as shown in Figure I.1. As a consequence of RF, microwaves, and optics simply being different regimes of the same electromagnetic phenomena, there is a gradual transition between these regimes. The continuity of these regimes results in constant re-evaluation of the appropriate design strategies and trade-offs as device and circuit technology advances. For example, miniaturization of active and passive components often increases the frequencies at which lumped-element circuit models are sufficiently accurate, since by reducing component dimensions the time delay for propagation through a component is proportionally reduced. As a consequence, lumped-element components at "microwave" frequencies are becoming increasingly common in systems previously based on distributed elements due to significant advances in miniaturization, even though the operational frequencies remain unchanged. Component and circuit miniaturization also leads to tighter packing of interconnects and components, potentially introducing new parasitic coupling and distributed-element effects into circuits that could previously be treated using lumped-element RF models.

The comparable scales of components and signal wavelengths has other implications for the designer as well, since neither the ray-tracing approach from optics nor the lumped-element approach from RF circuit design are valid in this middle ground. In this regard, microwave engineering can also be considered to be "applied electromagnetic engineering," as the design of guided-wave structures such as waveguides and transmission lines, transitions between different types of transmission lines, and antennae all require analysis and control of the underlying electromagnetic fields.

Guided wave structures are particularly important in microwave circuits and systems. There are many different approaches to the implementation of guided-wave structures; a sampling of the more common options are shown in Figure I.3. Figure I.3a shows a section of coaxial cable. In this common cable type, the grounded outer conductor shields the dielectric and inner conductor from external signals and also prevents the signals within the cable from radiating. The propagation in this structure is controlled by the dielectric properties, the cross-sectional geometry, and the metal conductivity. Figure I.3b shows a rectangular waveguide. In this structure, the signal propagates in the free space within the structure, while the rectangular metal structure is grounded. Despite the lack of an analog to the center conductor in the coaxial line, the structure supports traveling-wave solutions to Maxwell's equations, and thus can be used to transmit power along its length. The lack of a center conductor does prevent the structure from providing any path for dc along its length. The solution to Maxwell's equations in the rectangular waveguide also leads to multiple eigenmodes, each with its own propagation characteristics (e.g., characteristic impedance and propagation constant), and corresponding cutoff frequency. For frequencies above the cutoff frequency, the mode propagates down the waveguide with little loss, but below the cutoff frequency the mode is

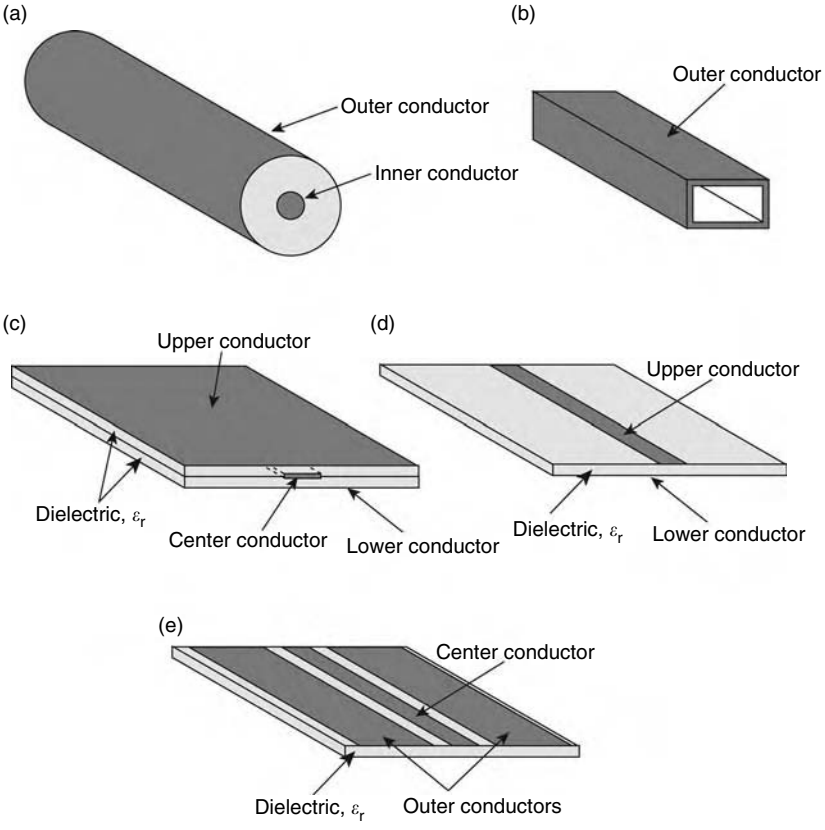


FIGURE I.3 Several common guided-wave structures. (a) coaxial cable, (b) rectangular waveguide, (c) stripline, (d) microstrip, and (e) coplanar waveguide.

evanescent and the amplitude falls off exponentially with distance. Since the characteristic impedance and propagation characteristics of each mode are quite different, in many systems the waveguides are sized to support only one propagating mode at the frequency of operation. While metallic waveguides of this type are mechanically inflexible and can be costly to manufacture, they offer extremely low loss and have excellent high-power performance. At W-band and above in particular, these structures currently offer much lower loss than coaxial cable alternatives. Figure I.3c through I.3e show several planar structures that support guided waves. Figure I.3c illustrates the stripline configuration. This structure is in some ways similar to the coaxial cable, with the center conductor of the coaxial line corresponding to the center conductor in the stripline, and the outer shield on the coaxial line corresponding to the upper and lower ground planes in the stripline. Figures I.3d and I.3e show two planar guided-wave structures often encountered in circuit-board and integrated circuit designs. Figure I.3d shows a microstrip configuration, while Figure I.3e shows a coplanar waveguide. Both of these configurations are easily realizable using conventional semiconductor and printed-circuit fabrication techniques. In the case of microstrip lines, the key design variables are the dielectric properties of the substrate, the dielectric thickness, and the width of the top conductor. For the coplanar waveguide case, the dielectric properties of the substrate, the width of the center conductor, the gap between the center and outer ground conductors, and whether or not the bottom surface of the substrate is grounded control the propagation characteristics of the lines.

For all of these guided-wave structures, an equivalent circuit consisting of the series concatenation of many stages of the form shown in Figure I.4 can be used to model the transmission line. In this equivalent circuit, the key parameters are the resistance per unit length of the line (R), the inductance per unit length (L), the parallel conductance per unit length of the dielectric (G), and the capacitance per unit length (C). Each of these parameters can be derived from the geometry and material properties of the line. Circuits of this form give rise to traveling-wave solutions of the form

$$V(z) = V_0^+ e^{-\gamma z} + V_0^- e^{\gamma z}$$

$$I(z) = \frac{V_0^+}{Z_0} e^{-\gamma z} - \frac{V_0^-}{Z_0} e^{\gamma z}$$

In these equations, the characteristic impedance of the line, which is the constant of proportionality between the current and voltage associated with a particular traveling-wave mode on the line, is given by $Z_0 = \sqrt{(R + j\omega L)/(G + j\omega C)}$. For lossless lines, $R = 0$ and $G = 0$, so that Z_0 is real; even in many practical cases the loss of the lines is small enough that the characteristic impedance can be treated as real. Similarly, the propagation constant of the line can be expressed as $\gamma = \alpha + j\beta = \sqrt{(R + j\omega L)(G + j\omega C)}$. In this expression, α characterizes the loss of the line, and β captures the wave propagation. For lossless lines, γ is pure imaginary, and thus α is zero. The design and analysis of these guided-wave structures is treated in more detail in Chapter 30 of the companion volume *RF and Microwave Applications and Systems* in this handbook series.

The distinction between RF and microwave engineering is further blurred by the trend of increasing commercialization and consumerization of systems using what have been traditionally considered to be microwave frequencies. Traditional microwave engineering, with its historically military applications,

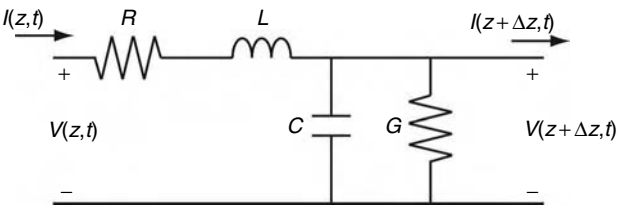


FIGURE I.4 Equivalent circuit for an incremental length of transmission line. A finite length of transmission line can be modeled as a series concatenation of sections of this form.

has long been focused on delivering performance at any cost. As a consequence, special-purpose devices intended solely for use in high performance microwave systems and often with somewhat narrow ranges of applicability were developed to achieve the required performance. With continuing advances in silicon microelectronics, including Si bipolar junction transistors, SiGe heterojunction bipolar transistors (HBTs) and conventional scaled CMOS, microwave-frequency systems can now be reasonably implemented using the same devices as conventional low-frequency baseband electronics. These advanced silicon-based active devices are discussed in more detail in the companion volume *RF and Microwave Passive and Active Technologies*, Chapters 16–19. In addition, the commercialization of low-cost III–V compound semiconductor electronics, including ion-implanted metal semiconductor field-effect transistors (MESFETs), pseudomorphic and lattice-matched high electron mobility transistors (HEMTs), and III–V HBTs, has dramatically decreased the cost of including these elements in high-volume consumer systems. These compound-semiconductor devices are described in Chapters 17 and 20–22 in the *RF and Microwave Passive and Active Technologies* volume of this handbook series. This convergence, with silicon microelectronics moving ever higher in frequency into the microwave spectrum from the low-frequency side and compound semiconductors declining in price for the middle of the frequency range, blurs the distinction between "microwave" and "RF" engineering, since "microwave" functions can now be realized with "mainstream" low-cost electronics. This is accompanied by a shift from physically large, low-integration-level hybrid implementations to highly-integrated solutions based on monolithic microwave integrated circuits (MMICs) (see Chapters 25–26 of this volume and Chapters 24–25 in the companion volume *RF and Microwave Passive and Active Technologies*). This shift has a dramatic effect not only on the design of systems and components, but also on the manufacturing technology and economics of production and implementation as well. A more complete discussion of the active device and integration technologies that make this progression possible is included in Section II of the companion volume *RF and Microwave Passive and Active Technologies* while modeling of these devices is described in Section III of this volume.

Aside from these defining characteristics of RF and microwave systems, the behavior of materials is also often different at microwave frequencies than at low frequencies. In metals, the effective resistance at microwave frequencies can differ significantly from that at dc. This frequency-dependent resistance is a consequence of the skin effect, which is caused by the finite penetration depth of an electromagnetic field into conducting material. This effect is a function of frequency; the depth of penetration is given by $\delta_s = (1/\sqrt{\pi f \mu \sigma})$, where μ is the permeability, f is the frequency, and σ is the conductivity of the material. As the expression indicates, δ_s decreases with increasing frequency, and so the electromagnetic fields are confined to regions increasingly near the surface as the frequency increases. This results in the microwave currents flowing exclusively along the surface of the conductor, significantly increasing the effective resistance (and thus the loss) of metallic interconnects. Further discussion of this topic can be found in Chapter 28 of the companion volume *RF and Microwave Applications and Systems* and Chapter 26 of the *RF and Microwave Passive and Active Technologies* volume in this handbook series. Dielectric materials also exhibit frequency-dependent characteristics that can be important. The permeability and loss of dielectrics arises from the internal polarization and dissipation of the material. Since the polarization within a dielectric is governed by the response of the material's internal charge distribution, the frequency dependence is governed by the speed at which these charges can redistribute in response to the applied fields. For ideal materials, this dielectric relaxation leads to a frequency-dependent permittivity of the form $\varepsilon(\omega) = \varepsilon_\infty + (\varepsilon_{dc} - \varepsilon_\infty)/(1 + j\omega\tau)$, where ε_{dc} is the low-frequency permittivity, ε_∞ is the high-frequency (optical) permittivity, and τ is the dielectric relaxation time. Loss in the dielectric is incorporated in this expression through the imaginary part of ε. For many materials the dielectric relaxation time is sufficiently small that the performance of the dielectric at microwave frequencies is very similar to that at low frequencies. However, this is not universal and some care is required since some materials and devices exhibit dispersive behavior at quite low frequencies. Furthermore, this description of dielectrics is highly idealized; the frequency response of many real-world materials is much more complex than this idealized model would suggest. High-value capacitors and semiconductor devices are among the classes of devices that are particularly likely to exhibit complex dielectric responses.

In addition to material properties, some physical effects are significant at microwave frequencies that are typically negligible at lower frequencies. For example, radiation losses become increasingly important

as the signal wavelengths approach the component and interconnect dimensions. For conductors and other components of comparable size to the signal wavelengths, standing waves caused by reflection of the electromagnetic waves from the boundaries of the component can greatly enhance the radiation of electromagnetic energy. These standing waves can be easily established either intentionally (in the case of antennae and resonant structures) or unintentionally (in the case of abrupt transitions, poor circuit layout, or other imperfections). Careful attention to transmission line geometry, placement relative to other components, transmission lines, and ground planes, as well as circuit packaging is essential for avoiding excessive signal attenuation and unintended coupling due to radiative effects.

A further distinction in the practice of RF and microwave engineering from conventional electronics is the methodology of testing and characterization. Due to the high frequencies involved, the impedance and standing-wave effects associated with test cables and the parasitic capacitance of conventional test probes make the use of conventional low-frequency circuit characterization techniques impractical. Although advanced measurement techniques such as electro-optic sampling can sometimes be employed to circumvent these difficulties, in general the loading effect of measurement equipment poses significant measurement challenges for debugging and analyzing circuit performance, especially for nodes at the interior of the circuit under test. In addition, for circuits employing dielectric or hollow guided-wave structures, voltage and current often cannot be uniquely defined. Even for structures in which voltage and current are well-defined, practical difficulties associated with accurately measuring such high-frequency signals make this difficult. Furthermore, since a dc-coupled time-domain measurement of a microwave signal would have an extremely wide noise bandwidth, the sensitivity of the measurement would be inadequate. For these reasons, components and low-level subsystems are characterized using specialized techniques.

One of the most common techniques for characterizing the linear behavior of microwave components is the use of *s*-parameters. While *z*-, *y*-, and *h*-parameter representations are commonly used at lower frequencies, these approaches can be problematic to implement at microwave frequencies. The use of *s*-parameters essentially captures the same information as these other parameter sets, but instead of directly measuring terminal voltages and currents, the forward and reverse traveling waves at the input and output ports are measured instead. While perhaps not intuitive at first, this approach enables accurate characterization of components at very high frequencies to be performed with comparative ease. For a two-port network, the *s*-parameters are defined by:

$$\begin{bmatrix} V_1^- \\ V_2^- \end{bmatrix} = \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix} \begin{bmatrix} V_1^+ \\ V_2^+ \end{bmatrix}$$

where the V^- terms are the wave components traveling away from the two-port, and the V^+ terms are the incident terms. These traveling waves can be thought of as existing on "virtual" transmission lines attached to the device ports. From this definition,

$$s_{11} = \left.\frac{V_1^-}{V_1^+}\right|_{V_2^+=0}$$

$$s_{12} = \left.\frac{V_1^-}{V_2^+}\right|_{V_1^+=0}$$

$$s_{21} = \left.\frac{V_2^-}{V_1^+}\right|_{V_2^+=0}$$

$$s_{22} = \left.\frac{V_2^-}{V_2^+}\right|_{V_1^+=0}$$

To measure the s-parameters, the ratio of the forward and reverse traveling waves on the virtual input and output transmission lines is measured. To achieve the $V_1^+ = 0$ and $V_2^+ = 0$ conditions in these expressions, the ports are terminated in the characteristic impedance, Z_0, of the virtual transmission lines. Although in principle these measurements can be made using directional couplers to separate the forward and reverse traveling waves and phase-sensitive detectors, in practice modern network analyzers augment the measurement hardware with sophisticated calibration routines to remove the effects of hardware imperfections to achieve accurate *s*-parameter measurements. A more detailed discussion of *s*-parameters, as well as other approaches to device and circuit characterization, is provided in Section I of this volume.

I.2 General Applications

The field of microwave engineering is currently experiencing a radical transformation. Historically, the field has been driven by applications requiring the utmost in performance with little concern for cost or manufacturability. These systems have been primarily for military applications, where performance at nearly any cost could be justified. The current transformation of the field involves a dramatic shift from defense applications to those driven by the commercial and consumer sector, with an attendant shift in focus from design for performance to design for manufacturability. This transformation also entails a shift from small production volumes to mass production for the commercial market, and from a focus on performance without regard to cost to a focus on minimum cost while maintaining acceptable performance. For wireless applications, an additional shift from broadband systems to systems having very tightly-regulated spectral characteristics also accompanies this transformation.

For many years the driving application of microwave technology was military radar. The small wavelength of microwaves permits the realization of narrowly-focused beams to be achieved with antennae small enough to be practically steered, resulting in adequate resolution of target location. Long-distance terrestrial communications for telephony as well as satellite uplink and downlink for voice and video were among the first commercially viable applications of microwave technology. These commercial communications applications were successful because microwave-frequency carriers (f_c) offer the possibility of very wide absolute signal bandwidths (Δf) while still maintaining relatively narrow fractional bandwidths (i.e., $\Delta f / f_c$). This allows many more voice and data channels to be accommodated than would be possible with lower-frequency carriers or baseband transmission.

Among the current host of emerging applications, many are based largely on this same principle, namely, the need to transmit more and more data at high speed, and thus the need for many communication channels with wide bandwidths. Wireless communication of voice and data, both to and from individual users as well as from users and central offices in aggregate, wired communication including coaxial cable systems for video distribution and broadband digital access, fiber-optic communication systems for long- and short-haul telecommunication, and hybrid systems such as hybrid fiber-coax systems are all designed to take advantage of the wide bandwidths and consequently high data carrying capacity of microwave-frequency electronic systems. The widespread proliferation of wireless Bluetooth personal-area networks and WiFi local-area networks for transmission of voice, data, messaging and online services operating in the unlicensed ISM bands is an example of the commoditization of microwave technology for cost-sensitive consumer applications. In addition to the explosion in both diversity and capability of microwave-frequency communication systems, radar systems continue to be of importance with the emergence of nonmilitary and nonnavigational applications such as radar systems for automotive collision avoidance and weather and atmospheric sensing. Radar based noncontact fluid-level sensors are also increasingly being used in industrial process control applications. Traditional applications of microwaves in industrial material processing (primarily via nonradiative heating effects) and cooking have recently been augmented with medical uses for microwave-induced localized hyperthermia for oncological and other medical treatments.

In addition to these extensions of "traditional" microwave applications, other fields of electronics are increasing encroaching into the microwave-frequency range. Examples include wired data networks based on coaxial cable or twisted-pair transmission lines with bit rates of over 1 Gb/s, fiber-optic communication systems with data rates well in excess of 10 Gb/s, and inexpensive personal computers and other digital systems with clock rates of well over 1 GHz. The continuing advances in the speed and capability of conventional microelectronics is pushing traditional circuit design ever further into the microwave-frequency regime. These advances have continued to push digital circuits into regimes where distributed circuit effects must be considered. While system- and board-level digital designers transitioned to the use of high-speed serial links requiring the use of distributed transmission lines in their designs some time ago, on-chip transmission lines for distribution of clock signals and the serialization of data signals for transmission over extremely high-speed serial buses are now an established feature of high-end designs within a single integrated circuit. These trends promise to both invigorate and reshape the field of microwave engineering in new and exciting ways.

I.3 Frequency Band Definitions

The field of microwave and RF engineering is driven by applications, originally for military purposes such as radar and more recently increasingly for commercial, scientific, and consumer applications. As a consequence of this increasingly diverse applications base, microwave terminology and frequency band designations are not entirely standardized, with various standards bodies, corporations, and other interested parties all contributing to the collective terminology of microwave engineering. Figure I.5 shows graphically the frequency ranges of some of the most common band designations. As can be seen from the complexity of Figure I.5, some care must be exercised in the use of the "standard" letter designations; substantial differences in the definitions of these bands exist in the literature and in practice. While the IEEE standard for radar bands [8] expressly deprecates the use of radar band designations for nonradar applications, the convenience of the band designations as technical shorthand has led to the use of these band designations in practice for a wide range of systems and technologies. This appropriation of radar band designations for other applications, as well as the definition of other letter-designated bands for other applications (e.g., electronic countermeasures) that have different frequency ranges is in part responsible for the complexity of Figure I.5. Furthermore, as progress in device and system performance opens up new system possibilities and makes ever-higher frequencies useful for new systems, the terminology of microwave engineering is continually evolving.

Figure I.5 illustrates in approximate order of increasing frequency the range of frequencies encompassed by commonly-used letter-designated bands. In Figure I.5, the dark shaded regions within the bars indicate the IEEE radar band designations, and the light cross-hatching indicates variations in the definitions by different groups and authors. The double-ended arrows appearing above some of the bands indicate other non-IEEE definitions for these letter designations that appear in the literature. For example, multiple distinct definitions of L, S, C, X, and K band are in use. The IEEE defines K band as the range from 18 to 27 GHz, while some authors define K band to span the range from 10.9 to 36 GHz, encompassing most of the IEEE's K_u, K, and K_a bands within a single band. Both of these definitions are illustrated in Figure I.5. Similarly, L band has two substantially different, overlapping definitions, with the IEEE definition of L band including frequencies from 1 to 2 GHz, with an older alternative definition of 390 MHz–1.55 GHz being found occasionally in the literature. Many other bands exhibit similar, though perhaps less extreme, variations in their definitions by various authors and standards committees. A further caution must also be taken with these letter designations, as different standards bodies and agencies do not always ensure that their letter designations are not used by others. As an example, the IEEE and U.S. military both define C, L, and K bands, but with very different frequencies; the IEEE L band resides at the low end of the microwave spectrum, while the military definition of L band is from 40 to 60 GHz. The designations (L–Y) in Figure I.5a are presently used widely in practice and the technical literature, with the newer U.S. military

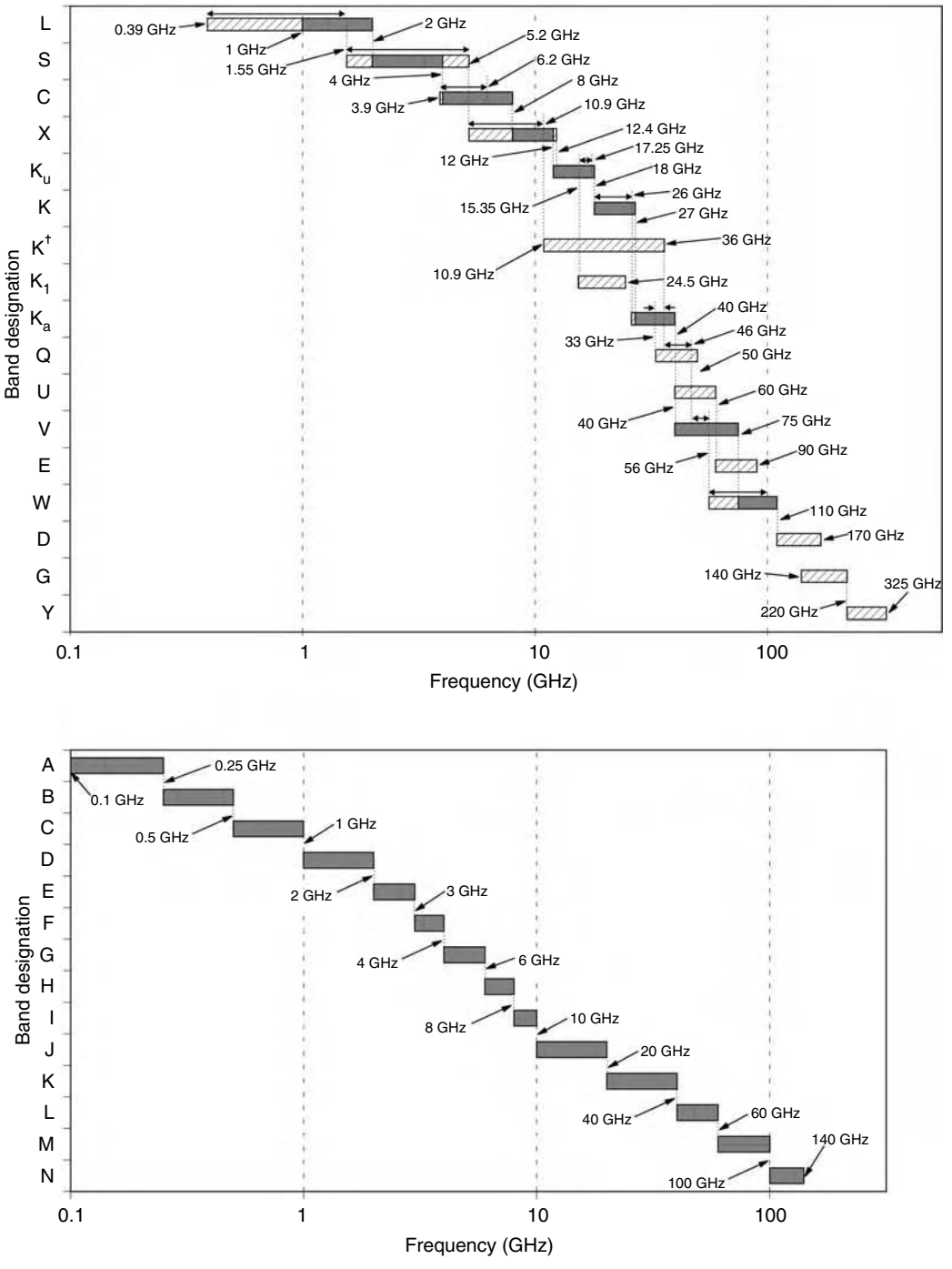


FIGURE I.5 Microwave and RF frequency band designations [1–7]. (a) Industrial and IEEE designations. Diagonal hashing indicates variation in the definitions found in literature; dark regions in the bars indicate the IEEE radar band definitions [8]. Double-ended arrows appearing above bands indicate alternative band definitions appearing in the literature, and $K^{\dagger}$ denotes an alternative definition for K band found in Reference [7]. (b) U.S. military frequency band designations [2–5].

designations (A–N) shown in Figure I.5b having not gained widespread popularity outside of the military community.

I.4 Overview of The RF and Microwave Handbook

The field of microwave and RF engineering is inherently interdisciplinary, spanning the fields of system architecture, design, modeling, and validation; circuit design, characterization, and verification; active and passive device design, modeling, and fabrication, including technologies as varied as semiconductor devices, solid-state passives, and vacuum electronics; electromagnetic field theory, atmospheric wave propagation, electromagnetic compatibility and interference; and manufacturing, reliability and system integration. Additional factors, including biological effects of high-frequency radiation, system cost, and market factors also play key roles in the practice of microwave and RF engineering. This extremely broad scope is further amplified by the large number of technological and market-driven design choices faced by the practitioner on a regular basis.

The full sweep of microwave and RF engineering is addressed in this three-volume handbook series. Section I of this volume features coverage of the unique difficulties and challenges encountered in accurately measuring microwave and RF devices and components, including linear and non-linear characterization approaches, load-pull and large-signal network analysis techniques, noise measurements, fixturing and high-volume testing issues, and testing of digital systems. Consideration of key circuits for functional blocks in a wide array of system applications is addressed in Section II, including low-level circuits such as low-noise amplifiers, mixers, oscillators, power amplifiers, switches, and filters, as well as higher-level functionalities such as receivers, transmitters, and phase-locked loops. Section III of this volume discusses technology computer-aided design (TCAD) and nonlinear modeling of devices and circuits, along with analysis tools for systems, electromagnetics, and circuits.

A companion volume in this handbook series, *RF and Microwave Applications and Systems*, features detailed discussion of system-level considerations for high-frequency systems. Section I of this companion volume focuses on system-level considerations with an application-specific focus. Typical applications, ranging from nomadic communications and cellular systems, wireless local-area networks, analog fiber-optic links, satellite communication networks, navigational aids and avionics, to radar, medical therapies, and electronic warfare applications are examined in detail. System-level considerations from the viewpoint of system integration and with focus on issues such as thermal management, cost modeling, manufacturing, and reliability are addressed in Section II of this volume in the handbook series, while the fundamental physical principles that govern the operation of devices and microwave and RF systems generally are discussed in Section III. Particular emphasis is placed on electromagnetic field theory through Maxwell's equations, free-space and guided-wave propagation, fading and multipath effects in wireless channels, and electromagnetic interference effects.

Comprehensive coverage of passive and active device technologies for microwave and RF systems is provided in a third companion volume in the handbook series, *RF and Microwave Passive and Active Technologies.* Passive devices are discussed in Section I of this volume, which includes coverage of radiating elements, cables and connectors, and packaging technology, as well as in-circuit passive elements including resonators, filters, and other components. The fundamentals of active device technologies, including semiconductor diodes, transistors and integrated circuits as well as vacuum electron devices, are discussed in Section II. Key device technologies including varactor and Schottky diodes, as well as bipolar junction transistors and heterojunction bipolar transistors in both the SiGe and III-V material systems are described, as are Si MOSFETs and III-V MESFETs and HEMTs. A discussion of the fundamental physical properties at high frequencies of common materials, including metals, dielectrics, ferroelectric and piezoelectric materials, and semiconductors, is provided in Section III of this volume in the handbook series.

References

1. Chang, K., Bahl, I., and Nair, V., *RF and Microwave Circuit and Component Design for Wireless Systems*, John Wiley & Sons, New York, 2002.
2. Collin, R. E., *Foundations for Microwave Engineering*, McGraw-Hill, New York, 1992, 2.
3. Harsany, S. C., *Principles of Microwave Technology*, Prentice Hall, Upper Saddle River, 1997, 5.
4. Laverghetta, T. S., *Modern Microwave Measurements and Techniques*, Artech House, Norwood, 1988, 479.
5. Misra, D. K., *Radio-Frequency and Microwave Communication Circuits: Analysis and Design*, John Wiley & Sons, New York, 2001.
6. Rizzi, P. A., *Microwave Engineering*, Prentice-Hall, Englewood Cliffs, 1988, 1.
7. *Reference Data for Radio Engineers*, ITT Corp., New York, 1975.
8. IEEE Std. 521-2002.

1

Overview of Microwave Engineering

Mike Golio
HVVi Semiconductor

1.1 Semiconductor Materials for RF and Microwave Applications

In addition to consideration of unique properties of metal and dielectric materials, the radio frequency (RF) and microwave engineer must also make semiconductor choices based on how existing semiconductor properties address the unique requirements of RF and microwave systems. Although semiconductor materials are exploited in virtually all electronics applications today, the unique characteristics of RF and microwave signals requires that special attention be paid to specific properties of semiconductors which are often neglected or of second-order importance for other applications. Two critical issues to RF applications are (a) the speed of electrons in the semiconductor material and (b) the breakdown field of the semiconductor material.

The first issue, speed of electrons, is clearly important because the semiconductor device must respond to high frequency changes in polarity of the signal. Improvements in efficiency and reductions in parasitic losses are realized when semiconductor materials are used which exhibit high electron mobility and velocity. Figure 1.1 presents the electron velocity of several important semiconductor materials as a

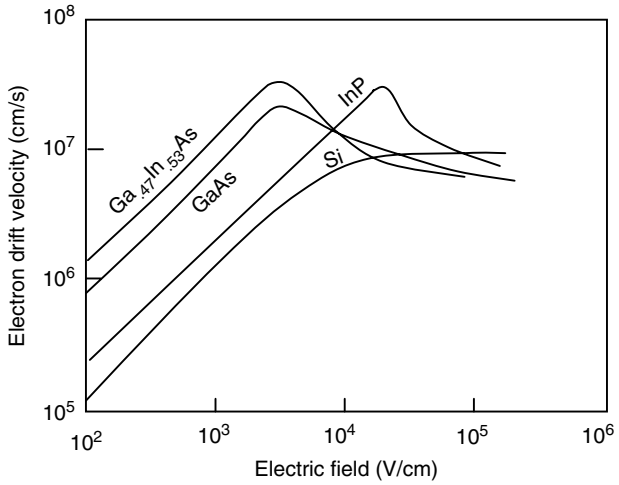


FIGURE 1.1 The electron velocity as a function of applied electric field for several semiconductor materials which are important for RF and microwave applications.

TABLE 1.1 Mobility and Breakdown Electric Field Values for Several Semiconductors Important for RF and Microwave Transmitter Applications

Property	Si	SiC	InP	GaAs	GaN
Electron mobility (cm^2/Vs)	1900	40–1000	4600	8800	1000
Breakdown field (V/cm)	3×10^5	20×10^4 to 30×10^5	5×10^5	6×10^5	$>10 \times 10^5$

function of applied electric field. The carrier mobility is given by

$$\mu_c = \frac{\nu}{e} \quad \text{for small values of } E \tag{1.1}$$

where ν is the carrier velocity in the material and E is the electric field.

Although Silicon is the dominant semiconductor material for electronics applications today, Figure 1.1 illustrates that III–V semiconductor materials such as GaAs, GaInAs, and InP exhibit superior electron velocity and mobility characteristics relative to Silicon. Bulk mobility values for several important semiconductors are also listed in Table 1.1. As a result of the superior transport properties, transistors fabricated using III–V semiconductor materials such as GaAs, InP, and GaInAs exhibit higher efficiency and lower parasitic resistance at microwave frequencies.

From a purely technical performance perspective, the above discussion argues primarily for the use of III–V semiconductor devices in RF and microwave applications. These arguments are not complete, however. Most commercial wireless products also have requirements for high yield, high volume, low cost, and rapid product development cycles. These requirements can overwhelm the material selection process and favor mature processes and high volume experience. The silicon high volume manufacturing experience base is far greater than that of any III–V semiconductor facility.

The frequency of the application becomes a critical performance characteristic in the selection of device technology. Because of the fundamental material characteristics illustrated in Figure 1.1, Silicon device structures will always have lower theoretical maximum operation frequencies than identical III–V device structures. The higher the frequency of the application, the more likely the optimum device choice will be a III–V transistor over a Silicon transistor. Above some frequency, f_{III-V}, compound semiconductor devices dominate the application space, with Silicon playing no significant role in the

microwave portion of the product. In contrast, below some frequency, f_{Si}, the cost and maturity advantage of Silicon provide little opportunity for III–V devices to compete. In the transition spectrum between these two frequencies Silicon and III–V devices coexist. Although Silicon devices are capable of operating above frequency f_{Si}, this operation is often gained at the expense of DC current drain. As frequency is increased above f_{Si} in the transition spectrum, efficiency advantages of GaAs and other III–V devices provide competitive opportunities for these parts. The critical frequencies, f_{Si} and f_{III-V} are not static frequency values. Rather, they are continually being moved upward by the advances of Silicon technologies—primarily by decreasing critical device dimensions.

The speed of carriers in a semiconductor transistor can also be affected by deep levels (traps) located physically either at the surface or in the bulk material. Deep levels can trap charge for times that are long compared to the signal period and thereby reduce the total RF power carrying capability of the transistor. Trapping effects result in frequency dispersion of important transistor characteristics such as transconductance and output resistance. Pulsed measurements as described in Section 1.4.4 (especially when taken over temperature extremes) can be a valuable tool to characterize deep level effects in semiconductor devices. Trapping effects are more important in compound semiconductor devices than in silicon technologies.

The second critical semiconductor issue listed in Table 1.1 is breakdown voltage. The constraints placed on the RF portion of radio electronics are fundamentally different from the constraints placed on digital circuits in the same radio. For digital applications, the presence or absence of a single electron can theoretically define a bit. Although noise floor and leakage issues make the practical limit for bit signals larger than this, the minimum amount of charge required to define a bit is very small. The bit charge minimum is also independent of the radio system architecture, the radio transmission path or the external environment. If the amount of charge utilized to define a bit within the digital chip can be reduced, then operating voltage, operating current, or both can also be reduced with no adverse consequences for the radio.

In contrast, the required propagation distance and signal environment are the primary determinants for RF signal strength. If 1 W of transmission power is required for the remote receiver to receive the signal, then reductions in RF transmitter power below this level will cause the radio to fail. Modern radio requirements often require tens, hundreds, or even thousands of Watts of transmitted power in order for the radio system to function properly. Unlike the digital situation where any discernable bit is as good as any other bit, the minimum RF transmission power must be maintained. A Watt of RF power is the product of signal current, signal voltage and efficiency, so requirements for high power result in requirements for high voltage, high current and high efficiency.

The maximum electric field before the onset of avalanche breakdown, *breakdown field*, is the fundamental semiconductor property that often limits power operation in a transistor. Table 1.1 presents breakdown voltages for several semiconductors that are commonly used in transmitter applications. In addition to Silicon, GaAs and InP, two emerging widebandgap semiconductors, SiC and GaN are included in the table. Interest from microwave engineers in these less mature semiconductors is driven almost exclusively by their attractive breakdown field capabilities. Figure 1.2 summarizes the semiconductor material application situation in terms of the power–frequency space for RF and microwave systems.

1.2 Propagation and Attenuation in the Atmosphere

Many modern RF and microwave systems are wireless. Their operation depends on transmission of signals through the atmosphere. Electromagnetic signals are attenuated by the atmosphere as they propagate from source to target. Consideration of the attenuation characteristics of the atmosphere can be critical in the design of these systems. In general, atmospheric attenuation increases with increasing frequency. As shown in Figure 1.3, however, there is significant structure in the atmospheric attenuation versus frequency plot. If only attenuation is considered, it is clear that low frequencies would be preferred for long range communications, sensor, or navigation systems in order to take advantage of the low attenuation of the atmosphere. If high data rates or large information content is required, however, higher frequencies

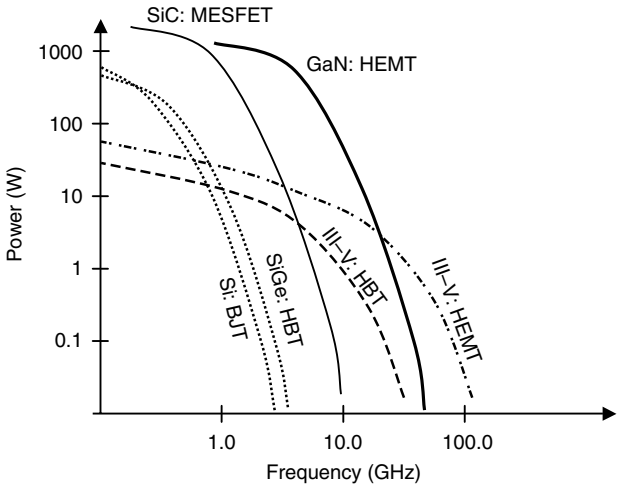


FIGURE 1.2 Semiconductor choices for RF applications are a strong function of the power and frequency required for the wireless application.

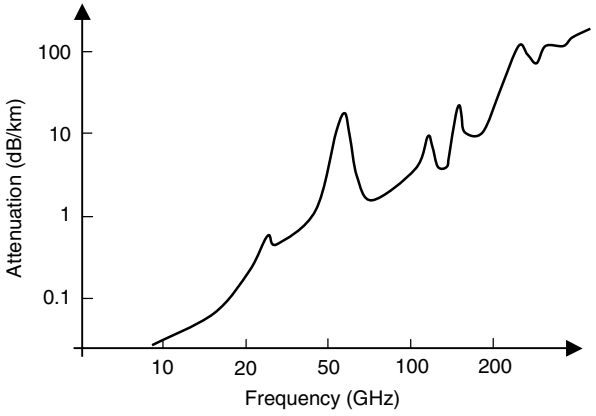


FIGURE 1.3 Attenuation of electromagnetic signals in the atmosphere as a function of frequency.

are needed. In addition to the atmospheric attenuation, the wavelengths of microwave systems are small enough to become effected by water vapor and rain. Above 10 GHz these effects become important. Above 25 GHz, the effect of individual gas molecules becomes important. Water and oxygen are the most important gases. These have resonant absorption lines at ∼23, ∼69, and ∼120 GHz. In addition to absorption lines, the atmosphere also exhibits "windows" that may be used for communication, notably at ∼38 and ∼98 GHz.

RF and microwave signal propagation is also affected by objects such as trees, buildings, towers, and vehicles in the path of the wave. Indoor systems are affected by walls, doors, furniture, and people. As a result of the interaction of electromagnetic signals with objects, the propagation channel for wireless communication systems consists of multiple paths between the transmitter and receiver. Each path will experience different attenuation and delay. Some transmitted signals may experience a deep fade (large attenuation) due to destructive multipath cancellation. Similarly, constructive multipath addition can produce signals of large amplitude. Shadowing can occur when buildings or other objects obstruct the line-of-site path between transmitter and receiver.

The design of wireless systems must consider the interaction of specific frequencies of RF and microwave signals with the atmosphere and with objects in the signal channel that can cause multipath effects.

1.3 Systems Applications

There are four important classes of applications for microwave and RF systems: communications, navigation, sensors, and heating. Each of these classes of applications benefits from some of the unique properties of high-frequency electromagnetic fields.

1.3.1 Communications

Wireless communications applications have exploded in popularity over the past decade. Pagers, cellular phones, radio navigation, and wireless data networks are among the RF products that consumers are likely to be familiar with. Prior to the growth of commercial wireless communications, RF and microwave radios were in common usage for communications satellites, commercial avionics communications, and many government and military radios. All of these systems benefit from the high frequencies that offer greater bandwidth than low frequency systems, while still propagating with relatively low atmospheric losses compared to higher frequency systems.

Cellular phones are among the most common consumer radios in use today. Analog cellular (first generation or 1G cellular) operates at 900 MHz bands and was first introduced in 1983. Second generation (2G) cellular using TDMA, GSM TDMA, and CDMA digital modulation schemes came into use more than 10 years later. The 2G systems were designed to get greater use of the 1.9 GHz frequency bands than their analog predecessors. Emergence of 2.5G and 3G systems operating in broader bands as high as 2.1 GHz is occurring today. These systems make use of digital modulation schemes adapted from 2G GSM and CDMA systems. With each advance in cellular phones, requirements on the microwave circuitry have increased. Requirements for broader bandwidths, higher efficiency and greater linearity have been coupled with demands for lower cost, lighter, smaller products, and increasing functionality. The microwave receivers and transmitters designed for portable cellular phones represent one of the highest volume manufacturing requirements of any microwave radio. Fabrication of popular cell phones has placed an emphasis on manufacturability and yield for microwave radios that was unheard of prior to the growth in popularity of these products.

Other microwave-based consumer products that are growing dramatically in popularity are the wireless local area network (WLAN) or Wi-Fi and the longer range WiMAX systems. These systems offer data rates more than five times higher than cellular-based products using bandwidth at 2.4, 3.5, and 5 GHz. Although the volume demands for Wi-Fi and WiMAX components are not as high as for cellular phones, the emphasis on cost and manufacturability is still critical to these products.

Commercial communications satellite systems represent a microwave communications product that is less conspicuous to the consumer, but continues to experience increasing demand. Although the percentage of voice traffic carried via satellite systems is rapidly declining with the advent of undersea fiber-optic cables, new video and data services are being added over existing voice services. Today satellites provide worldwide TV channels, global messaging services, positioning information, communications from ships and aircraft, communications to remote areas, and high-speed data services including internet access. Allocated satellite communication frequency bands include spectrum from as low as 2.5 GHz to almost 50 GHz. These allocations cover extremely broad bandwidths compared to many other communications systems. Future allocation will include even higher frequency bands. In addition to the bandwidth and frequency challenges, microwave components for satellite communications are faced with reliability requirements that are far more severe than any earth-based systems.

Avionics applications include subsystems that perform communications, navigation, and sensor applications. Avionics products typically require functional integrity and reliability that are orders of magnitude more stringent than most commercial wireless applications. The rigor of these requirements is matched or exceeded only by the requirements for space and/or certain military applications. Avionics must function in environments that are more severe than most other wireless applications as well. Quantities of products required for this market are typically very low when compared to commercial wireless applications, for example, the number of cell phones manufactured every single working day far exceeds the number of

aircraft that are manufactured in the world in a year. Wireless systems for avionics applications cover an extremely wide range of frequencies, function, modulation type, bandwidth, and power. Due to the number of systems aboard a typical aircraft, Electromagnetic Interference (EMI) and Electromagnetic Compatibility (EMC) between systems is a major concern, and EMI/EMC design and testing is a major factor in the flight certification testing of these systems. RF and microwave communications systems for avionics applications include several distinct bands between 2 and 400 MHz and output power requirements as high as 100 Watts.

In addition to commercial communications systems, military communication is an extremely important application of microwave technology. Technical specifications for military radios are often extremely demanding. Much of the technology developed and exploited by existing commercial communications systems today was first demonstrated for military applications. The requirements for military radio applications are varied but will cover broader bandwidths, higher power, more linearity, and greater levels of integration than most of their commercial counterparts. In addition, reliability requirements for these systems are stringent. Volume manufacturing levels, of course, tend to be much lower than commercial systems.

1.3.2 Navigation

Electronic navigation systems represent a unique application of microwave systems. In this application, data transfer takes place between a satellite (or fixed basestation) and a portable radio on earth. The consumer portable product consists of only a receiver portion of a radio. No data or voice signal is transmitted by the portable navigation unit. In this respect, electronic navigation systems resemble a portable paging system more closely than they resemble a cellular phone system. The most widespread electronic navigation system is GPS. The nominal GPS constellation is composed of 24 satellites in six orbital planes, (four satellites in each plane). The satellites operate in circular 20,200 km altitude (26,570 km radius) orbits at an inclination angle of 55°. Each satellite transmits a navigation message containing its orbital elements, clock behavior, system time, and status messages. The data transmitted by the satellite are sent in two frequency bands at 1.2 and 1.6 GHz. The portable terrestrial units receive these messages from multiple satellites and calculate the location of the unit on the earth. In addition to GPS, other navigation systems in common usage include NAVSTAR, GLONASS, and LORAN.

1.3.3 Sensors (Radar)

Microwave sensor applications are addressed primarily with various forms of radar. Radar is used by police forces to establish the speed of passing automobiles, by automobiles to establish vehicle speed and danger of collision, by air traffic control systems to establish the locations of approaching aircraft, by aircraft to establish ground speed, altitude, other aircraft and turbulent weather, and by the military to establish a multitude of different types of targets.

The receiving portion of a radar unit is similar to other radios. It is designed to receive a specific signal and analyze it to obtain desired information. The radar unit differs from other radios, however, in that the signal that is received is typically transmitted by the same unit. By understanding the form of the transmitted signal, the propagation characteristics of the propagation medium, and the form of the received (reflected) signal, various characteristics of the radar target can be determined including size, speed, and distance from the radar unit. As in the case of communications systems, radar applications benefit from the propagation characteristics of RF and microwave frequencies in the atmosphere. The best frequency to use for a radar unit depends upon its application. Like most other radio design decisions, the choice of frequency usually involves trade-offs among several factors including physical size, transmitted power, and atmospheric attenuation.

The dimensions of radio components used to generate RF power and the size of the antenna required to direct the transmitted signal are, in general, proportional to wavelength. At lower frequencies where

wavelengths are longer, the antennae and radio components tend to be large and heavy. At the higher frequencies where the wavelengths are shorter, radar units can be smaller and lighter.

Frequency selection can indirectly influence the radar power level because of its impact on radio size. Design of high power transmitters requires that significant attention be paid to the management of electric field levels and thermal dissipation. Such management tasks are made more complex when space is limited. Since radio component size tends to be inversely proportional to frequency, manageable power levels are reduced as frequency is increased.

As in the case of all wireless systems, atmospheric attenuation can reduce the total range of the system. Radar systems designed to work above about 10 GHz must consider the atmospheric loss at the specific frequency being used in the design.

Automotive radar represents a large class of radars that are used within an automobile. Applications include speed measurement, adaptive cruise control, obstacle detection, and collision avoidance. Various radar systems have been developed for forward-, rear-, and side-looking applications.

V-band frequencies are exploited for forward looking radars. Within V-band, different frequencies have been used in the past decade, including 77 GHz for U.S. and European systems, and 60 GHz in some Japanese systems. The choice of V-band for this application is dictated by the resolution requirement, antenna size requirement and the desire for atmospheric attenuation to insure the radar is short range. The frequency requirement of this application has contributed to a slow emergence of this product into mainstream use, but the potential of this product to have a significant impact on highway safety continues to keep automotive radar efforts active.

As in the case of communications systems, avionics and military users also have significant radar applications. Radar is used to detect aircraft both from the earth and from other aircraft. It is also used to determine ground speed, establish altitude, and detect weather turbulence.

1.3.4 Heating

The most common heating application for microwave signals is the microwave oven. These consumer products operate at a frequency that corresponds to a resonant frequency of water. When exposed to electromagnetic energy at this frequency, all water molecules begin to spin or oscillate at that frequency. Since all foods contain high percentages of water, electromagnetic energy at this resonant frequency interacts with all foods. The energy absorbed by these rotating molecules is transferred to the food in the form of heat.

RF heating can also be important for medical applications. Certain kinds of tumors can be detected by the lack of electromagnetic activity associated with them and some kinds of tumors can be treated by heating them using electromagnetic stimulation.

The use of RF/microwaves in medicine has increased dramatically in recent years. RF and microwave therapies for cancer in humans are presently used in many cancer centers. RF treatments for heartbeat irregularities are currently employed by major hospitals. RF/microwaves are also used in human subjects for the treatment of certain types of benign prostrate conditions. Several centers in the United States have been utilizing RF to treat upper airway obstruction and alleviate sleep apnea. New treatments such as microwave aided liposuction, tissue joining in conjunction with microwave irradiation in future endoscopic surgery, enhancement of drug absorption, and microwave septic wound treatment are continually being researched.

1.4 Measurements

The RF/microwave engineer faces unique measurement challenges. At high frequencies, voltages and currents vary too rapidly for conventional electronic measurement equipment to gauge. Conventional curve tracers and oscilloscopes are of limited value when microwave component measurements are needed. In addition, calibration of conventional characterization equipment typically requires the use

of open and short circuit standards that are not useful to the microwave engineer. For these reasons, most commonly exploited microwave measurements focus on the measurement of power and phase in the frequency domain as opposed to voltages and currents in the time domain.

1.4.1 Small Signal

Characterization of the linear performance of microwave devices, components and boards is critical to the development of models used in the design of the next higher level of microwave subsystem. At lower frequencies, direct measurement of *y*-, *z*-, or *h*-parameters is useful to accomplish linear characterization. As discussed in Chapter 1, however, RF and microwave design utilizes *s*-parameters for this application. Other small signal characteristics of interest in microwave design include impedance, VSWR, gain, and attenuation. Each of these quantities can be computed from two-port *s*-parameter data.

The *s*-parameters defined in Chapter 1 are complex quantities normally expressed as magnitude and phase. Notice that S_{11} and S_{22} can be thought of as complex reflection ratios since they represent the magnitude and phase of waves reflected from port 1 (input) and 2 (output), respectively. It is common to measure the quality of the match between components using the *reflection coefficient* defined as

$$\Gamma = |S_{11}| \tag{1.2}$$

for the input reflection coefficient of a two-port network, or

$$\Gamma = |S_{22}| \tag{1.3}$$

for the output reflection coefficient.

Reflection coefficient measurements are often expressed in dB and referred to as *return loss* evaluated as

$$L_{\text{return}} = -20\log(\Gamma). \tag{1.4}$$

Analogous to the reflection coefficient, both a forward and reverse *transmission coefficient* can be measured. The forward transmission coefficient is given as

$$T = |S_{21}| \tag{1.5}$$

while the reverse transmission coefficient is expressed

$$T = |S_{12}|\,. \tag{1.6}$$

As in the case of reflection coefficient, transmission coefficients are often expressed in dB and referred to as *gain* given by

$$G = 20\log(T). \tag{1.7}$$

Another commonly measured and calculated parameter is the *standing wave ratio* or the *voltage standing wave ratio* (VSWR). This quantity is the ratio of maximum to minimum voltage at a given port. It is commonly expressed in terms of reflection coefficient as

$$\text{VSWR} = \frac{1+\Gamma}{1-\Gamma}. \tag{1.8}$$

The vector network analyzer (VNA) is the instrument of choice for small signal characterization of high-frequency components. Figure 1.4 illustrates a one-port VNA measurement. These measurements

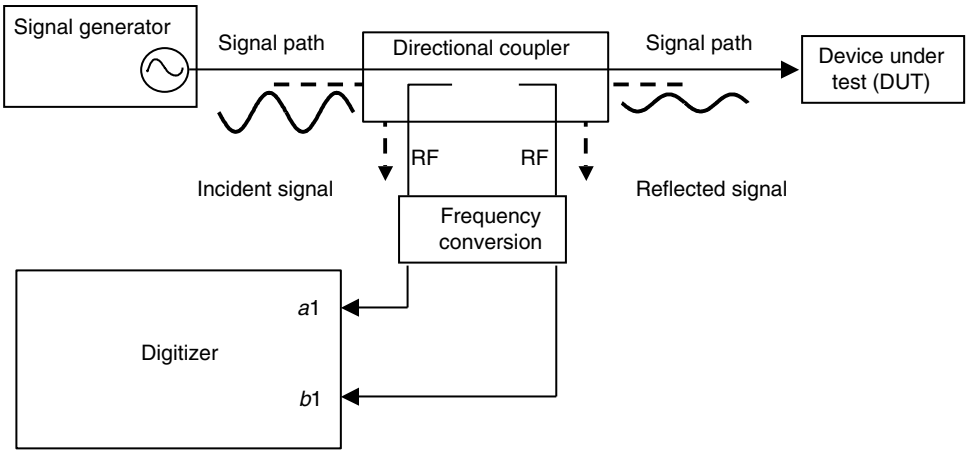


FIGURE 1.4 Vector network analyzer measurement configuration to determine *s*-parameters of a high-frequency device, component, or subsystem.

use a source with well-defined impedance equal to the system impedance and all ports of the device under test (DUT) are terminated with the same impedance. This termination eliminates unwanted signal reflections during the measurement. The port being measured is terminated in the test channel of the network analyzer that has input impedance equal to the system characteristic impedance. Measurement of system parameters with all ports terminated minimizes the problems caused by short-, open-, and test-circuit parasitics that cause considerable difficulty in the measurement of *y*- and *h*-parameters at very high frequencies. If desired, *s*-parameters can be converted to *y*- and *h*-parameters using analytical mathematical expressions.

The directional coupler shown in Figure 1.4 is a device for measuring the forward and reflected waves on a transmission line. During the network analyzer measurement, a signal is driven through the directional coupler to one port of the DUT. Part of the incident signal is sampled by the directional coupler. On arrival at the DUT port being measured, some of the incident signal will be reflected. This reflection is again sampled by the directional coupler. The sampled incident and reflected signals are then downconverted in frequency and digitized. The measurement configuration of Figure 1.4 shows only one-half of the equipment required to make full two-port *s*-parameter measurements. The *s*-parameters as defined in Chapter 1 are determined by analyzing the ratios of the digitized signal data.

For many applications, knowledge of the magnitude of the incident and reflected signals is sufficient (i.e., Γ is all that is needed). In these cases, the scalar network analyzer can be utilized in place of the VNA. The cost of the scalar network analyzer equipment is much less than VNA equipment and the calibration required for making accurate measurements is easier when phase information is not required. The scalar network analyzer measures reflection coefficient as defined in Equations 2.1 and 2.2.

1.4.2 Large Signal

Virtually all physical systems exhibit some form of nonlinear behavior and microwave systems are no exception. Although powerful techniques and elaborate tools have been developed to characterize and analyze linear RF and microwave circuits, it is often the nonlinear characteristics that dominate microwave engineering efforts. Nonlinear effects are not all undesirable. Frequency conversion circuitry, for example, exploits nonlinearities in order to translate signals from one frequency to another. Nonlinear performance characteristics of interest in microwave design include harmonic distortion, gain compression, intermodulation distortion (IMD), phase distortion, and adjacent channel power. Numerous other nonlinear phenomena and nonlinear figures-of-merit are less commonly addressed, but can be important for some microwave systems.

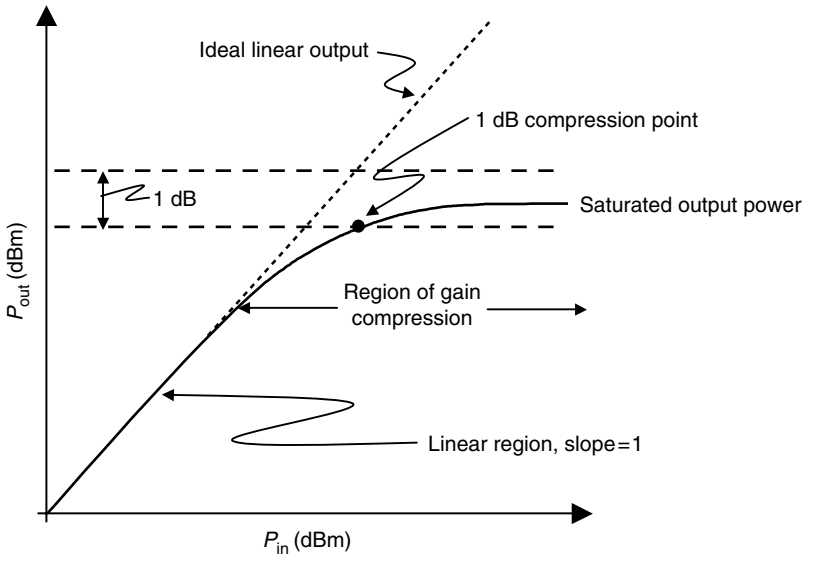


FIGURE 1.5 Output power versus input power at the fundamental frequency for a nonlinear circuit.

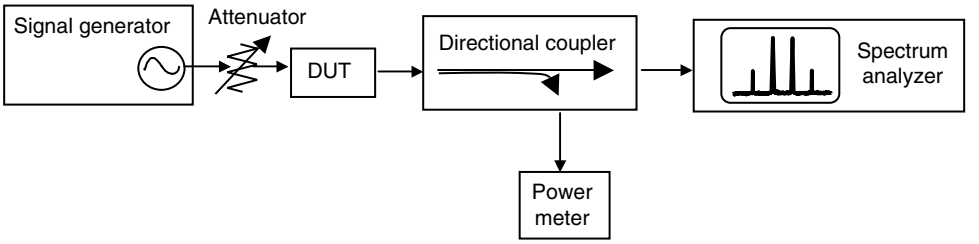


FIGURE 1.6 Measurement configuration to characterize gain compression and harmonic distortion. By replacing the signal generator with two combined signals at slightly offset frequencies, the configuration can also be used to measure intermodulation distortion.

1.4.2.1 Gain Compression

Figure 1.5 illustrates gain compression characteristics of a typical microwave amplifier with a plot of output power as a function of input power. At low power levels, a single frequency signal is increased in power level by the small signal gain of the amplifier ($P_{out} = G * P_{in}$). At lower power levels, this produces a linear P_{out} versus P_{in} plot with slope = 1 when the powers are plotted in dB units as shown in Figure 1.5. At higher power levels, nonlinearities in the amplifier begin to generate some power in the harmonics of the single frequency input signal and to compress the output signal. The result is decreased gain at higher power levels. This reduction in gain is referred to as *gain compression*. Gain compression is often characterized in terms of the power level when the large signal gain is 1 dB less than the small signal gain. The power level when this occurs is termed the *1*dB *compression point* and is also illustrated in Figure 1.5.

The microwave spectrum analyzer is the workhorse instrument of nonlinear microwave measurements. The instrument measures and displays power as a function of swept frequency. Combined with a variable power level signal source (or multiple combined or modulated sources), many nonlinear characteristics can be measured using the spectrum analyzer in the configuration illustrated in Figure 1.6.

1.4.2.2 Harmonic Distortion

A fundamental result of nonlinear distortion in microwave devices is that power levels are produced at frequencies which are integral multiples of the applied signal frequency. These other frequency

components are termed *harmonics* of the fundamental signal. Harmonic signal levels are usually specified and measured relative to the fundamental signal level. The harmonic level is expressed in dBc, which designates dB relative to the fundamental power level. Microwave system requirements often place a maximum acceptable level for individual harmonics. Typically third and second harmonic levels are critical, but higher-order harmonics can also be important for many applications. The measurement configuration illustrated in Figure 1.6 can be used to directly measure harmonic distortion of a microwave device.

1.4.2.3 Intermodulation Distortion

When a microwave signal is composed of power at multiple frequencies, a nonlinear circuit will produce IMD. The IMD characteristics of a microwave device are important because they can create unwanted interference in adjacent channels of a radio or radar system. The intermodulation products of two signals produce distortion signals not only at the harmonic frequencies of the two signals, but also at the sum and difference frequencies of all of the signal's harmonics. If the two signal frequencies are closely spaced at frequencies f_c and f_m, then the IMD products located at frequencies $2f_c - f_m$ and $2f_m - f_c$ will be located very close to the desired signals. This situation is illustrated in the signal spectrum of Figure 1.7. The IMD products at $2f_c - f_m$ and $2f_m - f_c$ are third-order products of the desired signals, but are located so closely to f_c and f_m that filtering them out of the overall signal is difficult.

The spectrum of Figure 1.7 represents the nonlinear characteristics at a single power level. As power is increased and the device enters gain compression, however, harmonic power levels will grow more quickly than fundamental power levels. In general, the *n*th-order harmonic power level will increase at n times the fundamental. This is illustrated in the P_{out} versus P_{in} plot of Figure 1.8 where both the fundamental and the third-order products are plotted. As in the case of the fundamental power, third-order IMD levels will compress at higher power levels. IMD is often characterized and specified in terms of the *third-order intercept point*, IP3. This point is the power level where the slope of the small signal gain and the slope of the low power level third-order product characteristics cross as shown in Figure 1.8.

1.4.2.4 Phase Distortion

Reactive elements in a microwave system give rise to time delays that are nonlinear. Such delays are referred to as *memory effects* and result in *AM–PM distortion* in a modulated signal. AM–PM distortion creates

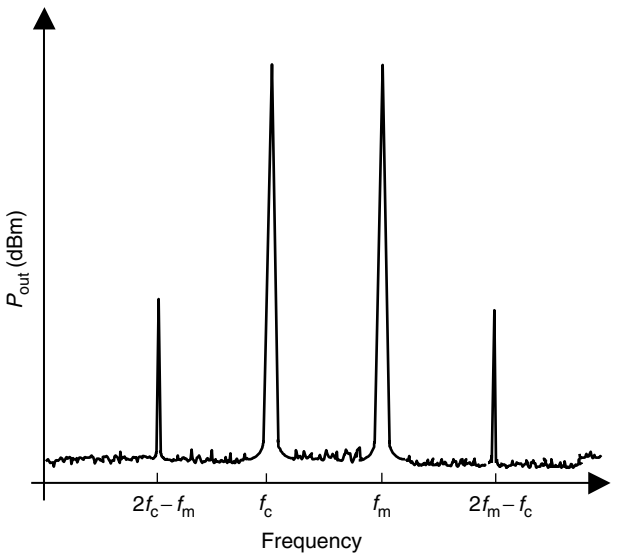


FIGURE 1.7 An illustration of signal spectrum due to intermodulation distortion from two signals at frequencies f_c and f_m.

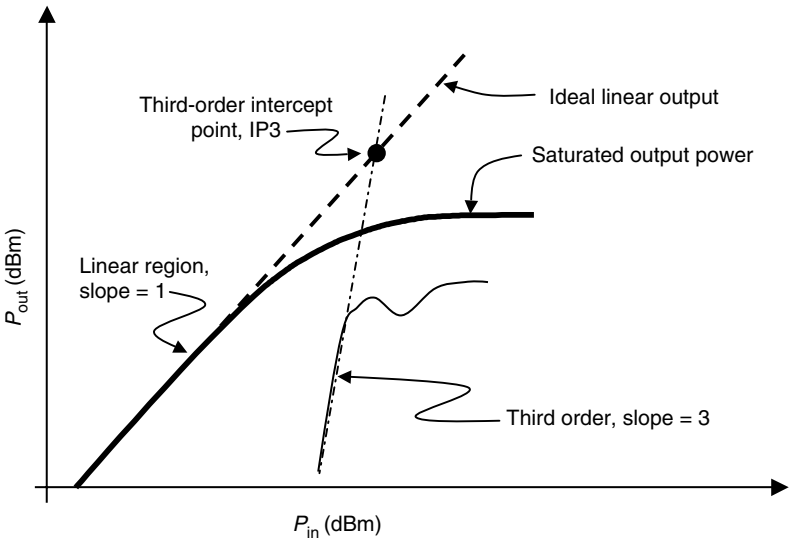


FIGURE 1.8 Relationship between signal output power and intermodulation distortion product levels.

sidebands at harmonics of a modulating signal. These sidebands are similar to the IMD sidebands, but are repeated for multiple harmonics. AM–PM distortion can dominate the out-of-band interference in a radio. At lower power levels, the phase deviation of the signal is approximately linear and the slope of the deviation, referred to as the modulation index, is often used as a figure-of-merit for the characterization of this nonlinearity. The *modulation index* is measured in degrees per volt using a VNA. The phase deviation is typically measured at the 1 dB compression point in order to determine modulation index. Because the VNA measures power, the computation of modulation index, k_ϕ, uses the formula

$$k_\phi = \frac{\Delta\Phi(P_{1\text{dB}})}{2Z_0\sqrt{P_{1\text{dB}}}} \tag{1.9}$$

where $\Delta\Phi(P_{1\text{dB}})$ is the phase deviation from small signal at the 1 dB compression point, Z_0 is the characteristic impedance of the system and $P_{1\text{dB}}$ is the 1 dB output compression point.

1.4.2.5 Adjacent Channel Power Ratio

Amplitude and phase distortion affect digitally modulated signals resulting in gain compression and phase deviation. The resulting signal, however, is far more complex than the simple one or two carrier results presented in Sections 1.4.2.2 through 1.4.2.4. Instead of IMD, *adjacent channel power ratio* (ACPR) is often specified for digitally modulated signals. ACPR is a measure of how much power leaks into adjacent channels of a radio due to the nonlinearities of the digitally modulated signal in a central channel. Measurement of ACPR is similar to measurement of IMD, but utilizes an appropriately modulated digital test signal in place of a single tone signal generator. Test signals for digitally modulated signals are synthesized using an *arbitrary waveform generator*. The output spectrum of the DUT in the channels adjacent to the tested channel are then monitored and power levels are measured.

1.4.2.6 Error Vector Magnitude

Adjacent channel power specifications are not adequate for certain types of modern digitally modulated systems. *Error vector magnitude* (EVM) is used in addition to, or instead of adjacent channel power for these systems. EVM specifications have already been written into system standards for GSM, NADC, and PHS, and they are poised to appear in many important emerging standards.

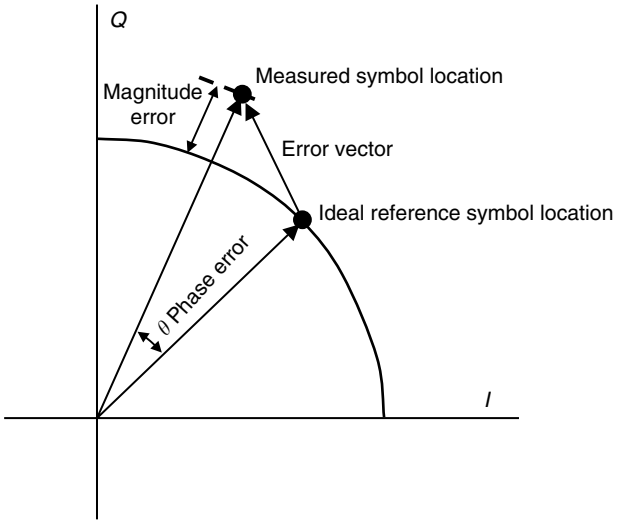


FIGURE 1.9 *I*–*Q* diagram indicating the error vector for EVM measurements.

The EVM measurement quantifies the performance of a radio transmitter against an ideal reference. A signal sent by an ideal transmitter would have all points in the *I*–*Q* constellation fall precisely at the ideal locations (i.e., magnitude and phase would be exact). Nonideal behavior of the transmitter, however, causes the actual constellation points to fall in a slightly scattered pattern that only approximates the ideal *I*–*Q* location. EVM is a way to quantify how far the actual points are from the ideal locations. This is indicated in Figure 1.9.

Measurement of EVM is accomplished using a vector signal analyzer (VSA). The equipment demodulates the received signal in a similar way to the actual radio demodulator. The actual *I*–*Q* constellation can then be measured and compared to the ideal constellation. EVM is calculated as the ratio of the root mean square power of the error vector to the RMS power of the reference.

1.4.3 Noise

Noise is a random process that can have many different sources such as thermally generated resistive noise, charge crossing a potential barrier, and generation–recombination (G–R) noise. Understanding noise is important in microwave systems because background noise levels limit the sensitivity, dynamic range and accuracy of a radio or radar receiver.

1.4.3.1 Noise Figure

At microwave frequencies noise characterization involves the measurement of noise power. The noise power of a linear device can be considered as concentrated at its input as shown in Figure 1.10. The figure considers an amplifier, but the analysis is easily generalized to other linear devices.

All of the amplifier noise generators can be lumped into an equivalent noise temperature with an equivalent input noise power per Hertz of $N_e = kT_e$, where k is Boltzmann's constant and T_e is the equivalent noise temperature. The noise power per Hertz available from the noise source is $N_S = kT_S$ as shown in Figure 1.10. Since noise limits the system sensitivity and dynamic range, it is useful to examine noise as it is related to signal strength using a signal-to-noise ratio (SNR). A figure-of-merit for an amplifier, *noise factor* (F), describes the reduction in SNR of a signal as it passes through the linear device illustrated in Figure 1.10. The noise factor for an amplifier is derived from the figure to be

$$F = \frac{\mathrm{SNR_{IN}}}{\mathrm{SNR_{OUT}}} = 1 + \frac{T_e}{T_S} \tag{1.10}$$

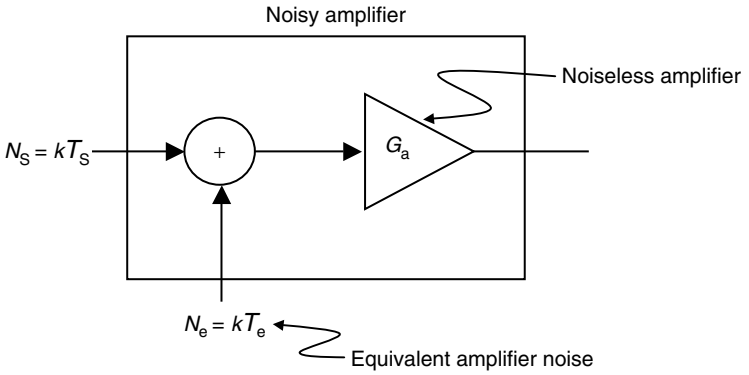


FIGURE 1.10 System view of amplifier noise.

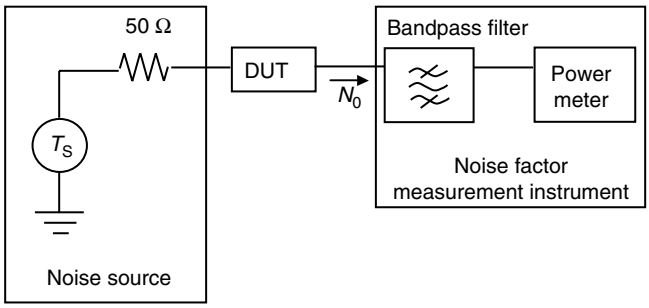


FIGURE 1.11 Measurement configuration for noise factor measurement.

Device noise factor can be measured as shown in Figure 1.11. To make the measurement, the source temperature is varied resulting in variation in the device noise output, N_0. The device noise contribution, however, remains constant. As T_S changes the noise power measured at the power meter changes providing a method to compute noise output.

In practice, the noise factor is usually given in decibels and called the *noise figure*,

$$\mathrm{NF} = 10 \log F \tag{1.11}$$

1.4.3.2 Phase Noise

When noise is referenced to a carrier frequency it modulates that carrier and causes amplitude and phase variations known as phase noise. Oscillator phase modulation (PM) noise is much larger than amplitude modulation (AM) noise. The phase variations caused by this noise result in *jitter* which is critical in the design and analysis of digital communication systems.

Phase noise is most easily measured using a spectrum analyzer. Figure 1.12 shows a typical oscillator source spectrum as measured directly on a spectrum analyzer. Characterization and analysis of phase noise is often described in terms of the power ratio of the noise at specific distances from the carrier frequency. This is illustrated in Figure 1.12.

1.4.4 Pulsed *I* –*V*

Although most of the measurements commonly utilized in RF and microwave engineering are frequency domain measurements, pulsed measurements are an important exception used to characterize

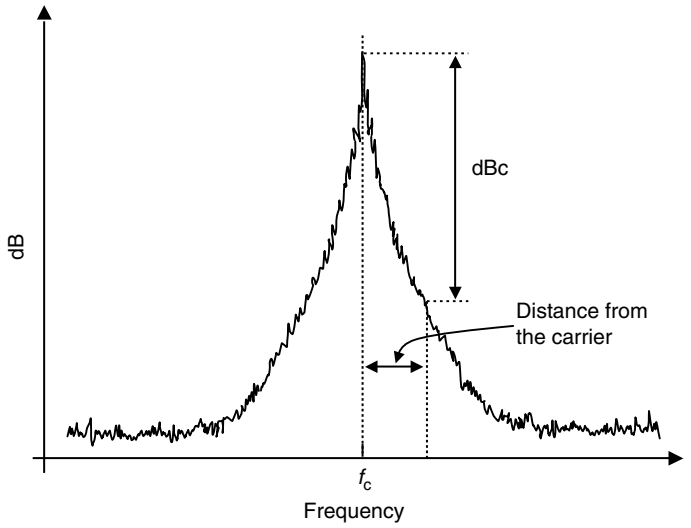


FIGURE 1.12 Typical phase noise spectrum observed on a spectrum analyzer.

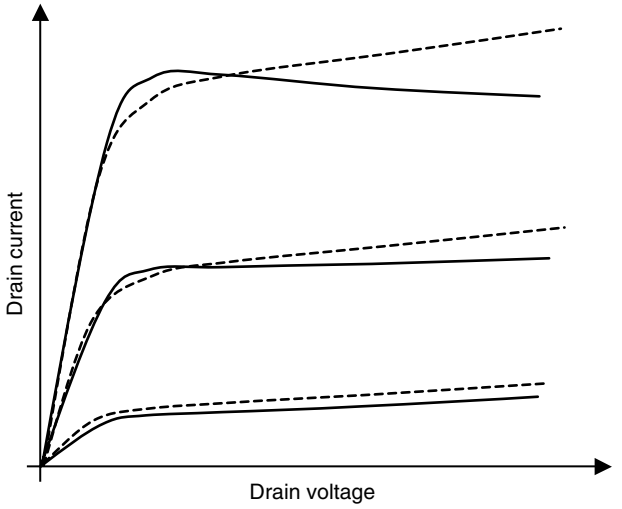


FIGURE 1.13 Pulsed I–V characteristics of a microwave FET. Solid lines are DC characteristics while dashed lines are pulsed.

high-frequency transistors. At RF and microwave frequencies, mechanisms known as *dispersion effects* become important to transistor operation. These effects reveal themselves as a difference in I–V characteristics obtained using a slow sweep as opposed to I–V characteristics obtained using a rapid pulse. The primary physical causes of I–V dispersion are thermal effects and carrier traps in the semiconductor. Figure 1.13 illustrates the characteristics of a microwave transistor under DC (solid lines) and pulsed (dashed lines) stimulation. In order to characterize dispersion effects, pulse rates must be shorter than the thermal and trapping time constants that are being monitored. Typically, for microwave transistors, that requires a pulse on the order of 100 ns or less. Similarly, the quiescent period between pulses should be long compared to the measured effects. Typical quiescent periods are on the order of 100 ms or more. The discrepancy between DC and pulsed characteristics is an indication of how severely the semiconductor traps and thermal effects will impact device performance.

Another use for pulsed I–V measurement is the characterization of high power transistors. Many high power transistors (greater than a few dozen Watts) are only operated in a pulsed mode or at a bias level far below their maximum currents. If these devices are biased at higher current levels for a few milliseconds, the thermal dissipation through the transistor will cause catastrophic failure. This is a problem for transistor model development, since a large range of I–V curves—including high current settings—is needed to extract an accurate model. Pulsed I–V data can provide input for model development while avoiding unnecessary stress on the part being characterized.

1.5 Circuits and Circuit Technologies

Figure 1.14 illustrates a generalized radio architecture that is typical of the systems used in many wireless applications today. The generalized diagram can apply to either communications or radar applications. In a wired application, the antenna of Figure 1.14 can be replaced with a transmission line. The duplexer of Figure 1.14 will route signals at the transmission frequency from the PA to the antenna while isolating that signal from the low noise amplifier (LNA). It will also route signals at the receive frequency from the antenna to the LNA. For some systems, input and output signals are separated in time instead of frequency. In these systems, an RF switch is used instead of a duplexer. Matching elements and other passive frequency selective circuit elements are used internally to all of the components shown in the figure. In addition, radio specifications typically require the use of filters at the ports of some of the components illustrated in Figure 1.14.

A signal received by the antenna is routed via the duplexer to the receive path of the radio. An LNA amplifies the signal before a mixer downconverts it to a lower frequency. The downconversion is accomplished by mixing the received signal with an internally generated local oscillator (LO) signal. The ideal receiver rejects all unwanted noise and signals. It adds no noise or interference and converts the signal to a lower frequency that can be efficiently processed without adding distortion.

On the transmitter side, a modulated signal is first upconverted and then amplified by the PA before being routed to the antenna. The ideal transmitter boosts the power and frequency of a modulated signal to that required for the radio to achieve communication with the desired receiver. Ideally, this

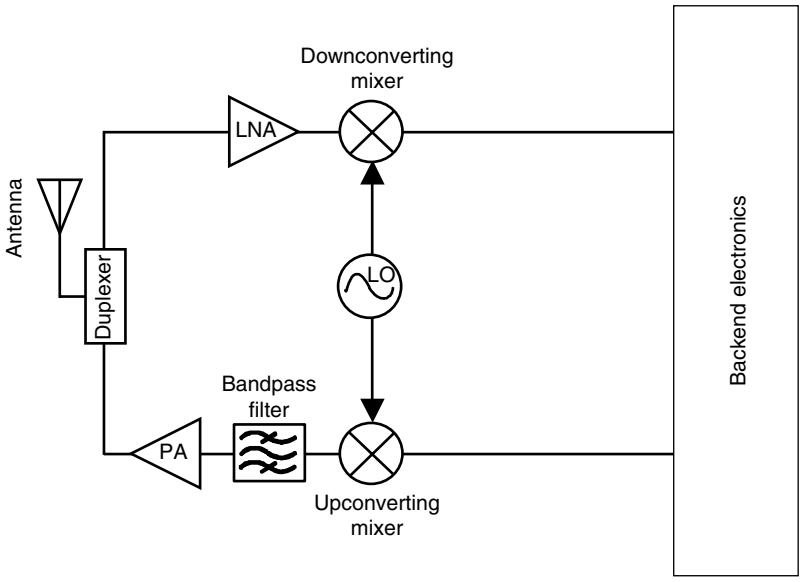


FIGURE 1.14 Generalized microwave radio architecture illustrating the microwave components in both the receiver and transmitter path.

process is accomplished efficiently (minimum DC power requirements) and without distortion. It is especially important that the signal broadcast from the antenna include no undesirable frequency components.

To accomplish the required transmitter and receiver functions, RF and microwave components must be developed either individually or as part of an integrated circuit. The remainder of this section will examine issues related to individual components that comprise the radio.

1.5.1 Low Noise Amplifier

The LNA is often most critical in determining the overall performance of the receiver chain of a wireless radio. The noise figure of the LNA has the greatest impact of any component on the overall receiver noise figure and receiver sensitivity. The LNA should minimize the system noise figure, provide sufficient gain, minimize nonlinearities, and assure stable 50 Ω impedance with low power consumption. The two performance specifications of primary importance to determine LNA quality are gain and noise figure.

In many radios, the LNA is part of a single chip design that includes a mixer and other receiver functions as well as the LNA. In these applications, the LNA may be realized using Silicon, SiGe, GaAs or another semiconductor technology. Si BJTs and SiGe HBTs dominate the LNA business at frequencies below a couple of GHz because of their tremendous cost and integration advantages over compound semiconductor devices. Compound semiconductors are favored as frequency increases and noise figure requirements decrease. For applications that require extremely low noise figures, cooled compound semiconductor HEMTs are the favored device.

1.5.2 Power Amplifier

A PA is required at the output of a transmitter to boost the signal to the power levels necessary for the radio to achieve a successful link with the desired receiver. PA components are almost always the most difficult and expensive part of microwave radio design. At high power levels, semiconductor nonlinearities such as breakdown voltage become critical design concerns. Thermal management issues related to dissipating heat from the RF transistor can dominate the design effort. Efficiency of the amplifier is critical, especially in the case of portable radio products. PA efficiency is essential to obtain long battery lifetime in a portable product. Critical primary design specifications for PAs include output power, gain, linearity, and efficiency.

For many applications, PA components tend to be discrete devices with minimal levels of on-chip semiconductor integration. The unique semiconductor and thermal requirements of PAs dictate the use of unique fabrication and manufacturing techniques in order to obtain required performance. The power and frequency requirements of the application typically dictate what device technology is required for the PA. At frequencies as low as 800 MHz and power levels of 1 Watt, compound semiconductor devices often compete with Silicon and SiGe for PA devices. As power and frequency increase from these levels, compound semiconductor HBTs and HEMTs dominate in this application. Vacuum tube technology is still required to achieve performance for some extremely high-power or high-frequency applications.

1.5.3 Mixer

A mixer is essentially a multiplier and can be realized with any nonlinear device. If at least two signals are present in a nonlinear device, their products will be produced at the device output. The mixer is a frequency translating device. Its purpose is to translate the incoming signal at frequency, f_{RF}, to a different outgoing frequency, f_{IF}. The LO port of the mixer is an input port and is used to *pump* the RF signal and create the IF signal.

Mixer characterization normally includes the following parameters:

- Input match (at the RF port)
- Output match (at the IF port)

- LO to RF leakage (from the LO to RF port)
- LO to IF leakage (from LO to IF port)
- Conversion Loss (from the RF port to the IF port)

The first four parameters are single frequency measurements similar to *s*-parameters S_{11}, S_{22}, S_{13}, and S_{23}. Conversion loss is similar to *s*-parameter S_{21}, but is made at the RF frequency at the input port and at the IF frequency at the output port.

Although a mixer can be made from any nonlinear device, many RF/microwave mixers utilize one or more diodes as the nonlinear element. FET mixers are also used for some applications. As in the case of amplifiers, the frequency of the application has a strong influence on whether Silicon or compound semiconductor technologies are used.

1.5.4 RF Switch

RF switches are control elements required in many wireless applications. They are used to control and direct signals under stimulus from externally applied voltages or currents. Phones and other wireless communication devices utilize switches for duplexing and switching between frequency bands and modes.

Switches are ideally a linear device so they can be characterized with standard *s*-parameters. Since they are typically bi-directional, $S_{21} = S_{12}$. Insertion loss (S_{21}) and reflection coefficients (S_{11} and S_{22}) are the primary characteristics of concern in an RF switch. Switches can be reflective (high impedance in the off state) or absorptive (matched in both on-and off-state).

The two major classes of technologies used to implement switches are PIN diodes and FETs. PIN diode switches are often capable of providing superior RF performance to FET switches but the performance can come at a cost of power efficiency. PIN diodes require a constant DC bias current in the on state while FET switches draw current only during the switching operation. Another important emerging technology for microwave switching is the micro-electro-mechanical systems (MEMS) switch. These integrated circuit devices use mechanical movement of integrated features to open and close signal paths.

1.5.5 Filter

Filters are frequency selective components that are central to the operation of a radio. The airwaves include signals from virtually every part of the electromagnetic spectrum. These signals are broadcast using various modulation strategies from TV and radio stations, cell phones, base stations, wireless LANs, police radar, and so on. An antenna at the front end of a radio receives all these signals. In addition, many of the RF components in the radio are nonlinear, creating additional unwanted signals within the radio. In order to function properly, the radio hardware must be capable of selecting the specific signal of interest while suppressing all other unwanted signals. Filters are a critical part of this selectivity. An ideal filter would pass desired signals without attenuation while suppressing signals at all other frequencies to elimination.

Although Figure 1.14 shows only one filter in the microwave portion of the radio, filters are typically required at multiple points along both the transmit and receive signal paths. Further selectivity is often accomplished by the input or output matching circuitry of amplifier or mixer components.

Filter characteristics of interest include the bandwidth or passband frequencies, the insertion loss within the passband of the filter, as well as the signal suppression outside of the desired band. The quality factor, *Q*, of a filter is a measure of how sharply the performance characteristics transition between passband and out-of-band behavior.

At lower frequencies, filters are realized using lumped inductors and capacitors. Typical lumped components perform poorly at higher frequencies due to parasitic losses and stray capacitances. Special manufacturing techniques must be used to fabricate lumped inductors and capacitors for microwave applications. At frequencies above about 5 or 6 GHz, even specially manufactured lumped element components are often incapable of producing adequate performance. Instead, a variety of technologies are exploited to accomplish frequency selectivity. Open- and short-circuited transmission line segments are

often realized in stripline, microstrip, or coplanar waveguide forms to achieve filtering. Dielectric resonators, small pucks of high dielectric material, can be placed in proximity to transmission lines to achieve frequency selectivity. Surface acoustic wave (SAW) filters are realized by coupling the electromagnetic signal into piezoelectric materials and tapping the resulting surface waves with appropriately spaced contacts. Bulk acoustic wave (BAW) filters make use of acoustic waves flowing vertically through bulk piezoelectric material. MEMS are integrated circuit devices that combine both electrical and mechanical components to achieve both frequency selectivity and switching.

1.5.6 Oscillator

Oscillators deliver power either within a narrow bandwidth, or over a frequency range (i.e., they are tunable). Fixed oscillators are used for everything from narrowband power sources to precision clocks. Tunable oscillators are used as swept sources for testing, FM sources in communication systems, and the controlled oscillator in a phase-locked loop (PLL). Fixed tuned oscillators will have a power supply input and the oscillator output, while tunable sources will have one or more additional inputs to change the oscillator frequency. The output power level, frequency of output signal and power consumption are primary characteristics that define oscillator performance. The quality factor, Q, is an extremely important figure-of-merit for oscillator resonators. Frequency stability (jitter) and tunability can also be critical for many applications.

The performance characteristics of an oscillator depend on the active device and resonator technologies used to fabricate and manufacture the component. Resonator technology primarily affects the oscillator's cost, phase noise (jitter), vibration sensitivity, temperature sensitivity, and tuning speed. Device technology mainly affects the oscillator maximum operating frequency, output power, and phase noise (jitter).

Resonator choice is a compromise of stability, cost, and size. Generally the quality factor, Q, is proportional to volume, so cost and size tend to increase with Q. Technologies such as quartz, SAW, yttrium-iron-garnet (YIG) and dielectric resonators allow great reductions in size while achieving high Q by using acoustic, magnetic, and dielectric materials, respectively. Most materials change size with temperature, so temperature stable cavities have to be made of special materials. Quartz resonators are an extremely mature technology with excellent Q, temperature stability, and low cost. Most precision microwave sources use a quartz crystal to control a high-frequency tunable oscillator via a PLL. Oscillator noise power and jitter are inversely proportional to Q^2, making high resonator Q the most direct way to achieve a low noise oscillator.

Silicon bipolar transistors are used in most low noise oscillators below about 5 GHz. Heterojunction bipolar transistors (HBTs) are common today and extend the bipolar range to as high as 100 GHz. These devices exhibit high gain and superior phase noise characteristics over most other semiconductor devices. For oscillator applications, CMOS transistors are poor performers relative to bipolar transistors, but offer levels of integration that are superior to any other device technology. Above several GHz, compound semiconductor MESFETs and HEMTs become attractive for integrated circuit applications. Unfortunately, these devices tend to exhibit high phase noise characteristics when used to fabricate oscillators. Transit time diodes are used at the highest frequencies where a solid-state device can be used. IMPATT and Gunn diodes are the most common types of transit-time diodes available. The IMPATT diode produces power at frequencies approaching 400 GHz, but the avalanche breakdown mechanism inherent to its operation causes the device to be very noisy. In contrast, Gunn diodes tend to exhibit very clean signals at frequencies as high as 100 GHz.

1.6 CAD, Simulation, and Modeling

The unique requirements of RF and microwave engineering establish a need for design and analysis tools that are distinct from conventional electrical engineering tools. Simulation tools that work well for a

digital circuit or computational system designer fail to describe RF and microwave behavior adequately. Component and device models must include detailed descriptions of subtle parastic effects not required for digital and low-frequency design. Circuit CAD tools must include a much wider range of components than for traditional electrical circuit design. Transmission line segments, wires, wire bonds, connector transitions, specialized ferrite, and acoustic wave components are all unique to microwave circuit design. In addition, the impact of the particular package technology utilized as well as layout effects must be considered by the microwave engineer. Electromagnetic simulators are also often required to develop models for component transitions, package parasitics, and complex board layouts.

In the microwave design environment, a passive component model for a single chip resistor requires a model that uses several ideal circuit elements. Shunt capacitances are required to model parasitic displacement currents at the input and output of the component. Ideal capacitors, inductors, or transmission line segments must be used internal to the desired component to model phase-shift effects. Nonideal loss mechanisms require the inclusion of additional resistor elements. Similar complexities are required for chip capacitors and inductors. The complexity is compounded by the fact that the ideal element values required to model a single component will change depending on how that component is connected to the circuit board and the kind of circuit board used.

Device models required for microwave and RF design are significantly more detailed than those required for digital or low-frequency applications. In digital design, designers are concerned primarily with two voltage–current states and the overall transition time between those states. A device model that approximately predicts the final states and timing is adequate for many of these applications. For analog applications, however, a device model must describe not only the precise I–V behavior of the device over all possible transitions, but also accurately describe second and third derivatives of those transitions for the model to be able to predict second-and third-order harmonics. Similar accuracy requirements also apply to the capacitance-voltage characteristics of the device. In the case of PA design, the device model must also accurately describe gate leakage and breakdown voltage—effects that are considered second order for many other types of circuit design. Development of a device model for a single microwave transistor can require weeks or months of detailed measurement and analysis.

Another characteristic that distinguishes RF microwave design is the significant use of electromagnetic simulation. A microwave circuit simulation that is completely accurate for a set of chips or components in one package may fail completely if the same circuitry is placed in another package. The transmission line-to-package transitions and proximity effects of package walls and lid make these effects difficult to determine and model. Often package effects can only be modeled adequately through the use of multi-dimensional electromagnetic simulations. Multilevel circuit board design also requires use of such simulations. Radiation effects can be important contributors of observed circuit behavior but cannot be captured without the use of electromagnetic simulation.

Because of the detail and complexity required to perform adequate RF and microwave design, the procedure used to develop such circuits and systems is usually iterative. Simple ideal-element models and crude models are first used to determine preferred topologies and approximate the final design. Nonideal parasitics are then included and the design is reoptimized. Electromagnetic simulators are exercised to determine characteristics of important transitions and package effects. These effects are then modeled and included in the simulation. For some circuits, thermal management can become a dominant concern. Modeling this behavior requires more characterization and simulation complexity. Even after all of this characterization and modeling has taken place, most RF and microwave circuit design efforts require multiple passes to achieve success.

References

1. Collin, R. E., *Foundations for Microwave Engineering*, McGraw-Hill, New York, 1992, p. 2.
2. Adam, S. F., *Microwave Theory and Applications*, Prentice Hall, Englewood Cliffs, NJ, 1969.
3. Halliday, D., Resnick, R. and Walker, J., *Fundamentals of Physics*, 7th ed., John Wiley & Sons, Hoboken, NJ, 2005.

4. Schroder, D. K., *Semiconductor Material and Device Characterization*, John Wiley & Sons, Hoboken, NJ, 1990.
5. Plonus, M. A., *Applied Electromagnetics*, McGraw-Hill, New York 1978.
6. Jonscher, A. K., *Dielectric Relaxation in Solids*, Chelsea Dielectric Press, London, 1983.
7. Kittel, C., *Introduction to Solid State Physics*, 3rd ed. John Wiley & Sons, Hoboken, NJ, 1967.
8. Tsui, J. B., *Microwave Receivers and Related Components*, Peninsula Publishing, Los Altos Hills, CA, 1985.
9. Pearce, C. W. and Sze, S. M., *Physics of Semiconductors*, John-Wiley & Sons, New York, 1981.
10. Bahl, I. and Bhartia P., *Microwave Solid State Circuit Design*, John Wiley & Sons, New York, 1988.
11. Lee, W. C. Y., *Mobile Communications Engineering*, McGraw-Hill Book Company, New York, 1982.
12. Jakes, W. C., Jr. (ed.), *Microwave Mobile Communications*, John Wiley & Sons, Inc., New York, 1974.
13. Evans, J., *Network Interoperability Meets Multimedia*, Satellite Communications, pp. 30–36, February 2000.
14. Laverghetta, T. S., *Modern Microwave Measurements and Techniques*, Artech House, Dedham, MA, 1989.
15. Ramo, S., Winnery, J. R., and Van Duzer, T., *Fields and Waves in Communications Electronics*, 2nd ed., John Wiley & Sons, New York, 1988.
16. Matthaei, G. L., Young, L., and Jones, E. M. T., *Microwave Filters, Impedance-Matching Networks, and Coupling Structures*, Artech House, Dedham, MA, 1980.
17. Ambrozy, A., *Electronic Noise*, McGraw-Hill, New York, 1982.
18. Watson, H. A., *Microwave Semiconductor Devices and their Circuit Applications*, McGraw-Hill, New York, 1969.
19. Vendelin, G., Pavio, A. M., and Rohde, U. L., *Microwave Circuit Design*, John Wiley & Sons, Hoboken, NJ, 1990.

I

Microwave Measurements

2

Linear Measurements

Ronald E. Ham
MW and RF Consulting

Microwave and radio frequency (RF) measurements can be classified in two distinct but sometimes overlapping categories: signal measurements and network measurements. Signal measurements include observation and determination of the absolute characteristics of waves and waveforms. This includes frequency and time characteristics that are intentionally imparted to a signal such as modulation by information or unintentional signal perturbations such as phase and amplitude noise. Network measurement determines the relative terminal and signal transfer characteristics of devices and systems with any number of ports. Networks such as modulators and communication paths are sometimes characterized as linear time variant paths requiring both signal and network measurements for characterization. Useful parameters for both signals and networks can be obtained in the time, frequency, and modulation domains.

2.1 Signal Measurements

Signal measurements are made in one or more of three measurement planes as illustrated in Figure 2.1. Time domain quantifies amplitude as a function of time. Frequency domain is a measure of spectral energy as a function of frequency. Modulation domain yields the time dependence of frequency.

The most common measurement at low frequencies is in the time domain where the amplitude of a signal waveform is observed with respect to time. The instrument used for this is an oscilloscope: basically a voltmeter that either records or displays amplitude as time changes. At RF frequencies conventional direct measuring or direct digitizing instruments with greatly enhanced high-frequency circuitry are used; however, microwave frequency response requires the use of sampled data circuitry.

Observing the amplitude of the signal over a small fraction of the total bandwidth and moving the center frequency of the small relative bandwidth over the frequency range of the signal, the spectral components of the signal are separated and quantified. This measurement is normally made with a swept frequency spectrum analyzer.

Determining the frequency of a signal versus time is a modulation domain measurement. If the signal is sampled over a long period of time the stability can be obtained. Sampling over a very short time with respect to the period of the signal can yield phase variation information such as modulation or phase noise.

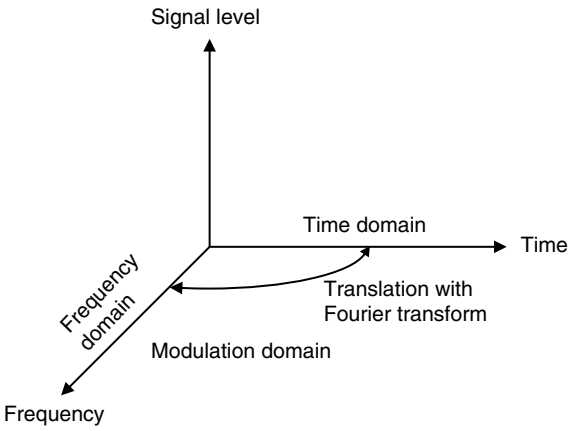


FIGURE 2.1 Three different types of measurements characterize signals.

Digital instrumentation is blurring the differentiation among instruments that have historically functioned for specific measurements. Rapid digitization of analog quantities with analog-to-digital converters that continue to improve in performance enables data to be manipulated in any measurement plane. Data obtained in the time domain can be moved to the frequency domain by mathematical manipulation with the Fourier transform. Likewise, time domain information may be derived from complex frequency domain data via the inverse Fourier transform. If the signal to be measured is too high in frequency, sampled data and frequency conversion techniques can often be implemented to move the spectrum of interest to ranges where digitally based instrumentation can be used.

2.1.1 Time Domain

A basic low-frequency oscilloscope records the amplitude of a signal with respect to time. An analog oscilloscope writes the amplitude of the signal on the y-axis of a display while advancing the time of the measurement on the x-axis of the display. If the amplitude is digitally sampled at discrete times the information can be stored, processed, and displayed as desired as in a digital storage oscilloscope.

Observation of RF and microwave signals with an analog or digital oscilloscope is limited by the speed of response of the front end instrument circuits and in the analog case by the display bandwidth. Building such instruments for operation beyond a gigahertz is difficult and expensive. For observing very high speed waveforms signal sampling techniques are incorporated.

A sampling oscilloscope, Figure 2.2, measures the value of a waveform at a particular time and stores or displays the data point. At a later time, usually a period of the high-frequency signal plus some small increased time, subsequent points are sampled. If the sampling can be performed fast enough the entire waveform shape can eventually be recreated from the sampled data.

The signal zero crossing in Figure 2.3a is used to trigger a fast ramp as shown in Figure 2.3b. The fast ramp is restarted at every signal trigger event. The fast ramp voltage is compared with the display unit slow sweep ramp. When the two are equal, the signal is sampled, held until the next sample, and displayed (Figure 2.3c). By taking one sample per period of the high frequency repetitive signal and delaying the relative sample point from period to period, the waveform is eventually captured. In this example, the display unit operates at one-tenth of the signal frequency.

The waveform that can be measured by the high-speed sampling technique must be recurring. This makes capturing a one time or low duty cycle occurrence such as some ultrawide bandwidth (UWB) microwave signals very difficult and, even if points in a gigahertz waveform can be captured, this does not mean that one cycle of the waveform can be captured without repetitive synchronous measurements.

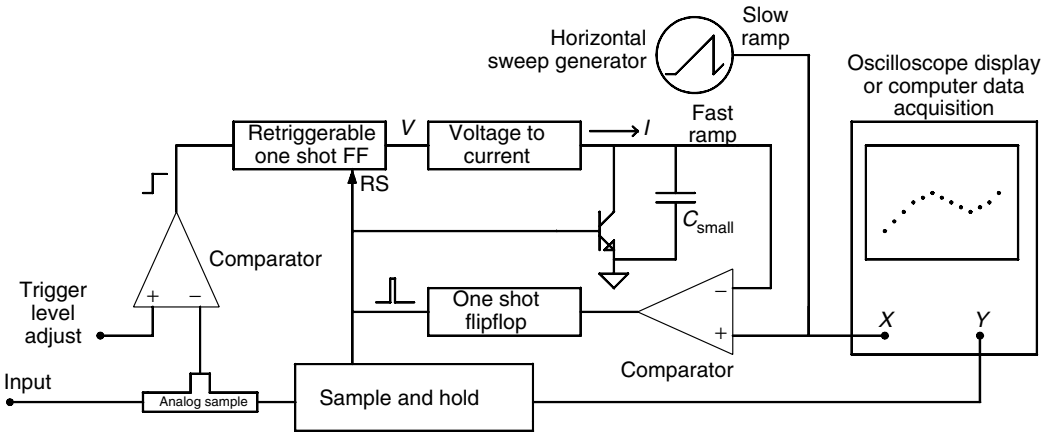


FIGURE 2.2 Simplified sampling oscilloscope.

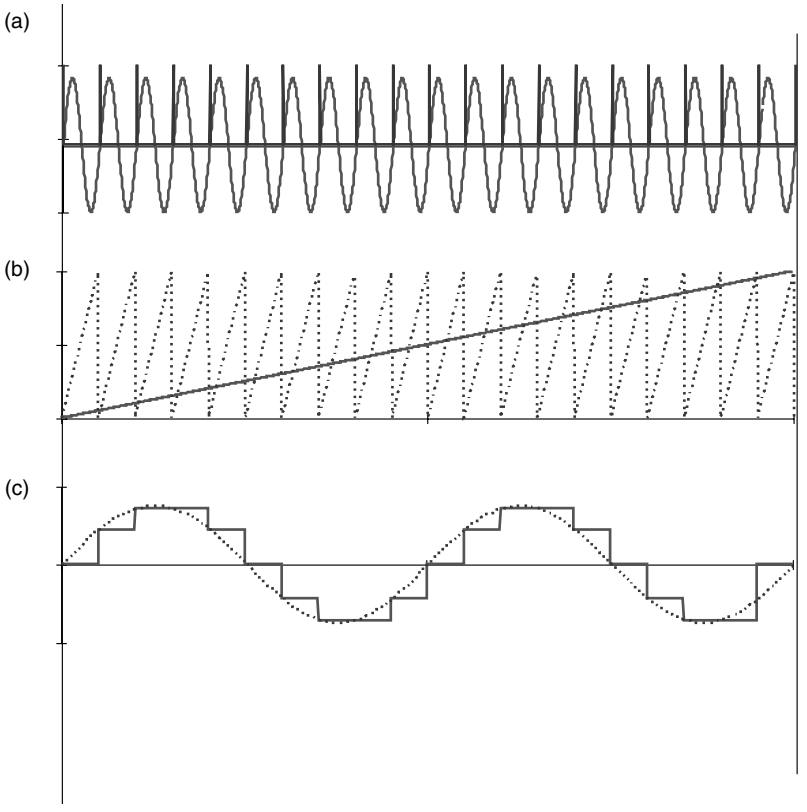
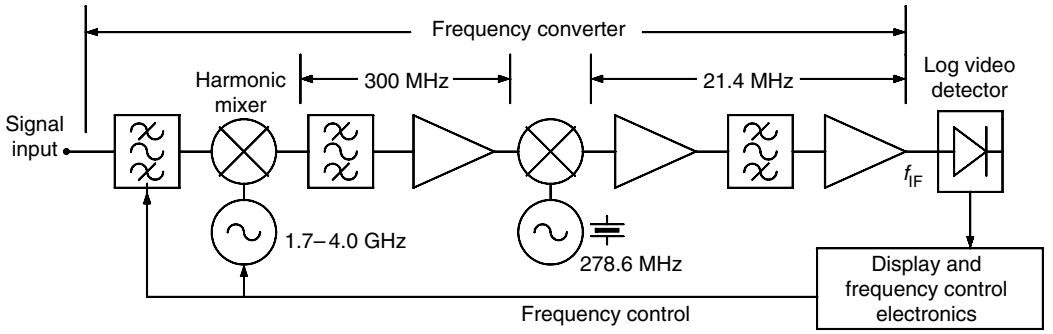


FIGURE 2.3 Primary waveforms in a sampling oscilloscope. (a) Input and zero crossing triggers, (b) fast display ramp and the slow sampling ramp, and (c) output waveform versus the input waveform.

If the oscilloscope display is replaced with a digital oscilloscope or digital storage oscilloscope then the instrument becomes a very versatile digital sampling oscilloscope (DSO). Various implementations of the DSO including multiple channel units enable very complex signal and network measurements well into the microwave spectrum.



Oscillator harmonic number(n)	Effective oscillator frequency (GHz)	Tuned RF frequency (GHz)
1	1.7–4.0	2.0–4.3
2	3.4–8.0	3.7–8.3
3	5.1–12.0	5.4–12.4
4	6.8–14.0	7.1–14.3
5	8.5–20.0	8.8–20.3

FIGURE 2.4 Simplified block diagram of a microwave spectrum analyzer.

2.1.2 Frequency Domain

A spectrum analyzer [1] is used to make frequency domain measurements of complex signals and signals with spectral characteristics that vary with time. This is basically a swept frequency filter with a detector to determine the signal amplitude within the bandwidth of the filter and some means of displaying or storing the measured information. To increase the selectivity and dynamic range of such a basic instrument, heterodyne conversions are used. Figure 2.4 is the block diagram of a typical multiple frequency conversion microwave spectrum analyzer.

The first intermediate frequency (IF) is chosen to permit a front end filter to eliminate the image from the first mixer. In this case, 300 MHz is chosen because the tunable filter, usually a YIG device, will have considerable attenuation at the image frequency 600 MHz away from the desired signal. The second IF is chosen because reasonably selective filters can be constructed to enable resolving signal components that are close to each other. In addition, detector and signal processing components, such as digital signal processors (DSPs) can be readily implemented at the lower frequency.

Because the normal frequency range required from a microwave spectrum analyzer is many octaves wide, multiple first conversion oscillators are required; however, this is an extremely expensive approach. Spectrum analyzers use a harmonic mixer for the first conversion and the first filter is tuned to eliminate the products that would be received due to the undesired harmonics of the conversion oscillator. Note the list of harmonic numbers (n) and the resulting tuned frequency of the example analyzer. As the harmonic number increases the sensitivity of the analyzer decreases because the harmonic mixer efficiency decreases with increasing n.

There is a significant trade-off between how much frequency resolution can be obtained and how fast a segment of the spectrum may be swept. Referring to Figure 2.5, as the narrowband filter determining the resolution bandwidth is effectively swept in frequency across signals, the filter must dwell on a given signal for enough time for the filter output to settle to a steady state value. The settling time is approximately the inverse of the bandwidth. If the spectrum to be analyzed is n times the filter bandwidth, the spectrum can not be accurately measured in less than n times the filter settling time.

The most important spectrum analyzer specifications are as follows:

1. Frequency tuning range—to include all of the frequency components of the signal to be measured.

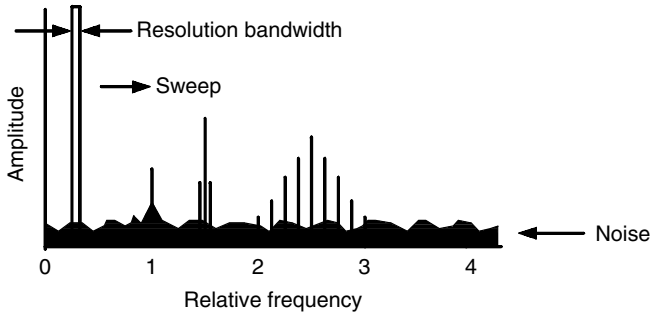


FIGURE 2.5 Swept frequency spectrum analyzer display.

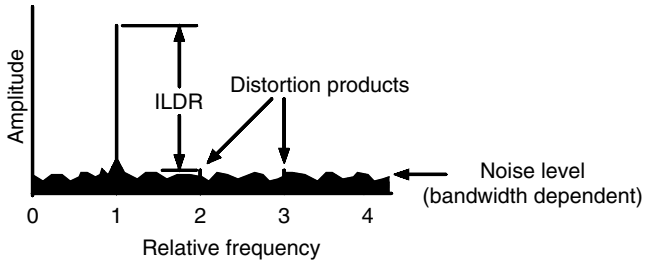


FIGURE 2.6 Instantaneous linear dynamic range.

2. Frequency accuracy and stability—to be more stable and accurate than the signal to be measured.
3. Sweep width—the band of frequencies over which the unit can sweep without readjustment.
4. Resolution bandwidth—narrow enough to resolve different spectral components of the signal.
5. Sensitivity and/or noise figure—to observe very small signals or small parts of large signals. Alternatively, SINAD (the ratio of signal + noise + distortion to noise + distortion) can be used to characterize sensitivity in an environment with distortion products due to multiple relatively high level signals.
6. Sweep rate—maximum sweep rate is established by the settling time of the filter that sets the resolution bandwidth.
7. Dynamic range—the difference between the largest and smallest signal the analyzer can measure without readjustment.
8. Instantaneous linear dynamic range—the difference between the largest signal the analyzer can measure and the measurement floor as defined by system noise and the nonlinear products of the large signals. (Referring Figure 2.6, this is a very important system specification for automated spectrum measurements.)
9. Phase noise—a signal with spectral purity greater than that of the analyzer conversion oscillators cannot be characterized.

Spectrum analyzers using other than swept frequency techniques are increasingly useful, especially for the measurement of complex signal modulation. For example, high speed sampling methods used with DSPs calculating the fast Fourier Transform (FFT) are readily implemented and are often referred to as vector signal analyzers (VSA); however, the speed of operation of the logic circuits limits the upper frequency of operation. This is a common method of IF demodulation and the useable frequency will move upward with semiconductor development.

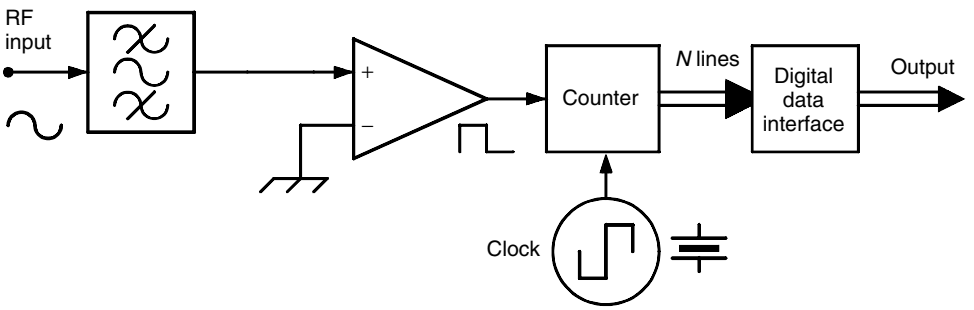


FIGURE 2.7 Frequency counter.

2.1.3 Modulation Domain

Modulation domain measurements yield the frequency of a signal as a function of time. Two examples of useful modulation domain data are the instantaneous frequency of a phase-locked oscillator as the loop settles and the pulse repetition rate of a fire control radar as it goes from search mode (low pulse repetition frequency or PRF) to lock-and-fire mode (high PRF). Longer duration modulation domain measurements include the frequency stability of an oscillator over an extended period of time.

The simplest modulation domain instrument is the frequency counter as illustrated in Figure 2.7. The clock provides a gating signal to the counter so that the zero crossings of the RF input signal can be accurately counted, yielding the signal frequency.

The number of measurements that must be made on a signal over a specified period of time is a function of the stability and the modulation placed on the signal. The exact measurement of the frequency of a stable and spectrally pure signal is normally made a few times per second. The frequency counter integrates the number of signal zero crossings of the repetitive waveform over the sampling period; hence, perturbations of phase, frequency, and amplitude are not observed other than as changes in sampling period data. By making very short period measurements relative to the frequency of the signal being measured, statistical manipulation of the sequential data can yield information about signal purity. This can be correlated to phase noise in a stable oscillator source.

Frequency prescalers and direct counting circuits are available well into the microwave frequency range. At high microwave frequencies, counters use conversion oscillators and mixers to heterodyne the signal down in frequency to where it can be directly counted. Microprocessor controllers and knowledge of the frequency of the conversion oscillators enable an exact signal frequency to be calculated.

A modulation domain analyzer [2] establishes the exact time at which a desired event occurs and catalogs the time. A simplified analyzer is shown in Figure 2.8. The event captured in a phase-locked oscillator measurement is the zero crossing of the oscillator output voltage. For a radar test it is the leading edge of each pulse. From this information the event frequency is calculated.

Various other modulation domain analyzers can be made with instantaneous frequency correlators and frequency discriminators. Figure 2.9 is a circuit widely used for measuring instantaneous frequency. If the delay line is of the order of a fraction of a wavelength of the measured signal, the circuit recovers broadband information such as frequency modulation. If the delay line is many wavelengths long the circuit is very sensitive to small variations in phase and is useful for narrowband measurements such as oscillator phase noise.

The demodulation and analysis of modern complex modulations often requires the use of specialized equipment specifically tailored to the modulation scheme; however, by rapid real-time digitization of a signal along with its quadrature component, all of the information is available to demodulate and analyze the signal. Figure 2.10 is the simplified block diagram for a generalized signal demodulator or VSA that can be used for both spectral analysis and modulation analysis. Basically, the unit is a "software radio" that can be programmed to output any presentation of data contained within the digitized IF stream over the period of time that can be processed in the DSP. For example, very fine resolution

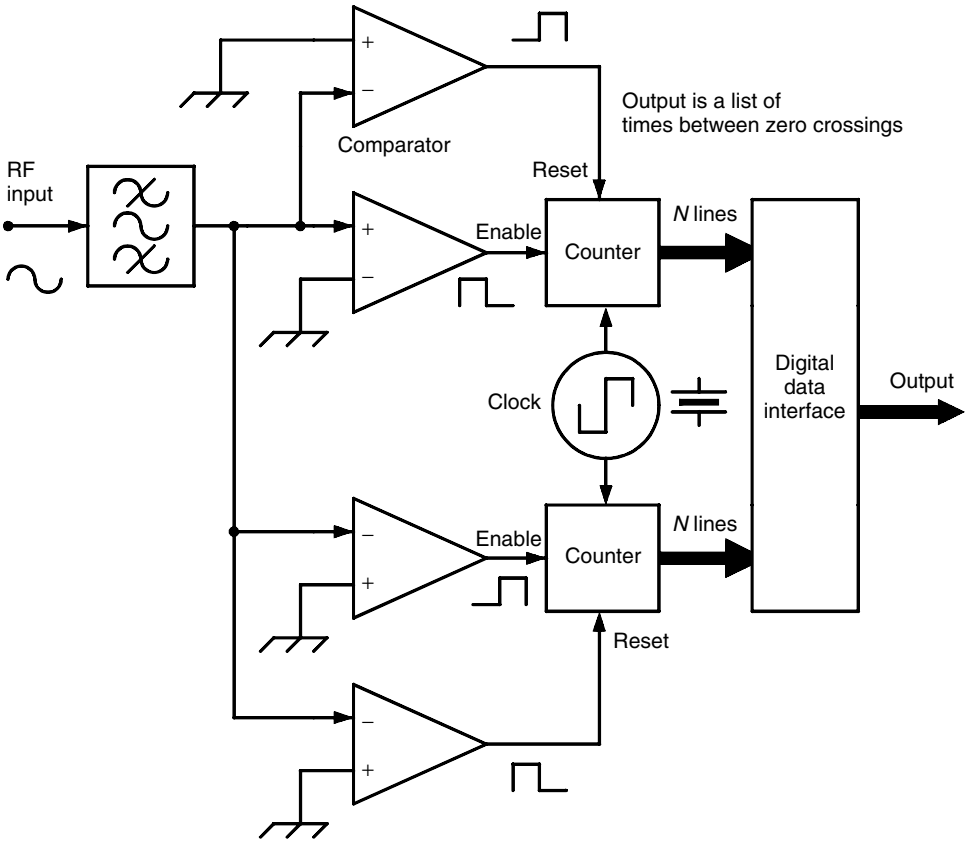


FIGURE 2.8 Modulation domain analyzer period measurements yield frequency as a function of time.

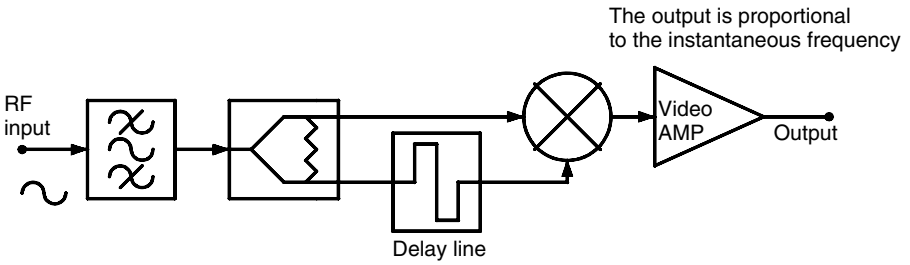


FIGURE 2.9 Instantaneous frequency and phase discriminator.

instantaneous frequency components or complex quadrature amplitude and phase modulation can be analyzed.

2.2 Network Measurements

Low-frequency circuit design and performance evaluation is based upon the measurement of voltages and currents. Knowing the impedance level at a point in a circuit to be the ratio of voltage to current, a voltage or current measurement can be used to calculate power. By measuring voltage and current as a complex quantity, yielding complex impedances, this method of circuit characterization can be used at relatively high frequencies even with the limitations of nontrivial values of circuit capacitive and inductive

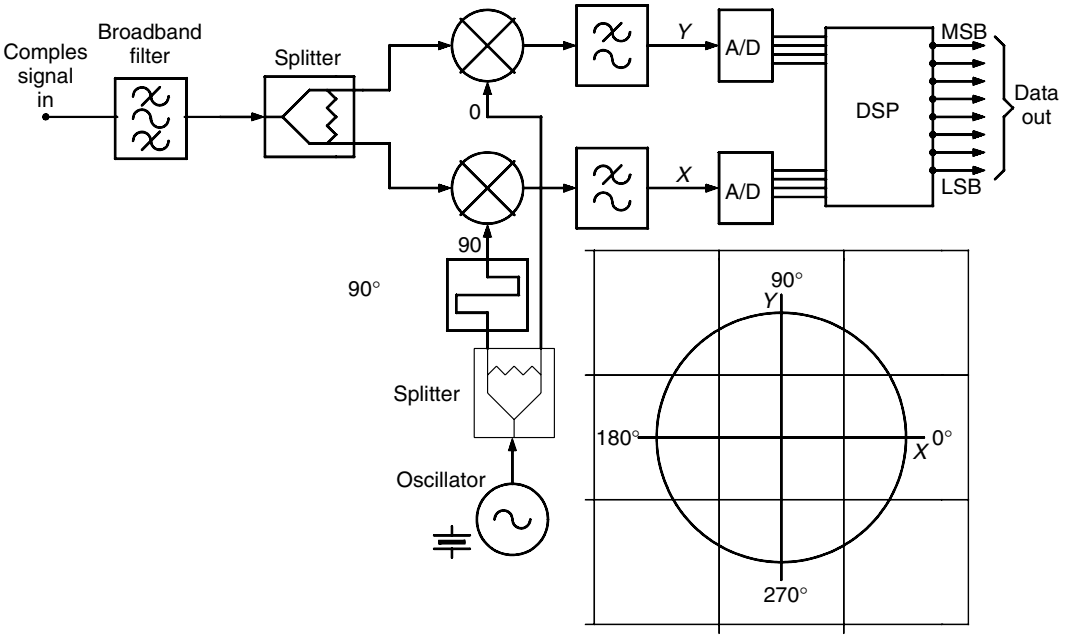


FIGURE 2.10 Modulation analyzer.

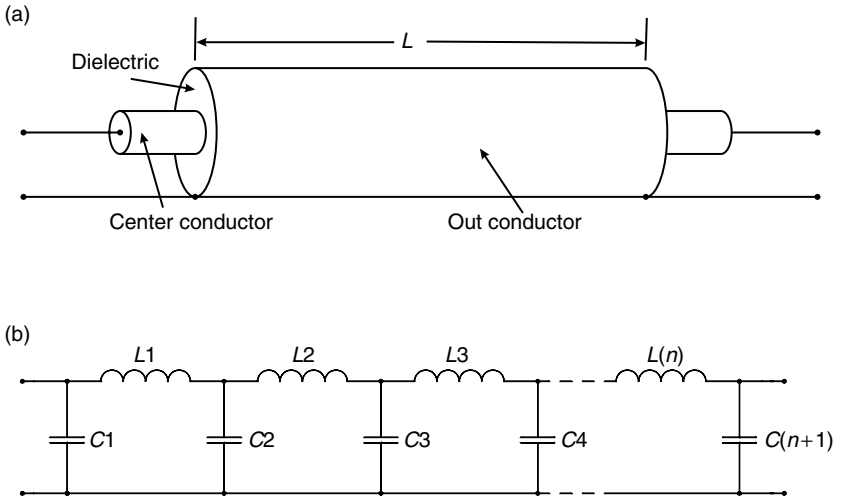


FIGURE 2.11 Examples of transmission lines. (a) Coaxial and (b) lumped element.

parasitics. When the parasitics can no longer be treated as lumped elements, distributed circuit concepts must be used.

A simple transmission line such as the coaxial line in Figure 2.11a can, if physically very small in all dimensions with respect to a wavelength, be modeled as a lumped-element circuit as shown in Figure 2.11b; however, as the size of the line increases relative to the wavelength, it becomes necessary to use an extremely complex lumped-element model or to use the transmission line equations for the distributed line. The concept of a transmission line accounts for the transformation of impedances between circuit points and for the time delay between points that must be considered when the circuit size approaches a significant

fraction of a wavelength of the frequency being measured; hence, RF and microwave measurements are primarily based upon transmission line concepts and measurements.

The basic quantities measured in high-frequency circuits are power, impedance, port-to-port transfer functions of *n*-port devices, frequency, and noise [3,4].

2.2.1 Power

Microwave power cannot be readily detected with equipment used at lower frequencies such as voltmeters and oscilloscopes [5]. The RF and microwave utility of these instruments is limited by circuit parasitics and the resultant limited frequency response. Central to all microwave measurements is the determination of the microwave power available at ports in the measurement circuit. To facilitate measurements a characteristic impedance or reference resistance is assumed. The instruments used to measure microwave and RF power typically have 50 Ω input and output impedance at the frequency being measured.

Diode detectors sense the amplitude of a signal by using the nonlinear voltage versus current characteristics of various types of diodes as illustrated in Figure 2.12. By establishing the input impedance of a diode detector, the power of a signal at a test port can be measured. The diode detector allows current to pass through the diode when the diode is forward biased and prevents current from flowing when the diode is reverse biased. The average of the current flow when forward biased results in a DC output from the lowpass RC filter that is proportional to the amplitude of the input voltage. Note that as the diode

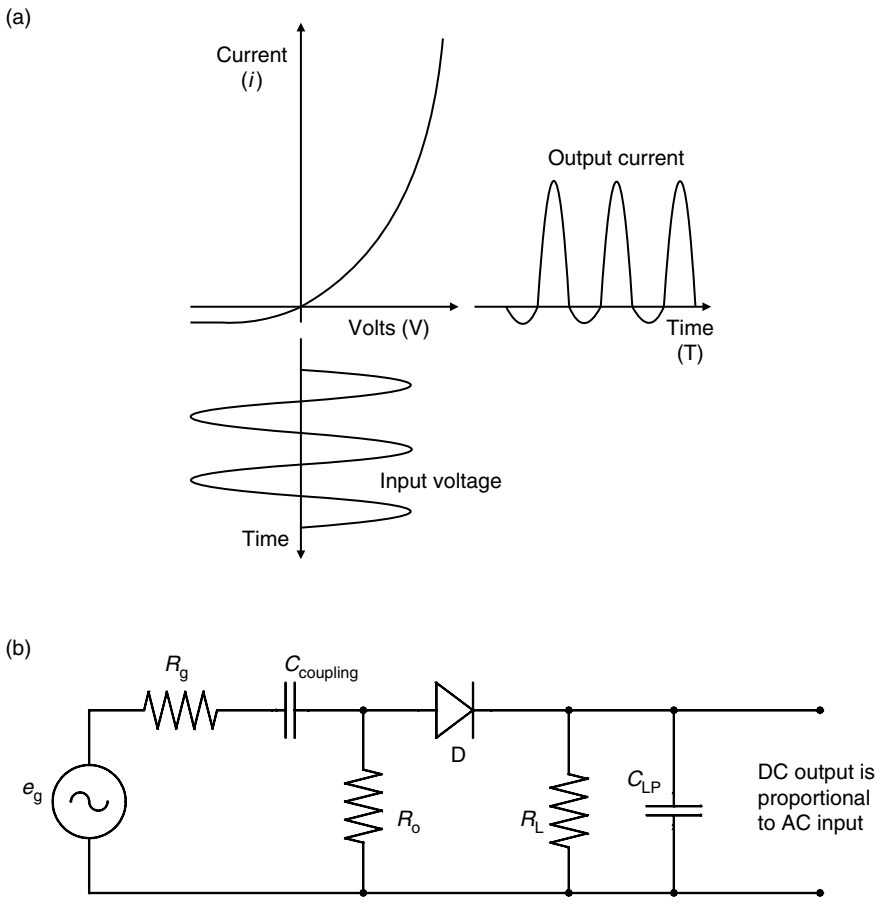


FIGURE 2.12 Diode detector. (a) Diode detector waveforms and (b) diode detector circuit.

junction area must be small to minimize the parasitic junction capacitance that would short the signal across the diode, the load resistor must be of a relatively large value to minimize the diode current; therefore, the impedance seen looking into the diode detector is established primarily by the resistor placed across the detector input. If the input voltage is less than that where the diode current becomes linearly proportional to the input voltage, the diode is in a predominantly square law region and the voltage out of the detector is proportional to the input power in decibels. This square law range typically extends over a 50 dB range from −60 to −10 dBm in a 50 Ω system. Diodes are used in the linear range up to about 10 dBm. The one significant disadvantage of the diode detector is the temperature sensitivity of the diode. The diode detector response can be very fast but it cannot easily be used for accurate power measurement.

The most accurate RF and microwave power measuring devices are thermally dependent detectors. These detectors absorb the power and by either measuring the change in the detector temperature or the change in the resistance of the detecting device with a change in temperature, the power absorbed by the detector can be accurately determined.

The primary thermally dependent detectors are the bolometer and the thermistor. They are placed across the transmission media as a matched impedance termination. A bridge as shown in Figure 2.13a can be used to detect a change in the resistance of the bolometer. To increase the detector sensitivity, two units can be placed in parallel for the RF/microwave signal and in series for the change in DC resistance as shown in Figure 2.13b. Unfortunately, this basic circuit can also be used as a thermometer, as the measured values change with the detector element temperature; therefore, an identical pair of bolometer detectors is normally placed in close thermal proximity but only one of the detectors is used to detect signal power. The other detector is used to detect environmental temperature changes so that the difference in temperature change is due to the signal power absorbed in the upper detector.

Figure 2.13c is a simplified example of a self-balancing bridge circuit that illustrates the operation of a temperature-dependent detector. To maintain a constant impedance looking into the bolometer elements, a bias current is passed through the elements to increase their temperature above operational ambient and to reduce the total series resistance of the bolometer head to 200 Ω (for a 50 Ω input impedance seen at the junction of B1 and B2). The resistance of the detectors is compared to a fixed resistance R_2 in the bridge. The bridge error as sensed by the difference amplifier is used to adjust V_{bridge} and hence the bias current in the bolometers to maintain the bolometer pair resistance at 200 Ω. The bias energy that must be removed from the detector to maintain a constant resistance is equal to the amount of signal energy absorbed by the detector. Microwave power into the bolometers is proportional to the reduction in current through R_3 relative to the current with no input signal. For example, if R_1 is equal to R_3 then the input power is the difference in the power dissipated in R_2 without an input signal and the power dissipated in R_2 with an input signal.

This example does not have temperature compensation; however, using an identical circuit for the second bolometer element pair in Figure 2.13b facilitates removal of temperature drift by measuring the difference in current between the two element pairs. The amount of microwave power into the bolometer head is then the square of the difference in the two currents times 200 Ω.

2.2.2 Impedance

Consider a very simple transmission line, two parallel pieces of wire spaced at a uniform distance and in free space, as shown in Figure 2.14. A DC voltage with a source resistance R_g and a series switch is connected to one end of the transmission line and a resistor R_L is placed across the other end.

First, let the length of the wires be zero. Close the switch. If the load resistor R_L is equal to the source resistance R_g, the condition necessary for maximum power transfer from a source to a load, then the voltage across the load R_L is $e_g/2$. This is the voltage that will be measured from a signal generator when the output is terminated in its characteristic impedance, commonly called Z_o. The signal power from the signal generator, and also the maximum available power from the generator, is then e_g^2/R_g. If R_L is a short circuit the output voltage is zero. If R_L is an open circuit the output voltage is two times $e_g/2$ or e_g.

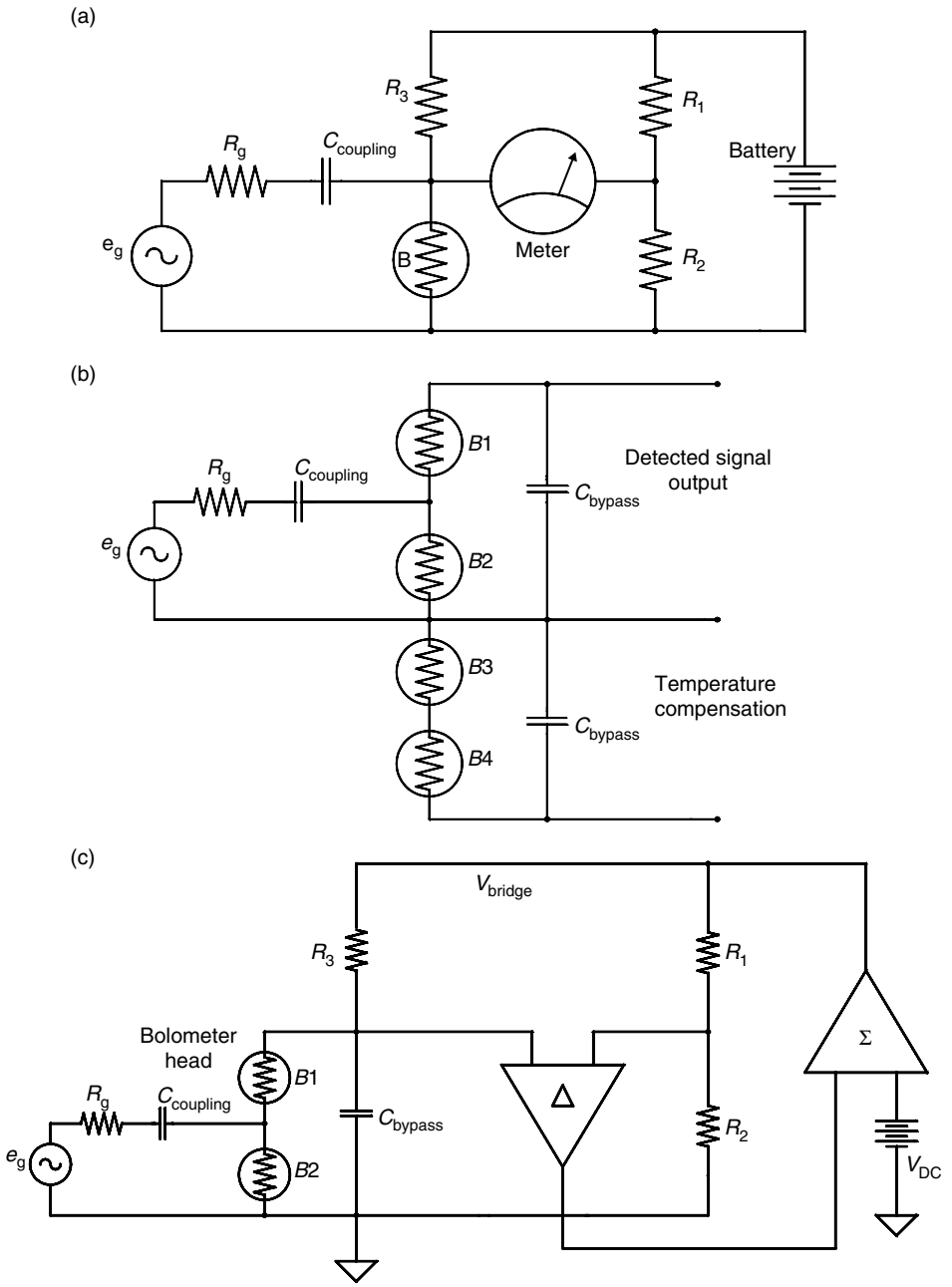


FIGURE 2.13 Thermally dependent detector circuits. (a) Bolometer in a bridge circuit, (b) temperature-compensated bolometer head, and (c) self-balancing bridge circuit.

Referencing the plots of voltage along the transmission line, let the line have a length, L. When the switch is closed, a traveling wave of voltage moves toward the load resistor at the speed of light, c. At time t, the wave has moved down the line a distance ct. A wave of current travels with the wave of voltage. If the characteristic impedance of this parallel transmission line is Z_o and the load resistance is equal to Z_o, then the current traveling with the voltage wave has a value at any point along the line of the value of the voltage at that point divided by Z_o. For this special case, when the wave reaches the load resistor all of

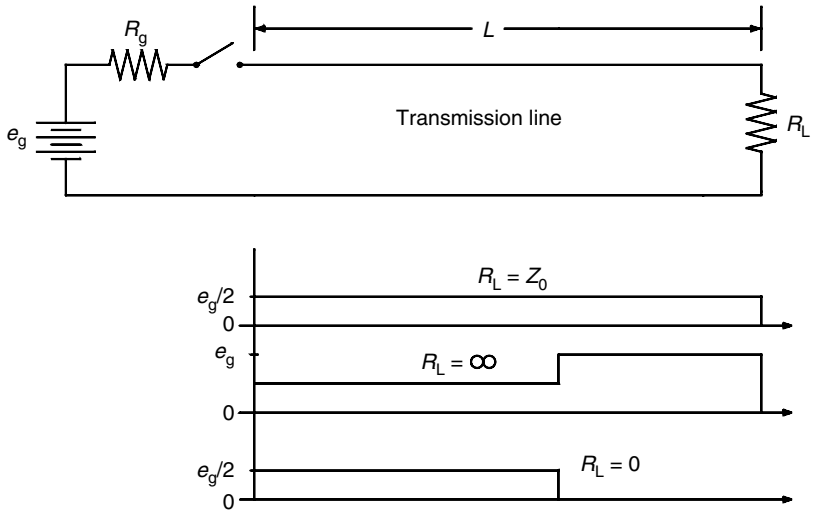


FIGURE 2.14 Switched DC line voltage at time > length/velocity for various impedances at the end of the line.

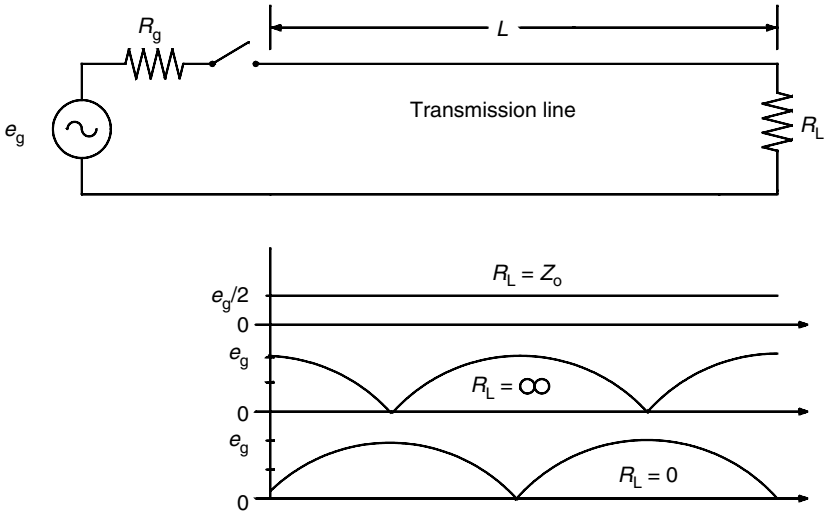


FIGURE 2.15 Waveforms on the line for a sinusoidal source and various impedances at the end of the line.

the energy in the wave is dissipated in the resistor; however, if the resistor is not equal to Z_o there is energy in the wave that must go someplace as it is not dissipated in the load resistor.

This mismatch between the characteristic impedance of the line and the terminating load resistor results in a reflected wave that travels back towards the voltage source. If the load resistor is a short circuit, the voltage at the end of the line must equal zero at all times. The only way for this to occur is for the reflected voltage at the end of the wire to be equal to -1 times the incident voltage at that same point. If the load is an open circuit, the reflected voltage will be exactly equal to the incident voltage; hence the sum of the incident and reflected voltages will be twice the value of the incident voltage at the end of the line. Note the similarity of these three cases to those of the zero length line.

Now replace the DC voltage source and switch with a sinusoidal voltage source as in Figure 2.15. The voltages shown are the RMS values of the vector sum of the incident and reflected waves. As the source voltage varies, the instantaneous value of the sinusoidal voltage between the wires travels down the wires.

The ratio of the traveling voltage wave to the traveling current wave is the characteristic impedance of the transmission line. If the terminating impedance is equal to the line characteristic impedance, there is no wave reflected back towards the generator; however, if the termination resistance is any value other than Z_o there is a reflected wave. If R_L is a real impedance and greater than Z_o the reflected wave is 180° out of phase with the incident wave. If R_L is a real impedance and is less than Z_o the reflected wave is in phase with the incident wave. The amount of variation of R_L from Z_o determines the magnitude of the reflected wave. If the termination is complex, the phase of the reflected wave is neither zero nor 180°.

Assuming the generator impedance R_s to be equal to the line characteristic impedance, so that a reflected wave incident on the generator does not cause another reflected wave, sampling the voltage at any point along the transmission line will yield the vector sum of the incident and reflected waves. With a matched impedance ($R_L = Z_o$) termination the magnitude of the AC voltage along the line is a constant. With a short circuit termination the voltage magnitude at the load will be zero and, moving back towards the generator, the voltage one-half of a wavelength from the end of the line will also be zero. With an open circuit there is a voltage maximum at the end of the line and a minimum on the line one-quarter of a wavelength back toward the generator.

The complex reflection coefficient Γ is the ratio of the reflected wave to the incident wave; hence it has a magnitude ρ between 0 and 1 and an angle θ between $+180°$ and $-180°$. The reflection coefficient as a function of the measured impedance Z_L with respect to the measurement system characteristic impedance Z_o is

$$\Gamma = \frac{Z_L - Z_o}{Z_L + Z_o} = \rho(\sin\theta + j\cos\theta)$$

2.2.2.1 Time-Domain Reflectometry

Switched voltage sources were used to illustrate the time and impedance characteristics of a transmission line in the previous section. In actual fact, referencing Figure 2.16, substituting a fast pulse or step generator for the switched voltage source and observing the sum of the forward and reflected waves at a sampling point close to the voltage waveform generator, yields a quantitative display of impedance and discontinuities along the signal path.

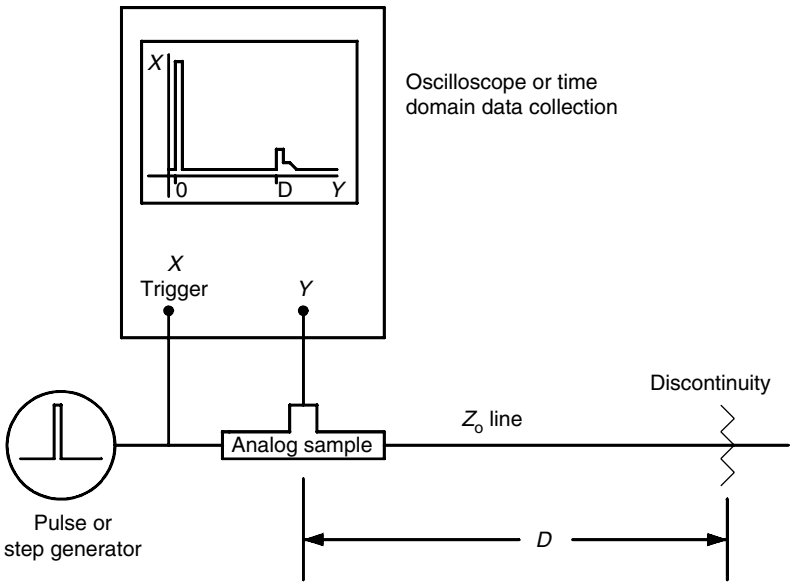


FIGURE 2.16 Time-domain reflectometer.

If the velocity of propagation of the transmission line is known, or is measured by comparing the electrical signal return time from a short circuit at the end of a known physical length of transmission line, then short circuits, open circuits, and any values in between can be observed and quantified in both position and value. If the discontinuities occur over a very short distance then a very fast test signal rise time must be used. To measure changes in impedance that affect gigahertz propagation (over lengths that are a significant fraction of a wavelength), these frequencies must exist in the Fourier transform of the time-domain waveform. That is, the probing signal waveform must have a rise time in the nanosecond range and the signal sampling instrument must be capable of functioning at high speed. Sampling oscilloscopes and multiple signal samples (repeated cycles from the pulse generator) are typically used in the microwave frequency range.

Time-domain reflectometry (TDR) is normally used only on passive circuits, as the voltages involved can be disruptive or even fatal to active circuitry. The frequency domain equivalent to TDR uses the inverse Fourier transform on a multioctave return loss measurement made with a vector network analyzer (VNA). While yielding virtually the same measurement information, but with increased sensitivity and dynamic range, the frequency-domain measurement probes the circuit under test with only a relatively small RF voltage; however, the frequency-domain equipment is much more complex and costly.

2.2.2.2 Slotted Line

Determination of the relative locations of the voltage and current minima and maxima along the line, or similarly the determination of the magnitude of waves traveling towards and away from the load impedance, is the basis for the measurement of RF and microwave impedance. The most basic instrument used for making this measurement is the slotted line. The slotted line is a transmission line with a slit in the side that enables a probe to be inserted into the transmission mode electromagnetic field as shown in Figure 2.17. A diode detector placed within the sliding probe provides a DC voltage that is proportional to the magnitude of the field in the slotted line. As the probe is moved along the line the minimum and maximum field positions and magnitudes can be determined. The ratio of the maximum field magnitude to the minimum field magnitude is the standing wave ratio (SWR). SWR is normally stated as a scalar quantity and is

$$\text{SWR} = \frac{1+\rho}{1-\rho}$$

Before placing an unknown impedance at the measurement terminal of the slotted line, the line is calibrated with a short circuit. This establishes a measurement plane at the short circuit. Any measurement made after calibrating with this reference short is made at the plane of the short circuit. A phase reference is located at the position on the slotted line of a minimum voltage measurement. The distance between two minimum voltage measurement locations is one-half of a wavelength at the measurement frequency.

If the short circuit is replaced with an open circuit the minimum voltage locations along the line are shifted by one-quarter of a wavelength. The difference between the phases of a reflected wave of an open

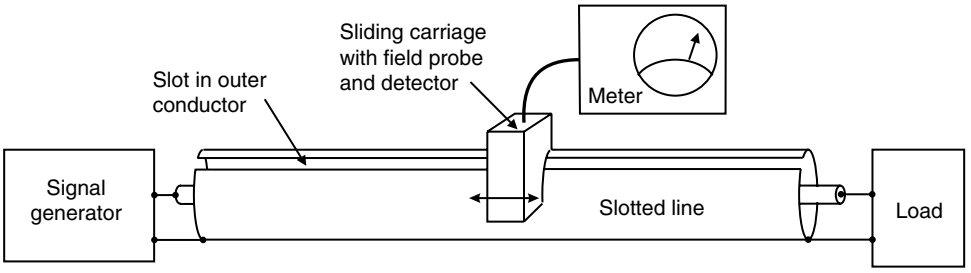


FIGURE 2.17 A slotted line is used to measure the impedance of an unknown load.

and a short circuit is 180°; hence, the distance between two minimum measurements represents 360° of phase shift in the reflected wave. Note that it is very difficult to use an open circuit for a reference at high frequencies because fringing and radiated fields at the end of the transmission line result in phase and amplitude errors in the reflected wave.

The impedance to be measured now replaces the calibrating short circuit. The new minimum voltage location is found by moving the detector carriage along the slotted line. The distance the minimum voltage measurement moves from the short circuit reference location is ratioed to 180° at a quarter of a wavelength shift. (e.g., a minimum shift of one-eighth wavelength results from a reflection coefficient phase shift of 90°). This is the phase difference between the forward and reflected waves on the transmission line. Either way the minimum moves from the short circuit calibrated reference point is a shift from 180° back toward 0°. If the shift is toward the load then the actual phase of the reflection coefficient is −180° plus the shift. If the shift is towards the generator from the reference point, the actual phase of the reflection coefficient is 180° minus the shift.

The best method of visualizing complex impedances as a function of the complex reflection coefficient is the Smith Chart [6–8]. A simplified Smith Chart is shown in Figure 2.18. The distance from the center of the chart towards the outside of the circle is the reflection coefficient ρ. The minimum value of ρ is 0 and the maximum value is 1. If there is no reflection the impedance is resistive and equal to the characteristic impedance of the transmission line or slotted line. If the reflected wave is equal to the incident wave the reflection coefficient is one and the impedance lies on the circumference of the circle. If the angle of the reflection coefficient is 0° or 180°, the impedance is real and lies along the central axis. Reflection coefficients with negative angles have capacitive components in the impedance and those with positive angles have inductive components.

2.2.2.3 Directional Coupler

Slotted lines must be on the order of a wavelength long. In addition, they do not lend themselves to computer controlled or automatic measurements. Another device for measuring the forward and reflected

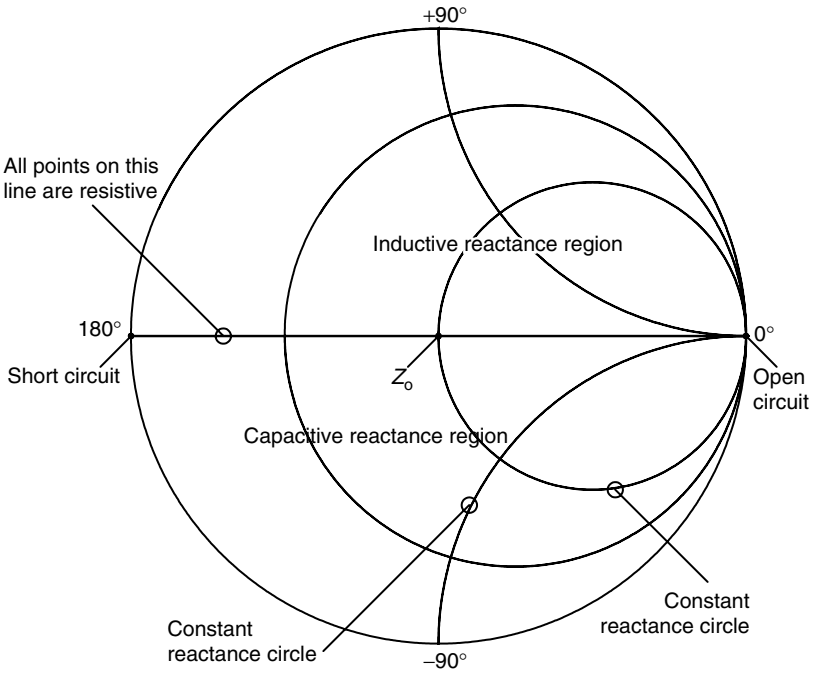


FIGURE 2.18 The Smith Chart is a plot of all nonnegative real impedances.

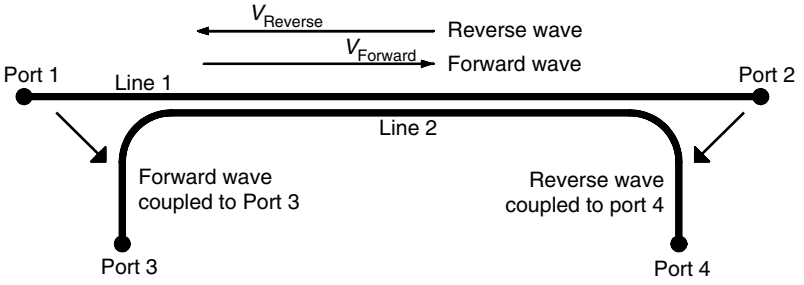


FIGURE 2.19 A directional coupler separates forward and reverse waves on a transmission line.

waves on a transmission line is the directional coupler [9]. Physically this is a pair of open transmission lines that are placed close enough for the fields generated by a propagating wave in one line to couple to the other line, hence inducing a proportional wave in the second line. The coupler is a four-port device. Referencing Figure 2.19, a wave propagating to the right in line one couples to line 2 and propagates to the left. A wave propagating in line 1 to the left couples to line 2 and propagates to the right; therefore, the outputs from ports 3 and 4 are proportional to the forward and reverse wave propagating in line 1.

The primary specifications for a coupler are its useful frequency range, the attenuation of the coupled wave to the coupled ports (coupling), and the attenuation of a signal traveling in the opposite direction to the desired signal at the desired signal's coupled port (directivity). For example, a 10 dB coupler with a 10 dBm signal propagating in the forward direction in line 1 will output a 1 dBm signal at port 3 and 9 dBm at port 2. If the directivity of the coupler is 30 dB there will also be a −40 dBm signal resulting from the forward wave at the reverse wave port, port 4. If the forward wave is properly terminated with the system impedance there will be no reverse wave on line 1; hence, there will not be an output at port 4 due to a reverse wave.

Note that power must be conserved through the coupler. Therefore if in the example above 1 dBm is coupled from the forward signal in line 1 to port 3, there will be only a 9 dBm output from port 2. This power must be taken into account in the measurement. The greater the attenuation to the coupled ports, the less the correction will be. Normally 20 or 30 dB couplers are used so the correction is minimal and, in many cases, small enough to be ignored.

By measuring the power from the forward and reverse coupled ports the magnitude of the reflection coefficient and the SWR can be calculated. Typically, the most common indication of the quality of the power match of a device being measured is the attenuation of the reflected wave or the return loss (RL) of the incident wave. This is

$$\text{RL (dB)} = 10 * \log_{10}\left(\frac{P_{\text{Forward}}}{P_{\text{Reverse}}}\right)$$

As power is proportional to voltage squared, when the termination resistance is equal on all ports, the RL can also be expressed as a voltage ratio.

$$\text{RL (dB)} = 10 * \log_{10}\left(\frac{V^2_{\text{Forward}}/Z_o}{V^2_{\text{Reverse}}/Z_o}\right) = -20 * \log_{10}\left(\frac{V_{\text{Reverse}}}{V_{\text{Forward}}}\right) = -20 * \log_{10}(\rho)$$

Hence the RL is the magnitude of the reflection coefficient ρ in decibels.

2.2.2.4 Resistive Bridge

The directional coupler is functionally equivalent to a bridge circuit, the primary difference being that the only losses in the transmission line coupler are from parasitics and can be designed to be very small.

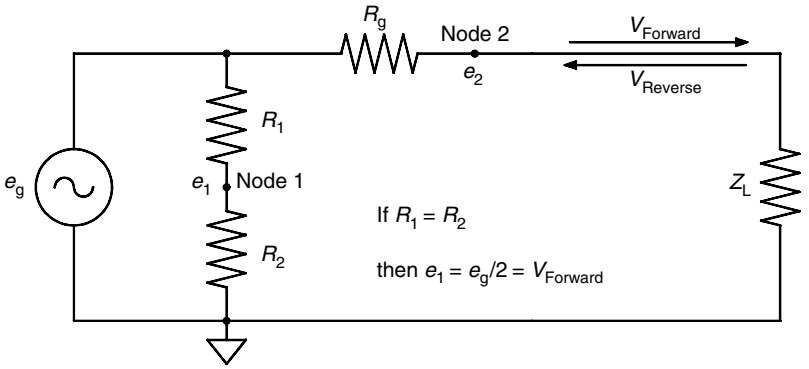


FIGURE 2.20 A resistive bridge can be used to measure the reverse wave on a transmission line.

Referring to Figure 2.20, the voltage drop across R_g when Z_L equals R_g is $e_g/2$ or one-half of the open circuit generator voltage. As the reflected voltage from a matched load is zero, the forward wave $V_{Forward}$ is one-half of the open circuit generator voltage, $e_g/2$. The voltage e_2 is the sum of the forward and reflected voltage so the reflection coefficient is

$$\Gamma = \frac{V_{Reverse}}{V_{Forward}} = \frac{e_2 - (e_g/2)}{e_g/2} = \frac{2e_2}{e_g} - 1$$

By placing a series circuit of two equal resistors across e_g, node 1 has a voltage of $e_g/2$. The voltage between node 1 and node 2 then is equal to the reflected wave. This is the standard resistive bridge circuit. Placing the balanced side of a broadband balanced to unbalanced transformer (a balun) between node 1 and node 2 and measuring the output of the unbalanced side of the balun yields the value of the reflected wave. The bridge is calibrated by using reference loads with known reflection coefficients. Note that if the complex voltage on both ends of R_g is measured (using sampling methods at higher frequencies) then the reflection coefficient Γ and impedance Z_L can be readily calculated. For this type measurement it is not necessary to know the source impedance of the signal generator as the measurement reference impedance is set by R_g.

2.2.3 Network Analyzers

General RF and microwave network analyzers measure ratios of forward, reflected, and incident waves where all measurements are relative to one of the wave variables, typically the incident signal from a signal source. These wave ratios define the scattering parameters (*s*-parameters) of the device being measured. A source with a well-defined impedance equal to the system impedance Z_o is used and all ports of the device under test (DUT) are terminated with the same impedance. The output port being measured is terminated in the test channel of the network analyzer that has an input impedance equal to the system characteristic impedance. Measurement of system parameters with all ports terminated minimizes the problems caused by short circuit, open circuit, and test circuit parasitics that cause considerable difficulty in the measurement of *Y*- and *h*-parameters at very high frequencies. *s*-Parameters can be converted to *Y*- and *h*-parameters.

Figure 2.21 illustrates a two-port DUT. If the generator is connected to port 1 and a matched load to port 2, the incident wave to the DUT is $V_{Forward}$ to port 1 or V_1^+. A wave reflected from the device back to port 1 is V_1^-. A signal traveling through the DUT and toward port 2 is V_2^-. Any reflection from the load (zero if it is truly a matched load) is $V_{Reverse}$ to port 2 or V_2^+. The *s*-parameters are defined in terms of

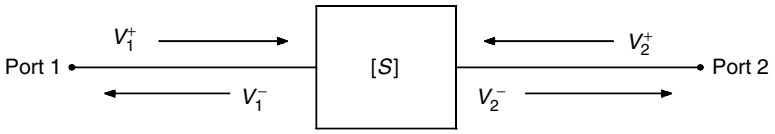


FIGURE 2.21 *s*-Parameters are defined by forward and reverse voltage waves.

these voltage waves:

$$s_{11} = \frac{V_1^-}{V_1^+} = \text{Input terminal reflection coefficient, } \Gamma_1$$

$$s_{12} = \frac{V_2^-}{V_1^+} = \text{Forward gain or loss}$$

By moving the signal generator to port 2 and terminating port 1, the other two-port *s*-parameters are measured:

$$s_{12} = \frac{V_1^-}{V_2^+} = \text{Reverse gain or loss}$$

$$s_{22} = \frac{V_2^-}{V_2^+} = \text{Output terminal reflection coefficient, } \Gamma_2$$

The *S*-matrix is then

$$[S] = \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix}$$

where

$$\begin{bmatrix} V_1^- \\ V_2^- \end{bmatrix} = [S] \begin{bmatrix} V_1^+ \\ V_2^+ \end{bmatrix}$$

2.2.3.1 Scalar Network Analyzer

A scalar network analyzer, Figure 2.22, with resistor-loaded diode probes or power meters is used to measure scalar RL and gain. Diode detectors, either used in the square law range as power detectors or logarithmic amplifiers, are used in the analyzer to produce nominally a 50 dB dynamic range of measurement. A spectrum analyzer with a test signal generator that tracks the center frequency of the spectrum analyzer can be used as a scalar analyzer with up to 90 dB of dynamic range.

Gains and losses are calculated in scalar analyzers by adding and subtracting relative power levels in decibels. Note that this can only establish the magnitude of the reflection coefficient so that an absolute impedance cannot be measured. To establish the impedance of a device the phase angle of the reflected wave relative to the incident wave must be known. To measure the phase difference between the forward and reflected wave a phase meter or VNA is used.

2.2.3.2 Vector Heterodyne Network Analyzer

Accurate direct measurement of the phase angle between two signals at RF and microwave frequencies is difficult; therefore, most vector impedance analyzers down convert the signals using a common local oscillator. By using a common oscillator the relative phase of the two signals is maintained. The signal is ultimately converted to a frequency where rapid and accurate comparison of the two signals yields their phase difference. In these analyzers the relative amplitude information is maintained so that the amplitude measurements are also made at the low IF.

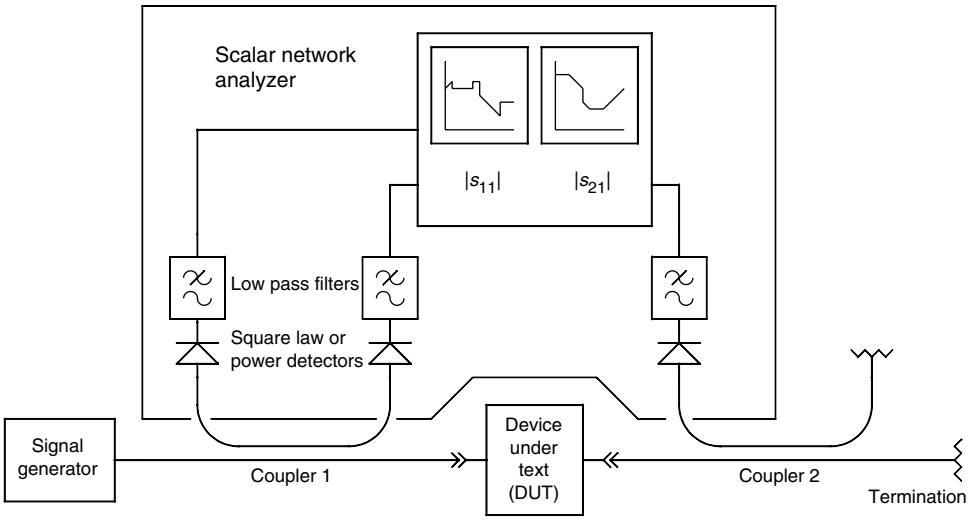


FIGURE 2.22 A scalar network analyzer can measure the magnitude of gain and return loss.

The VNA is a multi-channel phase coherent receiver with a tracking signal source. When interfaced with various power splitters and couplers the channels can measure forward, reverse, and transmitted waves. As the phase and amplitude information is available on each channel, parameters of the device being measured can be computed. The most common VNA configuration measures the forward and reflected waves to and from a two-port device. From these measurements the two-port scattering matrix can be computed.

The automatic VNA performs these operations under the supervision of a computer requiring the operator to input instructions relating to the desired data. The computer performs the routine "housekeeping."

The use of computers also facilitates extensive improvement in measurement accuracy by measuring known high-quality components, calculating nonideal characteristics of the measurement system, and applying corrections derived from these measurements to data from other devices. In other words, the accuracy of a known component can be transferred to the measurement accuracy of an unknown component.

With the measurement frequency accurately known and the phase and amplitude response measured and corrected, the inverse Fourier transform of the frequency domain yields the time-domain response. A very useful measurement of this type transforms the s_{11} frequency-domain data to a time-domain response with the same information as TDR; that is, deviations from the characteristic impedance can be seen over the length of the measured transmission media.

The simplified block diagram of a multichannel VNA using a two channel coherent receiver is shown in Figure 2.23. There are two channels fed from the test set. The inputs are converted first to a low IF such as 20 MHz and then to 100 kHz before being routed to phase detectors. A comb generator follows the first conversion oscillator and the oscillator is phase locked to the mixer output so the unit will frequency track the test source.

Multichannel software-defined radios (or coherent multichannel VSA) can also be used in the place of the heterodyne receiver. If four tracking channels are used there is no need for a switched test set as DUT input and output incident and reflected signals can be measured simultaneously. This also enables a VNA to be implemented to measure any number of ports.

Multiple methods of generating the conversion oscillator voltages are used. Low-RF frequency-analyzer signal generators commonly generate a test signal plus another output, that is, frequency offset by the desired IF frequency. This can be done with offset synthesizers or by mixing a common oscillator with a stable oscillator at the IF frequency and selecting the desired mixing product using phasing or filtering techniques.

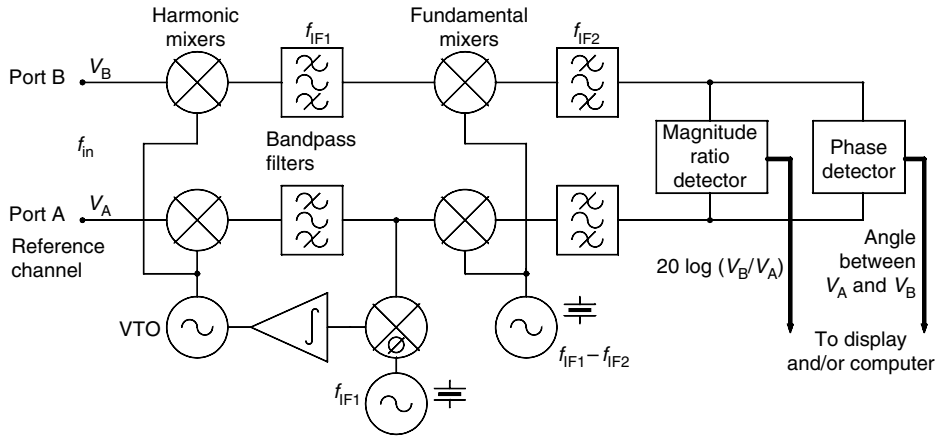


FIGURE 2.23 A vector network analyzer measures complex ratios.

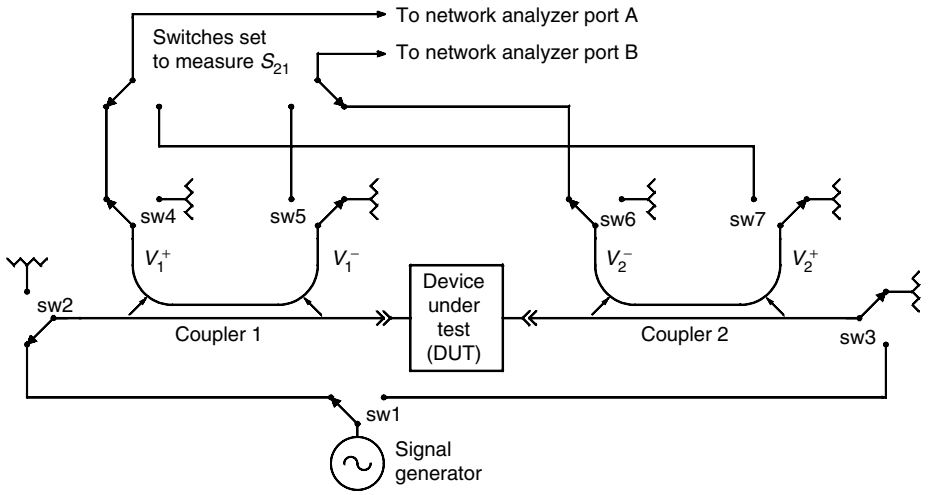


FIGURE 2.24 A two-port *s*-parameter test set can measure all four *s*-parameters without moving the DUT.

For microwave analyzers, because of the high cost of oscillators and the wide frequency coverage required, a more common method of generating conversion oscillators is to use a low-frequency oscillator and a very broadband frequency multiplier. A harmonic of the low-frequency conversion oscillator is offset by an oscillator equal to the IF frequency and the conversion oscillator is then phase locked to the reference channel of the VNA. The reference channel signal is normally the forward voltage wave derived from a directional coupler in an impedance measurement.

The outputs of the synchronous detectors supply the raw data to be converted to a format compatible with the computer. The processors then do corrections and manipulation of the data to the required output form.

The test set supplies the first mixer inputs with the sampled signals necessary to make the desired measurement, and there are many possible configurations. The most versatile is the two-port scattering matrix test set. This unit enables full two-port measurements to be made without the necessity of changing cable connections to the device. The simplified block diagram of a two-port *s*-parameter test set is shown in Figure 2.24. The RF/microwave input is switched between port 1 and port 2 measurements. In each case the RF is split into a reference and test channel. The reference channel is fed directly to a reference channel converter. The test channel feeds the DUT by way of a directional coupler. The coupler output

sampling the reflected power is routed to the test channel converter. Sampled components of incident and reflected power to both the input and output of the DUT are available for processing.

In a full two-port measurement multiple error terms can be identified, measured, and then used to translate the accuracy of calibration references to the measured data from the DUT. For example, if the load used is not ideal, there will be some reflection back into the DUT. If the source generator impedance is not ideal, any reflections from the input of the DUT back to the generator will result in a further contribution to the incident DUT voltage. The couplers are also nonideal and have phase and amplitude errors.

By measuring the full two-port *s*-parameters of a set of known references such as opens, shorts, matched loads, known lengths of transmission lines, and through and open-circuited paths, a system of equations can be derived that includes the error terms. If eight error terms are identified then eight equations with eight unknowns can be derived; although, because VNA measurements are ratios, only seven of the eight unknowns need be determined. The error terms can then be solved for and applied to the results of the measurement of an unknown two-port device to correct for measurement system deviations from the ideal.

Networks with any number of ports can be characterized with *s*-parameters measured on a two-port VNA. When two ports are connected to the test set for measurement, the other ports are terminated in the system characteristic impedance. The other ports are then measured, in turn, to complete the square scattering matrix.

2.2.3.3 Vector Six-Port Network Analyzer

A combination of couplers and power dividers, having 0°, 90°, and 180° phase differences in their output signals can be used to construct a circuit with multiple outputs where the power from the outputs can be used in a system of *n* equations with *n* unknowns. An example of this circuit is shown in Figure 2.25.

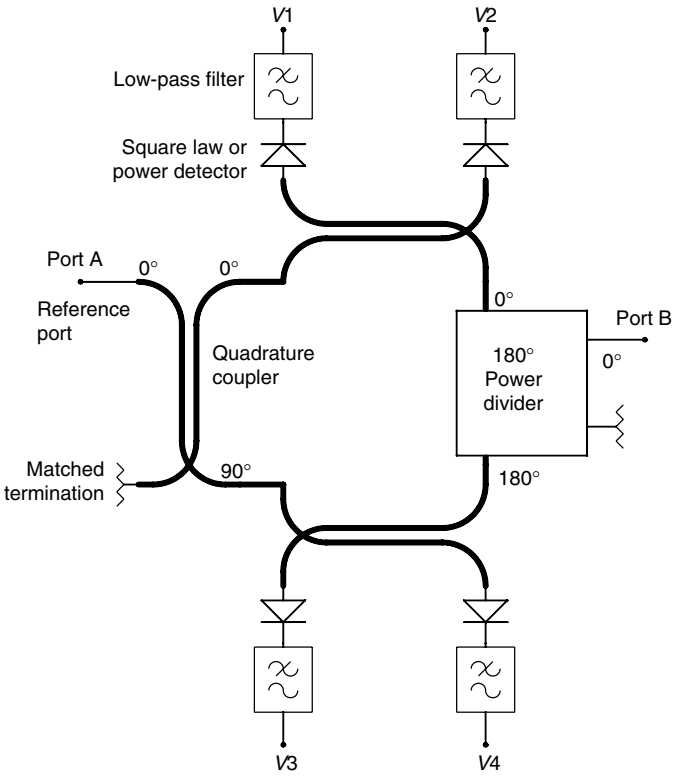


FIGURE 2.25 A six-port network can be used as a narrow-band vector analyzer.

In a properly designed circuit, among the solutions to the system of equations will be the magnitudes and relative phase of the forward and reflected wave. The optimum number of ports for such a device is six; hence, a passive six-port device with diode or power detectors on four of the ports can be used as a vector impedance analyzer [10].

The six-port analyzer has limited bandwidth, usually no more than an octave, because the couplers and power dividers [11] have the same limitation in frequency range to maintain the required amplitude and phase characteristics; however, the low cost of the six-port analyzer makes it attractive for narrow band and built-in test applications.

Typically, measurement test set deviations from the ideal are even more prevalent with the six-port analyzer than for the frequency converting VNA; therefore, use of known calibration elements and the application of the resultant error correction terms is very important for the six-port VNA. The derivation of the error terms and their application to measurement correction is virtually the same for the two analyzers.

References

1. M. Engelson and F. Telewski, *Spectrum Analyzer Theory and Applications*, Artech House, Dedham, MA, 1974.
2. Agilent Technologies, Inc., *Operating Reference Manual for HP 53310A Modulation Domain Analyzer*, Agilent, Santa Clara, CA, 1992.
3. S. F. Adam, *Microwave Theory and Applications*, Prentice-Hall, Englewood Cliffs, NJ, 1969.
4. T. S. Laverghetta, *Modern Microwave Measurements and Techniques*, Artech House, Dedham, MA, 1989.
5. J. G. Webster (ed.), *Wiley Encyclopedia of Electrical and Electronics Engineering*, vol. 13, John Wiley & Sons, NY, 1999, pp. 84–90.
6. P. H. Smith, Transmission Line Calculator, *Electronics*, vol. 12, Jan. 1939, p. 29.
7. F. E. Terman, *Electronic and Radio Engineering*, McGraw-Hill, New York, 1955, p. 100.
8. S. Ramo, J. R. Winnery, and T. Van Duzer, *Fields and Waves in Communications Electronics*, 2nd Edn., John Wiley & Sons, NY, 1988, pp. 229–238.
9. G. L. Matthaei, L. Young, and E. M. T. Jones, *Microwave Filters, Impedance-Matching Networks, and Coupling Structures*, Artech House, Dedham, MA, 1980, pp. 775–842.
10. G. F. Engen, A (Historical) Review of the Six-Port Measurement Technique, *IEEE Transactions on Microwave Theory and Technique*, vol. 45, no. 12, Dec. 1997, pp. 2414–2417.
11. P. A. Rizzi, *Microwave Engineering: Passive Circuits*, Prentice Hall, Englewood Cliffs, NJ, 1988, pp. 367–404.

3

Network Analyzer Calibration

Joseph Staudinger
Freescale Semiconductor, Inc.

Vector network analyzers (VNA) find very wide application as a primary tool in measuring and characterizing circuits, devices, and components. They are typically applied to measure small signal or linear characteristics of multi-port networks at frequencies ranging from RF to beyond 100 GHz (submillimeter in wavelength). Although current commercial VNA systems can support such measurements at much lower frequencies (a few Hz), higher frequency measurements pose significantly more difficulties in calibrating the instrumentation to yield accurate results with respect to a known or desired electrical reference plane. For example, characterization of many microwave components is difficult since the devices cannot easily be connected directly to VNA-supporting coaxial or waveguide media. Often, the device under test (DUT) is fabricated in a noncoaxial or waveguide medium and thus requires fixturing and additional cabling to enable an electrical connection to the VNA (Figure 3.1). The point at which the DUT connects with the measurement system is defined as the DUT reference plane. It is generally the point where it is desired that measurements be referenced. However, any measurement includes not only that of the DUT but contributions from the fixture and cables as well. Note that with increasing frequency, the electrical contribution of the fixture and cables becomes increasingly significant. In addition, practical limitations of the VNA in the form of limited dynamic range, isolation, imperfect source/load match, and other imperfections contribute systematic error to the measurement. To lessen the contribution of systematic error, remove contributions of cabling and fixturing, and therefore enhance measurement accuracy, the VNA must first be calibrated though a process of applying and measuring standards in lieu of the DUT.

Basic measurements consist of applying a stimulus and then determining incident, reflected, and transmitted waves. Ratios of these vector quantities are then computed via post processing yielding network scattering parameters (S-parameters). Most VNAs support measurements on one- and two-port networks, although equipment is commercially available that supports measurements on circuits with more than two ports as well as on differential networks.

3.1 VNA Functionality

A highly simplified block diagram illustrating the functionality of a vector network is provided in Figure 3.2. Generally, a VNA includes an RF switch such that the RF stimulus can be applied to either port 1 or 2, thereby allowing full two-port measurements without necessitating manual disconnection of

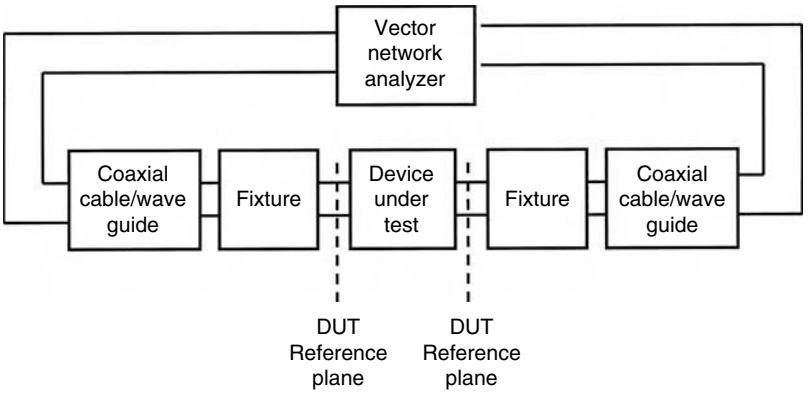


FIGURE 3.1 Typical measurement setup consisting of a device under test embedded in a fixture connected to the vector network analyzer with appropriate cables.

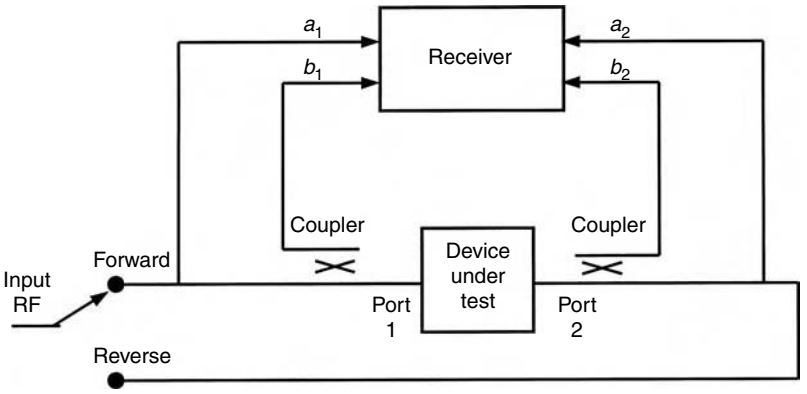


FIGURE 3.2 High simplified VNA block diagram.

the DUT and reversing connections. RF couplers attached at the input and output ports allow measuring reflected voltages. With the RF signal applied in the forward direction (i.e., to port 1), samples of the incident (a_1) and reflected signals at port 1 (b_1) are routed to the receiver. The transmitted signal b_2 reaching port 2 is also directed to the receiver. The receiver functions to downconvert these signals to a lower frequency, which enables digitization and post-processing. Assuming ideal source and load terminations such that a_2 is equal to zero, two scattering parameters can be defined:

$$S_{11} = \frac{b_1}{a_1} \text{ and } S_{21} = \frac{b_2}{a_1}.$$

In reverse operation, the RF signal is directed to port 2 and samples of signals a_2, b_2, and b_1 are directed to the receiver. Assuming ideal source and load terminations such that a_1 is equal to zero, the remaining two scattering parameters are defined:

$$S_{22} = \frac{b_2}{a_2} \text{ and } S_{12} = \frac{b_1}{a_2}.$$

3.2 Sources of Measurement Uncertainties

Sources of uncertainty or error in VNA measurements are primarily the result of systematic, random, and drift errors. The latter two effects tend to be unpredictable and therefore cannot be removed from the measurement. They are the results of factors such as system noise, connector repeatability, temperature variations, and physical changes within the VNA. Systematic errors, however, arise from imperfections within the VNA, are repeatable, and can be largely removed through a process of calibration. Of the three, systematic errors are generally the most significant, particularly at RF and microwave frequencies. In calibration, such errors are quantified by measuring characteristics of known devices (standards). Hence, once quantified, systematic errors can be removed from the resulting measurement. The choice of calibration standards is not necessarily unique. Selection of a suitable set of standards is often based on such factors as ease of fabrication in a particular medium, repeatability, and the accuracy to which the characteristics of the standard can be determined.

3.3 Modeling VNA Systematic Errors

A mathematical description of systematic errors is accomplished using the concept of error models. The error models are intended to represent the most significant systematic errors of the VNA system up to the reference plane—the electrical plane where standards are connected (Figure 3.1). Hence, contributions from cables and fixturing in the measurement, up to the reference plane, are accounted for as well.

A flow graph illustrating a typical error model for one-port reflection measurements is depicted in Figure 3.3. The model consists of three terms, E_{DF}, E_{RF}, and E_{SF}. The term S_{11M} represents the reflection coefficient measured by the receiver within the VNA. The term S_{11} represents the reflection coefficient of the DUT with respect to the reference plane (i.e., the desired quantity).

The three error terms represent various imperfections. Term E_{DF} accounts for directivity in that the measured reflected signal does not consist entirely of reflections caused by the DUT. Limited directivity of the coupler and other signal leakage paths result in other signal components vectorally combining with the DUT reflected signal. Term E_{SF} accounts for source match in that the impedance at the reference plane is not exactly the characteristic impedance (generally 50 Ω). Term E_{RF} describes frequency tracking imperfections between reference and test channels.

A flow graph illustrating a typical error model for two-port measurement, accounting for both reflection and transmission coefficients is depicted in Figure 3.4. The flow graph consists of both forward (RF signal applied to port 1) and reverse (RF signal applied to port 2) error models. The model consists of twelve terms, six each for forward and reverse paths. Three more error terms are included in addition to those shown in the one-port model (E_{LF}, E_{TF}, and E_{XF} for the forward path, and similarly E_{LR}, E_{TR}, and E_{XR} for the reverse path). As before, reflection as well as transmission coefficients measured by the receiver within the VNA are denoted with an *M* subscript (e.g., S_{21M}). The desired two-port S-parameters referenced with respect to port 1/2 reference

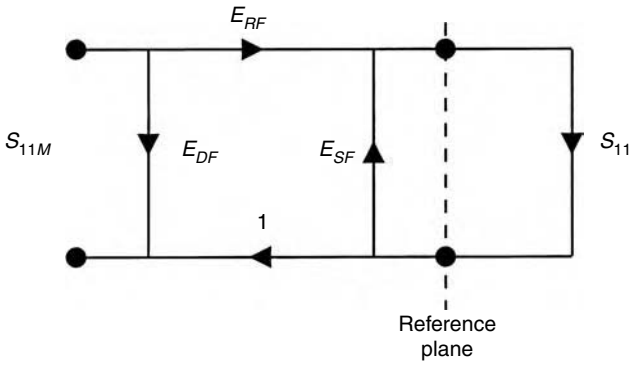


FIGURE 3.3 Typical one-port VNA error model for reflection coefficient measurements.

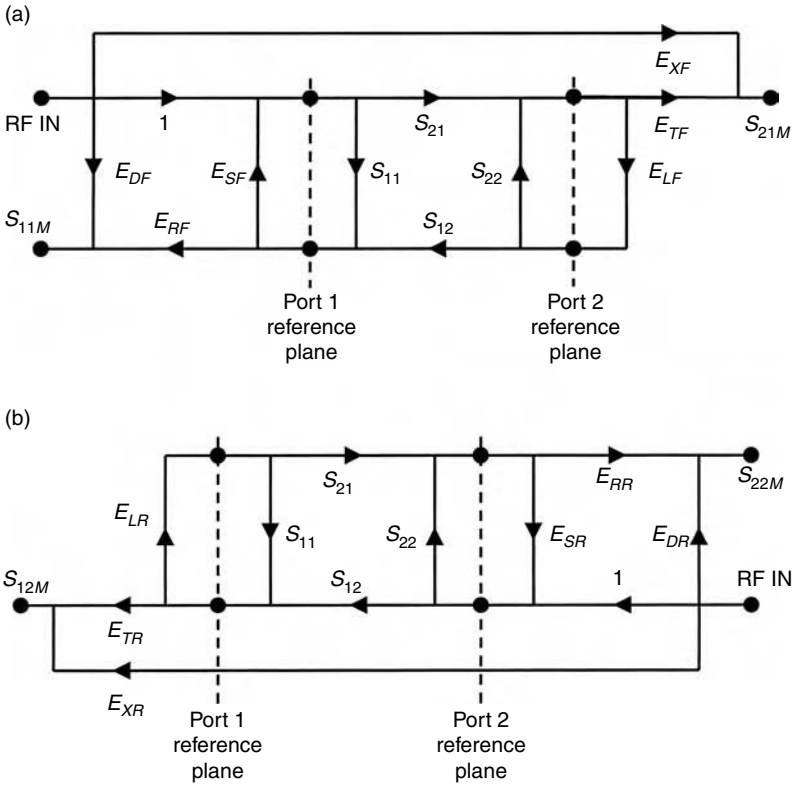


FIGURE 3.4 Typical two-port VNA error model: (a) forward model and (b) reverse model.

planes are denoted as S_{11}, S_{21}, S_{12}, and S_{22}. The transmission coefficients are ratios of transmitted and incident signals. Error term E_{LF} accounts for measurement errors resulting from an imperfect load termination. Term E_{TF} describes transmission frequency tracking errors. The term E_{XF} accounts for isolation in that a small component of the transmitted signal reaching the receiver is due to finite isolation where it reaches the receiver without passing through the DUT. The error coefficients for the reverse path are similarly defined.

3.4 Calibration

From the above discussion, it is possible to mathematically relate uncorrected scattering parameters measured by the VNA (S_{M}) to the above-mentioned error terms and the S-parameters exhibited by the DUT (S). For example, with the VNA modeled for one-port measurements as illustrated in Figure 3.3, the reflection coefficient of the DUT (S_{11}) is given by:

$$S_{11} = \frac{S_{11\mathrm{M}} - E_{\mathrm{DF}}}{E_{\mathrm{SF}}\left(S_{11\mathrm{M}} - E_{\mathrm{DF}}\right) + E_{\mathrm{RF}}}$$

Similarly, for two-port networks, DUT S-parameters can be mathematically related to the error terms and uncorrected measured S-parameters. DUT parameters S_{11} and S_{21} can be described as functions of $S_{11\mathrm{M}}$, $S_{21\mathrm{M}}$, $S_{12\mathrm{M}}$, $S_{22\mathrm{M}}$ and the six forward error terms. Likewise, S_{12} and S_{22} are functions of the four measured S-parameters and the six reverse error terms. Hence, when each error coefficient is known, the DUT S-parameters can be determined from uncorrected measurement.

Therefore, calibration is essentially the process of determining these error coefficients. This is accomplished by replacing the DUT with a number of standards whose electrical properties are known with respect to the desired reference plane (the reader is referred to [1–5] for additional information). Additionally, since the system is frequency dependent, the process is repeated at each frequency of interest.

3.5 Calibration Standards

Determination of the error coefficients requires the use of several standards, although the choice of which standards to use is not necessarily unique. Traditionally, short, open, load, and through (SOLT) standards have been applied, especially in a coaxial medium that facilitates their accurate and repeatable fabrication. Electrical definitions for ideal and lossless SOLT standards (with respect to port 1 and 2 reference planes) are depicted in Figure 3.5. Obviously, and especially with increasing frequency, it is impossible to fabricate standards such that they are (1) lossless and (2) exhibit the defined reflection and transmission coefficients at these reference planes. Fabrication and physical constants dictate some nonzero length of transmission line must be associated with each (Figure 3.6). Hence, for completeness, the characteristics of the transmission line must be (1) known, and (2) included in defining the parameters of each standard. Wave propagation is described as

$$V(z) = Ae^{-\gamma z} + Be^{\gamma z}$$

where γ is the propagation constant defined as

$$\gamma = \alpha + j\beta$$

Assuming the electrical length of the transmission line associated with the standards is short, losses become small and perhaps α can be neglected without significant degradation in accuracy. Alternatively, commercial VNA manufacturers often describe the transmission line in terms of a delay coefficient with a small resistive loss component. The open standard exhibits further imperfections since the electric field pattern at the open end tends to vary with frequency. The open-end effect is often described in terms of a frequency-dependent fringing capacitance (C_{Open}) expressed in terms of a polynomial expansion taking the form:

$$C_{\text{Open}} = C_0 + C_1F + C_2F^2 + C_3F^3 + \cdots$$

where C_0, C_1, … are coefficients and F is frequency.

The load termination largely determines forward and reverse directivity error terms (E_{DF} and E_{DR}). Considering the error models in Figures 3.3 and 3.4, with the load standard applied on port 1, forward directivity error takes the following form:

$$E_{DF} = S_{11M} - \frac{S_{11Load}\,E_{RF}}{1 - E_{SF}\,S_{11Load}}$$

where $S_{11\ \text{load}}$ is the actual reflection coefficient of the load standard. Ideally, the load standard should exhibit an impedance of Z_0 (characteristic impedance) and thus a reflection coefficient of zero (i.e., $S_{11\ \text{load}} = 0$) in which case E_{DF} becomes the measured value of S_{11} with the load standard connected to port 1. High quality coaxial-based fixed load standards exhibiting high return loss over broad bandwidths are generally commercially available, especially at RF and microwave frequencies. At higher frequencies and/or where the electrical performance of the fixed load terminations is inadequate, sliding terminations are employed. Sliding terminations use mechanical methods to adjust the electrical length of a transmission line associated with the load standard. Neglecting losses in the transmission line, the above expression

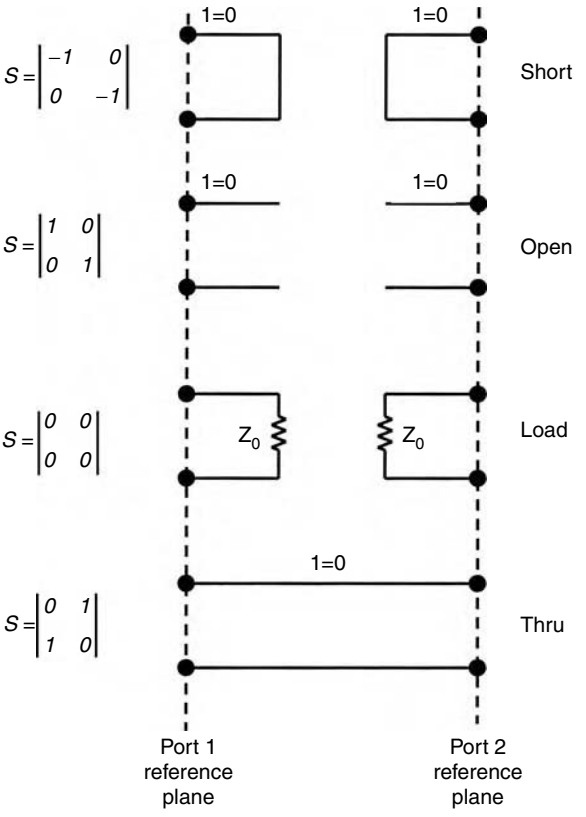


FIGURE 3.5 Electrical definition for lossless and ideal SOLT standards.

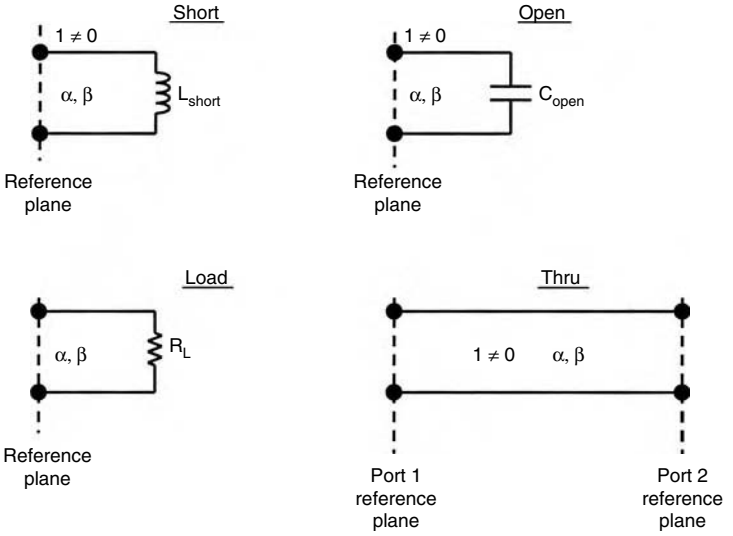


FIGURE 3.6 High frequency descriptions of SOLT standards generally consider nonzero length transmission lines, loss mechanisms, and fringing field effects associated with the open standard.

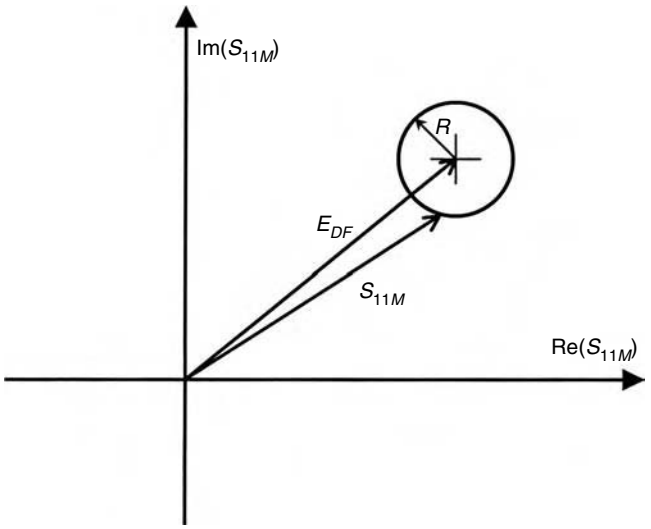


FIGURE 3.7 Characterizing directivity error terms using a sliding load termination.

forms a circle in the S_{11} measurement plane as the length of the transmission line is varied. The center of the circle defines error term E_{DF} (Figure 3.7).

Often it is desirable to characterize devices in noncoaxial media. For example, measuring the characteristics of devices, and circuits at the wafer level by connecting microwave probes directly to the wafer. Other situations arise where components cannot be directly probed but must be placed in packages with coaxial connectors and it is desirable to calibrate the fixture/VNA at the package/fixture interface. Although fabrication techniques favor SOLT standards in coax, it is difficult to realize them precisely in other media such as microstrip and hence non-SOLT standards are more appropriate. Presently, standards based on one or more transmission lines and reflection elements have become popular for RFICs and MMICs. Fundamentally, they are more suitable for MMICs and RFICs since they rely on fabricating transmission lines (in microstrip, for example), where the impedance of the lines can be precisely determined based on physical dimensions, metalization, and substrate properties. The TRL (thru, reflect, line) series of standards have become popular as well as variations of it such as LRM (line, reflect, match), and LRL (line, reflect, line) to name but a few. In general, TRL utilizes a short length thru (sometime assumed zero length), a highly reflective element, and a nonzero length transmission line. One advantage of this technique is that a complete electrical description of each standard is not necessary. However, each standard is assumed to exhibit certain electrical criteria. For example, the length of the thru generally must be known, or alternatively, the thru may in many cases be fabricated such that its physical dimensions approach zero length at the frequencies of interest and are therefore insignificant. The characteristic impedance of the line standard is particularly important in that it is the major contributor in defining the reference impedance of the measurement. Its length is also important. Lengths approaching either 0^{o} or multiples of 180^{o} (relative to the length of the thru) are problematic and lead to poor calibrations. The phase of the reflection standard is not critical, although its phase generally must be known to within one-quarter of a wavelength.

In the interest of reducing hardware cost, a series of VNAs are commercially available based on a receiver architecture containing three rather than four sampling elements. In four-sampling receiver architecture, independent measurements are made of a_1, b_1, a_2, and b_2. Impedance contributions of the internal switch that routes the RF stimulus to port 1 for forward measurements and to port 2 for reverse measurements can be accounted for during the calibration process. In a three-sampling receiver architecture, independent measurements are made on b_1, b_2 and on a combined a_1 and a_2. This architecture

is inherently less accurate than the former in that systematic errors introduced by the internal RF switch are not fully removed via TRL calibration, although mitigating this effect to some extent is possible [5]. However, it should be noted that this architecture provides measurement accuracy that is quite adequate for many applications.

References

1. Staudinger, J., A two-tier method of de-embedding device scattering parameters using novel techniques, Master Thesis, Arizona State University, May 1987.
2. Lane, R., De-Embedding Device Scattering Parameters, *Microwave J.,* Aug. 1984.
3. Fitzpatrick, J., Error Models For Systems Measurement, *Microwave J.*, May 1978.
4. Operating and Programming Manual for the HP8510 Network Analyzer, Hewlett Packard, Inc., Santa Rosa, CA.
5. Metzger, D., Improving TRL* Calibrations of Vector Network Analyzers, *Microwave J.,* May 1995.

4

Absolute Magnitude and Phase Calibrations*

Kate A. Remley
Paul D. Hale
Dylan F. Williams
National Institute of Standards and Technology

4.1 Absolute Calibration for Signal Measurements

In vector network analyzer (VNA) measurements, the magnitude and phase of a transmitted or reflected wave are measured relative to that of the incident wave. VNA measurements are made with a single frequency at a time, and the phase relationships between frequency components are not measured. Calibrations of VNAs are relative as well and are based on measurements of a collection of known impedance standards. Measurements such as these, based on single-frequency acquisition and relative calibrations, are useful for finding the linear response of systems such as electronically passive networks.

However, there are a number of measurement scenarios where we are interested in finding the magnitude and phase relationships between all frequency components in a signal simultaneously. Measurements of complex digitally modulated signals and the nonlinear response of an amplifier are two examples. For these measurements we use calibrations that provide the absolute magnitude and phase relationships for all of the measured frequency components. Calibrating instruments that can simultaneously acquire all frequency components of input or output signals generally involves finding and correcting for the absolute

* Publication of the U.S. government, not subject to copyright in the United States.

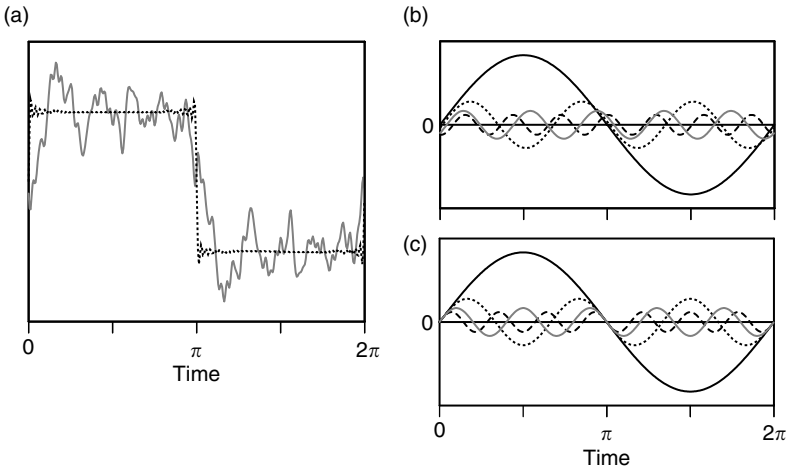


FIGURE 4.1 To illustrate the concept of absolute calibration, consider the measurement of a square wave and its Fourier series representation. The distorted (uncalibrated) waveform in (a) can be constructed from its Fourier series, the first few terms of which are shown in (b). When the measurement instrument is not properly calibrated, the sine waves are not correctly aligned. After proper calibration the sine waves are aligned as in (c), resulting in the more "square" waveform in (a).

magnitude and phase response of the instrument. When the response characterization is carried out from first principles, the calibration can be considered traceable to fundamental quantities.

The importance of an absolute calibration can be visualized by considering the first few terms of a Fourier series representation of signals such as those shown on the left side of Figure 4.1. We can think of the "square wave" as being the true input into the measurement instrument and the distorted waveform as the instrument measurement without calibration. The square wave can be represented by a superposition of sine waves that cross zero at times $t = 0$ and $t = \pi$. The distorted waveform, on the other hand, is composed of the same sinusoids but with slightly perturbed phases. The figure illustrates the importance of phase calibrations; if the instrument does not correctly measure the phases of each frequency component, the reconstructed time-domain waveform is in error.

Absolute calibrations are typically applied to time- and frequency-domain instruments that can measure the relationships between all frequency components of input or output signals. Time-domain measurement strategies sample the waveform and then use a discrete Fourier transform (DFT) to transform the waveform into the frequency domain for subsequent processing. Time-domain instruments can be categorized according to how the timing of the samples is achieved. Some time-domain instruments, such as sampling oscilloscopes, microwave transition analyzers (MTAs), and large-signal network analyzers (LSNAs) use forms of *equivalent-time sampling*. These instruments build up a waveform from samples collected over many cycles of the repeating signal. MTAs and LSNAs use *sampling down-conversion* [1,2] and require a strictly periodic input signal. Sampling oscilloscopes require a repetitive signal that may or may not be periodic. They use a programmed delay generator to determine the time interval between a trigger event and the time in the input waveform at which a sample is taken.

Instruments based on *real-time sampling* measure a block of time samples, where the time increment between samples is much shorter than any characteristic time feature of the signal. These instruments include real-time oscilloscopes, vector signal analyzers (VSAs), and real-time spectrum analyzers. Instruments using real-time sampling are often used to measure a single event or a long waveform with random modulation but can also be used to measure a periodic waveform.

Absolute calibrations can also be applied to instruments that use a more frequency-domain-like measurement strategy, such as the experimental systems described in References 3 and 4. These instruments have many characteristics in common with a VNA, but have a mechanism for calculating the phase

relationship between the different single-frequency measurements. Knowledge of this phase relationship enables these instruments to reconstruct a time-domain representation of the input signal.

4.2 Elements of Signal-Measurement Calibrations

As discussed below, absolute calibration of a signal measurement includes (a) determination of the instrument response and (b) removal of the instrument response from subsequent measurements to estimate the true measured signal. Two additional corrections may be required: first, when the instrument's impedance is different from the reference impedance of the measurement (usually 50 Ω), an additional calibration step may be necessary to correct for this impedance mismatch. Second, if the instrument's timebase is prone to significant errors, some form of correction may be required. In Section 4.2.1, we describe methods for finding and calibrating out the instrument response. In Section 4.2.2, we combine these methods with impedance mismatch correction. In Section 4.2.3, we describe techniques to correct for time-base errors, and discuss other considerations that must be kept in mind to obtain a rigorous calibration for measured signals.

4.2.1 Instrument Impulse Response

Whether we are using time- or frequency-domain measurement instruments, the output signal will consist of the measured signal distorted by the response of the instrumentation. Assuming the measurement system behaves in a linear and time-invariant fashion [5,6], time-domain instruments measure the convolution of the input signal with the instrument impulse response. In the frequency domain this is equivalent to the product of the signal's Fourier transform and the instrument's frequency response, which is the Fourier transform of the instrument's impulse response.

The first step in conducting an absolute calibration is to find the response of the instrument. We usually do this by measuring a well-characterized or calculable waveform. In the frequency domain, the response $h(\omega)$ of the instrument is found by taking the ratio of the measurement $m(\omega)$ to the known calibration waveform $s(\omega)$. Knowing the response, we can then correct subsequent measurements of other signals by eliminating the response from the measurement.

For instruments based on real-time acquisition, such as vector signal analyzers removal of the instrument response is often done internally in the time domain by applying a time-domain filter that approximates the inverse of the measured impulse response of the instrument. Such filters can be applied quickly and may be more convenient to implement in hardware or software than a frequency-domain filter.

When user-implemented post-processing techniques are used, removal of the instrument response is usually performed in the frequency domain by division, as described in Figure 4.2. When a frequency-domain output is desired over a finite bandwidth and the measurement and frequency response are relatively noise free, the calculated result is generally well behaved. However, when the measurement is

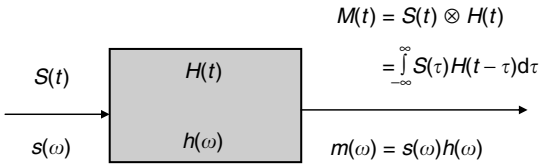


FIGURE 4.2 Absolute calibration for a linear, time-invariant measurement system. When determining the instrument response, a known calibration signal $s_c(\omega)$ is used as the input. The measured output of the instrument $m_c(\omega)$ is used to find the instrument response $h(\omega) = m_c(\omega)/s_c(\omega)$. When an unknown signal $s_u(\omega)$ is applied to the instrument, the measurement $m_u(\omega)$ is corrected for the imperfect response of the instrument by dividing out the instrument response $s_u(\omega) = m_u(\omega)/h(\omega)$. In most measurement systems, signals are actually measured only at discrete times and frequency-domain quantities are found at discrete frequencies.

noisy or the signal bandwidth approaches that of the instrument, removal of the system response can become unstable [7,8]. This can be particularly troublesome when a time-domain result is desired.

Commercially available signal-measurement instruments may have a factory-applied calibration or an internal calibration routine that uses a known calibration waveform. The manufacturer may specify a maximum interval between factory calibrations or the user is instructed to carry out an internal calibration if certain conditions have been met, such as an interval of time or change in temperature since the last calibration. The instrument's internal circuitry and the intended use of the instrument will also help determine how and when calibrations must be performed. For example, if an instrument is designed to provide the highest accuracy (one or two degrees of phase), calibrations may need to be carried out at each use. In the past, oscilloscopes have been assumed to be ideal and were operated without calibration. As these instruments find increasing use in quantitative signal-measurement applications, proper calibration becomes a necessity, as demonstrated in Section 4.3.

In the following paragraphs, we describe several calibration waveforms, the instruments to which they can be applied, and whether or not calibrations using these waveforms can presently be made traceable to fundamental physical quantities or are used as *transfer standards* (instruments that have been calibrated using another instrument whose measurements are traceable to fundamental quantities).

4.2.1.1 Calibration Waveforms

Calibration waveforms typically are those which have been measured by an instrument whose frequency response is much broader than the instrument we wish to calibrate and whose measurement uncertainty is known. In the following, we describe some waveforms that are commonly used to calibrate signal measurements. With the exception of sine waves, all the methods described include characterization of both the magnitude and the phase of the instrument response.

4.2.1.1.1 Swept-Frequency Sine Waves

The swept-sine or direct-substitution technique measures the absolute magnitude response of the test instrument when it is excited by sine waves of known power [9–11]. The frequency of the sine waves is swept over the desired frequency range and the instrument response found at each frequency. This method is relatively simple and can be made traceable to fundamental power standards at frequencies where they are available ($\sim$100 kHz to $>$50 GHz). The swept-sine technique can be applied to any of the instruments described above. Many manufacturers specify the bandwidth and response of their instruments using a swept-sine technique.

4.2.1.1.2 Pulses

Pulse generators that produce a fast impulse or step-like transition can create calibration waveforms that cover a broad frequency range. These pulses may or may not need to be periodic depending on the instrument we wish to calibrate. Periodic pulse trains are often used to calibrate instruments designed to measure periodic signals, as discussed in the section below on frequency combs.

For time-domain instruments such as sampling oscilloscopes utilized in the measurement of time-domain features of pulsed waveforms [12,13], we do not need to measure a periodic pulse train. Instead, we can calibrate the instrument using a step generator whose output signal has been characterized. Generally only one of the transitions of a step-generator waveform is used as a calibration waveform since the waveforms emanating from step generators usually exhibit one fast transition and one slow transition. Some national measurement institutes, including the National Institute of Standards and Technology (NIST) in the United States and the National Physical Laboratory (NPL), in the United Kingdom, have measurement services for characterizing the transition duration of fast step-like pulses.

Optoelectronic methods are used to generate some of the fastest calibration pulses available. NIST [14,15], NPL [16], and the German national measurement institute Physikalisch Technische Bundesanstalt [17] currently use a mode-locked laser that generates a periodic train of sub-picosecond optical impulses. A fast photodetector [18,19] converts the optical impulses to a train of electrical impulses with duration on the order of 1–5 ps. These pulses are then characterized by electro-optic sampling (EOS) [20]. Because EOS systems utilize on-wafer methods and involve fast electro-optic interactions, they have a system

impulse response on the order of 100 fs and have substantially more bandwidth than is needed to measure the impulses generated by the photodetector.

At NIST, the calibration waveform from the photodetector is found at a coaxial reference plane by correcting for impedance mismatch, dispersion, and loss in the EOS measurement system. Once calibrated, the portable photodetector can be used as a transfer standard to calibrate other signal-measurement instruments such as lightwave component analyzers (LCAs) [21], LSNAs, VSAs, and sampling oscilloscopes [22]. In turn, these instruments can be used as transfer standards themselves to calibrate more general signal-generation equipment, such as comb, pulse, and vector signal generators.

4.2.1.1.3 The Nose-to-Nose Calibration

The nose-to-nose calibration [23–25] is a variation of the pulse method described above. This method may be applied to sampling oscilloscopes having a balanced two-diode sampling architecture. In this case, the calibration waveform is a "kickout" pulse, generated each time the samplers on the oscilloscope are fired. The kickout pulse is assumed to have essentially the same shape as the impulse response of the oscilloscope itself. When the inputs of two like oscilloscopes are connected together (nose-to-nose) and triggered in such a way as to measure the kickout pulse with one of the oscilloscopes, this assumption allows deconvolution of the measured waveform to obtain the impulse response of the oscilloscopes. The method has the advantage of requiring very little custom equipment, and its sources of uncertainty have been studied in depth [9,24–26]. To obtain the highest possible accuracy with this technique, as with all measurements using sampling oscilloscopes, various corrections must be applied to the oscilloscope timebase [27]. While not reported in [22], the experiments there demonstrated that the nose-to-nose assumption of the proportionality of the kickout and impulse response can lead to significant errors at high frequencies. However, at lower frequencies the method is reliable.

4.2.1.1.4 Frequency Combs (Periodic Pulses)

As the name implies, a comb generator produces a set (or "comb") of discrete, harmonically related tones in the frequency domain, indicating a periodic waveform in the time domain. Generally, significant power is needed in each of the tones spanning the frequency range of the instrument we wish to calibrate. This is accomplished by ensuring that the periodic waveform contains a fast impulse or step-like transition.

Calibration waveforms from comb generators are often characterized using a calibrated oscilloscope. Once characterized, a comb generator can be used as a transfer standard to generate a calibration waveform that is useful for calibrating both time- and frequency-domain-measurement equipment. Often a comb generator is used to calibrate instruments such as LSNAs that measure periodic signals rich in harmonic content. The system described in References 3 and 4 uses a comb generator whose harmonics have a relatively narrow frequency spacing to find the phase relationships between measured frequency components across a wide frequency bandwidth.

4.2.1.1.5 Multisine Signals

A calibration approach useful for instruments that measure bandpass signals utilizes a multisine signal as the calibration waveform [28–30]. A multisine consists of a collection of simultaneously generated sine waves offset in frequency from each other by a frequency that is usually much smaller than the carrier frequency (kilohertz to megahertz versus hundreds of megahertz or gigahertz). Unlike pulsed waveforms, the magnitudes and/or phases of each sine wave component can be tailored to provide calibration-waveform signals with low peak-to-average power ratios. This enables the test instrument to measure a signal that lies in its linear operating range and is consistently above the noise floor. Multisines can be generated by a vector signal generator. If the vector signal generator is calibrated with an instrument traceable to fundamental quantities, such as a calibrated oscilloscope, the vector signal generator becomes a transfer standard and the multisine calibration can be considered traceable as well. This is a current topic of research [30].

4.2.1.2 Calculable Waveform Standards

Another method of finding the response of an instrument is to measure a device whose response to an electrical stimulus is known a priori from the physics of the device. We refer to waveforms that

are generated by a device whose output is known from fundamental physical principles as calculable calibration waveforms. If the phase relationship between the fundamental and the harmonics appearing at the output of the reference device is calculable, we can measure the reference device and use the measurement to correct measurements of other devices. A "golden diode" [31] is one example of such a device. The well-understood sinusoidal nonlinearity of an optical modulator has been used to measure the phase response of an optical receiver with good results [32]. The use of a superconducting transmission line as a phase reference standard has also been proposed [33,34].

Currently, calculable waveform standards have limited applicability because of the difficulty in manufacturing a physical artifact that is stable and whose behavior is understood with precision over a sufficient bandwidth, particularly a phase standard. However, some instrument manufacturers have had success building such artifacts into instrumentation designed for a specific use, such as a "golden mixer" built in to some VNA mixer measurement set-ups [35,36]. Verifying the accuracy of such devices and extending their usable frequency range are topics of current research at various labs.

4.2.2 Impedance Mismatch and Mismatch Correction

When calibrating signal-measurement instrumentation, great care is often taken to match the impedance of the measurement instrument and the impedance of the calibration-waveform signal generator to a reference impedance, typically 50 Ω. This not only helps to define the measurement, but reduces the effects of multiple reflections on the results. One way to achieve an impedance match is to isolate the test instrument from the waveform generator through use of matching networks or attenuators. For example, 50 Ω attenuators provide an impedance match to a 50 Ω signal generator, and any reflections that do occur are attenuated twice before they reach the circuit. However, the reduction in signal strength introduced by the attenuator also reduces the dynamic range of the calibration.

Another method is to use impedance mismatch correction techniques. Since the response of the test instrument and the known-signal generator are both assumed to be linear, we can construct an equivalent circuit model, as shown in Figure 4.3. This also allows us to calculate how the device will behave when placed in an arbitrary circuit. This wave-based representation is often used in microwave circuit descriptions [37], but is equivalent to the Thévenin- and Norton-equivalent circuit models commonly used at lower frequencies. References 38 and 39 give formulas for converting between the three representations.

When we talk about forward and reflected waves, we typically refer to complex frequency-domain forward and backward wave amplitudes a and b. Here we use the "pseudowaves" of Reference 40 with a different normalization. These pseudowaves correspond to the conventional power-normalized forward and backward wave amplitudes with a reference impedance of 50 Ω. These wave amplitudes have units of the square root of a watt, and are conventionally normalized so that the average power p transmitted across a reference plane is given by $p = 1/2(|a|^2 - |b|^2)$.*

The microwave flow diagram shown in Figure 4.3 describes the propagation of signals when a signal generator is connected directly to a measurement instrument, for example, a high-speed sampling oscilloscope. The generator is characterized by its forward-wave source amplitude b_g and its reflection coefficient Γ_g.

Following [6], we write the peak voltage v_g of the "forward voltage wave" associated with the wave amplitude b_g as $v_g = \sqrt{50\,\Omega}b_g$. Likewise, the relation between the peak voltage v_s that the oscilloscope measures and the wave a_s in Figure 4.3 is $v_s = \sqrt{50\,\Omega}a_s$. Note that the voltage v_g should not be confused with the total voltage at the generator's output port when the impedance of the load connected to the generator is not 50 Ω. This is because an imperfect load will reflect some of the energy incident on it back to the generator, and both the forward and backward waves b_1 and a_1 will then contribute to the total

*Reference 40 uses a root-mean-square (RMS) normalization in which the power p is given by $p = |a|^2 - |b|^2$, where a and b are the RMS pseudowaves. The normalized pseudowaves of reference 40 are related to those used here with $a = a/\sqrt{2}$ and $b = b/\sqrt{2}$.

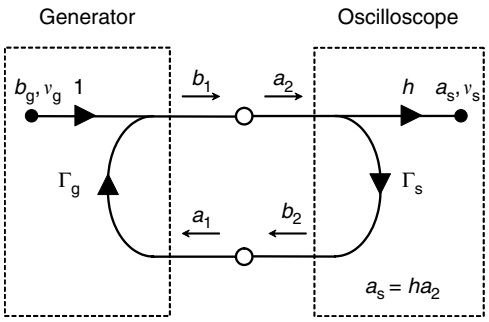


FIGURE 4.3 Microwave flow diagram describing the propagation of signals between a signal generator and a measurement instrument, for example, an oscilloscope. After [6].

voltage at the generator's output port. The multiple reflections between the generator and the oscilloscope are accounted for with a mismatch correction.

Following the arguments of Reference 6, the oscilloscope measures the voltage v_s in terms of the instrument's complex frequency response h and the vector reflection coefficients Γ_g of the generator and Γ_s of the oscilloscope as

$$v_s = \sqrt{50\,\Omega}\, a_s = \sqrt{50\,\Omega}\, h \frac{b_g}{1 - \Gamma_g \Gamma_s} = h \frac{v_g}{1 - \Gamma_g \Gamma_s}. \tag{4.1}$$

Rearranging this expression leads us to a solution for the impulse response of the instrument when excited by a known waveform in the presence of an impedance mismatch,

$$h = \frac{a_s}{b_g}(1 - \Gamma_g \Gamma_s) = \frac{v_s}{v_g}(1 - \Gamma_g \Gamma_s), \tag{4.2}$$

or the mismatched source signal as measured by a calibrated measurement instrument,

$$b_g = \frac{a_s}{h}(1 - \Gamma_g \Gamma_s), \tag{4.3}$$

or

$$v_g = \frac{v_s}{h}(1 - \Gamma_g \Gamma_s). \tag{4.4}$$

As mentioned earlier, our theory depends on the linearity and time invariance of the measurement system as discussed in Reference 6. In practice, the device being measured may be nonlinearly dependent on the load impedance presented by the measurement system, but this is typically a second-order effect. Furthermore, the impedance of the source may depend on the voltage being generated such as in a digital circuit. In this case, an attenuator may be required to reduce time-varying reflections to an acceptable level.

4.2.3 Other Considerations for Signal Measurements: Timebase Errors, Spectral Leakage, and Signal Level

Conducting an accurate calibration requires use of proper measurement set-up and techniques. We describe some of the most common and significant sources of error and give an overview of ways to minimize these errors in signal measurements and calibrations.

4.2.3.1 Timebase Errors

In the case of equivalent-time sampling oscilloscopes, timebase errors such as timebase distortion, jitter, and drift can distort the acquired time-domain waveform and may need to be corrected. Timebase distortion [41,42] is a repeatable systematic error in the time that each sample is taken, and can introduce spurious tones and spectral leakage. It can be accurately measured [41,42] and corrected [43]. Jitter [44] is a random, mean-zero variation in the time interval between samples. Through the 1990s, jitter has been treated as acting on averaged waveforms as a lowpass filter [45]. Recently, techniques have been developed that can simultaneously estimate and correct for timebase distortion and jitter [46]. Drift [47] is a correlated time shift of all the samples in a waveform, which changes with repeated measurement of the waveform and may have a mean different from zero. The drift in consecutive measurements can be estimated using cross-correlation for impulse-like waveforms or by detecting a level crossing in step-like waveforms. Once estimated, the time shifts can be corrected in the frequency domain or the time domain. When correcting timing errors in general, care must be taken to avoid cyclic boundary effects associated with the DFT.

4.2.3.2 Spectral Leakage

Calibrations and measurements using time-domain instruments such as sampling oscilloscopes, VSAs, and LSNAs are somewhat different from their swept-frequency counterparts, the VNA and the spectrum analyzer. It may be necessary to design measurements so that the times at which samples are acquired will provide information at the frequencies of interest once the DFT is taken. When a measurement of a periodic signal does not provide an integer multiple of radio frequency (RF) cycles or envelope periods, spectral leakage can occur, which spreads the signal's energy over frequencies other than those of interest. Acquisition parameters such as the time window over which data is acquired, the number of data points, and envelope period (for modulated signals) may all need to be taken into account to avoid spectral leakage. While some instruments perform the necessary set-up calculations in software, the user must sometimes provide and set these parameters manually.

For example, to ensure an integer number of RF cycles is acquired in an oscilloscope measurement, we set the time increment Δt such that Δt times the number of acquired points N is equal to an integer number M times the RF period T_{RF},

$$\Delta t = \frac{MT_{\mathrm{RF}}}{N}. \tag{4.5}$$

Similarly, for modulated signals we can specify a data-acquisition time window T_{opt} that corresponds to an integer M number of envelope periods T_{env} as

$$T_{\mathrm{opt}} = MT_{\mathrm{env}}. \tag{4.6}$$

For the case of a multisine signal whose frequency components are evenly spaced by an increment Δf, the envelope period is $1/\Delta f$ and the data-acquisition time is

$$T_{\mathrm{opt}} = \frac{M}{\Delta f}. \tag{4.7}$$

To avoid spectral leakage when acquiring N total points in T_{opt}, we should set the instrument for a span of

$$\mathrm{Span} = \frac{N}{T_{\mathrm{opt}}}. \tag{4.8}$$

These procedures are discussed in greater detail in Reference 48. Some instrument manufacturers provide internal algorithms to minimize spectral leakage automatically. Some instruments allow the user to utilize special windowing (filtering) functions for this purpose.

4.2.3.3 Optimal Input Signal Level

For instruments that acquire signals through sampling, providing an optimal signal level at the input port is especially important for accurate calibration or measurement. If a signal amplitude is too low, the instrument's own noise floor will interfere with accurate measurement, introducing a random electrical signal that mixes with the input signal. Conversely, a signal amplitude that is too high will overload the instrument's front-end electronics and create distortion. When acquiring time-domain waveforms it is particularly important to keep in mind that the peak power in the waveform may be much higher than its average power. A momentary peak with large amplitude can distort the measurement significantly. For example, a pulse or a multisine signal whose phases are all zero will have a large peak even though the average power in the signal is quite low.

4.3 Applications of Waveform Calibrations

4.3.1 RF Harmonic Phase Calibration Example: LSNA Measurement of a Square Wave

Large-signal network analyzers such as those in [4,49,50] measure the incident and reflected waves at the input and output ports of a device, circuit, or system at the RF fundamental frequency and its harmonics. As the name implies, one key feature of the LSNA is that the device, circuit, or system being tested may be measured in its large-signal operating state. This provides a realistic test environment for the device, circuit, or system and may be used to characterize the element's nonlinear behavior, if any.

Like a VNA, an LSNA provides a relative impedance calibration to remove the effects of the system hardware between the point where the measurement is made and the reference planes of the test structure. In addition to the relative calibration, the LSNA provides an absolute calibration that uses a power meter for magnitude calibration and a comb generator for phase calibration. Since this phase calibration utilizes the calibration-waveform techniques described in Section 4.2.1, we will focus on it here. As discussed in Section 4.2.1, the output of a comb generator is a sharp, periodic train of pulses that contain significant (measurable) energy at the fundamental and harmonics of the repetition frequency.

The LSNA phase calibrations are accomplished by driving the comb generator at the fundamental frequency of interest and then measuring the output of the comb generator (fundamental and harmonics) with the LSNA. Calibration coefficients are calculated by comparing the LSNA measurement to the known values of the waveform previously measured by a calibrated instrument such as an oscilloscope. In practice, comb generators are characterized at a limited set of frequencies, and so it is necessary to interpolate calibration data from these frequencies to the frequency grid of interest.

To illustrate the significance of the phase calibration on the measurement of an RF signal and its harmonics, we conducted an LSNA measurement of a square-wave signal that contained significant harmonic content up to about the twentieth harmonic of the square-wave fundamental frequency of 1 GHz. We first measured the output of the square-wave generator with a calibrated oscilloscope. Figure 4.4a and b show the time- and frequency-domain oscilloscope measurements, respectively. The dotted or dashed lines show the raw measured data, while the solid lines have both impulse response [22] and time-base correction [46] applied. Note the time-base discontinuity in the dotted line in Figure 4.4a at around 3 ns, which has been corrected in the calibrated waveform.

We then used the corrected oscilloscope measurement as our reference and compared the LSNA measurements to it. Figure 4.5a and b compare one-port phase-calibrated and uncalibrated LSNA measurements with the fully calibrated oscilloscope measurements. Figure 4.5a shows that the spectrum of the magnitude-corrected LSNA is comparable to that of the calibrated oscilloscope. The time-domain representation in Figure 4.5b demonstrates that the lack of a phase calibration can seriously degrade the measurement.

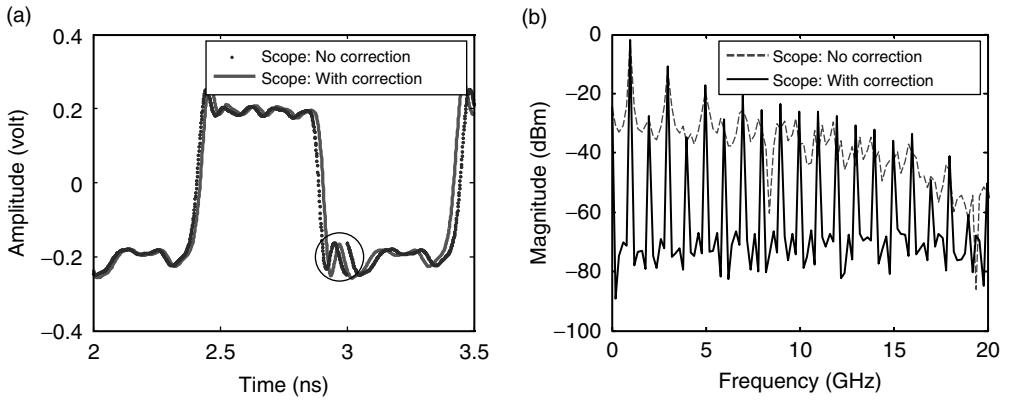


FIGURE 4.4 Oscilloscope measurement of a 1 GHz square-wave signal: (a) time domain and (b) frequency domain. The dotted lines are the raw data and the solid lines represent the calibrated oscilloscope measurement. The discontinuity in the uncorrected waveform circled in (a) is due to timebase distortion in the oscilloscope.

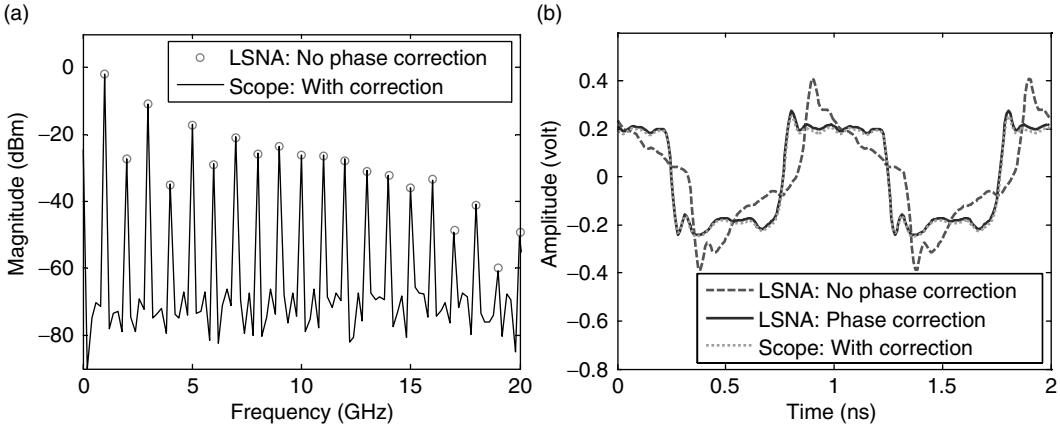


FIGURE 4.5 Magnitude-calibrated LSNA measurements compared to fully calibrated oscilloscope measurements. In (a), we see that the magnitudes compare well. In (b), LSNA measurements with and without phase calibration are compared to the oscilloscope measurement. The dashed line shows that the lack of an LSNA phase calibration can seriously degrade the measurement even though the magnitude spectrum is calibrated.

4.3.2 Oscilloscope Impulse Response Calibration Example: Mixer Reciprocity Measurement

We demonstrate the importance of oscilloscope calibration in a measurement of mixer reciprocity. Measurement of the up- and down-conversion transfer functions of microwave mixers is complicated, because the excitation and output frequencies are not the same. In fact, one common method for approximating the two transfer functions of a mixer is to measure the "round-trip" product of the up- and down-conversion by placing two mixers back-to-back (or to use a similar, but more accurate, three-mixer approach) [35,51–53]. The up- and down-conversion transfer functions of the mixer are found as the mean of the round-trip transfer function. While differences between up- and down-conversion cannot be distinguished using this method, it does eliminate the need to calibrate the absolute phase response of the measurement system at the up- and down-converted signal frequencies.

Even so, it has been known for some time that the up- and down-conversion transfer functions of microwave mixers are not generally equal [54]. To characterize the nonreciprocity of a microwave mixer,

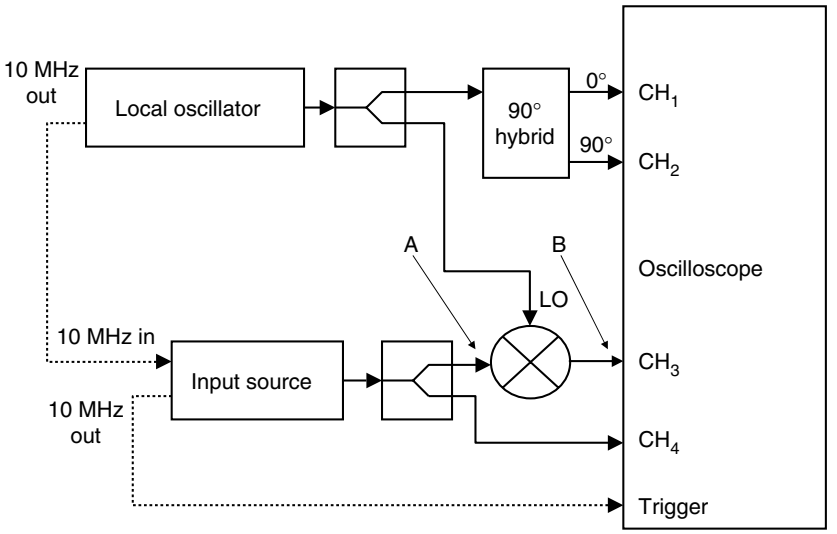


FIGURE 4.6 Set-up to measure the reciprocity of a microwave mixer by use of a calibrated oscilloscope. The phases of the signals at channels 3 and 4 are determined with respect to a fixed local oscillator phase, as measured on channels 1 and 2. From [57].

we need to measure the up- and down-conversion transfer functions separately in an absolute sense [54,55]. One method for measuring separately the absolute magnitude of the up- and down-conversion transfer functions of a mixer is to use a power-meter-corrected VNA [51,56]. This is a very accurate method for finding the magnitude of the transfer function, because the power meter is typically traceable to fundamental quantities. But this method cannot provide the phase.

To find both the magnitude and phase reciprocity of a microwave mixer, one option is to use an oscilloscope-based method [57]. The notion is simple: using a calibrated oscilloscope with multiple sampling heads, we can measure the input and output of a mixer simultaneously. To extract the phase transfer function, we align the simultaneously acquired signals relative to the local oscillator phase [57,58]. As we show, using a calibrated oscilloscope is critical in this application. In Reference 57, the impulse response of the oscilloscope was found by use of the calibration-waveform pulse generated by a photodetector, as described in Section 4.2.1. However, many of the other calibration-waveform techniques described above could have been used to calibrate the oscilloscope.

The measurement set-up used in [57] is shown in Figure 4.6. Channels 3 and 4 of the four-channel oscilloscope are used to measure the RF (or image) and IF of the mixer simultaneously. Channels 1 and 2 measure copies of the output of the local oscillator 90° out of phase with each other for correction of time-base distortion in a post-processing step [46]. Note that how one defines the mixer will depend on the application at hand. Here we included filter circuits at the IF and RF ports as part of the "mixer" in order to measure only the frequency components of interest. At particularly high frequencies, we may need to account for the mixer's internal interactions and with other circuit elements as well.

Figure 4.7a and b show the magnitudes of the mixer's up- and down-conversion transfer functions, respectively. The dots correspond to the oscilloscope measurement before the impulse response has been calibrated out, the x's correspond to the calibrated oscilloscope measurement, and the solid lines correspond to the power-meter calibrated VNA method. We see that measurements from the calibrated oscilloscope and the power-meter-calibrated VNA agree well, illustrating the importance of the oscilloscope calibration.

Figure 4.8 shows the calibrated-oscilloscope measurements of the phase up- and down-conversion. The error bars represent the uncertainty in the measurement. Using this measurement technique, we can clearly discern the nonreciprocity of the mixer.

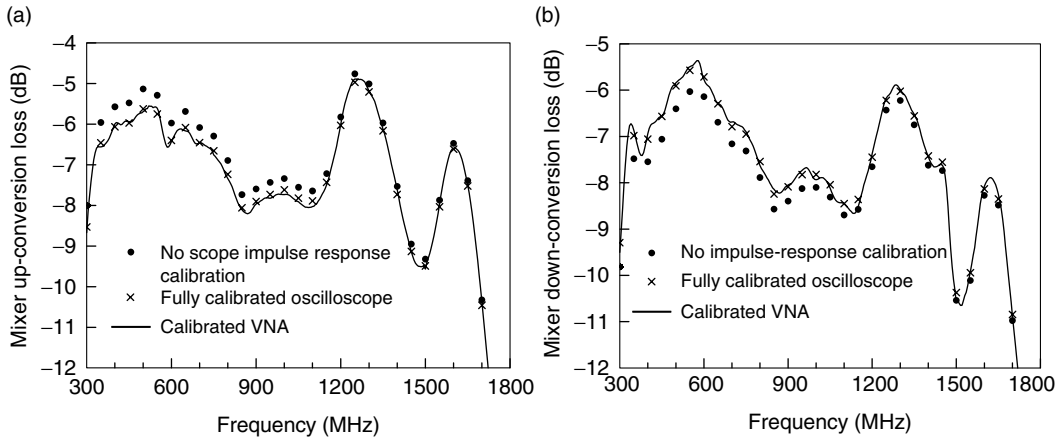


FIGURE 4.7 Measured magnitude response of a microwave mixer. (a) The up- and (b) the down-conversion processes. The solid lines represent power-meter calibrated VNA measurements. The symbols show measurements made with a sampling oscilloscope that was uncalibrated (dots) and calibrated (x's). Agreement is good between the calibrated VNA measurements and the calibrated oscilloscope measurements. The frequency axes are referred to the IF port. After [57].

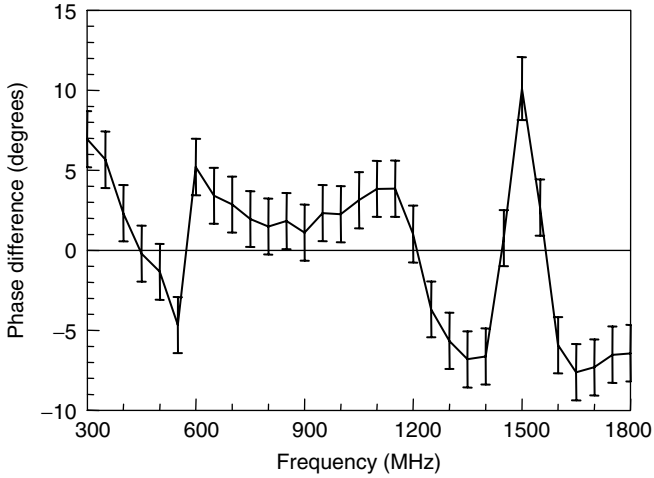


FIGURE 4.8 Phase difference of the mixer up- and down-conversion transfer functions measured with a digital sampling oscilloscope. The error bars represent the uncertainty in the measurement. The frequency axis is referred to the IF port. From [57].

4.3.2.1 Eye and Constellation Patterns

During the development of components used in digital systems, and in the qualification of digital communication systems, it is essential to be able to measure whether a signal is in one digital state or another. Measurements of digitally encoded signals are normally carried out at baseband, either directly (for digital signals) or once a digitally modulated signal has been downconverted from the RF (for telecommunication signals). Accurate decisions on whether a signal is in one digital state or another depends critically on the calibration of both the baseband instrument and the downconverter. Two common measurements for assessing system performance are eye patterns and constellation diagrams [59,60].

Eye patterns, so named because they look like a human eye, provide a useful time-domain representation of a measured binary signal. Eye patterns are typically acquired with an oscilloscope that is synchronized

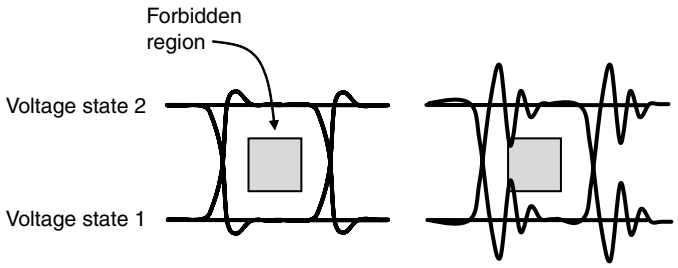


FIGURE 4.9 Schematic of a generic eye pattern measurement for a digital signal at baseband. Here, time is truncated to approximately one bit. The "Forbidden Region" is specified in a number of standards to determine pass/fail criteria for a particular component or to ensure error-free detection of a signal. The signal at right may not pass the eye diagram test since it enters the forbidden region.

with the system clock. Random, pseudorandom, or known-data sequences containing thousands (or millions) of bits may be acquired to get a good statistical representation of the signal. The bits are then overlayed on top of each other to show all possible trajectories from any state to any other state. The resulting time-domain waveform, gives insight into the sources of signal distortion in the system.

The eye pattern can be used to both qualitatively and quantitatively characterize the quality of the baseband data signal. For example, some standards specify an "eye mask" test with allowed and forbidden regions (see Figure 4.9). When samples enter the forbidden regions, that particular hardware component may be rejected in a pass/fail test because the probability of the system making an erroneous determination of the signal "digit" is unacceptably high. The mask test checks for errors in both signal timing and amplitude.

Traditionally, eye patterns have not been calibrated because they have been based on random samples of a waveform. This random approach eliminates any possibility of correcting the instrument response by use of standard DFT methods. Modern measurement techniques such as time-domain filtering in deep memory real-time oscilloscopes, the LSNA method described in Reference 61, and pattern lock methods used with sampling oscilloscopes now allow the sequential acquisition of a moderately large number of samples so that correction for the system response is possible.

Like eye patterns, constellation diagrams are used to develop, assess, and qualify components, and systems are used to detect digitally modulated signals. However, constellation diagrams are mainly used in evaluating components used in telecommunications systems where symbols are modulated in both magnitude and phase. Thus constellation diagrams are able to describe errors in measured symbols having multiple states that cannot be represented in an eye pattern. The ideal placement of the in-phase and quadrature voltage value of symbols for a given modulation type is shown as a dot on the I/Q plane, similar to the real and imaginary axes used in complex voltage representations. For example, Figure 4.10 shows the placement of symbols for a 16-state quadrature amplitude modulation (QAM) signal.

In practice, distortion introduced in the telecommunications channel, as well as phase noise, thermal noise, and nonlinearities in the system under test or in the measurement instrumentation will cause the measured values of the demodulated signals to be different from the ideal value. The difference between the ideal and measured symbols can be represented by an error vector, as shown in Figure 4.10. To efficiently calculate the error vector magnitude (EVM)—a common figure of merit in telecommunication system measurements—we can plot the ideal and measured symbols on a normalized graph [48].

As with the eye pattern, we can increase measurement confidence if we average over hundreds (or thousands) of samples in an EVM measurement. In both cases it is good practice to state the uncertainty—which depends on the number of samples taken—in conjunction with the measured results.

As discussed in Chapter 6, the goal of an EVM measurement is to assess distortion introduced by the telecommunication system we are testing. Proper calibration and instrument set-up are critical to separating distortion caused by instrumentation from that caused by the telecommunication system

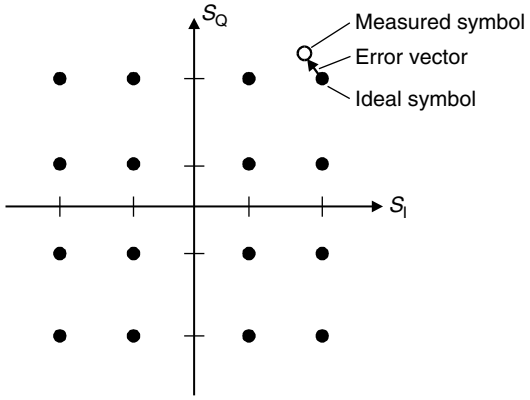


FIGURE 4.10 Constellation diagram for a 16-QAM signal. From [48].

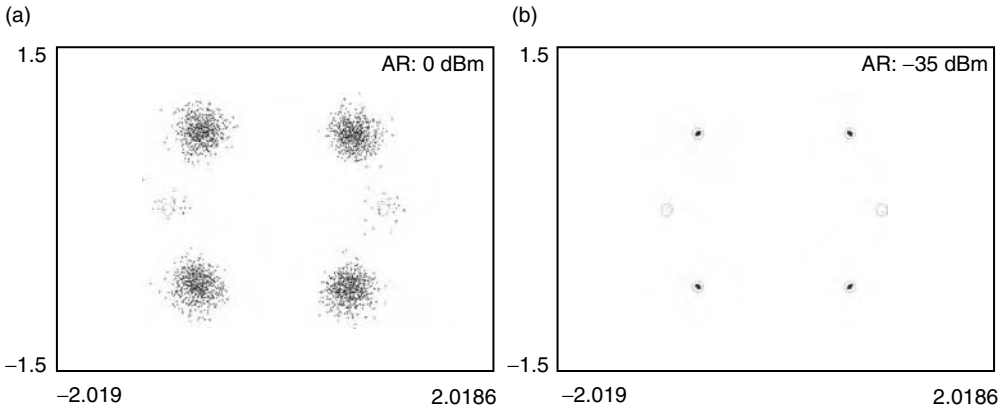


FIGURE 4.11 Illustration of uncorrected measurement-instrument-induced distortion on the constellation of a QPSK signal. Four QPSK information-payload symbol states are in quadrature with the graph's axes and two pilot symbol states are shown parallel to the horizontal axis. (a) The uncorrected measurement at left shows an increased spread on the I/Q plane for each symbol. The measurement in (b) shows the corrected signal measurement. The error vector magnitude in (a) would appear higher than it really is.

under test. The measured data in Figure 4.11a illustrates how increased spreading in a quadrature phase shift keyed (QPSK) constellation diagram can be caused by the use of uncalibrated data and a improper measurement set-up. Such spreading may cause the EVM to appear higher than that caused by the telecommunication system we are testing. Figure 4.11b illustrates a measurement of the same signal using correct instrument settings and calibrated data.

Note also that memory effects, distortion in some power-amplifier circuits arising from effects such as thermal heating, electron trapping, or bias circuit time constants, occur on a time-scale much longer than the RF period. The presence of memory effects can impede accurate determination of BER or EVM.

Calibration waveforms for instruments such as real-time signal analyzers and VSAs used in constellation diagram measurements include multisines and various pulsed waveforms. These calibrations correct for nonidealities in the instruments' subsystems, such as the mixers or samplers that convert an RF signal to baseband. These subsystems are generally quite stable over time, and so these calibrations are typically internal to the instrument and are conducted by the user or the manufacturer on an infrequent basis.

4.4 Conclusion

This chapter has dealt with methods for carrying out absolute calibrations for signal measurements. We focused on methods for finding the impulse response of instruments by use of known calibration waveforms. We described techniques for correcting the impedance mismatch between signal generators used in calibrations and the test instrument and discussed other factors that contribute to successful calibrated signal measurements. Examples illustrated the importance of absolute magnitude and phase calibration of instrumentation in measurement applications relating to RF, microwave, and high-speed digital signals.

Acknowledgments

The authors would like to acknowledge the contributions of NIST colleagues Tracy Clement, Chi-Ming Wang, Mike McKinley, Darryl Keenan, and Andrew Dienstfrey to this work.

References

1. J. Verspecht, "Return of the sampling frequency convertor" *62nd ARFTG Conf. Dig.*, Boulder, CO, Dec. 2002, pp. 155–164.
2. D. J. Ballo, J. A. Wendler, "The microwave transition analyzer—a new instrument architecture for component and signal analysis," *Hewlett-Packard J.*, vol. 43, no. 5, Oct. 1992, pp. 48–62.
3. P. Blockley, D. Gunyan, J. B. Scott, "Mixer-based, vector-corrected, vector signal/network analyzer offering 300kHz–20GHz bandwidth and traceable phase response," *2005 IEEE MTT-S Intl. Microwave Symp. Dig.*, June 12–17 2005, pp. 1497–1500.
4. P. Blockley, D. Gunyan, J. B. Scott, "Wide-bandwidth, vector-corrected measurement with high spurious-free dynamic range," *65th ARFTG Conf. Dig.*, Long Beach, CA, June 2005, pp.175–178.
5. D. F. Williams, K. A. Remley, D. C. DeGroot, "Nose-to-nose response of a 20-GHz sampling circuit," *54th ARFTG Conf. Dig.*, Atlanta, GA, Dec. 1999, pp. 64–70.
6. D. F. Williams, T. S. Clement, P. D. Hale, A. Dienstfrey, "Terminology for high-speed sampling-oscilloscope calibration," *68th ARFTG Conf. Dig.*, Broomfield, CO, Nov. 2006, pp. 9–14.
7. P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems*, Society for Industrial and Applied Mathematics, Philadelphia, 1997.
8. A. N. Tikhonov, V. Y. Arsenin, *Solutions of Ill-Posed Problems*, V. H. Winston & Sons: Washington, DC, 1977.
9. D. C. DeGroot, P. D. Hale, M. Vanden Bossche, F. Verbyst, J. Verspecht, "Analysis of interconnection networks and mismatch in the nose-to-nose calibration," *55th ARFTG Conf. Dig.*, Boston, MA, June 2000, pp. 116–121.
10. J. R. Juroshek, "A direct calibration method for measuring equivalent source mismatch," *Microwave J.*, vol. 40, Oct. 1997, pp. 106–118.
11. A. Dienstfrey, P. D. Hale, D. A. Keenan, T. S. Clement, D. F. Williams, "Minimum phase calibration of sampling oscilloscopes," *IEEE Trans. Microwave Theory Tech.*, vol. 54, no. 8, Aug. 2006, pp. 3197–3208.
12. W. L. Gans, "Dynamic calibration of waveform recorders and oscilloscopes using pulse standards," *IEEE Trans. Instrum. Meas.*, vol. 39, no. 6, Dec. 1990, pp. 952–957.
13. J. P. Deyst, N. G. Paulter, T. Daboczi, G. N. Stenbakken, T. M. Souders, "A fast-pulse oscilloscope calibration system," *IEEE Trans. Instrum. Meas.*, vol. 47, no. 5, Oct. 1998, pp. 1037–1041.
14. D. F. Williams, P. D. Hale, T. S. Clement, C.-M. Wang, "Uncertainty of the NIST electro-optic sampling system," *Natl. Inst. Stand. Technol. Tech. Note 1535*, 2005.
15. D. F. Williams, P. D. Hale, T. S. Clement, J. M. Morgan, "Calibrated 200-GHz waveform measurement," *IEEE Trans. Microwave Theory Tech.*, vol. 53, no. 4, Apr. 2005, pp. 1384–1389.
16. M. R. Harper, A. J. A. Smith, A. Basu, D. A. Humphreys, "Calibration of a 70 GHz oscilloscope," *Proceedings of the CPEM*, 2004, pp. 530–531.
17. S. Seitz, M. Bieler, M. Spitzer, K. Pierz, G. Hein, U. Siegner, "Optoelectronic measurement of the transfer function and time response of a 70 GHz sampling oscilloscope," *Meas. Sci. Tech.*, vol. 16, no. 10, Oct. 2005, pp. L7–L9.

18. J. E. Bowers, C. A. Burrus, "Ultrawide-band long-wavelength p-i-n photodetectors," *IEEE J. Lightwave Technol.*, vol. 5, no. 10, Oct. 1987, pp. 1339–1350.
19. K. Kato, "Ultrawide-band/high-frequency photodetectors," *IEEE Trans. Microwave Theory Tech.*, vol. 47, no. 7, July 1999, pp. 1265–1281.
20. B. H. Kolner, D. M Bloom, "Electrooptic sampling in GaAs integrated circuits," *IEEE J. Quantum Electron.*, vol. QE-22, no. 1, Jan. 1986, pp. 79–93.
21. P. D. Hale, D. F. Williams, "Calibrated measurement of optoelectronic frequency response," *IEEE Trans. Microwave Theory Tech.* vol. 51, no. 4, 2003, pp. 1422–1429.
22. T. S. Clement, P. D. Hale, D. F. Williams, C. M. Wang, A. Dienstfrey, D. A. Keenan, "Calibration of sampling oscilloscopes with high-speed photodiodes," *IEEE Trans. Microwave Theory Tech.*, vol. 54, no. 8, Aug. 2006, pp. 3173–3181.
23. K. Rush, S. Draving, J. Kerley, "Characterizing high-speed oscilloscopes," *IEEE Spectrum*, Sept. 1990, pp. 38–39.
24. J. Verspecht, K. Rush, "Individual characterization of broadband sampling oscilloscopes with a nose-to-nose calibration procedure," *IEEE Trans. Instrum. Meas.*, vol. 43, no. 2, Apr. 1994, pp. 347–354.
25. J. Verspecht, "Broadband sampling oscilloscope characterization with the 'nose-to-nose' calibration procedure: a theoretical and practical analysis," *IEEE Trans. Instrum. Meas.*, vol. 44, no. 6, Dec. 1995, pp. 991–997.
26. K. A. Remley, "The impact of internal sampling circuitry on the phase error of the nose-to-nose oscilloscope calibration," *Natl. Inst. Stand. Technol. Tech. Note 1528*, Aug. 2003.
27. P. D. Hale, T. S. Clement, K. J. Coakley, C. M. Wang, D. C. DeGroot, A. P. Verdoni, "Estimating magnitude and phase response of a 50 GHz sampling oscilloscope using the 'nose-to-nose' method," *55th ARFTG Conf. Dig.*, Boston, MA, June 2000, pp. 335–342.
28. A. Barel, Y. Rolain, "A microwave multisine with known phase for the calibration of narrowbanded nonlinear vectorial network analyzer measurements," *IEEE MTT-S Intl. Microwave Symp. Dig.*, vol. 4, June 7–12, 1998, pp. 1499–1502.
29. T. Van den Broeck, J. Verspecht, "Calibrated vectorial nonlinear-network analyzers," *IEEE MTT-S Intl. Microwave Symp. Dig.*, vol. 2, May 23–27, 1994, pp. 1069–1072.
30. K. A. Remley, P. D. Hale, D. I. Bergman, D. Keenan, "Comparison of multisine measurements from instrumentation capable of nonlinear system characterization," *66th ARFTG Conf. Dig.*, Washington, DC, Dec. 2005, pp. 34–43.
31. U. Lott, "Measurement of magnitude and phase of harmonics generated in nonlinear microwave two-ports," *IEEE Trans. Microwave Theory Tech.*, vol. 37, no. 10, Oct. 1989, pp. 1506–1511.
32. D. A. Humphreys, M. R. Harper, A. J. A. Smith, I. M. Smith, "Vector calibration of optical reference receivers using a frequency-domain method," *IEEE Trans. Instrum. Meas.*, vol. 54, no. 2, Apr. 2005, pp. 894–897.
33. J. C. Booth, S. A. Schima, J. A. Jargon, D. C. DeGroot, "Design and fabrication of a nonlinear phase reference device," *62nd ARFTG Conf. Dig.*, Boulder, CO, Dec. 2003, pp. 61–70.
34. J. C. Booth, K. Leong, S. A. Schima, J. A. Jargon, D. C. DeGroot, R. Schwall, "Phase-sensitive measurements of nonlinearity in high-temperature superconductor thin films," *IEEE Trans. Appl. Superconduct.*, vol. 15, no. 2, Part 1, June 2005, pp. 1000–1003.
35. *Amplitude and Phase Measurements of Frequency Translating Devices Using the HP 8510B Network Analyzer*, Product Note 8510-7, Hewlett-Packard Co., 1987.
36. *37000 Series Vector Network Analyzer: Measuring Frequency Conversion Devices*, Application Note, Anritsu Corp., April 1998.
37. G. F. Engen, *Microwave Circuit Theory*, Peter Peregrinus Ltd., London, UK, 1992.
38. D. M. Kerns, R. W. Beatty, *Basic Theory of Waveguide Junctions and Introductory Microwave Network Analysis*, Pergamon Press Inc., Oxford, 1967.
39. D. F. Williams, P. D. Hale, T. S. Clement, J. M. Morgan, "Mismatch corrections for electro-optic sampling systems," *56th ARFTG Conf. Dig.*, Broomfield, CO, Dec. 2000, pp. 141–145.
40. R. B. Marks, D. F. Williams, "A general waveguide circuit theory," *J. Res. Natl. Inst. Stand. Technol.*, vol. 97, no. 5, Sept.–Oct. 1992, pp. 533–562.

41. G. N. Stenbakken, J. P. Deyst, "Timebase nonlinearity determination using iterated sine-fit analysis," *IEEE Trans. Instrum. Meas.*, vol. 47, no. 5, Oct. 1998, pp. 1056–1061.
42. G. Vandersteen, Y. Rolain, J. Schoukens, "An identification technique for data acquisition characterization in the presence of nonlinear distortions and timebase distortions," *IEEE Trans. Instrum. Meas.*, vol. 50, no. 5, Oct. 2001, pp. 1355–1363.
43. Y. Rolain, J. Schoukens, G. Vandersteen, "Signal reconstruction for nonequidistant finite length sample sets: a 'KIS' approach," *IEEE Trans. Instrum. Meas.*, vol. 47, no. 5, Oct. 1998, pp. 1046–1052.
44. K. J. Coakley, C. M. Wang, P. D. Hale, T. S. Clement, "Adaptive characterization of jitter noise in sampled high-speed signals," *IEEE Trans. Instrum. Meas.*, vol. 52, no. 5, Oct. 2003, pp. 1537–1547.
45. W. L. Gans, "The measurement and deconvolution of time jitter in equivalent-time waveform samplers," *IEEE Trans. Instrum. Meas.*, vol. 32, no. 1, Mar. 1983, pp. 137–140.
46. P. D. Hale, C. M. Wang, D. F. Williams, K. A. Remley, J. Wepman, "Compensation of random and systematic timing errors in sampling oscilloscopes," *IEEE Trans. Instrum. Meas.*, vol. 55, no. 6, Dec. 2006, pp. 2146–2154.
47. K. J. Coakley, P. D. Hale, "Alignment of noisy signals," *IEEE Trans. Instrum. Meas.*, vol. 50, no. 1, Feb. 2001, pp. 141–149.
48. M. D. McKinley, K. A. Remley, M. Myslinski, J. S. Kenney, D. Schreurs, B. Nauwelaers, "EVM calculation for broadband modulated signals," *64th ARFTG Conf. Dig.*, Orlando, FL, Dec. 2004, pp. 45–52.
49. J. Verspecht, P. Debie, A. Barel, L. Martens, "Accurate on wafer measurements of phase and amplitude of the spectral components of incident and scattered voltage waves at the signal ports of a nonlinear microwave device," *IEEE MTT-S Intl. Microwave Symp. Dig.*, Orlando, FL, May 16–20, 1995, pp. 1029–1032.
50. *Large-signal network analyzer technology, Preliminary Product Overview*, Maury Microwave Corporation, May 2002.
51. J. Dunsmore, "Novel method for vector mixer characterization and mixer test system vector error correction," *IEEE MTT-S Intl. Microwave Symp. Dig.*, vol. 3, June 2002, pp. 1833–1836.
52. C. J. Clark, A. A. Moulthrop, M. S. Muha, C. P. Silva, "Transmission response measurements of frequency-translating devices using a vector network analyzer," *IEEE Trans. Microwave Theory Tech.*, vol. 44, no. 12, 1996, pp. 2724–2737.
53. D. R. Thornton, "A Simple VNA method for mixer conversion loss measurement," *Microwave J.*, vol. 40, no. 3, Mar. 1997, pp. 78–86.
54. H. C. Torrey, C. A. Whitmer, *Crystal Rectifiers,* McGraw-Hill, New York, 1948.
55. S. Maas, *Microwave Mixers*, Artech House, Boston, 1992.
56. B. Roth, D. Kother, M. Coady, T. Sporkmann, C. Sattler, "Applying a conventional VNA to nonlinear measurements without using frequency converting standards," in *3rd Int. Integr. Nonlinear Microwave Millimeterwave Circuits Workshop*, May 1994, pp. 243–252.
57. D. F. Williams, H. Khenissi, F. Ndagijimana, K. A. Remley, J. P. Dunsmore, P. D. Hale, C. M. Wang, T. S. Clement, "Sampling-oscilloscope measurement of a microwave mixer with single-digit phase accuracy," *IEEE Trans. Microwave Theory Tech.*, vol. 54, no. 3, Mar. 2006, pp. 1210–1217.
58. A. Cidronali, G. Loglio, J. Jargon, K. A. Remley, I. Magrini, D. DeGroot, D. Schreurs, K. C. Gupta, G. Manes, "RF and IF mixer optimum matching impedances extracted by large-signal vectorial measurements," in *Proc. 11th GaAs Symp.*, Munich, Germany, 2003, pp. 61–64.
59. P. Cochrane, "Digital system measurements," in *Microwave Measurements*, 2nd Edn., A. E. Bailey, Ed., Peter Peregrinus Ltd., London, UK, 1989.
60. S. W. Hinch, C. M. Miller, "Analysis of digital modulation on optical carriers," in *Fiber Optic Test and Measurement*, D. Derickson, Ed., Prentice Hall PTR, New Jersey, 1998.
61. J. B. Scott, J. Verspecht, B. Behnia, M. Vanden Bossche, A. Cognata, F. Verbeyst, M. L. Thorn, D. R. Scherrer, "Enhanced on-wafer time-domain waveform measurement through removal of interconnect dispersion and measurement instrument jitter," *IEEE Trans. Microwave Theory Tech.*, vol. 50, no. 12, Dec. 2002, pp. 3022–3028.

5

Noise Measurements

Alfy Riddle
Finesse, LLC

5.1 Fundamentals of Noise

5.1.1 Statistics

Noise is a random process. There may be nonrandom system disturbances we call noise, but this chapter will consider noise as a random process. Noise can have many different sources such as thermally generated resistive noise, charge crossing a potential barrier, and generation–recombination (G–R) noise [1]. The different noise sources are described by different statistics, the thermal noise in a resistor is a Gaussian process while the shot noise in a diode is a Poisson process. In the cases considered here, the number of noise "events" will be so large that all noise processes will have essentially Gaussian statistics and so be represented by the probability distribution in Equation 5.1

$$p(x) = 1/\left(2\pi\sigma^2\right) e^{-x^2/2\sigma^2} \tag{5.1}$$

The statistics of noise are essential for determining the results of passing noise through nonlinearities because the nonlinearity will change the noise distribution [2]. Noise statistics are useful even in linear networks because multiple noise sources will require correlation between the noise sources to find the total noise power. Linear networks will not change the statistics of a noise signal even if the noise spectrum is changed.

5.1.2 Bandwidth

The noise energy available from a hot resistor is given in Equation 5.2, where $h = 6.62 \times 10^{-34}$ J s, T is in degrees Kelvin, and $k = 1.38 \times 10^{-23}$ J/°K [1]. N is in joules, or watt-seconds, or W/Hz, which is noise power spectral density. For most of the microwave spectrum $hf \ll kT$ so Equation 5.2 reduces to Equation 5.3.

$$N = hf/\left(e^{hf/kT} - 1\right) \tag{5.2}$$

$$N \cong kT \qquad f \ll kT/h \tag{5.3}$$

The noise power available from the hot resistor will be the integration of this energy, or spectral density, over the measuring bandwidth as given in Equation 5.4.

$$P = \int_{f1}^{f2} N \, df \tag{5.4}$$

As the frequency increases, N reduces so the integration in Equation 5.4 will be finite even if the frequency range is infinite. Note that for microwave networks using cooled circuits, quantum effects can become important at relatively low frequencies because of the temperature-dependent condition in Equation 5.3. For a resistor at microwave frequencies and room temperature, N is independent of frequency so the total power available is simply $P = kT(f_2 - f_1)$, or $P = kTB$, as shown in Equation 5.5, where B is the bandwidth. Figure 5.1 shows a resistor with an available thermal power of kTB, which can be represented either as a series voltage source with $e_n^2 = 4\,kTRB$ or a shunt current source with $i_n^2 = 4\,kTB/R$, where the squared value is taken to be the mean-square value. At times it is tempting to represent e_n as $\sqrt{(4\,kTB)}$, but this is a mistake because e_n is a random variable, not a sinusoid. The process of computing the mean-square value of a noise source is important for establishing any possible correlation with any other noise source in the system [1,16]. Representing a noise source as an equivalent sinusoidal voltage can result in an error due to incorrect accounting of correlation.

$$P = kT\,B \qquad f \ll kT/h \tag{5.5}$$

When noise passes through a filter we must repeat the integration of Equation 5.4. Two useful concepts in noise measurement are noise power per hertz and equivalent noise bandwidth. Noise power per hertz is simply the spectral density of the noise, or N in the above equations because it has units of watt-seconds or joules. Spectral densities are also given in V^2/Hz and A^2/Hz. The equivalent noise bandwidth of a noise source can be found by dividing the total power detected by the maximum power detected per hertz, as shown in Equation 5.6.

$$B_e = P / \mathrm{Max}\{N\} \tag{5.6}$$

The noise equivalent bandwidth of a filter is especially useful when measuring noise sources with a spectrum analyzer. The noise equivalent bandwidth of a filter is defined by integrating its power transfer function, $|H(f)|^2$, overall frequency and dividing by the peak of the power transfer function, as shown in Equation 5.7.

$$B_e = \int_0^\infty |H(f)|^2 \, df \Big/ \mathrm{Max}\{|H(f)|^2\} \tag{5.7}$$

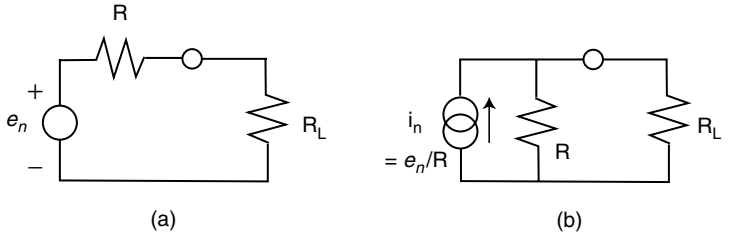


FIGURE 5.1 Equivalent thermal noise sources: (a) voltage and (b) current.

Power meters are often used with bandpass filters in noise measurements so that the noise power has a well-defined range. The noise equivalent bandwidth of the filter can be used to convert the noise power back to a power per hertz spectral density that is easier to use in computations and comparisons. As an example, a first-order bandpass filter has a B_e = ¼/2 B_{-3}, where B_{-3} is the –3 dB bandwidth. The noise equivalent bandwidth is greater than the 3 dB bandwidth because of the finite power in the filter skirts.

5.2 Detection

The most accurate and traceable measurement of noise power is by comparison with thermal standards [3]. In the everyday lab the second best method for noise measurement is a calibrated power meter preceded by a filter of known noise bandwidth. Because of its convenience, the most common method of noise power measurement is a spectrum analyzer. This most common method is also the most inaccurate because of the inherent inaccuracy of a spectrum analyzer and because of the nonlinear processes used in a spectrum analyzer for power estimation. As mentioned in the chapter on statistics, nonlinearities change the statistics of a noise source. For example, Gaussian noise run through a linear envelope detector acquires a Rayleigh distribution as shown in Figure 5.2.

The average of the Rayleigh distribution is not the standard deviation of the Gaussian, so a detector calibrated for sine waves will read about 1 dB high for noise. Spectrum analyzers also use a logarithmic amplifier that further distorts the noise statistics and accounts for another 1.5 dB of error. Many modern spectrum analyzers automatically correct for these nonlinear errors as well as equivalent bandwidth when put in a "Marker Noise" mode [4].

5.3 Noise Figure and *Y*-Factor Method

At high frequencies it is far easier to measure power flow than it is to measure individual voltage and current noise sources. All of a linear device's noise power can be considered as concentrated at its input as shown in Figure 5.3 [1].

We can lump all of the amplifier noise generators into an equivalent noise temperature with an equivalent input noise power per hertz of $N_a = kT_e$. As shown in Figure 5.3, the noise power per hertz available from the noise source is $N_S = kT_S$. In system applications the degradation of signal-to-noise ratio (SNR) is a primary concern. We can define a figure of merit for an amplifier, called the noise factor (F), which describes the reduction in SNR of a signal passed through the amplifier shown in Figure 5.3 [5]. The noise factor for an amplifier is derived in Equation 5.8

$$F = SNR_{IN}/SNR_{OUT} = S_{IN}/kT_S/\left(G_a S_{IN}/\left(G_a k\left(T_S + T_e\right)\right)\right) = \left(T_S + T_e\right)/T_S = 1 + T_e/T \qquad (5.8)$$

Equation 5.8 is very simple and only contains the amplifier equivalent temperature and the source temperature. *F* does vary with frequency and so is measured in a narrow bandwidth, or spot. Note that F is not a function of measurement bandwidth. Equation 5.8 also implies that the network is tuned for maximum available gain, which happens by default if all the components are perfectly matched to 50 Ω and used in a 50-Ω system.

Device noise factor can be measured with the setup shown in Figure 5.4 [1]. The *Y*-factor method takes advantage of the fact that as the source temperature is varied, the device noise output, N_O, varies yet the device noise contribution remains a constant. Figure 5.5 shows that as T_S changes the noise power measured as the power meter changes according to Equation 5.9

$$N_O\left(T_S\right) = \left(k\, T_S\, G_a + k\, T_e\, G_a\right)B \qquad (5.9)$$

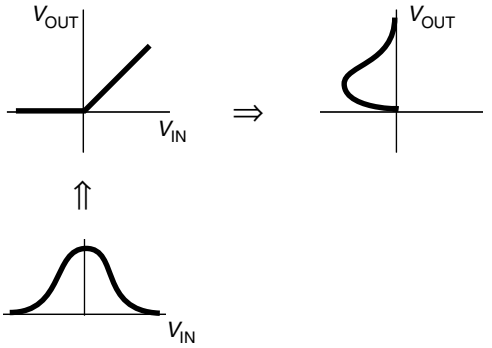


FIGURE 5.2 Nonlinear transformation of Gaussian noise.

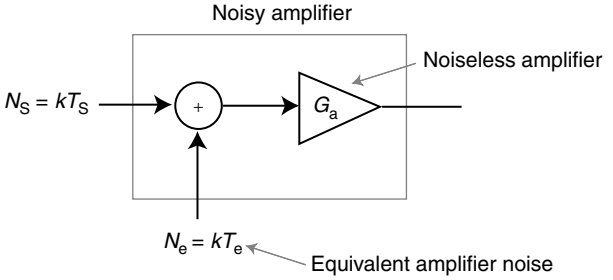


FIGURE 5.3 System view of amplifier noise.

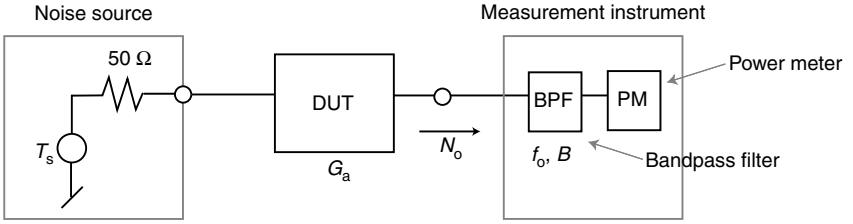


FIGURE 5.4 Test setup for Y-factor method.

The value of N_O ($T_S = 0$) gives the noise power of the device alone. By using two known values of T_S, a cold measurement at $T_S = T_{COLD}$, and a hot measurement at $T_S = T_{HOT}$, the slope of the line in Figure 5.5 can be derived. Once the slope is known, the intercept at $T_S = 0$ can be found by measuring $N_O(T_{COLD})$ and $N_O(T_{HOT})$. The room temperature, T_O, is also needed to serve as a reference temperature for the device noise factor, F. For a room temperature $F = 1 + T_e/T_O$, we can define $T_e = (F - 1)\ T_O$. The following equations derive the Y-factor method. Equation 5.10 is the basis for the Y-factor method.

$$Y = \frac{N_O(T_{HOT})}{N_O(T_{COLD})} = \frac{(k\,T_{HOT}\,G_a + (F-1)k\,T_O\,G_a)B}{(k\,T_{COLD}\,G_a + (F-1)k\,T_O\,G_a)B} = \frac{T_{HOT} + (F-1)\,T_O}{T_{COLD} + (F-1)\,T_O} \tag{5.10}$$

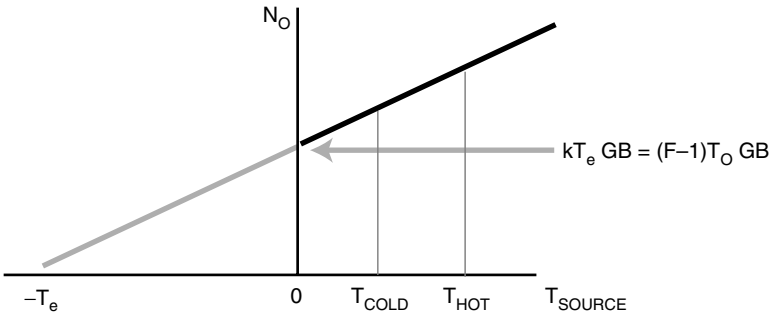


FIGURE 5.5 Output noise power versus source temperature.

Solving for F we get Equation 5.11 which can be solved for T_e as shown in Equation 5.12.

$$F = 1 + \frac{T_{HOT} - Y\,T_{COLD}}{(Y-1)T_O} = 1 + T_e/T_O \tag{5.11}$$

$$T_e = \frac{T_{HOT} - Y\,T_{COLD}}{(Y-1)} \tag{5.12}$$

Equation 5.12 can be rearranged to define another useful parameter known as the equivalent noise ratio, or ENR, of a noise source as shown in Equation 5.13 [1].

$$F = \frac{(T_{HOT}/T_O - 1) + Y(1 - T_{COLD}/T_O)}{(Y-1)} = \frac{ENR + Y(1 - T_{COLD}/T_O)}{(Y-1)} = \left.\frac{ENR}{(Y-1)}\right|_{T_{COLD} = T_O} \tag{5.13}$$

Note that when T_{COLD} is set to the reference temperature for *F*, which the IEEE gives as $T_O = 290°$ Kelvin, then the device noise factor has a simple relationship to both ENR and Y [1,3].

Practically speaking, the noise factor is usually given in decibels and called the noise figure, NF = 10 log *F*. While the most accurate noise sources use variable temperature loads, the most convenient variable noise sources use avalanche diodes with calibrated noise power versus bias current [6]. The diode noise sources usually contain an internal pad to reduce the impedance variation between on (hot) and off (cold) states. Also, the diodes come with an ENR versus frequency calibration curve.

5.4 Phase Noise and Jitter

5.4.1 Introduction

The noise we have been discussing was broadband noise. When noise is referenced to a carrier frequency it appears to modulate the carrier and so causes amplitude and phase variations in the carrier [7]. Because of the amplitude-limiting mechanism in an oscillator, oscillator phase-modulation (PM) noise is much larger than amplitude-modulation (AM) noise close enough to the carrier to be within the oscillator loop

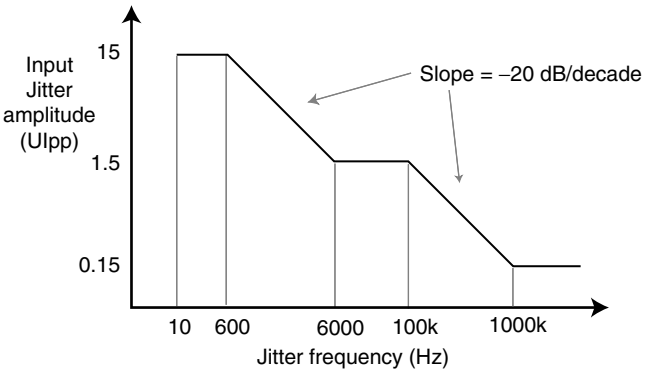


FIGURE 5.6 SONET category II jitter tolerance mask for OC-48.

bandwidth. The phase variations, caused by the noise at different offset, or modulation frequencies create a variance in the zero crossing time of the oscillator. This zero crossing variance in the time domain is called jitter and is critical in digital communication systems. Paradoxically, even though jitter is easily measured in the time domain and often defined in picoseconds it turns out to be better to specify jitter in the frequency domain as demonstrated by the jitter tolerance mask for an OC-48 SONET signal [8]. The jitter plot shown in Figure 5.6 can be translated into script *L* versus frequency using the equations in the following section [9].

Jitter is best specified in the frequency domain because systems are more sensitive to some jitter frequencies than others. Also, jitter attenuators, which are simply narrowband phase-locked loops (PLLs), have well-defined frequency-domain transfer functions that can be cascaded with the measured input jitter to derive the output jitter.

$$V_O(t) = V_C\left\{1 + m(t)\right\}\cos\left[\omega_C t + \beta(t)\right] \tag{5.14}$$

5.4.2 Mathematical Basics

Consider the time-domain voltage given in Equation 5.14. This signal contains both amplitude modulation, m(t), and phase modulation, β(t) [10]. If we let $m(t) = m_1(t)\cos(\omega_m t)$, $\beta(t) = \beta_1(t)\sin(\omega_m t)$, and we define $|\beta_1(t)| \ll 1$, then Equation 5.14 can be expanded into AM and PM sidebands as shown below:

$$\begin{aligned} V_O(t) \cong\ & V_C\cos\left[\omega_C t\right] \\ & + V_C m_1(t)/2\left\{\cos\left[\omega_C t + \omega_m t\right] + \cos\left[\omega_C t - \omega_m t\right]\right\} \\ & + V_C \beta_1(t)/2\left\{\cos\left[\omega_C t + \omega_m t\right] - \cos\left[\omega_C t - \omega_m t\right]\right\} \end{aligned} \tag{5.15}$$

We can let $m_1(t)$ and $\beta_1(t)$ be fixed amplitudes as when sinusoidal test signals are used to characterize an oscillator, or we can let $m_1(t)$ and $\beta_1(t)$ be slowly varying, with respect to ω_m, noise signals. The latter case gives us the narrowband Gaussian noise approximation, which can represent an oscillator spectrum when the noise signals are summed over all modulation frequencies, ω_m [7].

Several notes should be made here. First, as $\beta_1(t)$ becomes large the single sidebands of Equation 5.15 expand into a Bessel series that ultimately generates a flat-topped spectrum close into the average carrier frequency. This flat-topped spectrum is essentially the FM spectrum created by the large phase excursions

that result from the $1/f^3$ increase in phase noise at low modulating frequencies. Second, if $|m_1(t)| = |\beta_1(t)|$, and they are fully correlated, then by altering the phases between $m_1(t)$ and $\beta_1(t)$ we can cancel the upper or lower sideband at will. This second point also shows that a single sideband contains equal amounts of AM and PM, which is useful for testing and calibration purposes [11].

Oscillator noise analysis uses several standard terms such as AM spectral density, PM spectral density, FM spectral density, script L, and jitter [10,12]. These terms are defined in Equations 5.16 through 5.20. The AM spectral density, or $S_{AM}(f_m)$, shown in Equation 5.16 is derived by computing the power spectrum of m(t), given in Equation 5.14, with a 1-hertz-wide filter. S_{AM} is called a spectral density because it is on a 1-hertz basis. Similarly, the PM spectral density, or $S_\phi(f_m)$ in radians²/Hz, is shown in Equation 5.17 and is derived by computing the power spectrum of $\beta(t)$ in Equation 5.14. The FM spectral density, or $S_{FM}(f_m)$ in *Hz²/Hz*, is typically derived by using a frequency discriminator to measure the frequency deviations in a signal. Because frequency is simply the rate of change of phase, FM spectral density can be derived from PM spectral density as shown in Equation 5.18. Script L is a measured quantity usually given in dBc/Hz and best described by Figure 5.7. It is important to remember that the definition of script L requires the sidebands to be due to phase noise. Because script L is defined as a measure of phase noise it can be related to the PM spectral density as shown in Equation 5.19. Two complications arise in using script L. First, most spectrum analyzers do not determine if the sidebands are only due to phase noise. Second, the constant relating script L to S_ϕ is 2 if the sidebands are correlated and $\sqrt{2}$ if the sidebands are uncorrelated and the spectrum analyzer does not help in telling these two cases apart. Jitter is simply the rms value of the variation in zero crossing times of a signal compared with a reference of the same average frequency. Of course, the jitter of a signal can be derived by accumulating the phase noise as shown in Equation 5.20. In Equations 5.16 through 5.20 $\Im\{x\}$ denotes the Fourier transform of x [7]. Most of these terms can also be defined from the Fourier transform of the autocorrelation of m(t) or $\beta(t)$. The spectral densities are typically given in dB using a 1-Hz measurement bandwidth, abbreviated as dB/Hz.

$$S_{AM}(f_m) = \Im\{m(t)\}\,\Im\{m^*(t)\} = m_1^2/2\,\delta(f_m - f_a)\Big|_{m(t)=m_1\cos(\omega_a t)} \tag{5.16}$$

$$S_\phi(f_m) = \Im\{\beta(t)\}\,\Im\{\beta^*(t)\} = \beta_1^2/2\,\delta(f_m - f_a)\Big|_{\beta(t)=\beta_1\cos(\omega_a t)} \tag{5.17}$$

$$S_{FM}(f_m) = f_m^2\,S_\phi(f_m) \qquad \text{because } \omega_m = d\phi/dt \tag{5.18}$$

$$\text{script L}(f_m) = P_{SSB}(f_m)/\text{Hz}/P_C = S_\phi(f_m)/2\Big|_{\text{when phase noise has correlated sidebands}} \tag{5.19}$$

$$\text{jitter} = \sqrt{\left(\int_0^\infty S_\phi(f_m)\,df_m\right)} = \beta_1/\sqrt{2}\Big|_{\beta(t)=\beta_1\cos(\omega_m t)} \tag{5.20}$$

In the above equations f_m indicates the offset frequency from the carrier. In Equation 5.19 P_{SSB} is defined as phase noise, but often is just the noise measured by a spectrum analyzer close to the carrier frequency, and P_C is the total oscillator power. The jitter given in Equation 5.20 is the total jitter that results from a time-domain measurement. Jitter as a function of frequency, f_m, is just the square root of $S_\phi(f_m)$. Jitter as a function of frequency can be translated to various other formats, such as degrees, radians, seconds, and unit intervals (UIs), using Equation 5.21 [9].

$$\text{jitter} = \begin{Bmatrix} \beta_{RMS} \text{ in radians} \\ UI = \beta_{RMS}/2\pi \text{ in unit intervals} \\ UI/f_C = \beta_{RMS}/\omega_C \text{ in seconds} \\ 360\ UI = 360\beta_{RMS}/2\pi \text{ in degrees} \end{Bmatrix} \tag{5.21}$$

For most free-running oscillators the $1/f^3$ region of the phase noise dominates the jitter so integrating the $1/f^3$ slope gives UI $\approx f_a/2\ 10^{\text{scriptL}(f_a)/10}$ where f_a is any frequency on the $1/f^3$ slope and script L is in dBc/Hz. For PLL-based sources with large noise pedestals, a complete integration should be done.

5.4.3 Phase Noise Measurements

Phase noise is typically measured in one of three ways: spectrum analyzer, PLL, or transmission line discriminator [1,10,13–15]. The spectrum analyzer is the easiest method of measuring script $L(f_m)$ for any oscillator noisier than the spectrum analyzer reference source. Figure 5.7 shows a typical source spectrum. Care must be taken to make the resolution bandwidth, RBW, narrow enough to not cover a significant slope of the measured noise [1]. The spectrum analyzer cannot distinguish between phase and amplitude noise, so reporting the results as script L only holds where $S_\phi > S_{AM}$, which usually means within the $1/f^3$ region of the source. Spectrum analyzer measurements can be very tedious when the oscillator is noisy enough to wander significantly in frequency.

PLL-based phase noise measurement is used in most commercial systems [14]. Figure 5.8 shows a PLL-based phase noise test set. The reference oscillator in Figure 5.8 is phase locked to the device under test (DUT) through a low pass filter (LPF) with a cutoff frequency well below the lowest desired measurement frequency. This allows the reference oscillator to track the DUT and downconvert the phase noise sidebands without tracking the noise as well. The low-frequency spectrum analyzer measures the noise sidebands and arrives at a phase noise spectral density by factoring in the mixer loss or using a calibration tone [11]. A PLL system requires the reference source to be at least as quiet as the DUT. Another DUT can be used as a reference with the resulting noise sidebands increasing by 3 dB, but usually the reference is much quieter than the DUT so fewer corrections have to be made.

A transmission line frequency discriminator can provide accurate and high-resolution phase noise measurements without the need for a reference oscillator [13]. The discriminator resolution is proportional to the delay line delay, τ. The phase shifter is adjusted so that the mixer signals are in quadrature, which means the mixer DC output voltage is set to the internal offset voltage (approximately zero).

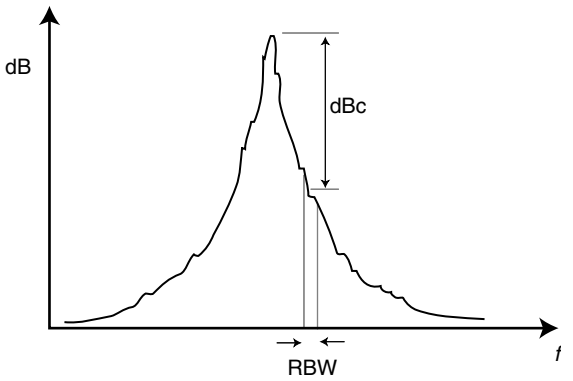


FIGURE 5.7 Typical measured spectrum on a spectrum analyzer.

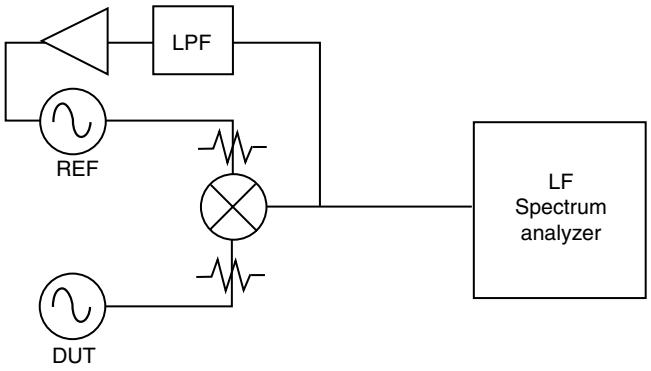


FIGURE 5.8 PLL phase noise measurement.

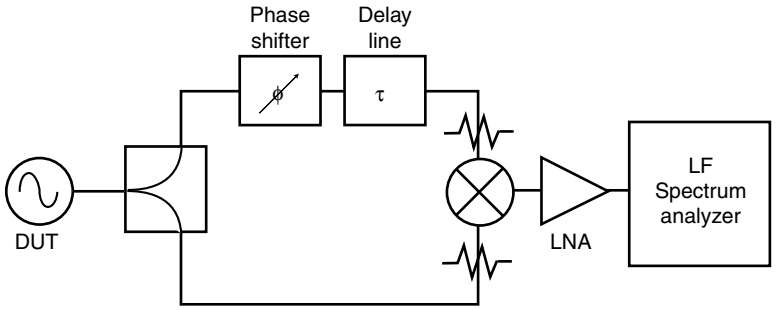


FIGURE 5.9 Transmission line discriminator.

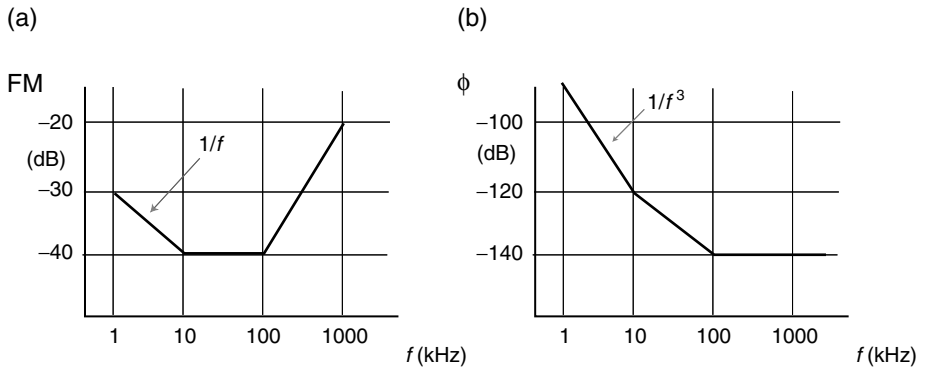


FIGURE 5.10 (a) FM and (b) phase noise spectral densities for the same device.

Transmission line discriminators can be calibrated with an offset source of known amplitude, as discussed previously, or with a source of known modulation sensitivity [11] (see Figure 5.9). The disadvantages of a transmission line discriminator are that high source output levels are required to drive the system (typically greater than 13 dBm), and the system must be retuned as the DUT drifts. Also, it is important to remember that the discriminator detects FM noise which is related to phase noise as given in Equation 5.18 and shown in Figure 5.10.

5.5 Summary

Accurate noise measurement and analysis must recognize that noise is a random process. While nonlinear devices will affect the noise statistics, linear networks will not change the noise statistics. Noise statistics are also important for analyzing multiple noise sources because the correlation between the noise sources must be considered. At very high frequencies it is easier to work with noise power flow than individual noise voltage and current sources, so methods such as the *Y*-factor technique have been developed for amplifier noise figure measurement. Measuring oscillator noise mostly involves the phase variations of a source. These phase variations can be represented in the frequency domain as script L, or in the time domain as jitter. Several techniques of measuring source phase noise have been developed which trade off accuracy for cost and simplicity.

References

1. Ambrozy, A., *Electronic Noise*, McGraw-Hill, New York, 1982.
2. Papoulis, A., *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, New York, 1965.
3. Haus, H.A., IRE Standards on Methods of Measuring Noise in Linear Twoports, *Proc. IRE*, 60–68, Jan. 1960.
4. Staff, *HP 8560 E-Series Spectrum Analyzer User's Guide*, Hewlett-Packard, Dec. 1997.
5. Friis, H.T., Noise Figures of Radio Receivers, *Proc. IRE*, 419–422, July 1944.
6. Pastori, W.E., A Review of Noise Figure Instrumentation, *Microwave J.,* 50–60, Apr. 1983.
7. Carlson, B.A., *Communication Systems*, McGraw-Hill, New York, 1975.
8. ANSI, Telecommunications-Synchronous Optical Network (SONET)- Jitter at Network Interfaces, *ANSI T1.105.03-1994*, 1994.
9. Adler, J.V., Clock Source Jitter: A Clear Understanding Aids Clock Source Selection, *EDN*, 79–86, Feb. 18, 1999.
10. Ondria, J.G., A Microwave System for Measurement of AM and FM Noise Spectra, *IEEE Trans. MTT*, 767–781, Sept. 1968.
11. Buck, J.R., and Healey, D.J. III, Calibration of Short-Term Frequency Stability Measuring Apparatus, *Proc. IEEE*, 305–306, Feb. 1966.
12. Blair, B.E., *Time and Frequency: Theory and Fundamentals*, U.S. Dept. Commerce, NBS Monograph 140, 1974.
13. Schielbold, C., Theory and Design of the Delay Line Discriminator for Phase Noise Measurement, *Microwave J.,* 103–120, Dec. 1983.
14. Harrison, D.M., Howes, M.J., and Pollard, R.D., The Evaluation of Phase Noise in Low Noise Oscillators, *IEEE MTT-S Digest*, 521–524, 1987.
15. Staff, *Noise Measurements Using the Spectrum Analyzer, Part One: Random Noise*, Tektronix, Beaverton, 1975.
16. Haus, H.A. and Adler, R.B., *Circuit Theory of Linear Noisy Networks*, MIT Press, Cambridge, MA, 1959.

6

Nonlinear Microwave Measurement and Characterization

J. Stevenson Kenney
Georgia Institute of Technology

While powerful methods have been developed to analyze complex linear circuits, it is unfortunate that almost all physical systems exhibit some form of nonlinear behavior. Often the nonlinear behavior of a microwave circuit is detrimental to the signals that pass through it. Such is the case with distortion within a microwave power amplifier. In some cases nonlinearities may be exploited to realize useful circuit functions, such as frequency translation or detection. In either case, methods have been devised to characterize and measure nonlinear effects on various signals. These effects are treated in this chapter and include:

- Harmonic Distortion
- Gain Compression
- Intermodulation Distortion

- Phase Distortion
- Adjacent Channel Interference
- Error Vector Magnitude

Many of the above characterizations are different manifestations of nonlinear behavior for different types of signals. For instance, both analog and digital communication systems are affected by *intermodulation distortion* (IMD). However, these effects are usually measured in different ways. Nevertheless, some standard measurements are used as figures of merit for comparing the performance to different circuits. These include

- Output Power at 1 dB Gain Compression
- Third Order Intercept Point
- Spurious Free Dynamic Range
- Noise Power Ratio
- Spectral Mask Measurements

This chapter treats the characterization and measurement of nonlinearities in microwave circuits. The concentration will be on standard techniques for analog and digital communication circuits. For more advanced techniques, the reader is advised to consult the references at the end of this chapter.

6.1 Mathematical Characterization of Nonlinear Circuits

To analyze the effects of nonlinearities in microwave circuits, one must be able to describe the input–output relationships of signals that pass through them. Nonlinear circuits are generally characterized by input–output relationships called *transfer characteristics*. In general, any memoryless circuit described by transfer characteristics that does not satisfy the following definition of a *linear* memoryless circuit is said to be *nonlinear:*

$$v_{out}(t) = Av_{in}(t), \tag{6.1}$$

where v_{in} and v_{out} are the input and output time-domain waveforms and A is a constant independent of time. Thus, one form of a nonlinear circuit has a transfer characteristic of the form

$$v_{out}(t) = g\left[v_{in}(t)\right]. \tag{6.2}$$

The form of $g(v)$ will determine all measurable distortion characteristics of a nonlinear circuit. Special cases of nonlinear transfer characteristics include

- *Time invariant*: g does not depend on t
- *Memoryless*: g is evaluated at time t using only values of v_{in} at time t

6.1.1 Nonlinear Memoryless Circuits

If a transfer characteristic includes no integrals, differentials, or finite time differences, then the instantaneous value at a time t depends only on the input values at time t. Such a transfer characteristic is said to be *memoryless*, and may be expressed in the form of a power series

$$g(v) = g_0 + g_1 v + g_2 v^2 + g_3 v^3 + \ldots \tag{6.3a}$$

where g_n are real-valued, time-invariant coefficients. Frequency domain analysis of the output signal $v_{out}(t)$ where $g(v)$ is expressed by Equation 6.3a yields a Fourier series, whereby the harmonic components are governed by the coefficients G_n. If $v_{in}(t)$ is a sinusoidal function at frequency f_c with amplitude V_{in}, then the output signal is a harmonic series of the form

$$v_{out}(t) = G_0 + G_1 V_{in} \cos(2\pi f_c t) + G_2 V_{in}^2 \cos(4\pi f_c t) + G_3 V_{in}^3 \cos(6\pi f_c t) + \ldots \tag{6.3b}$$

The coefficients, G_n are functions of the coefficients g_n, and are all real. The extent that the coefficients g_n are nonzero is called the *order* of the nonlinearity. Thus, from Equation 6.3b, it is seen that an nth order system will produce harmonics of nth order of amplitude $G_n V_{in}^n$.

6.1.2 Nonlinear Circuits with Memory

As described in Equation 6.3a, $g(v)$ is said to be *memoryless* because the output signal at a time t depends only on the input signal at time t. If the output depends on the input at times different from time t, the nonlinearity is said to have *memory*. A nonlinear function with a finite memory (i.e., a *finite impulse response*) may be described as

$$v_{out}(t) = g\left[v_{in}(t), v_{in}(t-\tau_1), v_{in}(t-\tau_2), \ldots v_{in}(t-\tau_n)\right]. \tag{6.4}$$

The largest time delay, τ_n, determines the length of the memory of the circuit. *Infinite impulse response* nonlinear systems may be represented as functions of integrals and differentials of the input signal

$$v_{out}(t) = g\left[v_{in}(t), \int_{-\infty}^{t} v_{in}(\tau)d\tau, \frac{\partial^n v_{in}}{\partial t^n}\right]. \tag{6.5}$$

The most general characterization of a nonlinear system is the *Volterra Series* [1]. Consider a linear circuit that is stimulated by an input signal $v_{in}(t)$. The output signal $v_{out}(t)$ is then given by the convolution with the input signal $v_{in}(t)$ and the *impulse response* $h(t)$. Unless the impulse response takes the form of the *delta function* $\delta(t)$, the output $v_{out}(t)$ depends on values of the input $v_{in}(t)$ at times other than t, that is, the circuit is said to have memory.

$$v_{out}(t) = \int_{-\infty}^{\infty} v_{in}(\tau)h(t-\tau)d\tau. \tag{6.6a}$$

Equivalently, in the frequency domain,

$$V_{out}(f) = V_{in}(f)H(f). \tag{6.6b}$$

In the most general case, a nonlinear circuit with reactive elements can be described using a Volterra series, which is said to be a power series with *memory*.

$$v_{out}(t) = g_0 + \int_{-\infty}^{\infty} v_{in}(\tau)g_1(t-\tau)d\tau + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} v_{in}(t-\tau_1)v_{in}(t-\tau_2)g_2(\tau_1,\tau_2)d\tau_1 d\tau_2 + \ldots \tag{6.7a}$$

An equivalent representation is obtained by taking the n-fold Fourier transform of Equation 6.7a

$$V_{out}(f_1, f_2, \ldots) = G_0\delta(f_1) + G_1(f_1)V_{in}(f_1) + G_2(f_1, f_2) + \ldots \tag{6.7b}$$

Notice that the Volterra series is applicable to nonlinear effects on signals with discrete spectra (i.e., a signal consisting of a sum of sinusoids). For instance, the DC component of the output signal is given by $g_0 = G_0$, while the fundamental component is given by $G_1(f_1)V_{in}(f_1)$, where G_1 and V_{in} are the Fourier transforms of the impulse response g_1 and v_{in}, respectively, evaluated at frequency f_1. The higher order terms in the Volterra series represent the harmonic responses and intermodulation response of the circuit.

Fortunately, extraction of high order Volterra series representations of nonlinear microwave circuits is rarely required to gain useful information on the deleterious and/or useful effects of distortion on common signals. Such simplifications often involve considering the circuit to be memoryless, as in Equation 6.3a,b, or having finite order, or having integral representations, as in Equation 6.5.

6.2 Harmonic Distortion

A fundamental result of the distortion of nonlinear circuits is that they generate frequency components in the output signal that are not present in the input signal. For sinusoidal inputs, the salient characteristic is *harmonic distortion,* whereby signal outputs consist of integer multiples of the input frequency.

6.2.1 Harmonic Generation in Nonlinear Circuits

As far as microwave circuits are concerned, the major characteristic of a nonlinear circuit is that the frequency components of the output signal differ from those of the input signal. This is readily seen by examining the output of a sinusoidal input from Equation 6.8.

$$\begin{aligned} v_{out}(t) &= g_0 + g_1 A\cos(2\pi f_c t) + g_2 A^2\cos^2(2\pi f_c t) + g_3 A^3\cos^3(2\pi f_c t) + \cdots \\ &= g_0 + \frac{g_2 A^2}{2} + \left(g_1 A + \frac{3g_3 A^3}{4}\right)\cos(2\pi f_c t) + \frac{ag_2 A^2}{2}\cos(4\pi f_c t) + \frac{g_3 A^3}{4}\cos(6\pi f_c t) + \cdots \end{aligned} \tag{6.8}$$

It is readily seen that, along with the *fundamental* component at a frequency of f_c, there exists a DC component, and *harmonic* components at integer multiples of f_c. The output signal is said to have acquired *harmonic distortion* as a result of the nonlinear transfer characteristic. This is illustrated in Figure 6.1. The function represented by $g(v)$ is that of an ideal limiting amplifier. The net effect of the terms are summarized in Table 6.1.

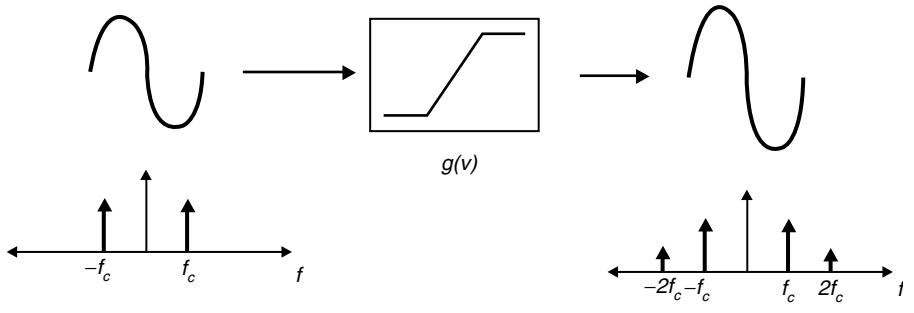


FIGURE 6.1 Effects of a nonlinear transfer characteristic on a sinusoidal input: harmonic distortion.

TABLE 6.1 Effect of Nonlinearities on Carrier Term by Term

Term	Amplitude	Qualitative Effect
DC	$g_0 + g_2A^2/2$	Small offset added due to RF detection
Fundamental	$20\log(g_1A + 3\,g_3A^3/4)$	Amplitude changed due to compression
2nd Harmonic	$40\log(g_2A^2/2)$	2:1 slope on P_{in}/P_{out} curve
3rd Harmonic	$60\log(g_3A^3/4)$	3:1 slope on P_{in}/P_{out} curve

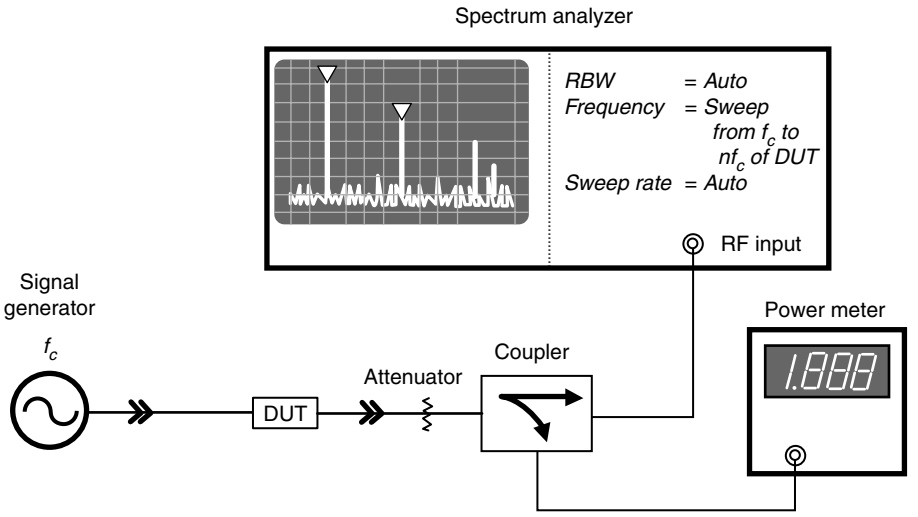


FIGURE 6.2 Setup used to measure harmonic distortion. Because harmonic levels are a function of output amplitude, a power meter is needed to accurately characterize the harmonic distortion properties.

6.2.2 Measurement of Harmonic Distortion

While instruments are available at low frequencies to measure the total harmonic distortion (THD), the level of each harmonic is generally measured individually using a spectrum analyzer. Such a setup is shown in Figure 6.2.

Harmonic levels are usually measured in a relative manner by placing a marker on the fundamental signal and a delta marker at the *n*th harmonic frequency. When measured in this mode, the harmonic level is expressed in dBc, which designates dB relative to carrier (i.e., the fundamental frequency) level. While it is convenient to set the spectrum analyzer sweep to include all harmonics of interest, it may be necessary to center a narrow span at the harmonic frequency in order to reduce the *noise floor* on the spectrum analyzer. An attenuator may be needed to protect the spectrum analyzer from overload. Note that the power level present at the spectrum analyzer input includes all harmonics, not just the ones displayed on the screen. Finally, it is important to note that spectrum analyzers have their own nonlinear characteristics that depend on the level input to the instrument. It is sometimes difficult to ascertain whether measured harmonic distortion is being generated within the device or with the test instrument. One method to do this is to use a step attenuator at the output of the device and step up and down. If distortion is being generated with the spectrum analyzer, the harmonic levels will change with different attenuator settings.

6.3 Gain Compression and Phase Distortion

A major result of changing impedances in microwave circuits is signal gain and phase shift that depend on input amplitude level. A change in signal gain between input and output may result from signal

clipping due to device current saturation or cutoff. Insertion phase may change because of nonlinear resistances in combination with a reactance. Though there are exceptions, signal gain generally decreases with increasing amplitude or power level. For this reason, the *gain compression* characteristics of microwave components are often characterized. Phase distortion may change either way, so it is often described as *phase deviation* as a function of amplitude or power level.

6.3.1 Gain Compression

Referring back to Equation 6.8, it is seen that, in addition to harmonic distortion, the level fundamental signal has been modified beyond that dictated by the linear term, g_1. This effect is described as *gain compression* in that the gain of the circuit becomes a function of the input amplitude A. Figure 6.3 illustrates this result. For small values of A, the g_1 term will dominate, giving a 1:1 slope when the output power is plotted against the input power on a log (i.e., dB) scale. Note that the power level of the nth harmonic plotted in like fashion will have an n:1 slope.

Gain compression is normally measured on a *bandpass* nonlinear circuit [2]. Such a circuit is illustrated in Figure 6.4. It is interesting to note that an ideal limiting amplifier described by Equation 6.9 when heavily overdriven at the input will eventually produce a square wave at the output, which is filtered by the bandpass filter. Note that the amplitude of the fundamental component of a square wave is at a level of 4/¼ times, or 2.1 dB greater than the amplitude of the square wave set by the clipping level.

$$v_{out}(t) = \begin{cases} g_1 v_{in}(t) & v_{out} < v_{\lim} \\ v_{\lim} & otherwise \end{cases} \tag{6.9}$$

For a general third-order nonlinear transfer characteristic driven by a sinusoidal input, the bandpass output is given by

$$v_{out}(t) = \left(g_1 A + \frac{3 g_3 A^3}{4} \right) \cos(2\pi f_c t) \tag{6.10}$$

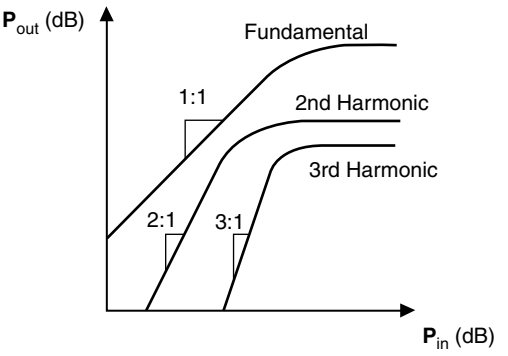


FIGURE 6.3 Output power vs. input power for a nonlinear circuit.

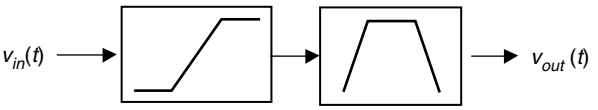


FIGURE 6.4 Bandpass nonlinear circuit.

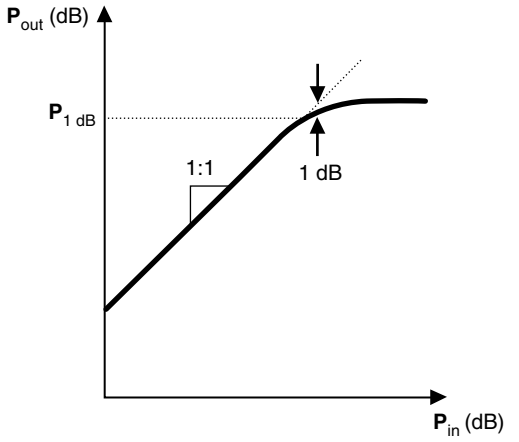


FIGURE 6.5 Gain compression of a bandpass nonlinear circuit. A figure of merit $P_{1\,\mathrm{dB}}$ is the output power at which the gain has been reduced by 1 dB.

A bandpass nonlinear circuit may be characterized by the power output at 1 dB gain compression, $P_{1\mathrm{dB}}$ as illustrated in Figure 6.5.

6.3.2 Phase Distortion

Nonlinear circuits may also contain reactive elements that give rise to *memory* effects. It is usually unnecessary to extract the entire Volterra representation of a nonlinear circuit with reactive elements if a few assumptions can be made. For bandpass nonlinear circuits with memory effects of time duration of the order of the period of the carrier waveform, a simple model is often used to describe the phase deviation versus amplitude:

$$v_{out}(t) = A(t)\cos\left\{2\pi f_c t + \Phi\left[A(t)\right]\right\}. \tag{6.11}$$

Equation 6.11 represents the AM–PM distortion caused by short-term memory effects (i.e., small capacitances and inductances in microwave circuits). The *effects* of AM–PM on an amplitude-modulated signal is illustrated in Figure 6.6.

For the case of input signals with small deviations of amplitude ΔA, the phase deviation may be considered linear, with a proportionality constant k_ϕ as seen in Figure 6.6. For a sinusoidally modulated input signal, an approximation for small modulation index FM signals may be utilized. One obtains the following expression for the output signal:

$$\begin{aligned} v_{out}(t) &= \left[A + \Delta A\sin\left(2\pi f_m t\right)\right]\cos\left\{2\pi f_c t + k_\phi A + k_\phi \Delta A\sin\left(2\pi f_m t\right)\right\} \\ &\approx A\cos\left(2\pi f_c t + k_\phi A\right)\sum_{n=0}^{\infty} J_n\left(k_\phi \Delta A\right)\cos\left(2n\pi f_m t\right) \end{aligned} \tag{6.12}$$

where J_n is the nth order Bessel function of the first kind [3]. Thus, like amplitude distortion, AM–PM distortion creates sidebands at the harmonics of the modulating signal. Unlike amplitude distortion, these sidebands are not limited to the first sideband. Thus, AM–PM distortion effects often dominate the out-of-band interference beyond $f_c \pm f_m$ as seen in Figure 6.7.

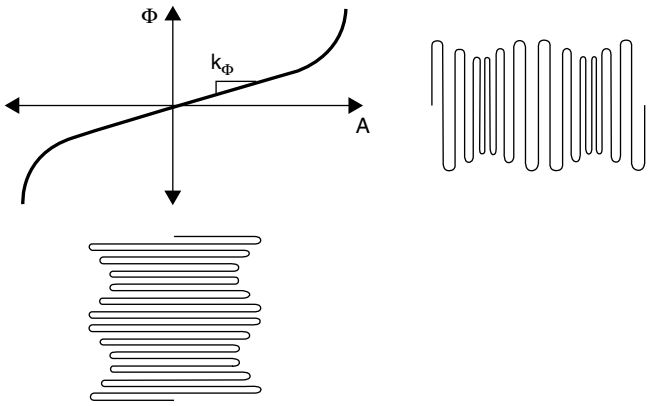


FIGURE 6.6 Effect of AM–PM distortion on a modulated signal. Input signal has AM component only. Output signal has interrelated AM and FM components due to the AM–PM distortion of the circuit.

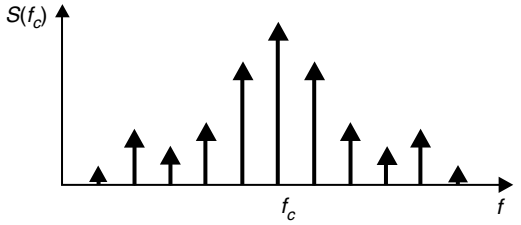


FIGURE 6.7 Output components of an amplitude-modulated signal distorted by AM–PM effects.

The FM modulation index k_ϕ may be used as a figure of merit to assess the impact of AM–PM on signal with small amplitude deviations. The relative level of the sidebands may be calculated from Equation 6.12. It must be noted that two sidebands nearest to the carrier may be masked from the AM components of the signal, but the out-of-band components are readily identified.

6.3.3 Measurement of Gain Compression and Phase Deviation

For bandpass components where the input frequency is equal to the output frequency, such as amplifiers, gain compression and phase deviation of a nonlinear circuit are readily measured with a *network analyzer* in power sweep mode. Such a setup is shown in Figure 6.8. $P_{1\,\mathrm{dB}}$ is easily measured using delta markers by placing the reference marker at the beginning of the sweep (i.e., where the DUT is not compressed), and moving the measurement marker where $\Delta\mathrm{Mag}(S_{21}) = -1$ dB. Sweeping at too high a rate may affect the readings. The sweep must be slow enough so that steady-state conditions exist in both the thermal case and the DC bias network within the circuit. Sweeper retrace may also affect the first few points on the trace. These points must be neglected when setting the reference marker.

The FM modulation index is often estimated by measuring the phase deviation at 1 dB gain compression $\Delta\Phi(P_{1\,\mathrm{dB}})$

$$k_\phi \approx \frac{\Delta\Phi\left(P_{1dB}\right)}{2Z_0\sqrt{P_{1dB}}}. \tag{6.13}$$

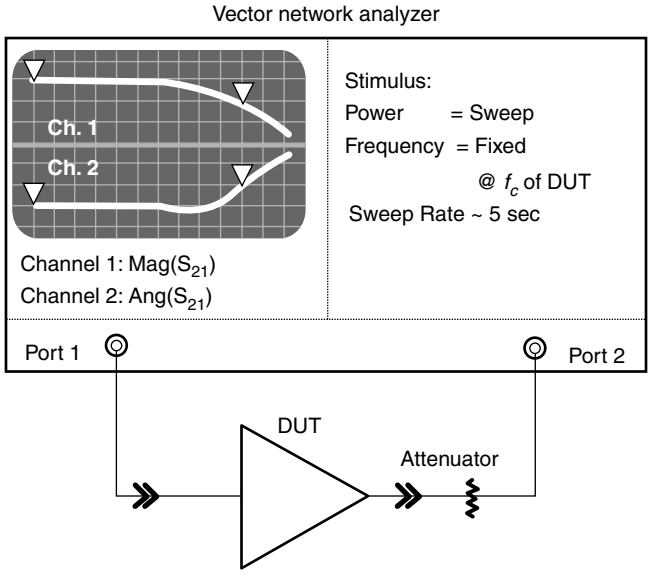


FIGURE 6.8 Setup used to measure gain compression and AM–PM.

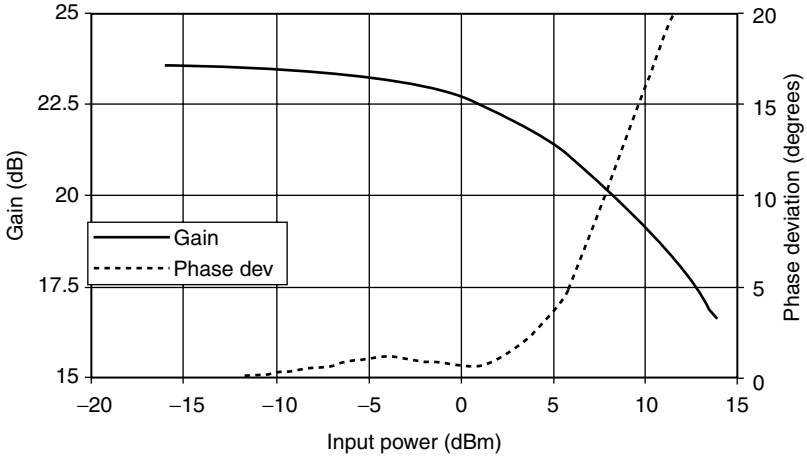


FIGURE 6.9 Measured gain compression and AM–PM of a 0.5 W 1960 MHz GaAs MESFET power amplifier IC using an HP8753C Vector Network Analyzer in power sweep mode.

For circuits such as mixers, where the input frequency is not equal to the output frequency, gain compression may be measured using the network analyzer with the measurement mode setup for frequency translation. The operation in this mode is essentially that of a *scalar network analyzer*, and all phase information is lost. AM–PM effects may be measured using a *spectrum analyzer* and fitting the sideband levels to Equation 6.12.

The gain compression and phase deviation of a GaAs power amplifier is shown in Figure 6.9. $P_{1\,\mathrm{dB}}$ for this amplifier is approximately 23 dBm or 0.5 W. The phase deviation $\Delta\Phi$is not constant from low power to $P_{1\,\mathrm{dB}}$. Nevertheless, as a figure of merit, the modulation index k_ϕ may be calculated from Equation 6.13 to be 0.14°/*V*. Notice that for higher power levels, the amplifier is well into compression, and the phase deviation occurs at a much higher slope than k_ϕ would indicate.

6.4 Intermodulation Distortion

When more than one frequency component is present in a signal, the distortion from a nonlinear circuit is manifested as IMD [4]. The IMD performance of microwave circuits is important because it can create unwanted interference in adjacent channels. While *bandpass* filtering can eliminate much of the effects of harmonic distortion, intermodulation distortion is difficult to filter out because the IMD components may be very close to the carrier frequency. A common figure of merit is *two-tone* intermodulation distortion.

6.4.1 Two-Tone Intermodulation Distortion

Consider a signal consisting of two sinusoids

$$v_{in}(t) = A\cos(2\pi f_1) + A\cos(2\pi f_2). \tag{6.14}$$

Such a signal may be represented in a different fashion by invoking well-known trigonometric identities.

$$v_{in}(t) = A\cos(2\pi f_c)\cos(2\pi f_m), \tag{6.15a}$$

where

$$f_c = \frac{f_1 + f_2}{2} \quad \text{and} \quad f_m = \left|\frac{f_1 - f_2}{2}\right| \tag{6.15b}$$

Applying such a signal to a memoryless nonlinearity as defined in Equation 6.3, one obtains the following result:

$$v_{out} = \left[\left(g_1 A + \frac{3g_3 A^3}{4}\right)\cos(2\pi f_m t) + \frac{3g_3 A^3}{4}\cos(6\pi f_m t)\right]\cos(2\pi f_c t). \tag{6.16}$$

Thus, it is seen that the IMD products near the input carrier frequency are simply the odd-order harmonic distortion products of the modulating *envelope*. This is illustrated in Figure 6.10.

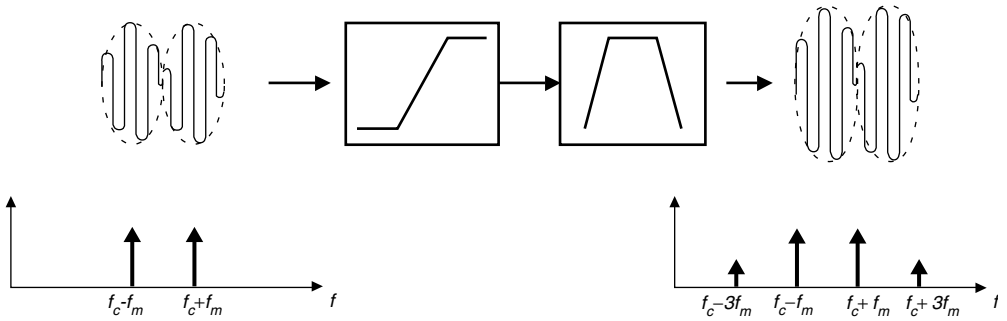


FIGURE 6.10 Intermodulation distortion of a two-tone signal. The output bandpass signal contains the original input signal as well as the harmonics of the envelope at the sum and difference frequencies.

6.4.2 Third-Order Intercept Point

Referring to Figure 6.11, note that the output signal varies at a 1:1 slope on a *log-log* scale with the input signal, while the IMD products vary at a 3:1 slope. Though both the fundamental and the IMD products saturate at some output power level, if one were to extrapolate the level of each and find the intercept point, the corresponding output power level is called the *third order intercept point* (IP_3). Thus, if the IP_3 of a nonlinear circuit is known, the IMD level relative to the output signal level may be found from

$$IMD_{dBc} = 2\left(P_{out,dBm} - IP_{3,dBm}\right). \tag{6.17}$$

It must be noted that 3rd order IMD is only dominant for low levels of distortion (<10 dB below P_{1dB}). At higher levels, 5th and higher order IMD effects can also produce sidebands at the 3rd order frequency. The net result is that the relative IMD level will change at a rate greater than 2:1 compared to carrier level. Care should be taken to avoid extrapolating IP_3 from points where this may be occurring. Another point of caution is AM–PM effects. In theory, the sidebands produced by phase modulation are in quadrature with those produced by AM distortion, and thus should add directly to the IMD power. However, the author's experience has shown that these AM–PM products can be rotated in phase and thus vector added to the AM sidebands. Since the FM sidebands are antiphase, one FM sideband adds constructively to the AM sidebands, while the other adds destructively. The net effect is an imbalance in the IMD levels from lower to higher sideband frequencies. Most specifications of IMD level will measure the worst case of the two.

For a limiting amplifier, an often used rule-of-thumb may be derived that predicts a relationship between P_{1dB} and IP_3 [4].

$$IP_3 = P_{1dB} + 9.6\ dB. \tag{6.18}$$

While this may not be rigorously relied upon for every situation, it is often accurate within ±2 dB for small-signal amplifiers and class-A power amplifiers.

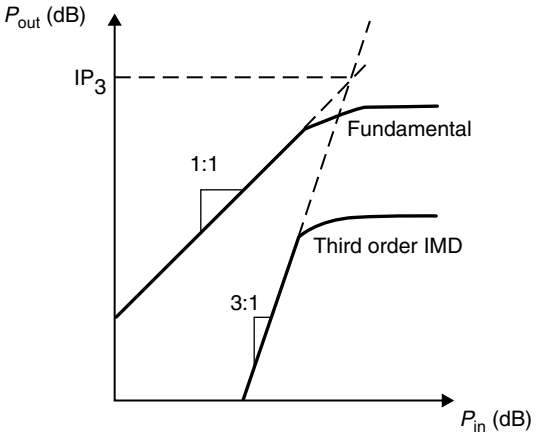


FIGURE 6.11 Relationship between signal output power and intermodulation distortion product power levels. Extrapolating the trends, a figure of merit called the *third-order intercept point* (IP_3) is obtained.

6.4.3 Dynamic Range

Because IMD generally increases with increasing signal levels, IP_3 may be used to establish the dynamic range of a system. The signal level at which the IMD level meets the noise floor is defined at the spurious free dynamic range (*SFDR*) [5]. This is illustrated in Figure 6.12.

The *SFDR* of a system with gain *G* may be derived from IP_3 and the *noise figure* (*NF*)

$$SFDR = \frac{2IP_3 - 2\left[10\log\left(kT_{eq}B\right) + NF + G\right]}{3}, \tag{6.19}$$

where *k* is Boltzman's constant, T_{eq} is the equivalent input noise temperature, and *B* is the bandwidth of the system.

6.4.4 Intermodulation Distortion of Cascaded Components

The question often arises when two components are cascaded of what effect the driving stage IMD has on the total IMD. This is shown in Figure 6.13. To the degree that the IMD products produced by the *n*th stage are uncorrelated with those of the $n+1$ stage, the output IMD may be calculated as the power addition of third-order IMD levels (P_3) with levels adjusted accordingly for gain.

$$\left(P_3\right)_{n+1} = 10\log\left[10^{\left[\left(P_3\right)_n + G_1\right]/10} + 10^{\left[3\left(P_1\right)_{n+1} - 2\left(IP_3\right)_{n+1}/10\right]}\right]. \tag{6.20}$$

6.4.5 Measurement of Intermodulation Distortion

IMD is normally measured with two-signal generators and a *spectrum analyzer*. Such a setup is shown in Figure 6.14. Care must be taken to isolate the signal generators, as IMD may result from one output mixing with signal from the opposing generator. The carrier levels should be within 0.5 dB of each other for accurate IMD measurements. Also, it is usually recommended that a power meter be used to get an accurate reading of output power level from the DUT. Relative IMD level is measured by placing a reference marker on one of the two carrier signals, and placing a delta marker at either sideband. Finally, the input level must be maintained well below the input IP_3 of the spectrum analyzer to insure error-free reading of the DUT.

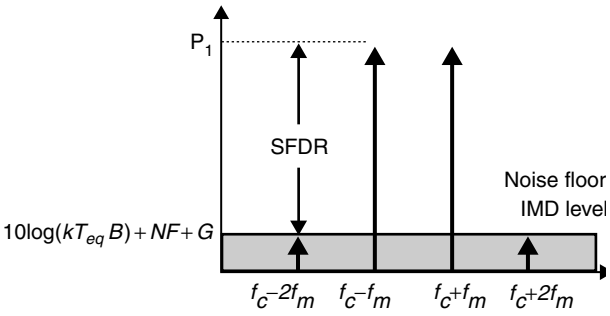


FIGURE 6.12 An illustration of spurious-free dynamic range, which defines the range of signal levels where the worst case signal-to-noise ratio is defined by the noise floor of the system, rather than the IMD level.

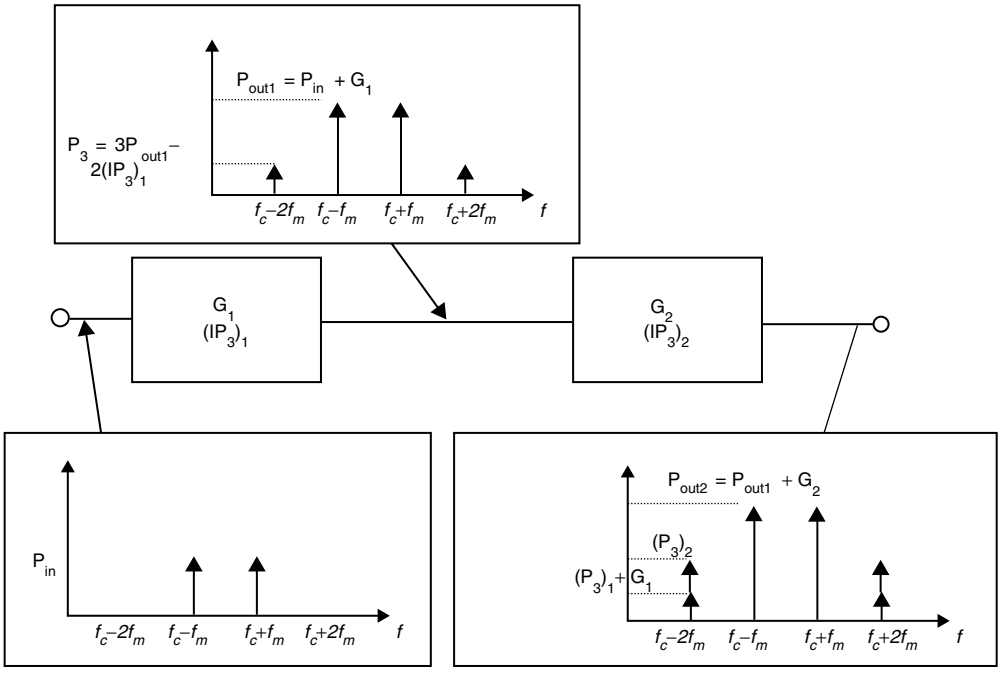


FIGURE 6.13 Effect of cascaded IMD levels. The IMD from the first stage may be power added to those of the second stage with levels adjusted for the gain of the stage.

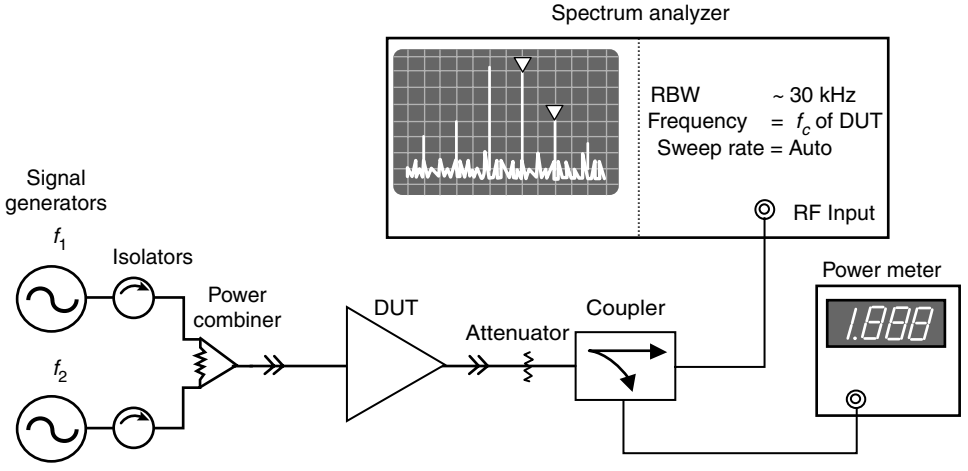


FIGURE 6.14 Setup used to measure two-tone intermodulation distortion.

6.5 Multicarrier Intermodulation Distortion and Noise Power Ratio

While two-tone intermodulation distortion serves to compare the linearity of one component to another, in many applications, a component will see more than two carriers in the normal operation of a microwave system. Thus direct measurement of multitone IMD is often necessary to insure adequate carrier-to-interference level within a communication system.

6.5.1 Peak-to-Average Ratio of Multicarrier Signals

The major difference between two-tone signals and multitone signals is the *peak-to-average* (*pk/avg*) power ratio [6]. From Equation 6.14, it is clear that the average power of a two-tone signal is equal to the sum of powers from the individual carriers. However, from Equation 6.15, one may derive that the *peak envelope power* (PEP) is four times the level of the individual carriers. Thus, it is said that the *pk/avg* ratio of a two-tone signal is a factor of 2, or 3 dB. From inductive reasoning, it is then clear that the *pk/avg* ratio of an *n*-tone signal is

$$pk/avg = 10\log(n). \tag{6.21}$$

While the absolute peak of a multicarrier is dependent only on the number of carriers, the probability distribution of the *pk/avg* ratio depends on the modulation. Figure 6.15 shows the difference between 16 phase-aligned tones, and 16 carriers with randomly modulated phases. In general, multiple modulated signals encountered in communication systems will mimic the behavior of random phase modulated sinusoids. Phase aligned sinusoids may be considered a worst case condition. As the number of carriers increases, and if their phases are uncorrelated, the Central Limit Theorem predicts that the distribution of *pk/avg* approaches that of white Gaussian random noise [7]. The latter signal is treated in the next section.

6.5.2 Noise–Power Ratio

For many systems, including those that process multicarrier signals, white Gaussian noise is a close approximation for the real-world signals. This is a result of the *Central Limit Theorem,* which states that the probability distribution of a sum of a large number of random variables will approach the Gaussian

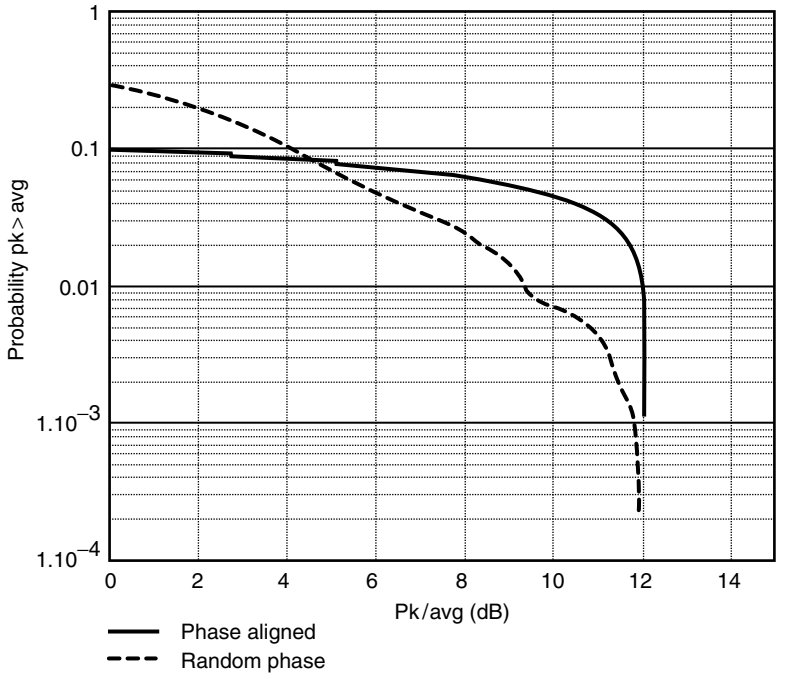


FIGURE 6.15 Distribution of *peak-to-average* ratio of a phase-aligned 16-carrier signal and a random-phase 16-carrier signal. The *y*-axis shows the probability that the signal exceeds a power level above average on the *x*-axis. While both signals ultimately have the same pk/avg ratio, their distributions are much different.

distribution, regardless of the distributions of the individual signals.[7] One metric that has been employed to describe the IMD level one would expect in a dense multicarrier environment is the *noise power ratio* (NPR). This concept is illustrated in Figure 6.16.

6.5.3 Measurement of Multitone IMD and NPR

Thus it is clear that power ratings for components must be increased for peak power levels given by Equation 6.21. Furthermore, two-tone intermodulation distortion may not be indicative of IMD of multitone signals. Measurement over various power levels is the only way to accurately predict multitone IMD. Figure 6.17 illustrates a setup that may be used to measure multitone intermodulation.

The challenge in measuring NPR is creating the signal. It is clear that, to get an accurate indication of IMD performance, the signal bandwidth must not exceed the bandwidth of the device under test. Furthermore, to measure NPR, one must notch out the noise power over a bandwidth approximating one channel *BW*. As an example, an NPR measurement on a component designed for North American Digital Cellular System (IS-136) ideally would produce a 25-MHz wide noise source with one channel of bandwidth equal to 30 kHz. The *Q* of a notch filter to produce such a signal would be in excess of 25,000. Practical measurements employ filters with *Q*s around 1000, and are able to achieve more than 50 dB of measurement range. Such a setup is shown in Figure 6.18.

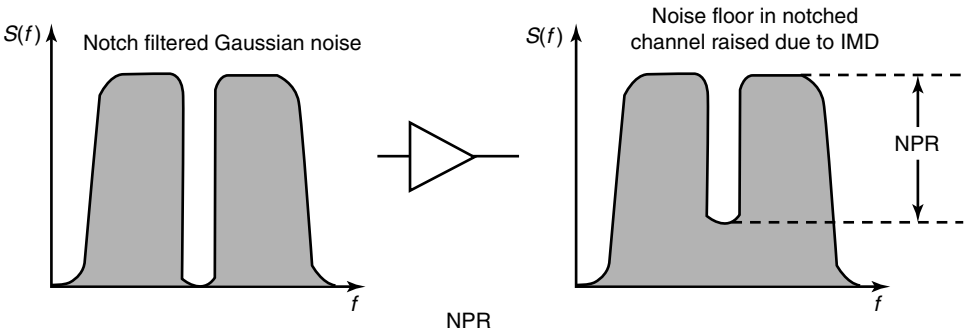


FIGURE 6.16 An illustration of NPR. NPR is essentially a measure of the carrier-to-interference level experienced by multiple carriers passing through a nonlinear component.

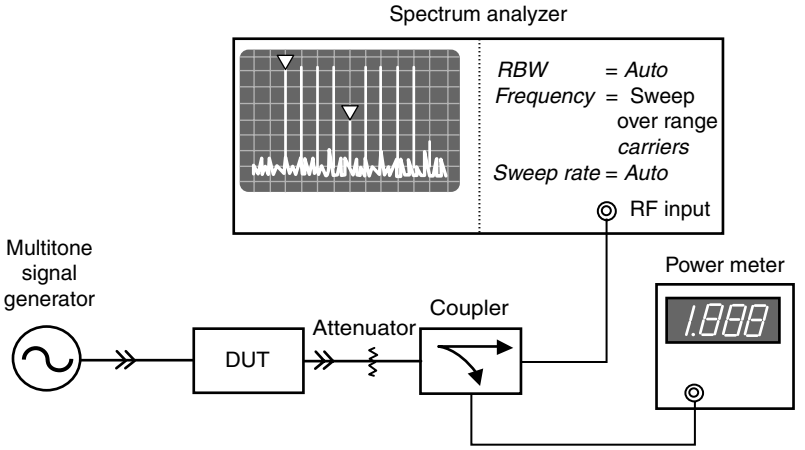


FIGURE 6.17 Measurement setup for multitone IMD. Tones are usually spaced equally, with the middle tone deleted to allow measurement of the worst-case IMD.

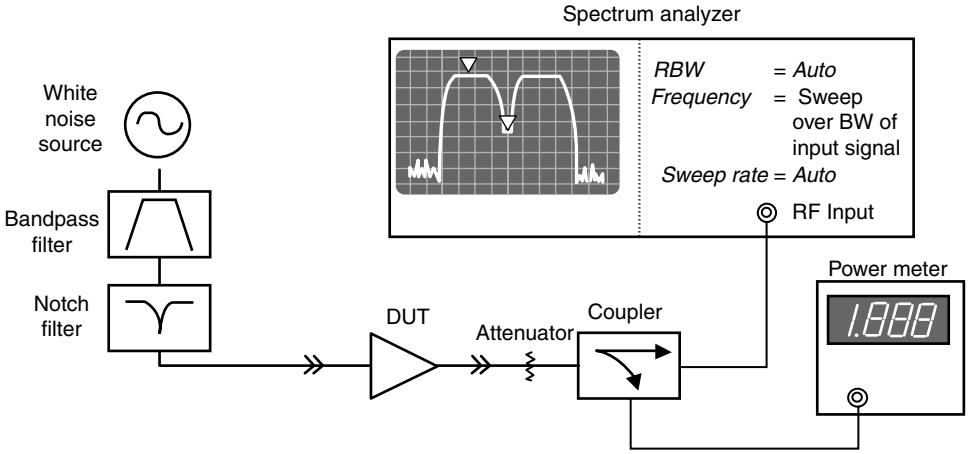


FIGURE 6.18 Noise Power Ratio measurement setup. The rejection of the notch filter should be at least 10 dB below the NPR level to avoid erroneous measurement.

6.6 Distortion of Digitally Modulated Signals

While standard test signals such as a two-tone or band-limited Gaussian noise provide relative figures of merit of the linearity of a nonlinear component, they cannot generally insure compliance with government or industry system-compatibility standards. For this reason, methods have been developed to measure and characterize the intermodulation distortion of the specific digitally modulated signals used in various systems. Table 6.2 summarizes the modulation formats for North American digital cellular telephone systems [8,9].

6.6.1 Intermodulation Distortion of Digitally Modulated Signals

Amplitude and phase distortion affect digitally modulated signals the same way they affect analog modulated signals: gain compression and phase deviation. This is readily seen in Figure 6.19. Because both amplitude and phase modulation are used to generate digitally modulated signals, they are often expressed as a *constellation* plot, with the in-phase component $I = A\cos\phi$ envelope plotted against the quadrature component $Q = A\sin\phi$. The instantaneous power envelope is given by

$$P(t) = I(t)^2 + Q(t)^2 = A(t)^2. \tag{6.22}$$

When the envelope is clipped and/or phase rotated, the resulting IMD is referred to as *spectral regrowth*. Figure 6.20 shows the effect of nonlinear distortion on a digitally modulated signal. The out-of-band products may lie in adjacent channels, thus causing interference to other users of the system. For this reason, the IMD of digitally modulated signals are often specified as *adjacent channel power ratio* (ACPR).[10]

ACPR may be specified in a number of ways, depending on the system architecture. In general, ACPR is given by

$$ACPR = \frac{I_{adj}}{C_{ch}} = \frac{\int_{f_o - B_{adj}}^{f_o - B_{adj}} S(f)\,df}{\int_{-B_{ch}/2}^{B_{ch}/2} S(f)\,df}, \tag{6.23}$$

TABLE 6.2 Modulation Formats for North American Digital Cellular Telephone Systems

Standard	Multiple Access Mode	Channel Power Output	Modulation	Channel Bandwidth
IS-136 [8]	TDMA	+28 dBm	¼/4-DQPSK	30 kHz
IS-95 [9]	CDMA	+28 dBm	OQPSK	1.23 MHz

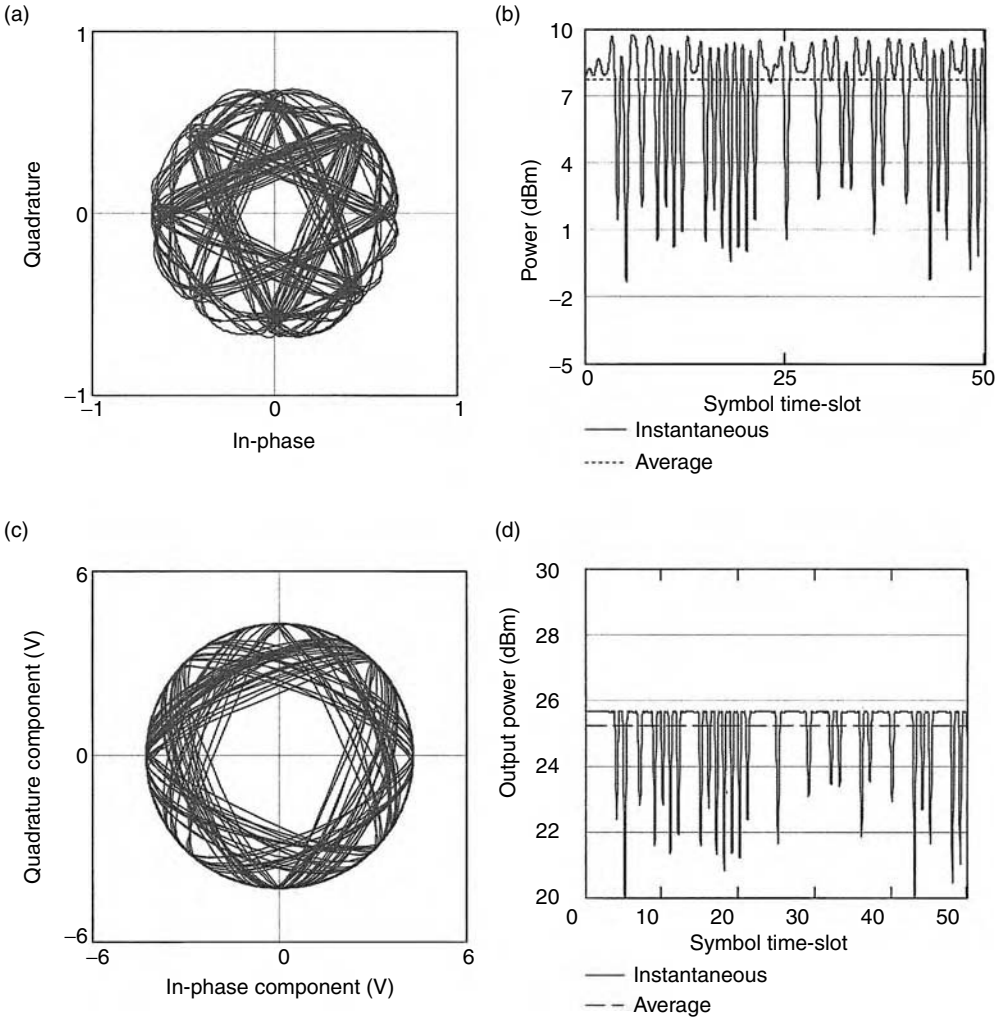


FIGURE 6.19 Effect of amplitude and phase distortion on digitally modulated signals. (a) Shows a ¼/4 DQPSK signal constellation, and its associated power envelope in (b). When such a signal is passed through a nonlinear amplifier, the resulting envelope is clipped (d), and portions of the constellation are rotated (c).

where I_{adj} is the total interference power in a specified adjacent channel bandwidth, B_{adj} at a given frequency offset f_o from the carrier frequency, and C_{ch} is the channel carrier power in the specified channel bandwidth B_{ch}. Note that the carrier channel bandwidth may be different from the interference channel bandwidth because of regulations enforcing interference limits between different types of systems. Furthermore, the interference level may be specified in more than one adjacent channel. In this case, the

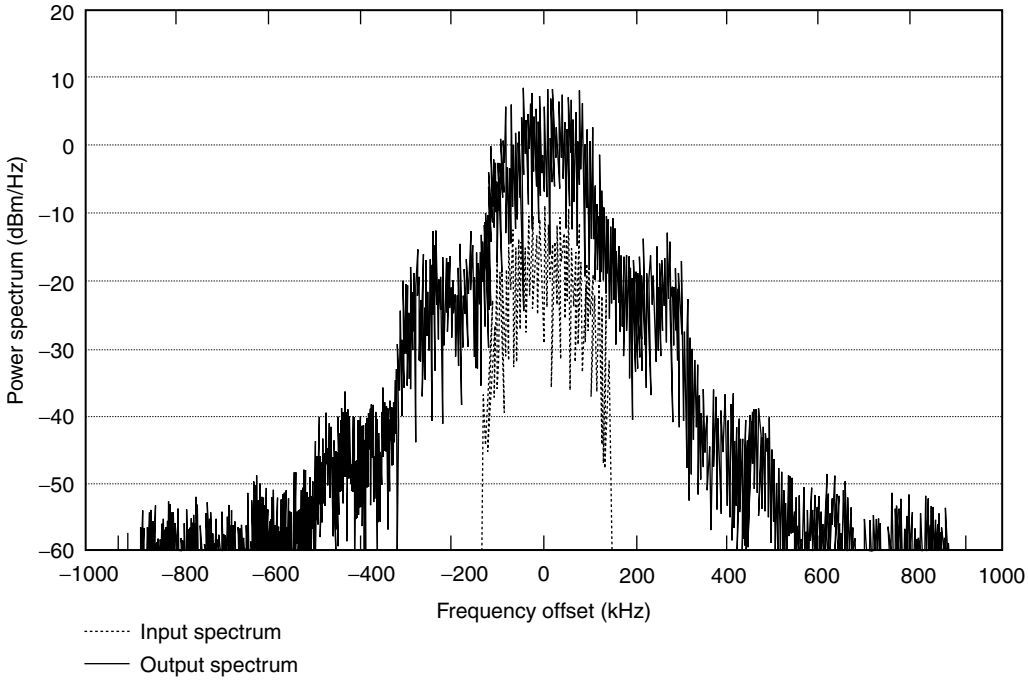


FIGURE 6.20 Effect of nonlinear distortion on a digitally modulated signal. The lower power input signal to a power amplifier has a frequency spectrum that is well contained within a specified channel bandwidth. IMD due to nonlinear distortion creates out-of-band products that may fall within the adjacent channels, causing interference to other users of the system.

TABLE 6.3 ACPR and EVM Specifications for Digital Cellular Subscriber Equipment

Standard	ADJ. CH. PWR	ALT. CH. PWR	EVM
IS-136	–26 dBc/30 kHz at ± 30 kHz	–45 dBc/30 kHz at ±60 kHz	12.5%
IS-95	–42 dBc/30 kHz at > ± 885 kHz	–54 dBc/30 kHz at > ±1.98 MHz	23.7%

specification is referred to as the *alternate channel power ratio*. Table 6.3 shows ACPR specifications for various digital cellular standards.

In addition to the out-of-band interference due to the intermodulation distortion in-band interference will also result from nonlinear distortion. The level of the in-band interference is difficult to measure directly because it is superimposed on the channel spectrum. However, when the signal is demodulated, errors in the output *I–Q* constellation occur at the sample points. This is shown in Figure 6.21. Because the demodulator must make a decision as to which symbol (i.e., which constellation point) was sent, the resulting errors in the *I–Q* vectors may produce a false decision, and hence cause *bit errors.*

There are two methods to characterize the level *I–Q* vector error: *error vector magnitude* (EVM), and a quality factor called the ρ-factor. [11] Both EVM and ρ-factor provide an indication of signal distortion, but they are calculated differently. EVM is the *rms* sum of vector errors divided by the number of samples.

$$EVM = \frac{1}{n}\sqrt{\sum_{n}\left[\left|I(t_n) - S_{In}\right|^2 + \left|Q(t_n) - S_{Qn}\right|^2\right]}, \tag{6.24}$$

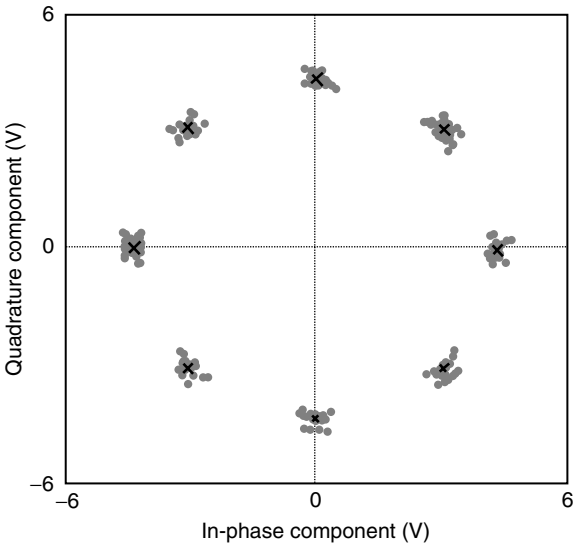


FIGURE 6.21 Errors in the demodulated *I*–*Q* constellation may result from the in-band IMD products. The *rms* summation of errors from the desired location (given by the × markers) give the *error vector magnitude* of the signal distortion.

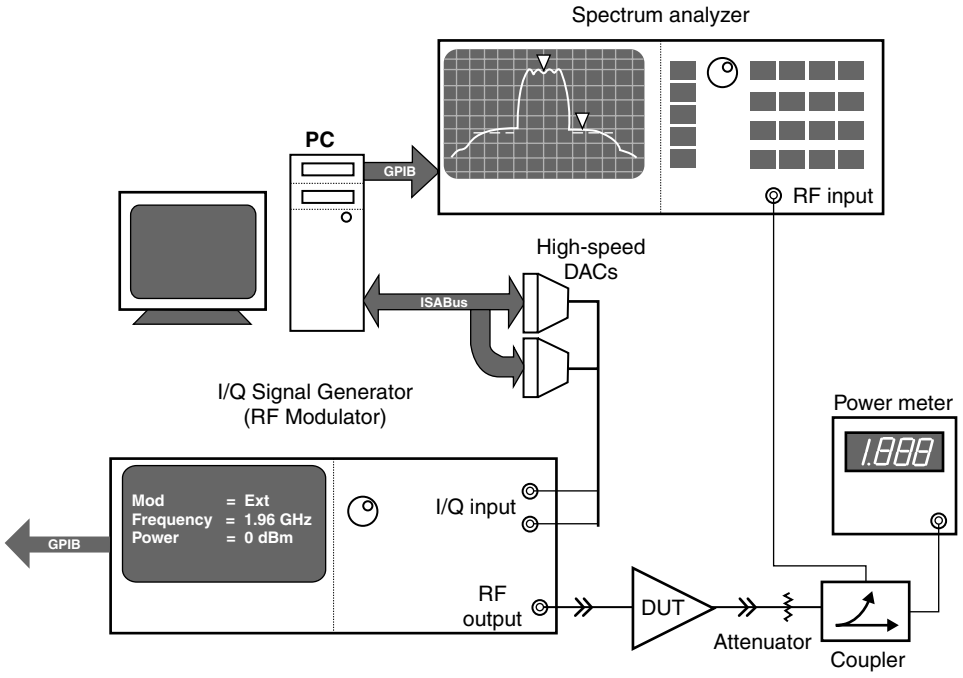


FIGURE 6.22 Measurement setup for ACPR. Waveforms are created using PCs or specialized arbitrary waveform generators. In either case, the baseband waveform must be upconverted to the center frequency of the DUT.

where the *I*–*Q* sample points at the *n*th sample windows are given by $I(t_n)$ and $Q(t_n)$, and the *n*th symbol location point in-phase and quadrature components are given by S_{In} and S_{Qn} respectively.

Whereas EVM provides an indication of *rms* % error of the signal envelope at the sample points, ρ-factor is related to the waveform quality of a signal. It is related to EVM by

$$\rho = \frac{1}{1 - EVM^2} \tag{6.25}$$

6.6.2 Measurement of ACPR, EVM, and ρ-Factor

ACPR may be measured using a setup similar to those for measuring IMD. The major difference involves generating the test signal. Test signals for digitally modulated signals must be synthesized according to system standards using an *arbitrary waveform generator* (AWG), which generates *I*- and *Q*-baseband envelopes. In the most basic form, these are high speed digital-to-analog converters (DACs). The files used to generate the envelope waveforms may be created using commonly available mathematics software, and are built in many commercially available AWGs. The *I*- and *Q*-baseband envelopes are fed to an RF modulator to produce a modulated carrier at the proper center frequency.

In the case of CDMA standards, deviations between test setups can arise from different selections of Walsh codes for the traffic channels. While a typical CDMA downlink (base station to mobile) signal has a *pk*/*avg* of approximately 9.5 dB, it has been shown that some selections of Walsh codes can result in peak-to-average ratios in excess of 13 dB [12]. Measurement of EVM is usually done with a *vector signal analyzer* (VSA). This instrument is essentially a receiver that is flexible enough to handle a variety of frequencies and modulation formats. Specialized software is often included to directly measure EVM or ρ-factor for well-known standards used in microwave radio systems.

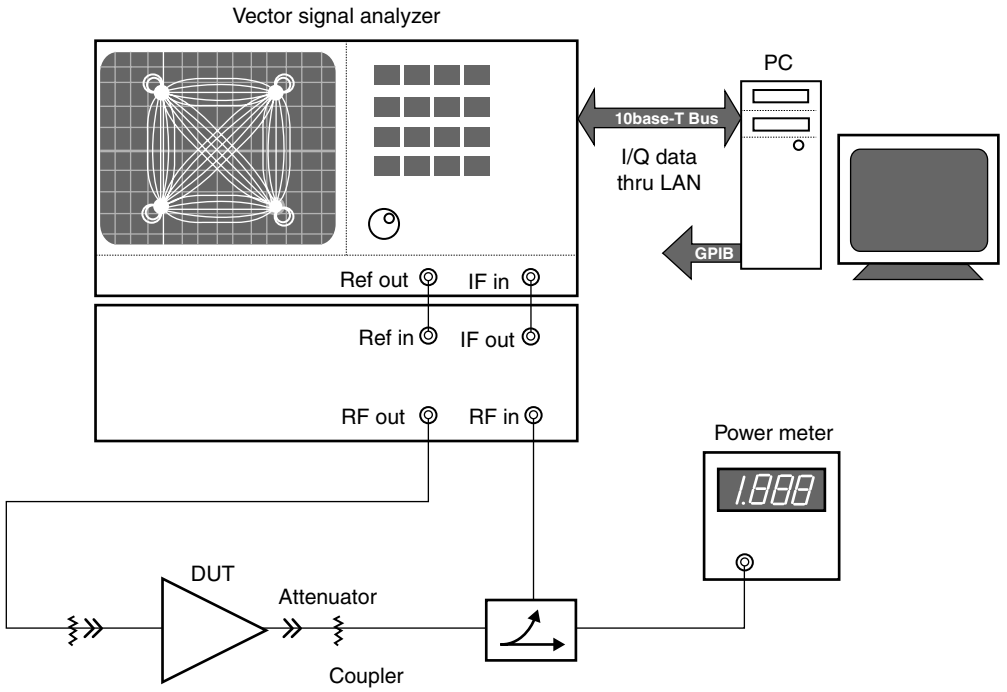


FIGURE 6.23 Setup for measuring EVM. The VSA demodulates the *I*–*Q* waveform and calculates the deviation from ideal to calculate EVM and ρ-factor as given in Equations 6.24 and 6.25, respectively.

6.7 Summary

This section has treated characterization and measurement techniques for nonlinear microwave components. Figures of merit were developed for such nonlinear effects as harmonic level, gain compression, and intermodulation distortion. While these offer a basis for comparison of the linearity performance between like components, direct measurement of adjacent channel power and error vector magnitude are preferred for newer wireless systems. Measurement setups for the above parameters were suggested in each section. For more advanced treatment, the reader is referred to the references at the end of this section.

References

1. Maas, S.A., *Nonlinear Microwave Circuits*, Artech House, Boston, 1988.
2. Blachman, N.M., Band-pass nonlinearities, *IEEE Trans. Information Theory,* IT-10, 162–64, April, 1964.
3. Andrews, L.C., *Special Functions of Mathematics for Engineers,* 2nd ed., McGraw-Hill, New York, 1992, chap. 6.
4. Cripps, S.C., *RF Power Amplifiers for Wireless Communications,* Artech House, Boston, 1999, chap. 7.
5. Carson, R.S., *Radio Concepts: Analog*, John Wiley & Sons, New York, 1990, chap. 10.
6. Kenney, J.S., and Leke, A., Design considerations for multicarrier CDMA base station power amplifiers, *Microwave J.,* 42, 2, 76–86, February, 1999.
7. Papoulis, A., *Probability, Random Variables, and Stochastic Processes,* 3rd ed., McGraw-Hill, New York, 1991, chap. 8.
8. IS-136 Interim Standard, Cellular System Dual-Mode Mobile Station — Base Station Compatibility Standards, Telecommunications Industry Assoc.
9. IS-95 Interim Standard, Mobile Station — Base Station Compatibility Standard for Dual-Mode Wideband Spread Spectrum Cellular Systems, Telecommunications Industry Assoc.
10. Kenney, J.S., and Leke, A., Power amplifier spectral regrowth for digital cellular and PCS applications, *Microwave J.*, 38, 10, 74–92, October 1995.
11. Lindsay, S.A., Equations derive error-vector magnitude, *Microwaves & RF,* April, 1995, 158–67.
12. Braithwaite, R.N., Nonlinear amplification of CDMA waveforms: an analysis of power amplifier gain errors and spectral regrowth, *Proc. 48th Annual IEEE Vehicular Techn. Conf.*, 2160–66, 1998.

7

Theory of High-Power Load-Pull Characterization for RF and Microwave Transistors

John F. Sevic
Maury Microwave Corporation

In both portable and infrastructure wireless systems the power amplifier (PA) often represents the largest single source of power consumption in the radio. While the implications of this are obvious for portable applications, manifested as talk-time, it is also important for infrastructure applications due to thermal management, locatability limitations, and main power limitations. Significant effort is devoted toward developing high-performance RF and microwave transistors and circuits to improve power amplifier efficiency. In the former case, an accurate and repeatable characterization tool is necessary to evaluate the performance of the transistor. In the latter case, it is necessary to determine the source and load impedance for the best trade-off in overall performance. Load-pull is presently the most common technique, and arguably the most useful for carrying out these tasks. In addition, load-pull is also necessary for large-signal model development and verification.

Load-pull as a design tool is based on measuring the performance of a transistor at various source and/or load impedances and fitting contours, in the gamma-domain, to the resultant data; measurements at various bias and frequency conditions may also be done. Several parameters can be superimposed over each other on a Smith chart and trade-offs in performance established. From this analysis, optimal source and load impedances are determined.

Load-pull can be classified by the method in which source and load impedances are synthesized. Since the complex ratio of the reflected to incident wave on an arbitrary impedance completely characterizes the impedance, along with a known reference impedance, it is convenient to classify load-pull by how the reflected wave is generated.

The simplest method to synthesize an arbitrary impedance is to use a stub tuner. In contrast to early load-pull based on this method, contemporary systems fully characterize the stub tuner a priori, precluding the need for determining the impedance at each load-pull state [1]. This results in a significant reduction in time and increases the reliability of the system. This method of load-pull is defined as passive-mechanical. Passive-mechanical systems are capable of presenting approximately 50:1 VSWR, with respect to 50 Ω, and are capable of working in very high power environments. Repeatability is better than –60 dB. Maury Microwave and Focus Microwave each develop passive-mechanical load-pull systems [2,3]. For high-power applications, for example, >100 W, the primary limitation of passive-mechanical systems is self-heating of the transmission line within the tuner, with the resultant thermally induced expansion perturbing the line impedance.

Solid-state phase-shifting and attenuator networks can also be used to control the magnitude and phase of a reflected wave, thereby effecting an arbitrary impedance. This approach has been pioneered by ATN Microwave [4]. These systems can be based on a lookup table approach, similar to the passive-mechanical systems, or can use a vector network analyzer for real-time measurement of tuner impedance. Like all passive systems, the maximum VSWR is limited by intrinsic losses of the tuner network. Passive-solid-state systems, such as the ATN, typically exhibit a maximum VSWR of 20:1 with respect to 50 Ω. These systems are ideally suited for medium power applications and noise characterization (due to the considerable speed advantage over other types of architectures).

Tuner and fixture losses are the limiting factor in achieving a VSWR in excess of 50:1 with respect to 50 Ω. This would be necessary not only for characterization of high-power transistors, but also low-power transistors at millimeter-wave frequencies, where system losses can be significant. In these instances, it is possible to synthesize a reflected wave by sampling the wave generated by the transistor traveling toward the load, amplifying it, controlling its magnitude and phase, and reinjecting it toward the transistor. Systems based on this method are defined as active load-pull. Although in principle active load-pull can be used to create very low impedance, the power necessary usually limits the application of this method to millimeter-wave applications [5,6]. Because active load-pull systems are capable of placing any reflection coefficient on the port being pulled (including reflections greater than unity) these systems can be very unstable and difficult to control. Instability in a high-power load-pull system can lead to catastrophic failure of the part being tested.

The present chapter is devoted to discussing the operation, setup, and verification of load-pull systems used for characterization of high-power transistors used in wireless applications. While the presentation is general in that much of the discussion can be applied to any of the architectures described previously, the emphasis is on passive-mechanical systems. There are two reasons for limiting the scope. The first reason is that passive-solid-state systems are usually limited in the maximum power incident on the tuners, and to a lesser extent, the maximum VSWR the tuners are capable of presenting. The second reason is that currently there are no active load-pull systems commercially available. Further, it is unlikely that an active load-pull system would be capable of practically generating the sub-1 Ω impedances necessary for characterization of high-power transistors.

The architecture of the passive-mechanical system is discussed first, with a detailed description of the necessary components for advanced characterization of transistors, such as measuring input impedance and ACPR [7]. Vector network analyzer (VNA) calibration, often overlooked, and the most important element of tuner characterization, is presented next. Following this, tuner, source, and load characterization methods are discussed. Fixture characterization methods are also presented, with emphasis on use of prematching fixtures to increase tuner VSWR. Finally, system performance verification is considered.

7.1 System Architecture for High-Power Load-Pull

Figure 7.1 shows a block diagram of a generalized high-power automated load-pull system, although the architecture can describe any of the systems discussed in the previous section. Subharmonic and harmonic tuners are also included for characterization of out-of-band impedances [8]. The signal sample ports are used to measure the incident and reflected voltage waves at the source–tuner interface and the

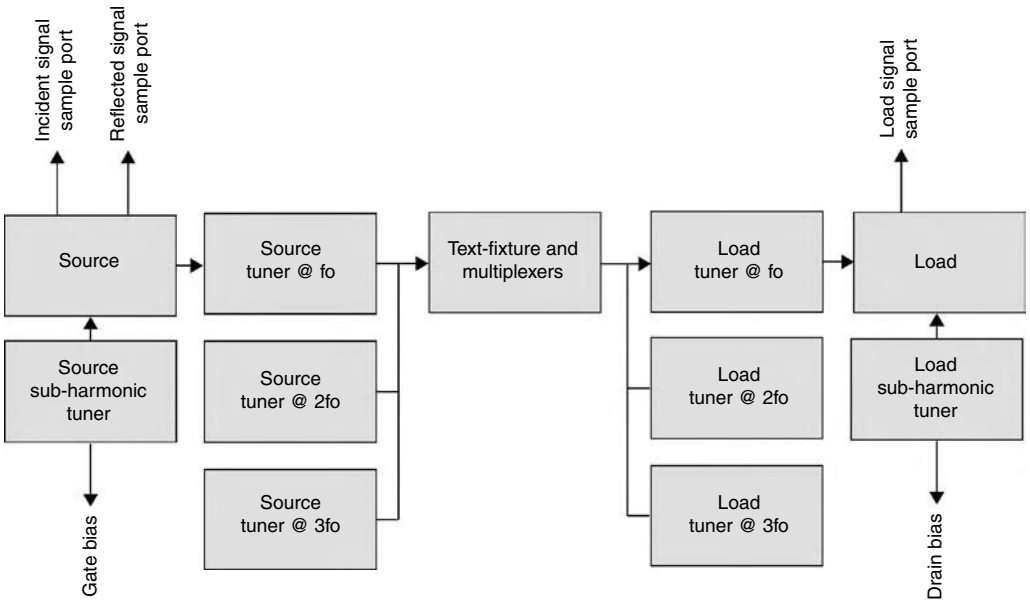


FIGURE 7.1 Block diagram of a generalized high-power load-pull system, illustrating the source, tuners, test-fixture, and load. The incident, reflected, and load signals are sampled at the three sampling points shown. Also shown, though not necessary, are harmonic and sub-harmonic tuners.

incident voltage wave at the load. The signals at each of these ports are applied to the equipment necessary to make the measurements the user desires. Each of these blocks is described subsequently.

The source block of Figure 7.1 usually includes all of the components necessary for generating the signal, leveling its power, providing gate/base bias for the device under test, and providing robust sampling points for the measurement equipment. Figure 7.2 shows the details of a typical source block. For flexibility and expediency in applying arbitrarily modulated signals, an arbitrary waveform generator and vector signal source are shown. The signal is typically created using MATLAB, and can represent not only digitally modulated signals, but also the more conventional two-tone signal. The signal is applied to a reference PA, which must be characterized to ensure that it remains transparent to the DUT; for high-power applications this is often a 50–100 W, PA.

Following the reference PA is a low-pass filter to remove harmonics generated from the source and/or reference PA. Next are the sampling points for the incident and reflected waves, which is done with two distinct directional couplers. Since the source tuner may present a high reflection, a circulator to improve directivity separates each directional coupler; the circulator also protects the reference PA from reflected power. The circulator serves to present a power-invariant termination for the source tuner, the impedance of which is critical for sub 1 Ω load-pull. The bias-tee is the last element in the source block, which is connected to the gate/base bias source via a low-frequency tuner network for subharmonic impedance control. Since the current draw of the gate/base is typically small, remote sensing of the power supply can be done directly at the bias-tee.

Although components within the source block may have type-N or 3.5 mm connectors, interface to the source tuner is done with an adapter to an APC 7 mm connector. This is done to provide a robust connection and to aid in the VNA characterization of the source block. Depending on the measurements that are to be made during load-pull, a variety of instruments may be connected to the incident and reflected sample ports, including a power meter and VNA. The former is required for real-time leveling and the latter for measuring the input impedance to the DUT [9].

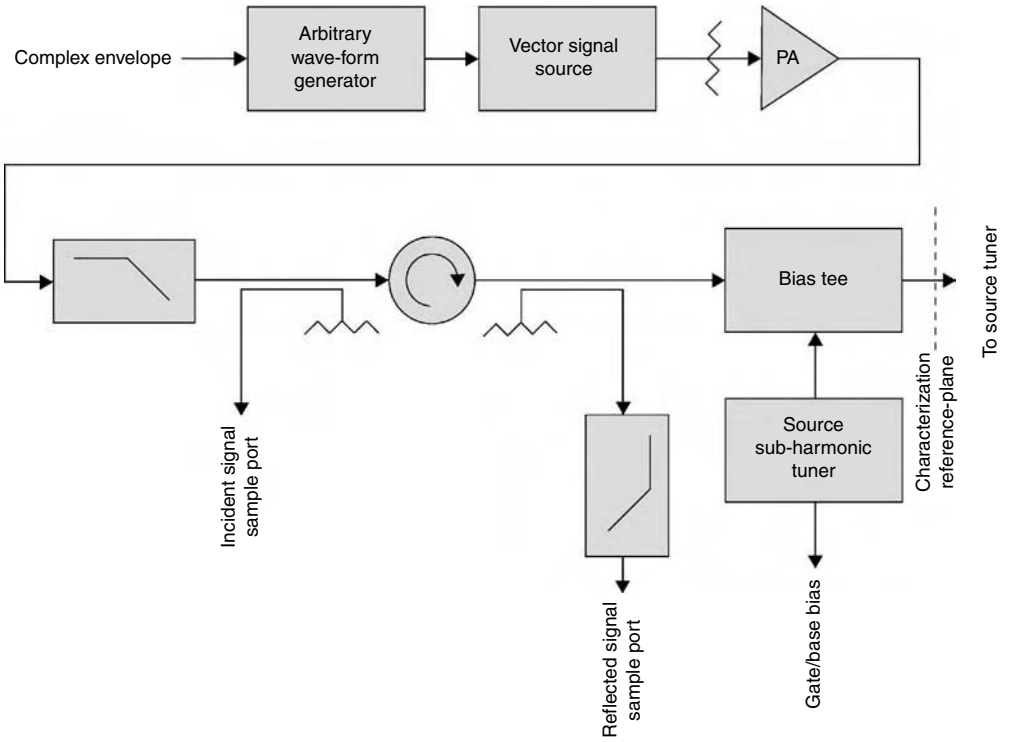


FIGURE 7.2 Detail of the source portion of Figure 7.1.

The load block of Figure 7.1 usually includes a port for sampling the load signal of the DUT and the padding and filtering necessary to interface the load signal to a power sensor. Figure 7.3 shows the details of a typical load block. The bias-tee comes first. Although remote-sense can be sampled here, in situations where significant current is required, the remote-sense should be sampled directly on the DUT test fixture. For a load-pull system capable of 100 W average power, the attenuator following the bias-tee should be appropriately rated and exhibit at least 30 dB attenuation.

The load signal is sampled at a directional coupler after the high-power pad. A spectrum analyzer is often connected at this port, and it may be useful to use a low coupling factor, for example, –30 dB, to minimize the padding necessary in front of the spectrum analyzer. This results in an optimal dynamic range of the system for measuring ACPR. Following the directional coupler is a low-pass filter, to remove harmonics,* which is followed by another attenuator. This attenuator is used to improve the return loss of the filter with respect to the power sensor. As with the source block, interface to the load tuner and power sensor are done with APC 7 mm connectors to improve robustness and power-handling capability.

The DUT test-fixture is used to interface the source and load tuners to a package. For cost and package de-embedding reasons, it is useful to standardize on two or three laboratory evaluation packages. For hybrid circuit design, it is useful to design a test fixture with feeds and manifolds identical to those used in hybrid to mitigate de-embedding difficulties. The collector/drain side of the test fixture should also have a sampling port for remote sensing of the power supply.

*Although a filter is not necessary, characterization of a DUT in significant compression will result in the average power detected by the power sensor including fundamental and harmonic power terms. When the DUT is embedded into a matching network, the matching network will usually attenuate the harmonics; thus, inclusion of the low-pass filter more closely approximates the performance that will be observed in practice.

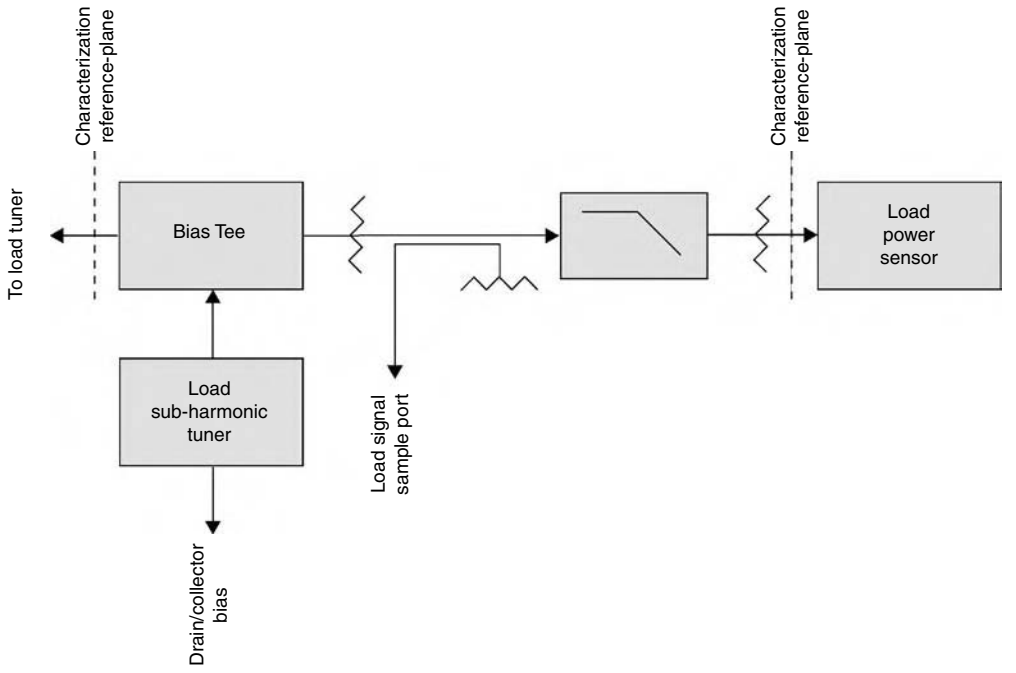


FIGURE 7.3 Detail of the load portion of Figure 7.1.

After the load-pull system has been assembled, it is recommended that the maximum expected power be applied to the system and changes in impedance be measured due to tuner self-heating. This may be significant where average powers exceed 100 W or peak powers exceed several hundred watts. Any impedance change will establish the upper power limit of the system with respect to impedance accuracy.

7.2 Characterization of System Components

Each of the blocks described in the previous section must be characterized using s-parameters in order for a load-pull system to function properly. In this section, the characterization procedure for each of the sections of Figure 7.1 is described, with emphasis on calibration of the VNA and the characterization of the transistor test fixture. Two-tier calibration and impedance renormalization are considered for characterizing quarter-wave prematching test fixtures.

7.2.1 Vector Network Analyzer Calibration Theory

Due to the extremely low impedances synthesized in high-power load-pull, the VNA calibration is the single most important element of the characterization process. Any errors in the measurement or calibration, use of low quality connectors, for example, SMA or type-N, or adoption of low-performance calibration methods, for example, SOLT, will result in a significant reduction in accuracy and repeatability. Only TRL calibration should be used, particularly for tuner and fixture characterization. Use of high-performance connectors is preferred, particularly APC 7 mm, due to its repeatability, power handling capability, and the fact that it has a hermaphroditic interface, simplifying the calibration process.

Vector network analysis derives its usefulness from its ability to characterize impedance based on ratio measurements, instead of absolute power and phase measurements, and from its ability to characterize and remove systematic errors due to nonidealities of the hardware. For a complete review of VNA architecture and calibration theory, the reader is encouraged to review notes from the annual ARFTG Short-Course given in November of each year [10,11].

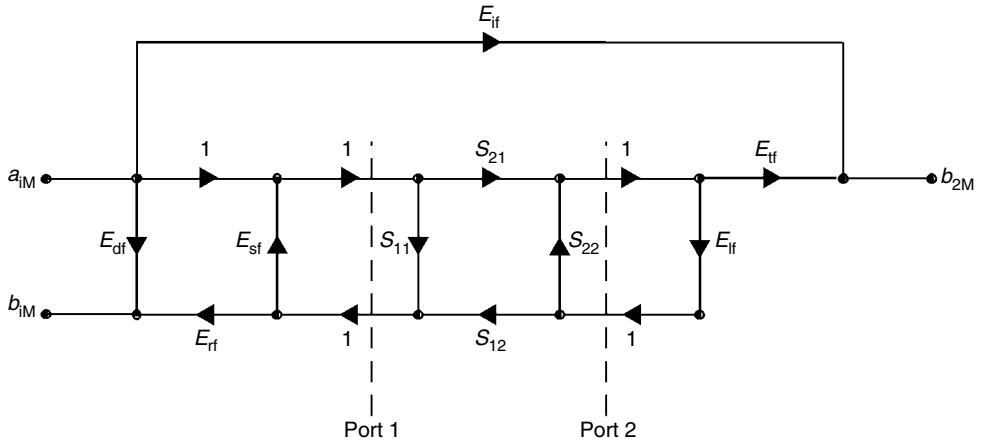


FIGURE 7.4 Signal-flow graph of the forward direction of a typical VNA.

Figure 7.4 shows a signal-flow graph of the forward direction of a common VNA architecture, where six systematic error terms are identified. An identical flow-graph exists for the reverse direction, with six additional error terms. Consider the situation where it is required to measure an impedance that exhibits a near total reflection, such as a load tuner set for 1 Ω. Assuming a 50 Ω reference impedance, nearly all of the incident power is reflected back toward the VNA, along with a phase shift of 180°. Consider what happens when the reflected wave is sampled at the VNA, denoted as b_{1M} in Figure 7.4. If there is any re-reflection of the reflected wave incident at the VNA, an error will occur in measuring the actual impedance of the load. The ability of a VNA to minimize this reflected power is characterized by its residual source match, which is the corrected source impedance looking into the VNA. The uncorrected source impedance looking into the VNA is characterized by the E_{sf} term in the flow graph of Figure 7.4.

Continuing with this example, Figure 7.5 shows a plot of the upper bound on apparent load impedance versus the residual source match (with respect to a reference impedance of 50 Ω and an actual impedance of 1 Ω). For simplicity, it is assumed that the residual source match is in phase with the reflected signal. Also shown are typical residual source match performance numbers for an HP 8510C using an HP 8514B test set. From this graph it is clear that use of low-performance calibration techniques will result in latent errors in any characterization performed using a DUT with reflection VSWR near 50:1. Using a 3.5 mm SOLT calibration can result in nearly 20% uncertainty in measuring impedance. Note that TRL*, the calibration method available on low-cost VNAs, offers similar performance to 3.5 mm SOLT, due to its inability to uniquely resolve the test-set port impedances. This limitation is due to the presence of only three samplers instead of four, and does not allow switch terms to be measured directly. For this reason, it is recommended that three-sampler architectures not be used for the characterization process.

Similar arguments can be made for the load reflection term of Figure 7.4, which is characterized by the residual load match error term. Identical error terms exist for the reverse direction too, so that there are a total of four error terms that are significant for low impedance VNA calibration.

TRL calibration requires a thru line, a reflect standard (known only within λ/4), and a delay-line. The system reference impedances will assume the value of the characteristic impedance of the delay-line, which if different from 50 Ω, must be appropriately renormalized back to 50 Ω [12–15]. TRL calibration can be done in a variety of media, including APC 7 mm coaxial waveguide, rectangular/cylindrical waveguide, microstrip, and stripline. Calibration verification standards, which must be used to extract the residual error terms described above, are also easily fabricated. Figure 7.6 shows the residual forward source and load match response of an APC 7 mm calibration using an HP 8510C with an HP 8514B test

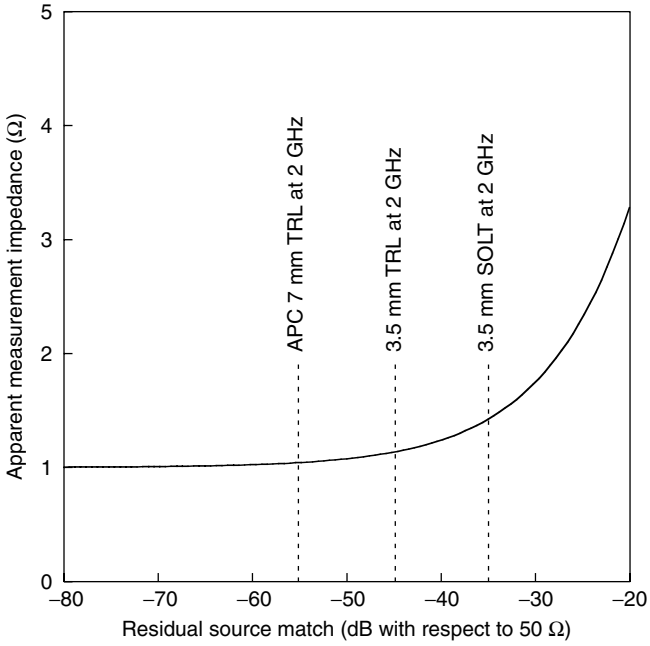


FIGURE 7.5 The influence of residual source match on the ability of a VNA to resolve a 1 Ω impedance with a 50 Ω reference impedance. The calibration performance numbers are typical for an HP 8510C with an 8514B test-set operating a 2 GHz.

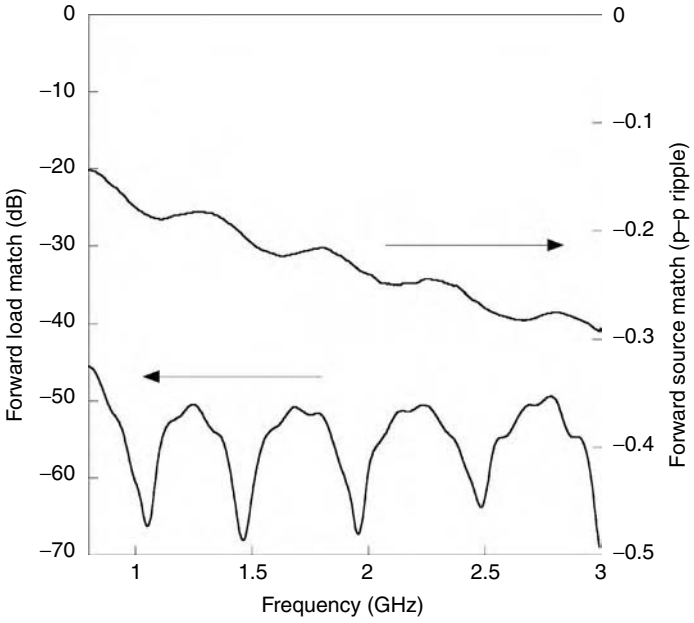


FIGURE 7.6 Typical response of an APC 7 mm TRL calibration using an offset-short and delay-line to extract source match and load match, respectively. The data were taken from an HP 8510C with an HP 8514B test set.

set. These were obtained with a 30 cm offset-short airline and 30 cm delay-line, respectively [16,17,18]. The effective source match is computed from the peak–peak ripple using

$$E_{sf} = 10 * \log_{10}\left[\frac{1-10^{-\frac{p-p\ ripple}{20}}}{1+10^{-\frac{p-p\ ripple}{20}}}\right] \tag{7.1}$$

where it is seen that better than –53 dB source match is obtained across the band. Due to finite directivity, 6 dB must be subtracted from the plot showing the delay-line response, indicating that better than –56 dB load match is obtained except near the low end of the band. Calibration performance such as that obtained in Figure 7.6 is necessary for accurate tuner and fixture characterization, and is easily achievable using standard TRL calibration.

For comparison purposes, Figures 7.7 and 7.8 show forward source and load match for 3.5 mm TRL and SOLT calibration, respectively. Here it is observed that the source match of the 3.5 mm TRL calibration has significantly degraded with respect to the APC 7 mm TRL calibration and the 3.5 mm SOLT calibration has significantly degraded with respect to the 3.5 mm TRL calibration.

Proper VNA calibration is an essential first step in characterization of any component used for high-power load-pull characterization and is particularly important for tuner and fixture characterization. All VNA calibrations should be based on TRL and must be followed by calibration verification to ensure that the calibration has been performed properly and is exhibiting acceptable performance, using the results of Figure 7.6 as a benchmark. Averaging should be set to at least 64. Smoothing should in general be turned off in order to observe any resonances that might otherwise be obscured. Although APC 7 mm is recommended, 3.5 mm is acceptable when used with a TRL calibration kit. Under no circumstances should type-N or SMA connectors be used, due to phase repeatability limitations and connector reliability limitations.

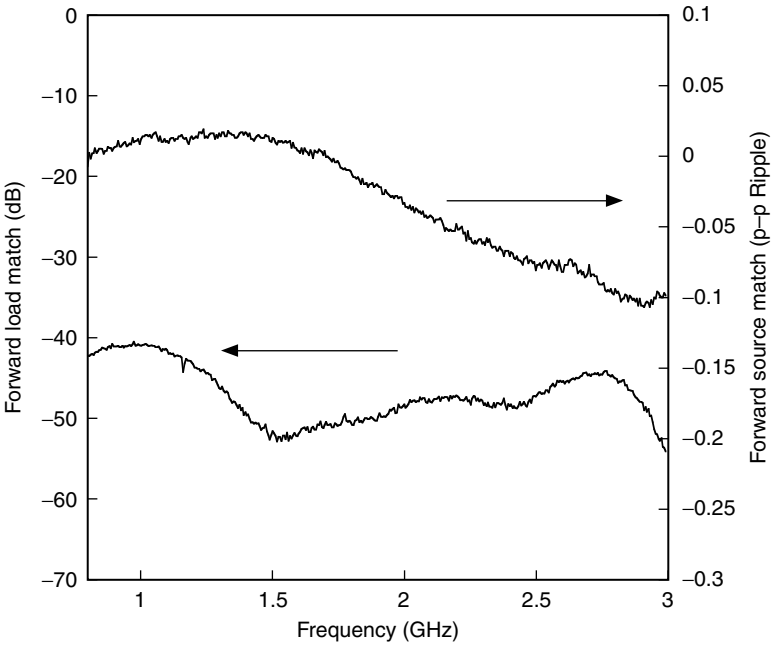


FIGURE 7.7 Typical response of a 3.5 mm TRL calibration using an offset-short and delay-line to extract source match and load match, respectively. The data were taken from an HP 8510C with an HP 8514B test set.

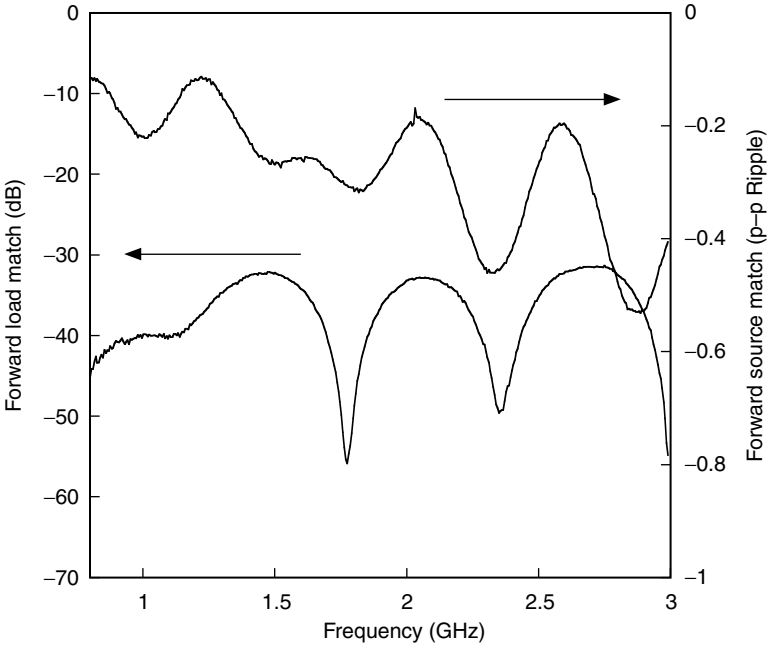


FIGURE 7.8 Typical response of a 3.5 mm SOLT calibration using an offset-short and delay-line to extract source match and load match, respectively. The data were taken from an HP 8510C with an HP 8514B test set.

7.2.2 s-Parameter Characterization of Tuners

Tuner characterization begins with proper calibration of the VNA, as described in the previous section. It is suggested at this point that any adapters on the tuner be serialized and alignment marks made to ensure that in the event of removal, they can be replaced in their original positions. Replacement of an adapter, for any reason, will require a new tuner characterization. Tuners should be leveled using a bubble-level and should be positioned such that the VNA test-port cables are not flexed. Proper torquing of all connector interfaces is essential. Since the tuner files usually consist of a small number of frequencies with respect to the number of frequencies present in a typical VNA calibration, it is appropriate to increase the number of averages to 128 or 256.

It is generally most useful to characterize a tuner without any additional components attached, such as a bias-tee, in order to maintain maximum flexibility in the use of the tuner subsequent to the characterization. For tuners that are being characterized for the first time, it is recommended that they be fully evaluated for insertion loss, minimum and maximum VSWR, and frequency response to ensure they are compliant with the manufacturer's specifications.

After characterization the tuner file should be verified by setting the tuner for arbitrary impedances near the center and edge of the Smith Chart over 2¼ radians. The error should be less than 0.2% for magnitude and 0.1° for phase. Anything worse than this may indicate a problem with either the calibration (verify it again) or the tuner.

7.2.3 s-Parameter Characterization of System Components

Characterization of system components consists of creating one-port and two-port s-parameter files of the source block and load block, as shown in Figures 7.1 and 7.2, respectively. Each of these figures show suggested reference-planes for characterization of the network. Since the reflection coefficient of each port of the source and load blocks is in general small with respect to that exhibited by tuners, the VNA

calibration is not as critical* as it is for tuner characterization. Nevertheless, it is recommended to use the same calibration as used for the tuner characterization and to sweep a broad range of frequencies to eliminate the possibility of characterization in the future at new frequencies.

If possible, each component of the source and load blocks should be individually characterized prior to integration into their respective block. This is particularly so for circulators and high-current bias-tees, which tend to have limited bandwidth. The response of the source and load block should be stored for future reference and/or troubleshooting.

7.2.4 Fixture Characterization to Increase System VSWR

In the beginning of this chapter it was indicated that high-power load-pull may require source and load impedances in the neighborhood of 0.1 Ω. This does not mean that the DUT may require such an impedance as much as it is necessary for generating closed contours, which are useful for evaluation of performance gradients in the gamma domain. A very robust and simple method of synthesizing sub-1 Ω impedances is to use a quarter-wave pre-matching network characterized using numerically well-defined two-tier calibration methods. To date, use of quarter-wave pre-matching offers the lowest impedance, though it is limited in flexibility due to bandwidth restrictions. Recently, commercially available passive mechanical systems cascading two tuners together have been made available offering octave bandwidths, though they are not able to generate impedances as low as narrowband quarter-wave pre-matching. In this section, a robust methodology for designing and characterizing a quarter-wave pre-matching network capable of presenting 0.1 Ω at 2 GHz is described in References 16 and 18. It is based on a two-tier calibration with thin-film gold on alumina substrates (quarter-wave pre-matching networks on soft substrates are not recommended due to substrate variations and repeatability issues over time).

The theory of quarter-wave pre-matching begins with the mismatch invariance property of lossless networks [19]. Consider the quarter-wave line of characteristic impedance Z_{ref} shown in Figure 7.9. This line is terminated in a mismatch of $VSWR_{load}$ with an arbitrary phase. The reference impedance of $VSWR_{load}$ is Z_L. The mismatch invariance property of lossless networks shows that the input VSWR is identical to the load VSWR, but it is with respect to the quarter-wave transformed impedance of Z_L. Thus, the minimum achievable impedance, which is real valued, is the impedance looking into the quarter-wave line when it is terminated in Z_L divided by $VSWR_{load}$. This is expressed as

$$R_{in,min} = \frac{\frac{Z_{ref}^2}{Z_L}}{VSWR_{load}} \tag{7.2}$$

Suppose it is desired to synthesize a minimum impedance of 0.1 Ω, which might be required for characterizing high power PCS and UMTS LDMOS transistors. If a typical passive-mechanical tuner is capable of conservatively generating a 40:1 VSWR, then the input impedance of the quarter-wave line must be approximately 4 Ω, requiring the characteristic impedance of the quarter-wave line to be approximately

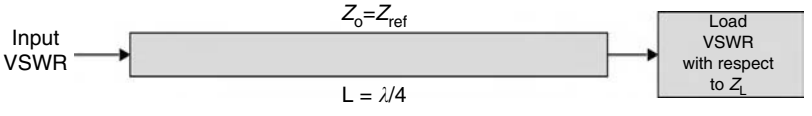


FIGURE 7.9 Network to describe the mismatch invariance property of lossless networks.

*If the magnitude of the reflection coefficient approaches the residual directivity of the VNA calibration, then errors may occur.

14 Ω, assuming a Z_L of 50 Ω. to the extent that the minimum impedance deviates from the ideal is directly related to fixture losses. Thus, the importance of using a low-loss substrate and metal system is apparent.

Full two-port characterization of each fixture side is necessary to reset the reference plane of each associated tuner. Several methods are available to do this, including analytical methods based on approximate closed-form expressions, full-wave analysis using numerical techniques, and employment of VNA error correction techniques [20,21,22]. The first method is based on approximations that have built-in uncertainty, as does the second method, in the form of material parameter uncertainty. The third method is entirely measurement based, and relies on well-behaved TRL error correction mathematics to extract a two-port characterization of each fixture half from a two-tier calibration. More importantly, using verification standards, it is possible to quantify the accuracy of the de-embedding, as described in the section on VNA calibration.

Using the error-box formulation of the TRL calibration, it is possible to extract the two-port characteristics of an arbitrary element inserted between two reference planes of two different calibrations [11]. The first tier of the calibration is usually done at the test-port cables of the VNA. The second tier of the calibration is done in the media that matches the implementation of the test fixture, which is usually microstrip. Figure 7.10 illustrates the reference-plane definitions thus described. The second tier of the calibration will have its reference impedance set to the impedance of the delay standard, which is the impedance of the quarter-wave line. Although there are many methods of determining the characteristic impedance of a transmission line, methods based on estimating the capacitance per unit length and phase velocity are well suited for microstrip lines [12,15]. The capacitance per unit length and phase velocity uniquely describe the quasi-TEM characteristic impedance as

$$Z_O = \frac{1}{v_p C} \tag{7.3}$$

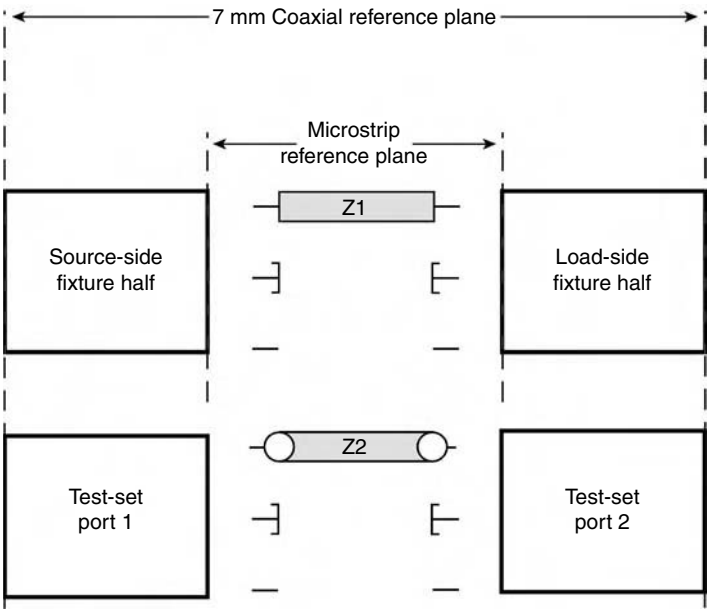


FIGURE 7.10 Reference-plane definitions for a two-tier calibration used for fixture characterization. The first tier is based on a TRL APC 7 mm calibration and the second tier is based on a microstrip TRL calibration.

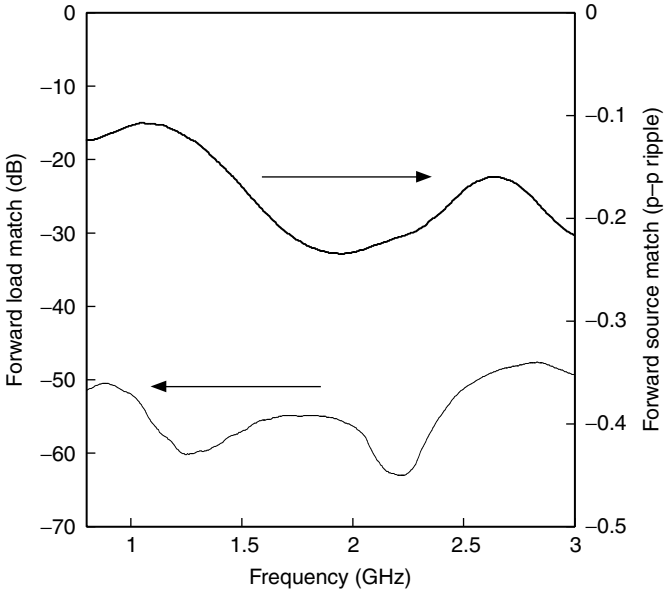


FIGURE 7.11 Microstrip TRL calibration using an offset-short and delay-line to extract source match and load match, respectively. The data were taken from an HP 8510C with an HP 8514B test set.

Once the characteristic impedance of the delay-line is known, the s-parameters can be renormalized to 50 Ω to make them compatible with the 50 Ω reference impedance that most automated load-pull systems use [2,3,15].

Figure 7.11 shows the forward source and load match of the second tier microstrip calibration used in the pre-matching fixture described in References 16 and 18. This fixture was intended to present 0.1 Ω at 2 GHz with extremely high accuracy. From the verification data, the resultant source match is better than –45 dB across the band and the resultant load match is better than –52 dB across the band. Comparing these results with Figure 7.5 shows that the uncertainty is very low.

A significant advantage of using a transforming network to increase system VSWR, whether it be a quarter-wave line or an additional cascaded tuner, is that the two-port characterization of each element is done at manageable impedance levels. Characterization of a tuner presenting a 50:1 VSWR in direct cascade of a quarter-wave pre-match network would result in a significant increase in measurement uncertainty since the VNA must resolve impedances near 0.1 Ω. Segregating the characterization process moves the impedances that must be resolved to the 1–2 Ω range, where the calibration uncertainty is considerably smaller.

The final step of the fixture verification process is to verify that the two-tier calibration has provided the correct two-port s-parameter description of each fixture half. Figure 7.12 shows each fixture half cascaded using the port definitions adopted by NIST Multical™ [15]. With microstrip, an ideal thru can be approximated by butting each fixture half together and making top-metal contact with a thin conductive film. When this is not possible, it is necessary to extract a two-port characterization of the thru. The cascaded transmission matrix is expressed as

$$\begin{bmatrix} A_{11} & B_{12} \\ C_{21} & D_{22} \end{bmatrix}_{cascade} = \begin{bmatrix} A_{11} & B_{12} \\ C_{21} & D_{22} \end{bmatrix}_{source} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}_{thru} \begin{bmatrix} A_{11} & B_{12} \\ C_{21} & D_{22} \end{bmatrix}_{load} \tag{7.4}$$

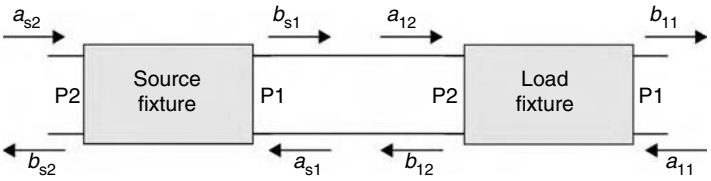


FIGURE 7.12 Port and traveling-wave definitions for cascading the source-fixture and load-fixture to examine the accuracy of the two-tier calibration fixture characterization.

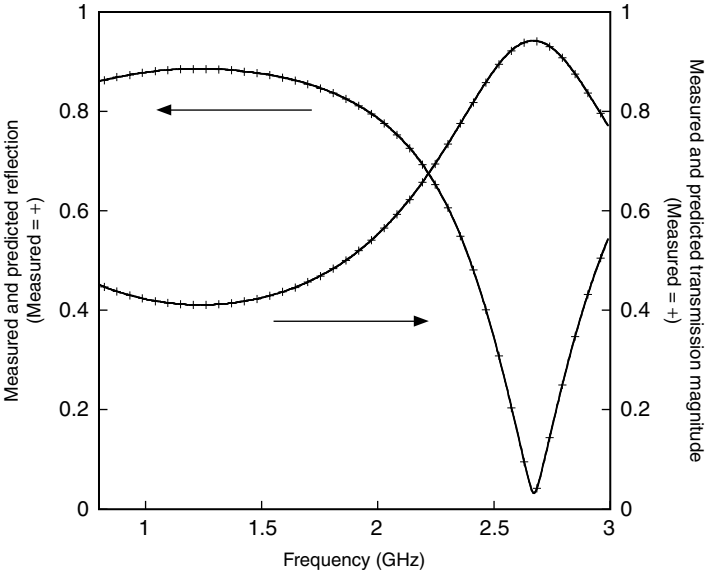


FIGURE 7.13 Forward reflection and transmission magnitude comparison of measured and cascaded fixture response. The error is so small the curves sit on top of each other.

where the middle matrix of the right-hand-side is the transmission matrix of a lossless zero phase-shift thru network. Converting the cascade transmission matrix back to s-parameter form yields the predicted response of the cascaded test-fixture, which can then be compared to the measurements of the cascade provided by the VNA.

Figure 7.13 shows the measured and predicted cascade magnitude response of a typical PCS quarter-wave pre-matching fixture based on an 11 Ω quarter-wave line; the phase is shown in Figure 7.14 [16,18]. The relative error across the band is less than 0.1%. This type of fixture characterization performance is necessary to minimize error for synthesizing sub-1 Ω impedances.

7.3 System Performance Verification

Just as verification of VNA calibration is essential, so too is verification of overall load-pull system performance essential. Performance verification can be done with respect to absolute power or with respect to power gain. The former is recommended only occasionally, for example when the system is assembled or when a major change is made. The latter is recommended subsequent to each power calibration. Each of the methods will be described in this section.

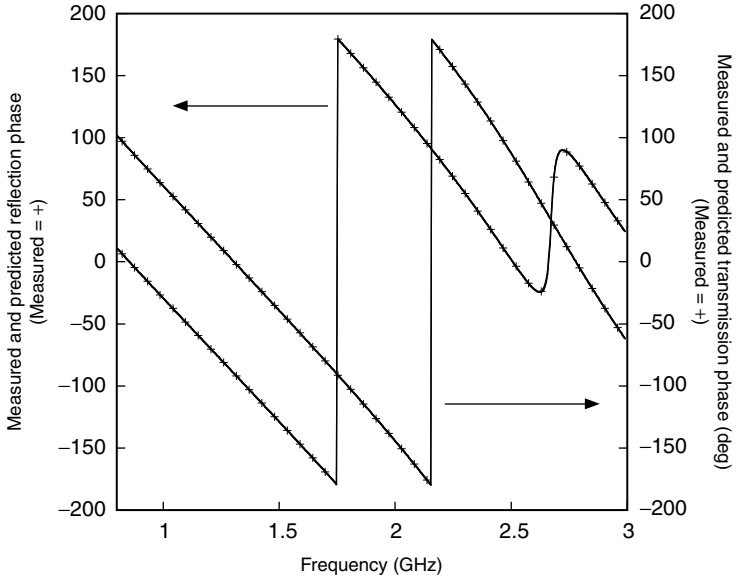


FIGURE 7.14 Forward reflection and transmission phase comparison of measured and cascaded fixture response. The error is so small the curves sit on top of each other.

Absolute power calibration is done by applying a signal to the source tuner via the source block of Figure 7.2. After appropriately padding a power sensor, it is then connected to DUT side of the source tuner and, with the tuners set for 1:1 transformation, the resultant power is compared to what the overall cascaded response is expected to be.

This procedure is repeated for the load tuner except that the signal is injected at the DUT side of the load tuner and the power sensor is located as shown in Figure 7.3. Splitting this verification in two steps assists in isolating any issues with either the source or load side. It is also possible to vary the impedance of each tuner and calculate what the associated available gain or power gain is, although this step is more easily implemented in the power-gain verification.

Power-gain verification starts with a two-port characterization of a known mismatch standard. The simplest way to implement this standard is to use one of the tuners, and then set the other tuner for the conjugate of this mismatch. In this case, the mismatch standard is an ideal thru similar to the one used in fixture verification described in the previous section. Since it is unlikely that both the source and load tuners would have identical impedance domains, the measured loss must be compensated to arrive at actual loss. To compensate for this, the mismatch loss is computed as

$$G_{mm} = 10\log_{10}\left[\frac{\left(1-\left|\Gamma_s\right|^2\right)\left(1-\left|\Gamma_l\right|^2\right)}{\left|1-\Gamma_s\Gamma_l\right|^2}\right] \tag{7.5}$$

where Γ_s and Γ_l are the source and load reflection coefficients, respectively, looking back into each tuner. Figure 7.15 shows a typical response of an entire cascade, including the quarter-wave prematching network. A transducer gain response boundary of ±0.1 dB is typical, and ±0.2 dB should be considered the maximum.

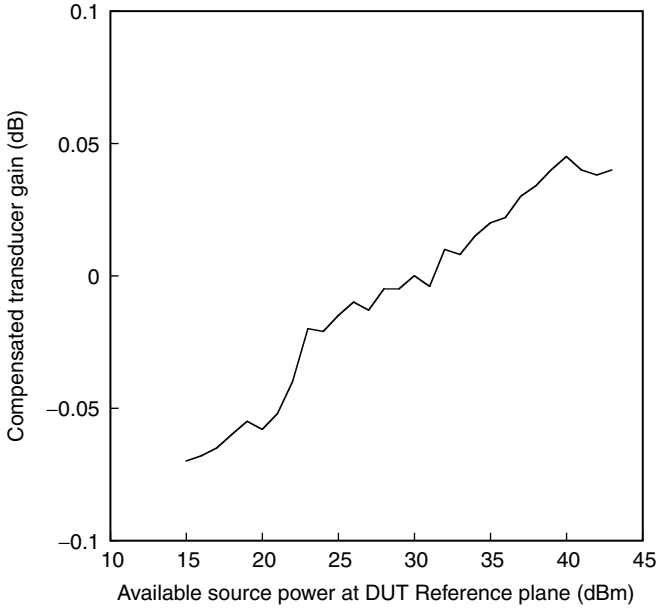


FIGURE 7.15 Measured transducer gain under the condition of conjugate match with mismatch loss compensation included.

7.4 Summary

Load-pull is a valuable tool for evaluating high-power RF and microwave transistors, designing PAs, and verifying large-signal model performance and validity domains. To enhance the reliability of the data that a load-pull system provides, it is essential that high performance VNA calibration techniques be adopted. Further, as emphasized in the present section, treating each section of the load-pull separately is useful from a measurement perspective and from a problem resolution perspective. In the former case, it was shown that measuring quarter-wave pre-matching networks and tuners separately reduces the uncertainty of the calibration. In the latter case, it was shown that characterization of each section individually allows its performance to be verified prior to integrating it within the entire system.

The central theme of this section has been the VNA and its associated calibration. Due to the extremely low impedances synthesized in high-power load-pull, the VNA calibration is the single most important element of the characterization process. Any errors or uncertainty encountered in the VNA calibration will be propagated directly into the load-pull characterization files and may result in erroneous data, particularly if system performance verification is not performed.

To present the sub-1 Ω impedances necessary for evaluation of high-power transistors, transforming networks are required. These can be implemented using an impedance transforming network, such as a quarter-wave line, or by cascading two tuners together. The former offers the highest VSWR at the expense of narrow bandwidth, while the latter is in general more flexible. In either case, high performance and reliable characterization methods are necessary to attain the best possible results for using load-pull as a verification and design tool.

Acknowledgments

Kerry Burger (Philips), Mike Majerus (Motorola), and Gary Simpson and John King (Maury Microwave) have, in many ways, influenced the content of this chapter. Their support and friendship are happily acknowledged.

References

1. J. M. Cusak et al., Automatic load-pull contour mapping for microwave power transistors, *IEEE Transactions on Microwave Theory and Techniques*, 1146–1152, Dec. 1974.
2. *Automated Tuner System User's Manual*, v.1.9, Maury Microwave Corporation, 1998.
3. *Computer Controlled Tuner System User's Manual*, v. 6.0, Focus Microwave Corporation, 1998.
4. *LP2 Automated Load-Pull System User's Manual*, ATN Microwave Corporation, 1997.
5. F. Larose, F. Ghannouchi, and R. Bosisio, A new multi-harmonic load-pull method for non-linear device characterization and modeling, *Digest of the IEEE International Microwave Symposium Digest*, 443–446, June 1990.
6. F. Blache, J. Nebus, P. Bouysse, and J. Villotte, A novel computerized multi-harmonic load-pull system for the optimization of high-efficiency operating classes in power transistors, *IEEE International Microwave Symposium Digest*, 1037–1040, June 1995.
7. J. Sevic, R. Baeten, G. Simpson, and M. Steer, Automated large-signal load-pull characterization of adjacent-channel power ratio for digital wireless communication system, *Proceedings of the 45th ARFTG Conference*, 64–70, Nov. 1995.
8. J. Sevic, K. Burger, and M. Steer, A novel envelope-termination load-pull method for the ACPR optimization of RF/microwave power amplifiers, *Digest of the IEEE International Microwave Symposium Digest*, 723–726, June 1998.
9. G. Simpson and M. Majerus, Measurement of large-signal input impedance during load-pull, *Proceedings of the 50th ARFTG Conference*, 101–106, Dec. 1997.
10. D. Rytting, ARFTG Short-Course: Network Analyzer Calibration Theory, 1997.
11. R. Marks, Formulation of the basic vector network analyzer error model including switch terms, *Proceedings of the 50th ARFTG Conference*, 115–126, Dec. 1997.
12. R. Marks and D. Williams, Characteristic impedance measurement determination using propagation measurement, *IEEE Microwave and Guided Wave Letters*, 141–143, June 1991.
13. G. Engen and C. Hoer, Thru-reflect-line: an improved technique for calibrating the dual six-port automatic network analyzer, *IEEE Transactions on Microwave Theory and Techniques*, 987–993, Dec. 1979.
14. R. Marks, A multi-line method of network analyzer calibration, *IEEE Transactions on Microwave Theory and Techniques*, 1205–1215, July 1990.
15. *MultiCal™ User's Manual*, v. 1.0, National Institute of Standards and Technology, 1997.
16. J. Sevic, A sub 1 Ω load-pull quarter-wave pre-matching network based on a two-tier TRL calibration, *Proceedings of the 52nd ARFTG Conference*, 73–81, Dec. 1998.
17. D. Balo, Designing and calibrating RF fixtures for SMT devices, *Hewlett-Packard 1996 Device Test Seminar*, 1996.
18. J. Sevic, A sub 1 Ω load-pull quarter-wave pre-matching network based on a two-tier TRL calibration, *Microwave Journal*, 122–132, Mar. 1999.
19. R. Collin, *Foundations for Microwave Engineering*, McGraw-Hill: New York, 1966.
20. B. Wadell, *Transmission Line Design Handbook*, Artech House: Boston, 1991.
21. *EM User's Manual*, v. 6.0, Sonnet Software, Inc., Liverpool, NY, 1999.
22. *HP 8510C User's Manual*, Hewlett-Packard Company, 1992.

8

Pulsed Measurements

Anthony E. Parker
Macquarie University

James G. Rathmell
The University of Sydney

Jonathan B. Scott
University of Waikato

8.1 Introduction

Pulsed measurements attempt to ascertain the radio-frequency (RF) behavior of transistors or other devices at an unchanging **bias condition**. A pulsed measurement of a transistor begins with the application of a bias to its terminals. After the bias has settled to establish a **quiescent condition**, it is perturbed with pulsed stimuli during which the change in terminal conditions, voltage and current, is recorded. Sometimes a RF measurement occurs during the pulse. The responses to the pulse stimuli quantify the behavior of the device at the established quiescent point. **Characteristic curves**, which show the relationship between terminal currents or RF parameters and the instantaneous terminal potentials, portray the behavior of the device.

Pulsed measurement of the characteristic curves is done using short pulses with a relatively long time between pulses to maintain a constant quiescent condition. The characteristic curves are then specific to the quiescent point used during the measurement. This is of increasing importance with the progression of microwave-transistor technology because there is much better consistency between characteristic curves measured with pulses and responses measured at high frequencies. When the behavior of the device is bias or rate dependent, pulsed measurements yield the high-frequency behavior because the bias point

remains constant during the measurement. Pulse techniques are useful for characterizing devices used in large-signal applications or for testing equipment used in pulse-mode applications. When measurements at high potentials would otherwise be destructive, a pulsed measurement can safely explore breakdown or high-power points while maintaining a bias condition in the safe-operating area (SOA) of the device. When the device normally operates in pulse mode, a pulsed measurement ascertains its true operation.

The response of most microwave transistors to high-frequency stimuli depends on their operating conditions. If these conditions change, the characteristic curves vary accordingly. This causes dispersion of the characteristic curves when measured with traditional curve tracers. That is, a change in characteristics over time or frequency. The operating condition when sweeping up to a measurement point is different to that when sweeping down to the same point. The implication is that any change in the operating conditions during the measurement will produce ambiguous characteristic curves.

Mechanisms collectively called **dispersion effects** contribute to dispersion in characteristic curves. These mechanisms involve thermal and electron-trapping phenomena, and impact ionization. Dispersion effects are sometimes referred to as memory effects or as dynamic bias. Usually they are slow acting at moderate bias levels, so while the operating conditions of the device affect them, RF stimuli do not. Even if the sequence of measurement precludes observation of dispersion, dispersion effects may still influence the resulting characteristic curves.

Pulsed measurements can be used to acquire characteristic curves that are less affected by dispersion effects. The strategy is to maintain a constant operating condition while measuring the characteristic curves. The pulses are normally short enough to be a signal excursion rather than a change in bias. The period between pulses is normally long enough for the quiescent condition of the device to recover from any perturbation that may occur during each pulse.

Pulse techniques cause less strain, so are suitable for extending the range of measurement into regions of high-power dissipation and electrical breakdown. Pulse techniques are also valuable for experiments in device physics and exploration of new devices and material systems at a fragile stage of development.

Stresses that occur when operating in regions of breakdown or overheating can alter the characteristic curves permanently. In many large-signal applications, there can be excursions into both of these regions for periods brief enough to avoid undue stress on the device. To analyze these applications, it is desirable to extend characteristic curves into the stress regions. That is, the measurements must extend as far as possible into regions that are outside the SOA of the device. This leads to another form of dispersion, where the characteristic curves change after a measurement at a point that stresses the device.

Pulsed measurements can extend to regions outside the SOA without stressing or damaging the device. If the pulses are sufficiently short, there is no permanent change in the characteristic curves. With pulses, the range of the measured characteristic curves can often extend to encompass completely the signal excursions experienced during the large-signal operation of devices.

In summary, pulsed measurements yield an extended range of characteristic curves for a device that, at specific operating conditions, correspond to the high-frequency behavior of the device. The following sections present the main principles of the pulse technique and the pulse-domain paradigm, which is central to the technique. The pulse-domain paradigm considers the characteristic curves to be a function of quiescent operating condition. Therefore, the basis for pulse techniques is the concept of measurements made in **isodynamic** conditions, which is effectively an invariable operating condition. A discussion is included of the requirements for an isodynamic condition, which vary with the transistor type and technology. There is also a review of pulsed measurement equipment and specifications in terms of cost and complexity, which vary with application. Finally, there is an examination of various pulsed measurement techniques.

8.2 Isothermal and Isodynamic Characteristics

For the analysis of circuit operation and the design of circuits, designers use transistor characteristics. The characteristics consist of characteristic curves derived from measurements or theoretical analysis. These give the relationship between the variable, but interdependent, terminal conditions and other information

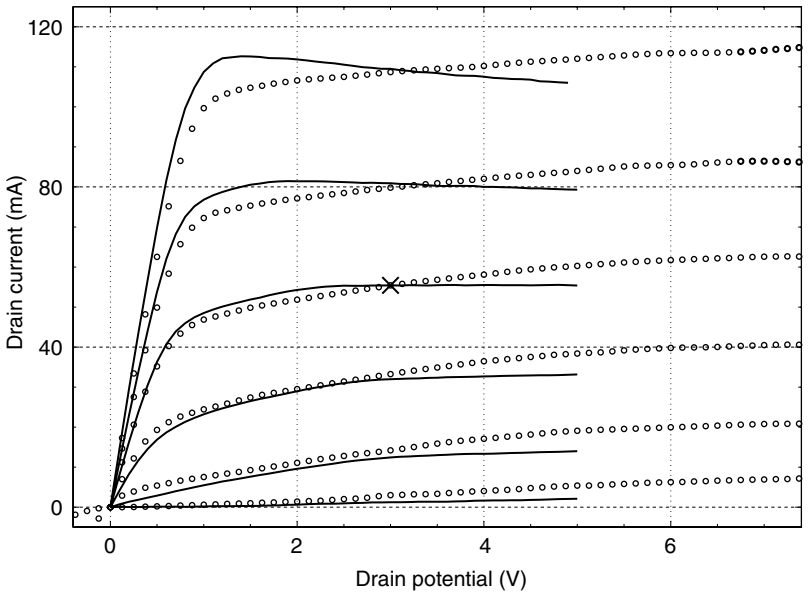


FIGURE 8.1 Characteristic curves for a MESFET. Shown are the DC characteristics (—) and the pulsed characteristics (∘), with 300 ns pulses separated by 200 ms quiescent periods, for the quiescent point $V_{DS} = 3.0$ V, $I_D = 55.4$ mA (×). Gate-source potential from −2.0 to +0.5 V in 0.5 V steps is the parameter.

that describes the behavior of the device. To be useful, the characteristics need to be applicable to the operating condition of the device in the circuit.

In all circuits, when there is no RF signal the device operates in a quiescent condition established by bias networks and power supplies. The **DC characteristics** are characteristic curves obtained with slow curve tracers, conventional semiconductor analyzers, or variable power supplies and meters. They are essentially data from a set of measurements at different bias conditions. Consequently, the quiescent operating point of a device is predictable with DC characteristics derived from DC measurements. Figure 8.1 shows a set of DC characteristics for a typical microwave MESFET. This figure also shows the very different set of **pulsed characteristics** for the same device made at the indicated quiescent point. The pulsed characteristics give the high-frequency behavior of the MESFET when biased at that quiescent point.

A clear example of a dispersion effect that causes the observed difference between the DC and pulsed characteristics is heating due to power dissipation. When the characteristics are measured at a slow rate (≈10 ms/point) the temperature of the device at each data point changes to the extent that it is heated by the power being dissipated at that point. Pulsed characteristics are determined at the constant temperature corresponding to the power dissipation of a single bias point. This measurement at constant temperature is one made in isothermal conditions. In general, device RF characteristics should be measured in a constant bias condition that avoids the thermal and other dispersion effects that are not invoked by a RF signal. Such a measurement is one made in isodynamic conditions.

8.2.1 Small-Signal Conditions

Devices operating in small-signal conditions give a nearly linear response, which can be determined by steady-state RF measurements made at the quiescent point. A network analyzer, operated in conjunction with a bias network, performs such a measurement in isodynamic conditions. Once the quiescent condition is established, RF measurements characterize the terminal response in terms of small-signal parameters, such as y-parameters. A different set of small-signal parameters is required for each quiescent condition.

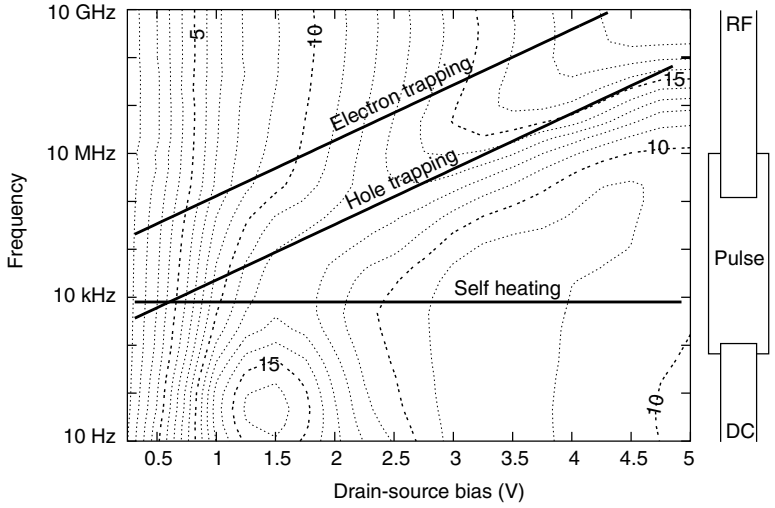


FIGURE 8.2 Frequency and drain-bias variation of intrinsic gain for a HEMT at a gate-bias of $V_{GS} = -0.8$ V. The characteristic frequencies of heating and charge trapping effects are indicated by the labeled lines. Also indicated are the typical frequency ranges applicable to DC, pulsed, and RF measurements.

It is not possible to correlate the small-signal parameters with the DC characteristics when there are dispersion effects. For example, as shown in Figure 8.2, the intrinsic gain (ratio of gate-drain transconductance to drain-source admittance) of a typical HEMT varies with frequency and bias [1]. For this device, the small-signal intrinsic gain varies little with frequency above about 1 MHz at low drain biases. The gain is easily determined from the ratios of the real parts of Y_{21} and Y_{22} measured with a network analyzer. For low frequencies, the gain can also be determined from the differences between pulsed characteristics near each quiescent point. It can be expected that data from short pulses, in the regime of 10 MHz, will give an **isodynamic characteristic** for this typical device at drain potentials below 2–3 V because the intrinsic gain remains constant as frequency increases. At higher drain bias dispersion effects are expected to influence fast pulsed characteristics. However, small-signal RF measurements can be used to probe isodynamic behavior.

With slow pulses, corresponding to lower frequencies, dispersion effects are expected to influence the characteristics. The dispersion effects are prominent at the slow 10–1000 Hz rate of curve-tracer operation, which is why dispersion is observed in curve-tracer measurements. True DC measurements usually require slower rates.

8.2.2 Device Dispersion

The behavior of a device can be considered in terms of an isodynamic characteristic, which is that seen by extremely fast pulses and extremely high frequencies. The meaning of "extreme" depends on the device type and its operating condition. Dispersion occurs when these characteristics are perturbed by heating, impact ionization, and trapping.

Thermal dispersion reduces the isodynamic current by a function of averaged power [2]. To quantify this, consider the relationship between the terminal current i_T [A], instantaneous power dissipation p_D [W] and the isodynamic current i_O [A]. The temperature dependence of drain current is well described by

$$i_T = i_O(1 - \delta \cdot p_D * h(t)). \tag{8.1}$$

The term δ [1/W] is a function of the thermal resistance of the device structure and the temperature dependence of the terminal current. The power dissipation is convolved with a thermal impulse

response $h(t)$. Measurements and simulation of a two-dimensional dynamic thermal model show that the impulse response corresponds to that of a sub-first order low-pass filter.

With isothermal conditions, the temperature rise remains constant during operation of the device. This occurs when the rate of change of power dissipation, due to signal components, is much faster than the response of $h(t)$ (which is shown in Figure 8.2 for a typical HEMT).

Impact ionization in field effect transistors (FETs) gives an additional drain current and trapped charge that adds a potential to the gate potential. The additional drain current is the product of the drain current and the impact-ionization rate [3]. The ionization rate is highly dependent on electric field and hence drain-source potential. The frequency response varies considerably with drain bias, as shown in Figure 8.2, and rolls-off at much less than the 20 dB/decade of a first-order low-pass filter. There is very little, or no, ionization below a critical drain potential, which causes a kink in the drain-current characteristics.

Carriers in the channel contribute to electron trapping that also affects to the gate potential at a bias-dependent characteristic frequency, as shown in Figure 8.2.

The dispersive processes discussed above are evident, to varying degrees, in all FETs. Trapping has strong drain-bias dependence and weak gate-bias dependence. Although impact ionization is possible in all FETs, it usually occurs at or above breakdown in low-breakdown HEMTs and MESFETs. Thus it is not usually observed below breakdown potentials in MESFETs, low-noise InP, and GaN devices. Thermal dispersion is common to all devices, including HBTs.

8.2.3 Large-Signal Conditions

Transistors operating in large-signal conditions operate with large signal excursions that can extend to limiting regions of the device. Large-signal limits, such as clipping due to breakdown or due to excessive input currents, can be determined from an extended range of characteristic curves. Steady-state DC measurements are confined to operating regions in the SOA of the device. It is necessary to extrapolate to breakdown and high-power conditions, which may prompt pushing the limits of measurements to regions that cause permanent, even if nondestructive, damage. The stress of these measurements can alter the characteristics and occurs early in the cycle of step-and-sweep curve tracers, which leads to incorrect characteristic curves in the normal operating region. The observed dispersion occurs in the comparison of the characteristics measured over moderate potentials, measured before and after a stress.

Pulsed measurements extend the range of measurement without undue stress. Figure 8.3 shows characteristic curves of a HEMT that encompasses regions of breakdown and substantial forward-gate potential. The diagram highlights points in these regions, which are those with significant gate current. The extended characteristics are essential for large-signal applications to identify the limits of signal excursion. The pulsed characteristics in the stress regions are those that would be experienced during a large-signal excursion because the measurement is made in isodynamic conditions set by the operating point. There is little correlation between these and an extrapolation from DC characteristics, because the stress regions are significantly affected by bias conditions and temperature.

8.2.4 Pulsed Measurements

Dispersion effects in microwave devices generate a rich dynamic response to large signals and changing operating conditions. The dynamic behavior affects the DC and high-frequency characteristics. To observe the dynamic response, it is necessary to measure characteristics over a wide range of operating frequency (or time) conditions. Pulsed measurement techniques quantify various aspects of the dynamic behavior over time and bias conditions.

The pulsed-current/voltage (pulsed-I/V) characteristics are characteristic curves determined from a near-isodynamic measurement with short pulses separated by long relaxation periods at a specific quiescent point. Each quiescent point has its own pulsed-I/V characteristics, so a complete characterization of a device requires **pulsed-*I/V* measurements** over various quiescent points. For many devices and for specific operating conditions, dispersion effects do not influence each pulsed-I/V characteristic.

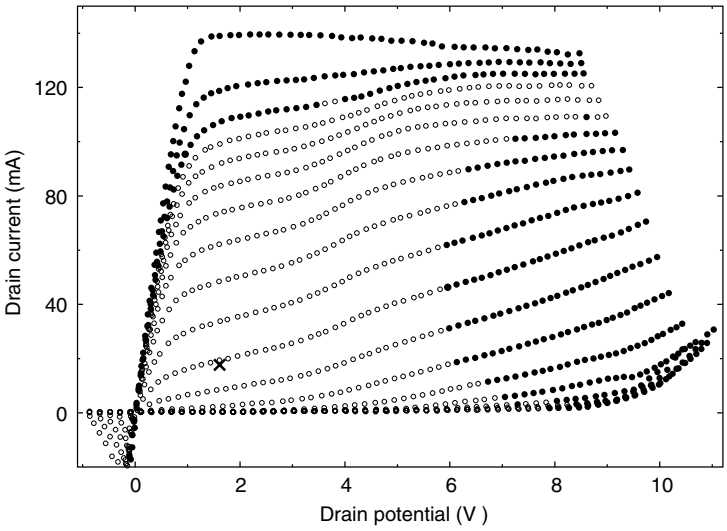


FIGURE 8.3 An example of the extended characteristic curves for an HEMT obtained with 500 ns pulses at 10 ms intervals for the quiescent point $V_{DS} = 1.6$ V, $I_D = 17.7$ mA (×). The solid points (●) are those for which the magnitude of gate current is greater than 1 mA. Gate-source potential from −3.0 to +1.5 V in 250 mV steps is the parameter.

However, dispersion will affect the variation between characteristics measured at different quiescent conditions.

The pulsed characteristics vary with pulse length. Short pulses produce isodynamic pulsed-I/V characteristics, and very long pulses produce DC characteristics. In this context, how short and how long depends on operation condition and device type. A time-domain pulsed measurement, performed by recording the variation of terminal conditions during and after a measured pulse, can trace the transition from isodynamic to DC behavior. The time constants of the dispersion effects are present in the time-domain characteristic. Note that the range of time-domain measurements is limited to the SOA for the long pulses used.

A collection of time-domain pulsed measurements from a bias to points on the DC curves produces **time-evolution characteristics** such as those in Figure 8.4. For each initial bias, each point on the characteristics is a transient response to a new bias. That is, the evolution over time, from a given bias point to the many bias points. Time-evolution characteristics exhibit a considerable dependence on the initial bias, so they reflect a different initial thermal and charge state, with all tending to the true DC curve.

Isodynamic characteristics depend on bias, temperature, and trapped charge. A change in bias will produce a change in the shape of the isodynamic characteristics as temperature and charge trapping respond. A **bias-evolution characteristic** shows the variation of isodynamic characteristics over time after the bias condition is changed [4]. A measurement of bias-evolution characteristics requires fast measurements of the characteristic, repeated over the time during which the bias state is changing. There is a class of applications, such as intermittently switched front-ends of wireless LAN and mobile telephone circuits, that operate at isodynamic frequencies, but are affected by bias-evolution of device characteristics. However, at certain bias conditions pulsed-I/V equipment can be too slow to capture isodynamic characteristics within the response times of temperature changes and charge trapping phenomena.

Time-evolution characteristics show a set of transitions from one bias point to many others. In contrast, bias-evolution characteristics show the change of isodynamic characteristics during the transition from one bias to another.

Isodynamic small-signal parameters are determined from **pulsed-RF measurements**. During the measurement pulse, a RF signal is applied and a pulsed vector network analyzer (VNA) determines the scattering parameters. The terminal potentials during each pulse are the **pulsed bias** for the RF measurement.

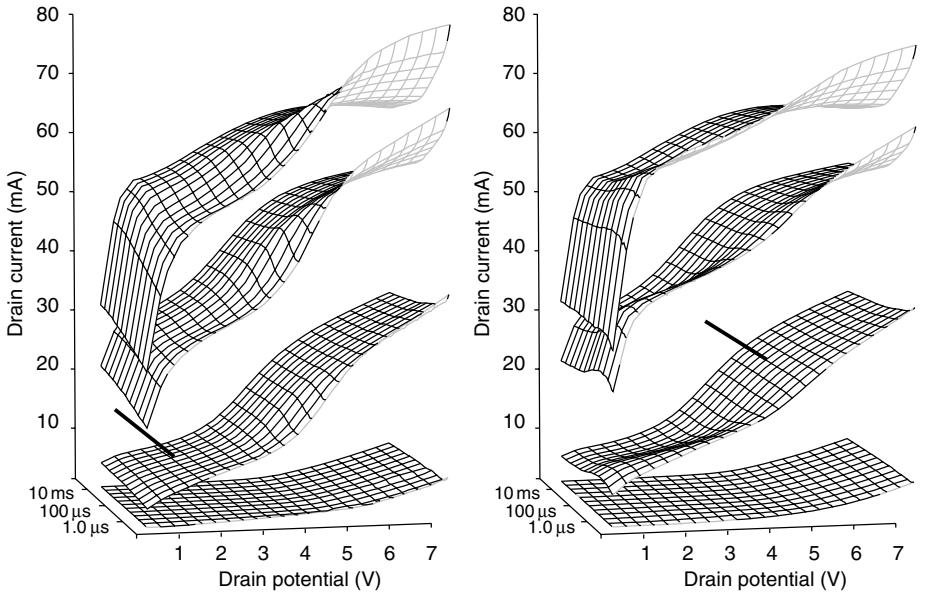


FIGURE 8.4 Time-evolution characteristics showing transients after a step change from the initial condition shown by solid lines (—). Each graph shows surfaces of drain current with gate-source potential from −1.5 to 0.0 V in 0.5 V steps as the parameter.

An operating point is established between pulses. Each operating point has its own set of pulsed-bias points and corresponding RF parameters. However, pulsed-RF measurements are relatively slow, so some bias-evolution of the operating point can occur at each pulsed bias. Pulsed-RF characteristics give small-signal parameters, such as reactance and conductance, as a surface function of terminal potentials. There is a small-signal parameter surface for each quiescent operating point and the dispersion effects only affect the variation of each surface with quiescent condition. Pulsed-RF measurements are also required for pulse-operated equipment, such as pulsed-radar transmitters, that have off-state quiescent conditions and pulse to an on-state condition that may be outside the SOA of the device.

Pulse timing and potentials vary with the measurement type. The periods required for isodynamic conditions and safe-operating potentials for various types of devices are discussed in the next section. The complexity and cost of pulse equipment, which also varies with application, are discussed in the subsequent section.

8.3 Relevant Properties of Devices

Three phenomena in active devices cause problems that are best addressed with pulsed measurements. These are the SOA constraint, thermal dependency of device characteristics, and dependency of device characteristics upon charge trapped in and around the device. The following discusses these phenomena and identifies devices in which they can be significant.

8.3.1 Safe-Operating Area

The idea of a SOA is simply that operating limits exist beyond which the device may be damaged. The SOA limits are generally bounds set by the following four mechanisms:

1. A maximum voltage, above which a mechanism such as avalanche breakdown can lead to loss of electrical control or direct physical alteration of the device structure.

2. A maximum power dissipation, above which the active part of the device becomes so hot that it is altered physically or chemically.
3. A maximum current, above which some part of the device like a bond wire or contact region can be locally heated to destruction.
4. A maximum current-time product, operation beyond which can cause physical destruction at local regions where adiabatic power dissipation is not homogeneous.

It is important to realize that damage to a device need not be catastrophic. The above mechanisms may change the chemical or physical layout of the device enough to alter the characteristics of the device without disabling it.

Pulsed-I/V measurements offer a way to investigate the characteristics of a device in areas where damage or deterioration can occur, because it is possible to extend the range of measurements under pulsed conditions, without harm. This is not a new idea—pulsed capability has been available in curve tracers for decades. These pulsed systems typically have pulses no shorter than a few milliseconds or a few hundred microseconds. However, shorter pulses allow further extension, and for modern microwave devices, true signal response may require submicrosecond stimuli.

There are time constants associated with SOA limitations. For example, the time constant for temperature rise can allow very high-power levels to be achieved for short periods. After that time, the device must be returned to a low-power condition to cool down. The SOA is therefore much larger for short periods than it is for steady DC conditions. Figure 8.5 shows successful measurement of a 140 μm^2 HBT well beyond the device SOA. The example shows a sequence of measurement sweeps with successively increasing maximum collector potential. There is no deterioration up to 7.5 V, which is an order of magnitude above that which would rapidly destroy the device under static conditions. The sweeps to a collector potential greater than 7.5 V alter the device so its characteristics have a lower collector current in subsequent sweeps. Shorter pulses may allow extension of this limit.

Different active devices are constrained by different SOA limits. For instance, GaN FETs are not usually limited by breakdown, whereas certain III–V HBTs are primarily limited by breakdown; silicon devices suffer more from a current-time product limit than do GaAs devices. Pulsed-I/V measurements provide a way for device designers to identify failure mechanisms, and a way for circuit designers to

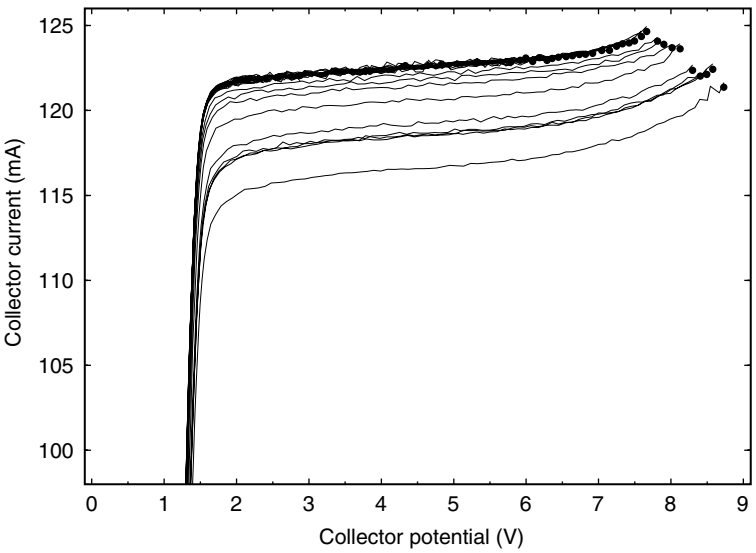


FIGURE 8.5 A single collector characteristic measured on a 140 μm^2 III–V HBT with sequentially increasing maximum voltage (shown by ●) applied in 1 μs pulses. Note the progressive deterioration above a certain instantaneous dissipation level.

obtain information about device characteristics in regions where signal excursions occur, but are outside the SOA.

8.3.2 Thermal Dispersion

GaAs devices, both FETs and HBTs, have greater thermal resistance than do their silicon counterparts. They tend to suffer larger changes in characteristics per unit change in junction temperature. Perhaps the first need for pulsed-I/V measurements arose with GaAs MESFETs due to the heating that occurs in simple DC measurement of these devices. Such a measurement turns out to be useless in high-frequency terms because each point in a measured DC characteristic is at a different temperature. This does not represent device characteristics in a RF situation where the temperature does not perceptibly change in each signal period. The sole utility of DC characteristics is to help predict quiescent circuit conditions.

A pulsed-I/V measurement can approach isothermal conditions, and can circumvent this problem. Figure 8.1, showing the DC and pulsed characteristics of a simple MESFET, exemplifies the difference. It is remarkable that the characteristics are for the same device.

Silicon devices, both FET and BJT, are relatively free of thermal dispersion effects, as are GaN FETs. The susceptibility of any given device, and the pulse duration and duty cycle required to obtain isothermal data, must be assessed on a case-by-case basis. Methods for achieving this are explored in the later discussion of measurement techniques.

8.3.3 Charge-Trapping

Temperature is not the only property of device operation that can give rise to dispersion. Charge trapped in the substrate or defects is particularly troublesome in FETs. Currents or junction potentials can control slow-responding charge states in the device structure. These phenomena are not as well understood as their thermal counterpart.

Exposing charge-trapping phenomena that may be influencing device performance is more difficult, but is still possible with an advanced pulsed-I/V system. One method is to vary the quiescent conditions between fast pulses, observing changes in the pulsed characteristic as quiescent fields and currents are varied independently, while holding power dissipation constant.

Figure 8.6 shows two pulsed characteristics measured with identical pulse stimulus regimes, but with different quiescent conditions. For low drain potentials, the pulsed drain current is high at the higher-power bias point, which is opposite to that expected from heating. For high drain potentials, there is a dramatic change in the shape of the characteristics consistent with impact ionization effects.

Charge-trapping dispersion is most prevalent in HEMTs, less so in HFETs and MESFETs. It has been recently reported in bipolar devices such as HBTs, but is ameliorated by their vertical structure. In many low-power FETs, impact ionization is not usually observed below breakdown potentials.

8.3.4 Time Constants

Avalanche effects can operate extremely quickly—much faster than any pulse system—so SOA measurements exhibit the thermal and charge-trapping time constants. Thermal effects typically have several time constants associated with them, each associated with the thermal capacity of some part of the system, from the active region to the external heat sink. The part with highest thermal resistance, usually the semiconductor, will dominate such that other time constants are insignificant. Heat generated in the active region of the device flows through the distributed thermal resistance/heat capacity structure of the semiconductor bulk. The distributed nature gives an impulse response $h(t)$ that is sub-first order [2]. Typical thermal time constants are of the order of hundreds of microseconds. The thermal dispersion of a device is dominated by the semiconductor when packaging and heat sinking have a significantly lower thermal resistance. The time constant is determined by the dimensions of the semiconductor. If packaging and heat sinking present a higher thermal resistance, then their thermal responses will be evident in the dispersion of the device. External heat sinks add long time constants, though anything above a few seconds

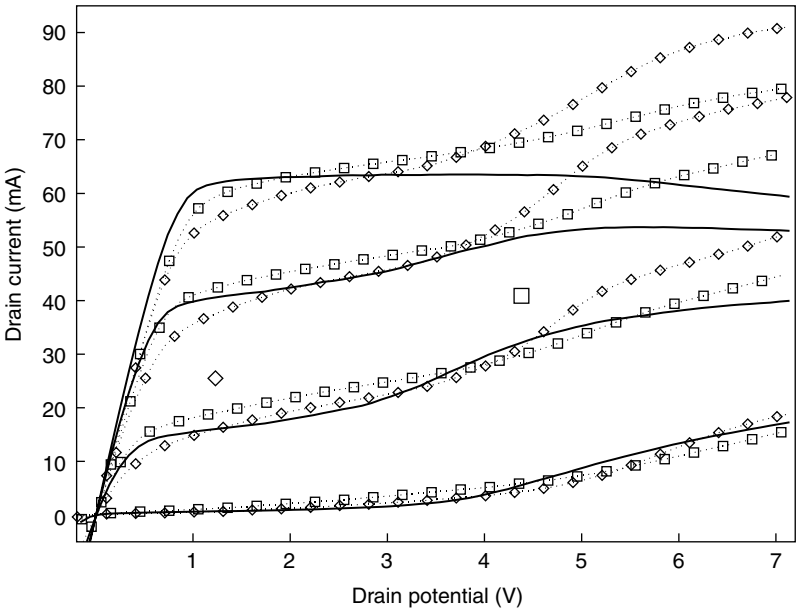


FIGURE 8.6 DC and pulsed characteristics at 100 ns for two initial operating conditions, with gate-source potential from −1.5 to 0.0 V in 0.5 V steps as the parameter. The graph shows two sets of pulse data covering the same terminal potentials as the DC data (—). Shown is pulse data (⋄) from initial point (◇) at $V_G = -0.75$ V and $V_{DS} = 1.2$ V, and data (□) from initial point (□) at $V_{GS} = -0.75$ V and $V_{DS} = 4.2$ V. This is the very fast, and very slow data of Figure 8.4.

could be disregarded or treated as environmental drift, since measurement or control of such external temperature is straightforward.

Charge-trapping time constants are more variable. They depend on the location of trapping in the band gap and are often exponentially dependent on bias potentials. Because of the wide variation of time constants, it is hard to know a priori what settings are appropriate for any measurement, let alone what capability ought to be specified in an instrument to make measurements. Figure 8.2 shows the variation with bias of charge-trapping that affects intrinsic gain. The time constants of a step to a bias are related to frequencies that affect the intrinsic gain at the new bias. The quiescent time for a pulse measurement should be longer than the slowest time constant at the quiescent bias point. The pulse time for isodynamic measurements needs to be faster than the time constant at the pulse point.

The time constants of trapping can be extremely long. There are reports of devices susceptible to disruption from charge stored apparently permanently, after the fashion of flash memory. Also, technologies in the early stages of development, such as recent GaN HEMTs, can have trapping sites with time constants of 100s of seconds [4].

8.3.5 Pulsed-*I*/*V* and Pulsed-RF Characteristics

Pulsed-I/V measurement is sometimes accompanied by pulsed-RF measurements. The RF equipment acquires the raw data during the pulse stimulus part of the measurement. Given that pulsed-I/V systems characterize devices in isodynamic conditions, the need for measurement at microwave frequencies, simultaneously with pulse stimuli, might be questioned. The problem is that it may not be possible to infer the reactive parameters for a given quiescent point from static S-parameters that are measured over a range of DC bias conditions. This is because of significant changes in RF behavior linked to charge-trapping or thermal dispersion effects.

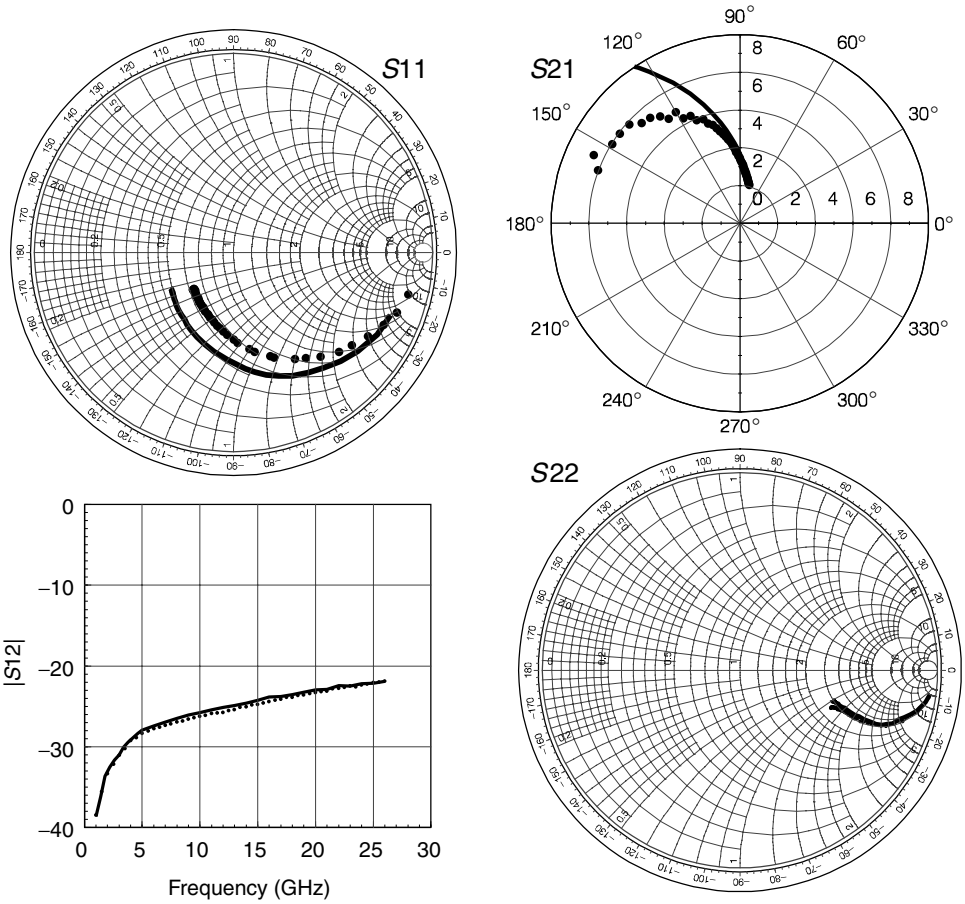


FIGURE 8.7 S-parameters measured at the same bias point with off-state and on-state quiescent conditions. The on-state parameters are from static, or DC measurements (—) and the off-state parameters are from measurements in a pulsed bias at the same point with off-state quiescent periods (•).

Figure 8.7 compares *S*-parameters of an HBT measured at a typical operating point (well within the SOA) using a DC bias and using a 1 μs pulsed bias at the same point with the device turned off between pulses. The differences, attributed to temperature, indicate the impact of dispersion effects on RF characteristics.

In addition, *S*-parameters cannot be gathered at bias points outside the SOA without pulse equipment. Pulse amplifiers often operate well beyond the SOA, so that a smaller, less expensive, device can be used. This is possible when the duration of operation beyond SOA is brief, but again, it is not possible to characterize the device with DC techniques. For many of these applications, pulsed-RF network analyzers have been developed. These can measure the performance of the transistor during its pulsed operating condition.

8.4 Pulsed Measurement Equipment

Pulsed measurement systems comprise subsystems for applying bias, pulse, and RF stimuli, and for sampling current, voltage, and RF parameters. Ancillary subsystems are included to synchronize system operation, provide terminations for the device under test (DUT), and store and process data. A simple

system can be assembled from individual pulse generators and data acquisition instruments. More sophisticated systems generate arbitrary pulse patterns and are capable of measurements over varying quiescent and pulse timing conditions. Pulsed-I/V systems can operate as stand-alone instruments or can operate in a pulsed-RF system to provide the pulsed bias.

8.4.1 System Architecture

The functional diagram of a pulsed measurement system, shown in Figure 8.8, includes both pulsed-I/V and pulsed-RF subsystems. Pulse and bias sources, voltage and current sampling blocks, and associated timing generators form the pulsed-I/V subsystem. A pulsed-RF source and mixer-based VNA form the pulsed-RF subsystem. The DUT is connected directly to the pulsed-I/V subsystem, or to bias networks that connect the pulsed-RF subsystem or RF terminations.

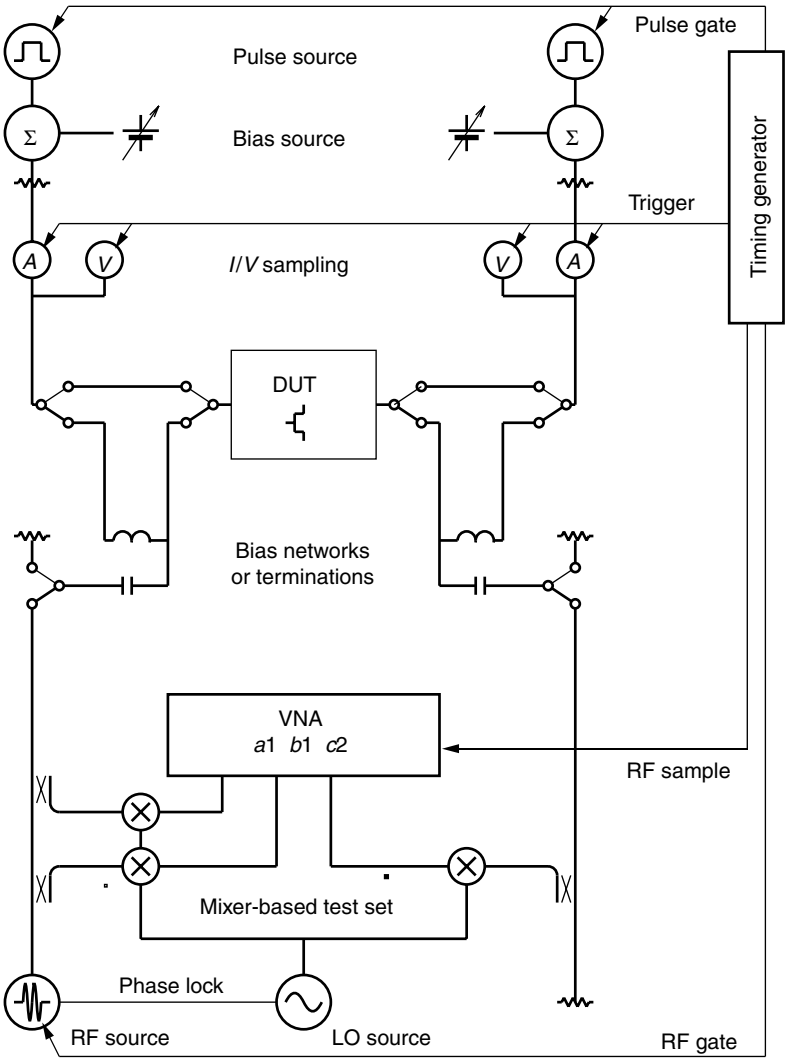


FIGURE 8.8 Simplified diagram of a generic pulsed measurement system. Alternative connections provide load terminations when there is no pulsed-RF test set or directly connect the pulsed-I/V subsystem to the DUT.

8.4.1.1 Pulsed-*I/V* System

Steady-state DC semiconductor parameter analyzers provide a source-monitor unit for each terminal of the DUT. The unit sources one of voltage or current while monitoring the other. In a pulsed measurement system, a pulsed voltage is added to a bias voltage and applied to the device. It is not practical to control the source potential within short pulse periods, so in order to ascertain the actual terminal conditions, both voltage and current are monitored. If a precise potential is required, then it is necessary to iterate over successive pulses, or to interpolate data from a range of pulsed measurements, or use longer pulse periods.

Simple systems use a pulse generator as the pulse source. Stand-alone pulse generators usually provide control of pulse and quiescent levels, so a single pulse point is measured during each test run. Such a system is easily assembled with pulse generators and is operated from their front panels. A single-point measurement mode is also employed by high-power pulses that deliver high current pulses by dumping charge from capacitors, which are pre-charged during the quiescent period.

Systems that measure several pulse points in sequence use computer-controlled arbitrary-function generators to provide pulse and quiescent potentials. The function generators are essentially digital memory delivering values to a digital-to-analog converter. Pulse values are stored in every second memory location and the quiescent value is stored in every other location. A timing generator then clocks through successive potentials at the desired pulse and quiescent time intervals. The quiescent potential is either simply delivered from the pulse generators or it is delivered from bench power supplies or other computer-controlled digital-to-analog converters. In the latter cases, a summing amplifier adds the pulse and quiescent potentials and drives the DUT. This architecture extends the pulse power capability of the system. Whereas the continuous rating of the amplifier dictates the maximum quiescent current delivered to the device, the pulse range extends to the higher transient current rating of the amplifier.

In most systems, either data-acquisition digitizers or digital oscilloscope channels sample current and voltage values. In a simple set up, an oscilloscope will display the terminal conditions throughout the pulse and the required data can be read on screen or downloaded for processing. Oscilloscope digitizers tend to have resolutions sufficient for displaying waveforms, but insufficient for linearity or wide dynamic range measurements. Data-acquisition digitizers provide wider dynamic range and ability to sample at specific time points on each pulse or throughout a measurement sequence. When several pulse points are measured in sequence, the digitizers record pulse data from each pulse separately or time-domain data from several points across each pulse. Either mode is synchronized by appropriate sampling triggers provided by a timing generator.

The position of the voltage and current sensors between the pulse source and the DUT is significant. There are transmission line effects associated with the cabling between the sensing points and the digitizers. The cable lengths and types of terminations will affect the transient response of, and hence the performance of, the pulse system. An additional complication is introduced when the DUT must be terminated for RF stability. A bias network is used but this introduces its own transient response to the measured pulses. For example, there is an initial 100 ns transient in many pulsed measurements that is generated by the bias network and is present when the DUT is replaced by a 50 Ω load.

Current is sensed by various methods, which trade between convenience and pulse performance. With a floating pulse source, a sense resistor in the ground return will give the total current delivered by the source. There is no common-mode component in this current sensor, so a single-ended digitizer input is usable. The current reading will include, however, transient components from the charging of capacitances associated with cables between the pulser and the DUT. Low-impedance cables can ameliorate this problem. Alternatively, Hall-effect/induction probes placed near the DUT can sense terminal current. These probes have excellent common-mode immunity but tend to drift and add their own transient response to the data. A stable measurement of current is possible with a series sense resistor placed in line near the DUT. This eliminates the effect of cable capacitance currents, but requires a differential input with very good common-mode rejection. The latter presents a severe limitation for short pulses because common-mode rejection degrades at high frequency.

Data collection and processing in pulse systems is different to that of slow curve tracers or semiconductor parameter analyzers. The latter usually measure over a predefined grid of step-and-sweep values. If the voltage grid is defined, then only the current is recorded. The user relies on the instrument to deliver the specified grid value. In pulse systems, a precise grid point is rarely reached during the pulse period. The pulse data, therefore, include measured voltage and current for each terminal. An important component in any pulse system is the interpretation process that recognizes that the pulse data do not lie on a regular grid of values. One consequence of this is that an interpolation process is required to obtain traditional characteristic curves.

8.4.1.2 Pulsed-RF System

Pulsed-RF test sets employ VNAs with a wide-band intermediate frequency (IF) receiver and an external sample trigger [5,6]. The system includes two RF sources and a mixer-based *S*-parameter test set. One source provides a continuous local oscillator signal for the mixers, while the other provides a gated RF output to the DUT. The local oscillator also provides a phase reference, so that a fast sample response is possible.

The pulsed bias must be delivered through bias networks that are essential for the pulsed-RF measurement. During a pulsed-I/V measurement, the RF source is disabled and the RF test set provides terminations for the DUT. Pulsed-RF measurements are made one pulse point at a time. With the pulsed bias applied, the RF source is gated for a specified period during the pulse and the network analyzer is triggered to sample the RF signals. The same pulse point is measured often enough for the analyzer to work through its frequency list and averaging requirements.

8.4.2 Technical Considerations

A trade between cost, complexity, and technical performance arises in the specification and assembly of pulsed measurement systems. Important considerations are pulse timing, measurement resolution and range, total time required for a measurement task, and the flexibility of the pulse sequencing.

8.4.2.1 Pulse Events

Pulsed measurement systems produce a continuous, periodic sequence of **pulse events**. The generic timing of each part of a pulse event is shown in Figure 8.9. Each pulse event provides a pulse stimulus period, T_{Pulse}, and a quiescent period, $T_{\text{Quiescent}}$. The total time for these two periods is the **pulse event period** and the inverse of this is the **pulse repetition frequency** (PRF) for continuous pulsing. Typical pulsed-I/V measurements use 200–500 ns pulses, and true DC measurements require periods of 100 ms or more. To achieve sub-100 ns pulses, usually the DUT is directly connected to a pulse generator to minimize transmission-line effects. Quiescent periods range from 10 μs to 1 s and often must be longer than 1 ms for isodynamic pulsed-I/V measurements. Research of immature technologies may require much greater than 1 s.

One or both terminals of the DUT may be pulsed. In some systems, the pulse width on the second terminal is inset relative to the first, by τ_{inset}, which gives some control over the trajectory of the initial pulse transient to avoid possible damage to the DUT. Often the gate pulse will be applied before the drain pulse is applied—an inset of 100 ns is typical. Sometimes it might be necessary for the drain pulse to lead the gate pulse in order to control the transition path over the I/V-plane. Thus, the parameter τ_{inset} might be positive or negative and might be different for leading and trailing pulse edges. In a simple system, it is most easily set to zero so that the terminal pulses are coincident.

Samples of current and voltage occur some time, τ_{P}, before the end of the pulse. Some systems gather a number, N_{P}, of samples over a period, T_{sample}, which may extend over the entire pulse if a time-domain transient response is measured. The number of samples is limited by the sampling rate and memory of the digitizers. A measurement of the quiescent conditions some time, τ_{Q}, before the start of each pulse may also be made.

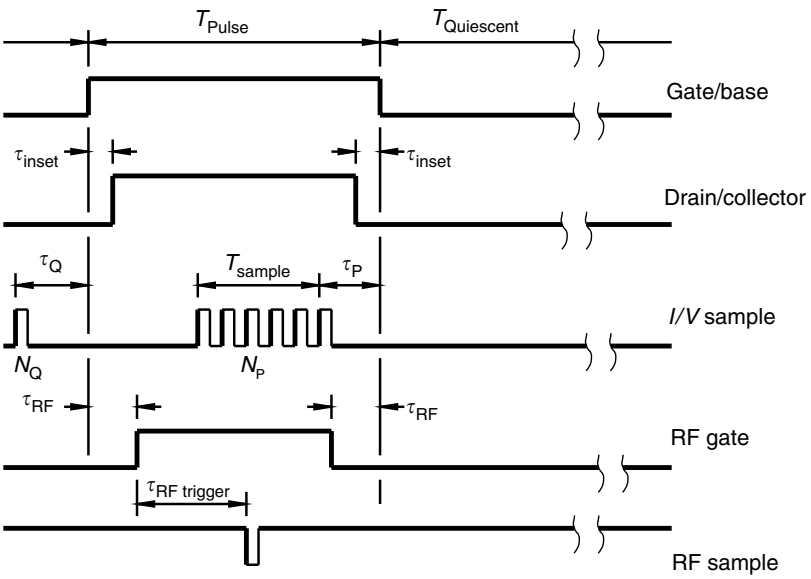


FIGURE 8.9 Generic timing diagram for each pulsed measurement pulse event.

For pulsed-RF measurements, the RF source is applied for a period that is inset, by τ_{RF}, within the pulsed bias. A RF trigger sequences sampling by the network analyzer. The RF source is disabled during pulsed-I/V measurements.

The above times would refer to the pulse event timing at the terminals of the DUT. Various instrument and cabling delays might require that these times be individually adjusted when referred to the pulse amplifiers and sample digitizers. Different signal paths for current and voltage digitizers might require separate triggers for these.

8.4.2.2 Measurement Cycles

A pulsed **measurement cycle** is a periodic repetition of a sequence of pulse events. A set of pulse events, required to gather device characteristics, is measured in one or more measurement cycles. With single-point measurements, there is only one pulse event in the sequence and a separate measurement cycle is required for each point in a set of characteristics. This is the case with pulsed-RF measurements, or with high-power pulsers, or with very-high-speed pulse generators. With arbitrary function generators, the measurement cycle is a sequence of pulse events at different points; so one cycle can measure several points on a pulsed-I/V characteristic.

To establish the bias condition, either the DC bias can be held for a **soak time** long enough for dispersions to settle, or the measurement cycles can be repeated for a **stabilizing period** to establish the bias condition that is the steady-state component of the repetition of pulse events. Then the cycle is continued while data are sampled. Typical soak times or stabilization periods can range from a few seconds to tens of seconds. These long times are required for initial establishment of stable operating conditions, whereas shorter quiescent periods are sufficient for recovery from pulse perturbations.

When several pulse points are measured in each cycle, the pulse stimulus is a steady-state repetition, so each pulse point has a well-known initial condition. Flexible pulse systems can provide an arbitrary initial condition within the cycle or use a pseudorandom sequencing of the pulse points. These can be used to assess the history dependence or isodynamic nature of the measurements. For example, it may be possible to precede a pulse point with an excursion into the breakdown region to assess short-term effects of stress on the characteristic.

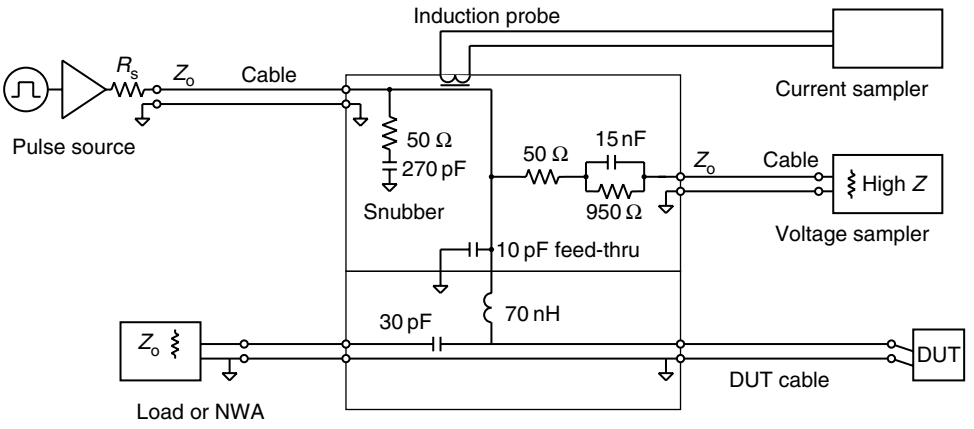


FIGURE 8.10 Schematic of a bias network that provides RF termination and pulsed bias feed with voltage and current measuring points.

8.4.2.3 Bias Networks

One of the most significant technical limitations to pulsed measurement timing is the bias network that connects the DUT to the pulse system. The network must perform the following:

- Provide RF termination for the DUT to prevent oscillations
- Pass pulsed-bias stimuli to the DUT
- Provide current and voltage sample points
- Control transients and overshoots that may damage the DUT

These are contradictory requirements that must be traded to suit the specific application. In general, the minimum pulse period is dictated by the bias network.

For very-fast pulsed measurements, less than 100 ns, the pulse generator is usually connected directly to the DUT [7]. The generator provides the RF termination required for stability, and current and voltage are sensed with a ground-return sense resistor and a high impedance probe, respectively. Pulsed-RF measurements are not contemplated with this arrangement.

Systems that are more flexible need a modified bias network similar to that shown in Figure 8.10. The DC-blocking capacitor must be small, so that it does not draw current for a significant portion of the pulsed bias, but must be large enough to provide adequate termination at RF frequencies. The isolating inductor must be small, so that it passes the pulsed bias, but must also be large enough to provide adequate RF isolation. In this example, the DUT is connected to a RF termination provided by a load or network analyzer. The DC-blocking capacitor, 30 pF, and isolating inductor, 70 nH, values are an order of magnitude smaller than those in conventional bias networks. The network provides a good RF path for frequencies above 500 MHz and does not significantly disturb pulses longer than 100 ns. Modifying the network to provide a RF path at lower frequencies will disturb longer pulses.

The pulsed bias is fed to the bias network in Figure 8.10 through a cable that will introduce transmission line transients. To control these, the source output impedance can provide line termination. Although this can provide significant protection from transients when fragile devices are being measured, it will limit the voltage and current range of the pulses. An alternative is to provide a termination at the bias network end of the cable with a series resistor–capacitor snubber. The values shown in this example are suitable for suppressing the 10 ns transients associated with a 1 m cable.

Voltage sampling in Figure 8.10 is through a frequency-compensated network that provides isolation between the RF path and the cable connected to the voltage sampling digitizer. Without this isolation, the capacitance of the cable would load the pulsed-bias waveform, significantly increasing its rise time. The voltage sample point should be as close as possible to the DUT to reduce the effect of the return pulse

reflected from the DUT. The network in this example sets a practical limit of about 15 cm on the length of the cable connecting the DUT to the bias network.

In general, bias networks that provide RF terminations or pulsed-RF capability will limit the accuracy of measurements in the first 100–200 ns of a pulse. With such an arrangement, the pulse source need not produce rise times less than 50 ns. Rather, shaped rising edges would be beneficial in controlling transients at the DUT.

Induction current probes introduce their own time constants to the measurement that is visible in the time-domain transient record. Current measurement with series sense resistors will ameliorate this, but will add to the output impedance of the pulse source. Usually a capacitance of a few picofarads is associated with the sense or bias network that will restrict the choice of resistance value for a specified rise time. Series-resistor sensing requires a floating differential amplifier operating over the range of pulse potentials. The common-mode gain of the amplifier will be higher for short-time intervals, so some of the step change in potential will be recorded as a current transient. Placing a sense resistor in the ground return is an alternative, but the transmission-line effects of the connection between the pulser and DUT need to be considered.

8.4.2.4 Measurement Resolution

Voltage and current ranges are determined by the pulse sources. Summing amplifiers provide a few hundred milliamps at 10–20 V. High-power, charge-dumping, pulsers provide several amps at 50 V. Current pulses are achieved with series resistors and voltage sources. These limit the minimum pulse time. For example, a 1 kΩ resistor may be used to set a base current for testing bipolar transistors. With 10 pF of capacitance associated with the bias network, the minimum rise time would be of the order of 10 μs. An isodynamic measurement would need to use short collector-terminal pulses that are inset within long base-terminal pulses.

There is no practical method for implementing current limiting within the short time frame of pulses other than the degree of safety afforded by the output impedance of the pulse source.

Measurement resolution is determined by the sampling digitizers and current sensors. Oscilloscopes provide 8-bit resolution with up to 11-bit linearity, which provides only 100 μA resolution in a 100 mA range. The 12-bit resolution, with 14-bit linearity, of high-speed digitizers may therefore be desirable. To achieve the high voltage or current resolutions, averaging is often required. Either the pulse system can repeat the measurement cycle to accumulate averages, or several samples in each pulse can be averaged.

8.4.2.5 Measurement Time

Measurement speed, in the context of production-line applications, is optimized with integrated systems that sequence several pulse points in each measurement cycle. As an example, acquiring 1000 pulse points with 1 ms quiescent periods, 500 ns pulse periods, and an averaging factor of 32, will necessarily require 32 s of pulsing. With a suitable stabilization period, and overhead in instrument setup and data downloading, this typical pulsed-I/V measurement can be completed in just less than one minute per quiescent point.

Single-point measurement systems have instrument setup and data downloading overhead at each pulse point. A typical 1000-point measurement usually requires substantially more than ten minutes to complete; especially when data communication is through GPIB controllers.

A pulsed-RF measurement is also slow because the network analyzer must step through its frequency list, and requires a hold-off time between RF sampling events. A typical pulsed-RF measurement with a 50-point frequency list, an averaging factor of 32 and only 100 pulse points, would take about half a minute to complete.

8.4.3 Commercial Measurement Systems

Figure 8.11 portrays graphically the areas covered in a frequency/signal level plane by various types of instruments used to characterize devices. The curve tracer, epitomized perhaps by the HP4145 and numerous analog predecessors made by companies such as Tektronix, covers the most basic measurement range.

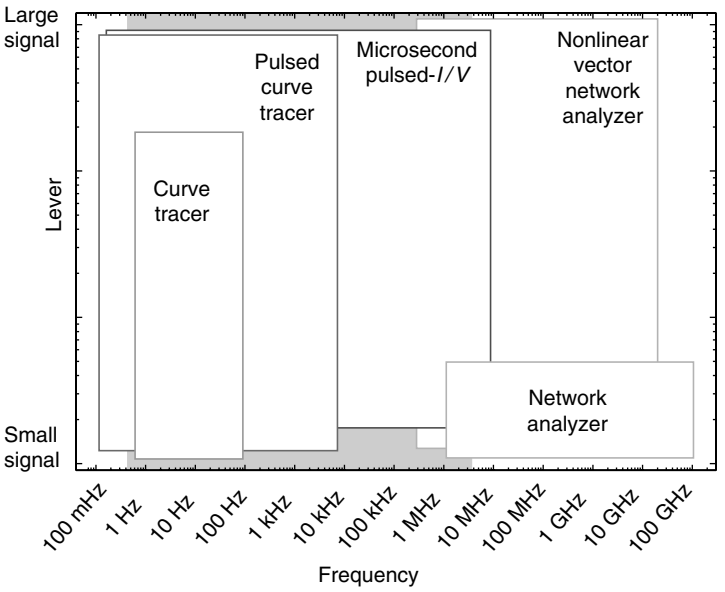


FIGURE 8.11 Relative position of various types of measurement equipment, including pulsed-I/V systems, in term of measurement frequency and signal level. The shaded area indicates the frequency range of typical dispersion effects in low-power FETs at moderate bias.

Beyond this range, instruments with some pulse capability, such as the HP4142/55/56 or Agilent E5260 series, offer very wide capability but this is still at speeds below that required for isodynamic characterization. Network analyzers reach millimeter-wave frequencies but perform small-signal measurements by definition. An emerging range of nonlinear VNAs capture harmonic information that can be transformed into large-signal time-domain and phase representation. Between these, pulsed-I/V systems have the advantage of large-signal capability and speed sufficient to give isodynamic characteristics.

The majority of pulsed measurements reported in the literature to date have been made with experimental equipment, or with systems under development. Some submicrosecond systems are commercially available [8–10]. These come with a range of options that require some assessment before purchase. This is partly a consequence of the immature nature of pulsed-I/V instrumentation (in comparison to conventional curve tracers), and partly a result of pulsed-I/V measurement being a more complicated problem.

Before selecting a system, it is useful to identify an intrinsic problem for pulsed measurements. The performance limit on pulsed-I/V systems is frequently the DUT connection network and the form of the stimulus, not the measurement system itself.

Network analyzers achieve very high-frequency resolution with a narrowband stimulus and receiver, which allows them to minimize noise and apply vector calibration techniques to eliminate parasitic disturbances. They define a measurement plane, behind which any fixed error is identified and eliminated by post-processing of the data. They can also allow the DUT to come to a steady-state during the measurement. Pulse systems conversely use a stimulus that contains many frequency components from the slow pulse repetition rate up to many times the fundamental component in the fast pulse. The measurement is both of wide bandwidth, and therefore noisy, and at high frequencies. Viewed in the time domain, the pulse width is limited by the charging of the unknown capacitance in the bias network, which can be minimized but not eliminated. For example, bias networks may contribute sufficient parasitic capacitance to limit pulsed measurements to 500 ns, or slower, with a pulse source impedance of 50 Ω. The situation is worse for current drive, and may be worse still, because of transients, for a voltage drive that does not

match transmission-line impedance. Thus, the system is infrequently limited by the minimum width of the pulse from the instrument, and some judgment needs to be exercised in each measurement setup.

8.5 Measurement Techniques

With flexible pulsed measurement systems, a wide range of measurements and techniques is possible. Consideration needs to be given to what is measured and the measurement procedures, in order to determine what the data gathered represent. The following sections discuss different aspects of the measurement process.

8.5.1 The Pulse-Domain Paradigm and Timing

A general pulsed-I/V plane can be defined as the grid of terminal voltages pulsed to, from a particular quiescent condition. For isodynamic pulsing, a separate pulsed characteristic would be measured for each quiescent condition.

At each pulse point on an I/V-plane, measurements can be characterized in terms of the following:

- The quiescent point pulsed from, defined by the established bias condition and the time this had been allowed to stabilize.
- The actual pulse voltages, relative to the quiescent voltage, the sequence of application of the terminal pulses, and possibly the voltage rise times, overshoot, and other transients.
- The position in time of sampling relative to the pulses.
- The type of measurements made; voltage and current at the terminals of the DUT, together with RF parameters at a range of frequencies.

Thus, if a number of quiescent conditions are to be considered, with a wide range of pulsed terminal voltages, a large amount of data will be generated. The time taken to gather this data can then be an important consideration. Techniques of overnight batch measurements may need to be considered, together with issues such as the stability of the measurement equipment. Equipment architecture can be categorized in terms of the applications to measurement over a generalized I/V-plane. Those that allow arbitrary pulse sequences within each measurement cycle enable an entire I/V-plane to be rapidly sampled. Systems intended for single pulses from limited quiescent conditions may facilitate precise measurement of a small region in the I/V-plane, but this is at the expense of speed and flexibility.

In the context of isodynamic pulsing, the most important consideration in interpreting the measured data is the sample timing. This is the time of current and voltage sampling relative to the application of the voltage pulses. As it is often information on time-dependent dispersion effects that is gathered, it is important to understand the time-placement of sampling relative to the time constants of these rate-dependent effects.

For an investigation of dispersion effects, time-domain pulse-profile measurements are used. Terminal currents and voltages are repeatedly sampled; from before the onset of an extended pulse, until after dispersion effects have stabilized. This can involve sampling over six orders of magnitude of time and hence produces large amounts of data. From such data, the time constants of dispersion effects can be extracted. From pulse-profile measurements of a range of pulse points, and from a range of initial conditions, the dependence of the dispersion effects upon initial and final conditions can be determined.

For isodynamic measurements, which are unaffected by dispersion, sampling must be done quickly after the application of the pulse, so that dispersion effects do not become significant. Additionally, the time since the application of the previous pulse, spent at the quiescent condition, must be long enough that there are no residual effects from this previous pulse. The device can then be considered to have returned to the same quiescent state. Generally, sampling must be done at a time, relative to pulse application, at least two orders of magnitude less than the time constants of the dispersion effects (for a less than

1% effect). Similarly, the quiescent time should be at least an order of magnitude greater than these time constants.

Note that for hardware of specific pulse and sampling-speed limitations, there may be some dispersion effects too fast for observation. Thus, this discussion refers to those dispersion time constants greater than the time-resolution of the pulse equipment. Alternative measurements, such as RF intermodulation and *S*-parameters can be considered for probing fast dispersions.

Quantification of suitable pulse width, sample time, and quiescent time can be achieved with reference to the time constants observed in a time-domain pulse profile. For example, for dispersion time constants in the 10–100 μs range, a pulse width of 1 μs with a quiescent time of 10 ms might be used. Sampling might be done 250 ns after pulse application, to allow time for bias network and cable transients to settle.

In the absence of knowledge of the applicable dispersion time constants, suitable pulse and quiescent periods can be obtained from a series of pulsed measurements having a range of pulse, sample, and quiescent periods. Observation of sampled current as a function of these times will reflect the dispersion effects present in a manner similar to that achievable with a time-domain pulse-profile measurement [11].

A powerful technique for verifying isodynamic timing is possible with measurement equipment capable of pulsing to points on the I/V-plane in a random sequence. If the quiescent time of pulse relaxation is insufficient, then the current measurement of a particular pulsed voltage will be dependent upon the particular history of previous pulse points. In conventional measurement systems, employing step-and-sweep sequencing whereby pulse points are swept monotonically at one terminal for a stepping of the other terminal, dispersion effects vary smoothly and are not obvious immediately. This is because adjacent points in the I/V-plane are measured in succession and, therefore, have similar pulse histories.

If, however, points are pulsed in a random sequence, adjacent points in the I/V-plane each have a different history of previous pulses. If pulse timing does not give isodynamic conditions, then the dispersion effects resulting from the pulse history will be evident in the characteristic curves. Adjacent points, having different pulse histories, will have dispersion effects of differing magnitude and hence markedly different values of current. This is observed in Figure 8.12, showing isodynamic and nonisodynamic

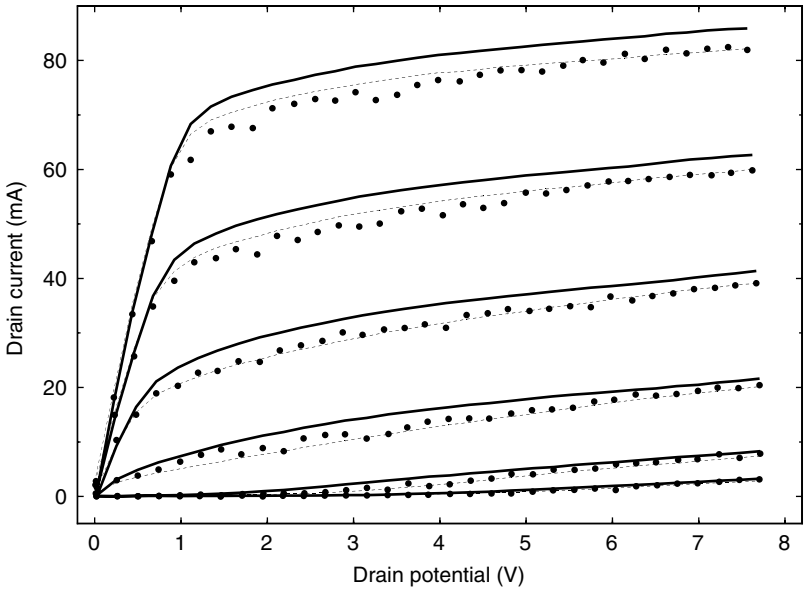


FIGURE 8.12 Characteristic curves for a MESFET measured with three different pulse sequences: a step-and-sweep with 1 μs pulses and 1 μs quiescent periods (- - - -), the same pulses sequenced in pseudorandom order (●), and an isodynamic measurement. The latter used 800 ns pulses with 1ms quiescent periods. The quiescent point is $V_{DS} = 0.0$ V, $V_{GS} = 0.0$ V and gate-source potential from −2.5 to +0.0 V in 0.5 V steps is the parameter.

measurement of the characteristics of a particular device. The nonisodynamic sets of characteristics were measured with the same pulse timing. One characteristic was measured by sweeping the drain-terminal pulse monotonically for different gate-terminal pulse settings. The other characteristic was measured as a random sequence of the same pulses. The smooth shape of former does not suggest dispersion effects. The apparently-noisy variation between adjacent points in the latter indicates history-dependent dispersion is in effect.

Thus, by random sequencing of the pulse points, isodynamic timing can be verified. To obtain isodynamic characteristics, shown in Figure 8.12, the quiescent relaxation time was increased and the pulse time reduced, until both curves became smooth and identical. That is, until there is no observable history-dependent dispersion.

8.5.2 General Techniques

Within the context of the pulse-domain paradigm discussed in the previous section, and the available equipment, a number of specific measurement techniques and issues arise. These are affected by the equipment limitations and influence the data gathered. A number of these techniques and issues are discussed here.

8.5.2.1 Establishing Quiescent Conditions

Prior to sampling pulsed data, a stable quiescent condition must be established. Normally this involves a step from zero bias to a desired bias point, which is a significant step change that may require a long time for resulting dispersions to settle. During this stabilization period, it is important that the pulse system holds off measurements or the data is discarded. There is the possibility of an evolution of bias if the pulse stimulus is applied immediately. This can be avoided by a soak time where the bias is held constant to establish the quiescent condition before starting the pulse stimulus.

8.5.2.2 Interpolation and Iteration

Often measurements are desired at a particular pulse point or specific grid of points. For a target pulse voltage, the actual voltage at the DUT at a certain time will usually be less. This results from various hardware effects such as amplifier output impedance and amplifier time constants, as well as cabling and bias network transients. Voltage drop across amplifier output impedance could be compensated for in advance with known current, but this current is being measured. This is why pulsed voltages need to be measured at the same time as the device currents.

If measurements are desired at specific voltage values, then one of two approaches can be used. First, over successive pulses, the target voltage values can be adjusted to iterate to the desired value. This necessarily involves a measurement control overhead and can require considerable time for many points. If the thermal noise implicit in using wide-bandwidth digitizers is considered, it is of dubious value to iterate beyond a certain point.

Alternatively, if a grid of pulse points is sampled, covering the range of points of interest, then the device characteristics at these particular points can be interpolated from the measured points. Without iteration, these measured points can be obtained quickly. A least-squares fit to a suitable function can then be used to generate characteristics at as many points as desired. Thus, provided the sampled grid is dense enough to capture the regional variation in characteristics, the data gathering is faster. The main concept is that it is more efficient to rapidly gather an entire I/V-plane of data and then post-process the data to obtain specific intermediate points.

8.5.2.3 Averaging

The fast pulses generally required for isodynamic measurement necessitates the use of wide-bandwidth digitizers. Voltage and current samples will then contain significant thermal noise. A least-squares fit to an assumed Gaussian distribution to an I/V-grid can be employed to smooth data. Alternatively, or additionally, averaging can be used.

Two types of averaging processes present themselves. The first process is to average multiple samples within each pulse. This assumes a fast digitizer and that there is sufficient time within the pulse before dispersion becomes significant. If dispersion becomes significant over the intrapulse period of sampling, then averaging cannot be employed unless some assumed model of dispersion is applied (a simple fitted time constant may suffice). An additional consideration with intrapulse averaging is that voltage value within a pulse cannot be considered constant. The measurement equipment providing the voltage pulse has nonzero output impedance and time constants. Thus, the actual voltage applied to the DUT will vary (slightly) during the voltage pulse. Consecutive samples within this pulse will then represent the characteristics for different voltage values. These are valid isodynamic samples if the sample timing is still below the time constants of dispersion effects. However, they could not be averaged unless the device current could be modeled as a linear function of pulsed voltages (over the range of voltage variation).

The second averaging process is to repeat each pulse point for as many identical measurements as required and average the results. Unlike intra-pulse averaging, this interpulse averaging will result in a linear increase in measurement time, in that a measurement cycle is repeated for each set of averaged values. Issues of equipment stability also need to be considered. Typically, both intra- and interpulse averaging might be employed. With careful application, averaging can provide considerable improvement in the resolution of the digitizers used, up to their limit of linearity.

8.5.2.4 Pseudorandom Sequencing

As previously discussed, randomizing the order of the sequence of pulse points can provide a means of verifying that the quiescent relaxation time is sufficient. It can also provide information on the dispersion effects present. In this, a sequence of voltage pulse values is determined for the specified grid of terminal values. These are first considered a sweeping of one terminal for a stepping of the other. To this sequence, a standard pseudo-randomizing process is used to resequence the order of application of pulses. As this is deterministic, for known randomizing processes, it is repeatable. This sequence is then applied to the DUT. Upon application of pulses, this random pulse sequence can help identify nonisodynamic measurement timing.

Additionally, if dispersion is present in the measured data, the known sequence of pulse points can provide information on history-dependent dispersion. With step-and-sweep sequencing of pulses, the prior history of each pulse is merely the similar adjacent pulse points. This represents an undersampling of the dispersion effects. With random sequencing, consecutive pulse points have a wide range of earlier points, providing greater information on the dispersion effects.

Thus, for the known sequence of voltage pulses and the nonisodynamic pulse timing, a model of the dispersion effects can be fitted. These can then be subtracted to yield isodynamic device characteristics. This, however, only applies to the longer time-constant effects and requires that the timing be close to that of the time constants of the dispersion effects.

8.5.2.5 Pulse Profile

In normal isodynamic pulsing, pulse widths are kept shorter than the time constants of applicable dispersion effects. Relaxation periods between pulses, at the quiescent condition, are longer than these times. Typically, pulse widths of 1 μs and quiescent periods of 100 ms would be suitable for many devices.

In a pulse-profile measurement, an extended pulse width of 0.1–1 s might be used, so that the dispersion effects can be observed. Dispersion with time constants greater than the pulse rise and settling time are then observable. Quiescent periods between these extended pulses still need to be long, so that subsequent pulses can be considered as being from the same bias condition.

Plotted on a logarithmic-time axis, the dispersion effects can be seen as a variation of device output current with time in Figure 8.13. Output current rises and falls at various times due to channel heating, electron trapping, and hole trapping related to impact ionization. Time constants of the amplifier driving the pulses might need to be deconvolved before identifying those of the DUT alone. From such a plot, it can first be identified where isodynamic conditions may apply (300–400 ns in Figure 8.13). That is, how soon after pulse application sampling can be done, if at all, before dispersion effects become

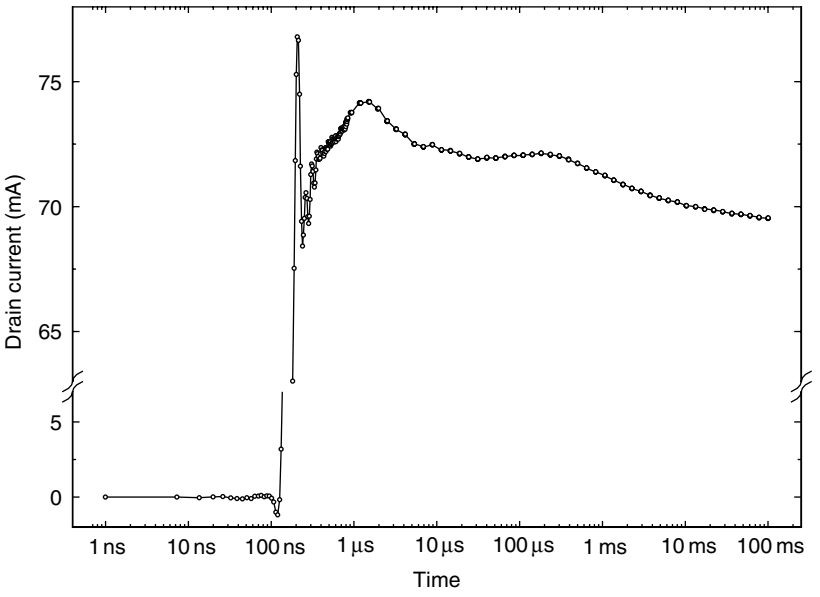


FIGURE 8.13 Transient response measured with eight repeated measurements at 50 ns intervals. Each repetition is shifted by 6.25 ns to give the composite response shown.

significant. How long these dispersion effects take to stabilize will indicate how long the quiescent periods of isodynamic pulsing need to be. Second, values for dispersion time constants can be extracted from the data, together with other parameters applicable to a particular dispersion model. In the case of Figure 8.13, there is evidence of heating around 0.5–50 ms, and charge-trapping 2–10 μs. The time-evolution data of Figure 8.4 is a set of repeated pulse profiles.

Note that, because the extended pulse widths of pulse-profile measurements are intended to bring into effect heating and dispersion, the range of pulse points on the I/V-plane must be restricted. With isodynamic pulsing, it is possible to pulse to voltages well outside the SOA of the DUT. This is because the short pulses do not invoke the time-dependent thermal and current damage of static conditions. With pulse-profile measurements, pulse widths extend to essentially static periods and so voltages must be restricted to the SOA for static measurements (although pulse-profile techniques could be used to observe destruction outside the SOA).

Equipment issues influence pulse-profile measurements in several ways. The first is pulse duration. Systems employing capacitor charge-dumping for pulsing will be limited in the length of time that they can hold a particular output voltage. The second is output rise and settling times. Bias network and cable transients and the response time of data measurement will limit the earliest time after pulse application for which valid samples can be taken. This, typically, might be of the order of 100 ns, although with restrictions on application might extend down to 1 ns. This necessarily affects the range of dispersion effects observable to those having time constants greater than perhaps an order of magnitude more than this minimum time resolution.

Digitizer speed and bandwidth are another major issue in pulse-profile measurements. A wide bandwidth is necessary so that sample values accurately reflect DUT conditions. In isodynamic pulsing, only one time point need be sampled, with a long time before the next pulse. With a pulse profile, it is desirable to repeatedly sample the pulse as fast as possible to observe variation with time. Sampling speed needs to be perhaps an order of magnitude faster than the time constant to be observed. Additionally, if bandwidth, jitter, and stability permit, an *equivalent time* sampling may be used. In this, repeated pulse-profile measurements are performed, with sample times relative to pulse onset shifted slightly with each successive pulse. As an example, a 20 MHz digitizer, sampling at 50 ns intervals, might be applied

to eight successive, identical pulses. Sampling is commenced 50 ns before the start of the first pulse, but offset an accumulating 6.25 ns on successive pulses. The sum of these then represents sampling at a rate of 160 MHz. This assumes the bandwidth of the digitizer input track-and-hold circuit is sufficient.

Sampling at a rate of 160 MHz generates a large amount of data when applied to a 1 s long pulse. However, as the dispersion processes to be observed tend to be exponential in effect over time, it is not necessary to continue sampling at this rate for the entire pulse profile. The sampling period needs to be less than 70% of the time constant to be observed, but typically sampling would be an order of magnitude faster for better amplitude resolution in noisy conditions. Thus, sampling may begin at 10 ns intervals for the first 100 ns, but then continue at every 100 ms towards the end of the 1 s pulse. Such logarithmic placement of sampling over the pulse is possible with digitizers that allow arbitrary triggering and systems that can generate arbitrary trigger signals. With such a system, sampling would be performed at a linear rate initially while requiring samples as fast as possible, reducing to a logarithmic spacing over time. For example, with a 20 MHz digitizer, sampling might be done every 50 ns for the first 1 μs, but then only ten samples per decade thereafter. This would give only 80 samples over a 1 s pulse, rather than the excessive 20 M samples from simple linear sampling. In this way, data can be kept to a manageable but adequate amount.

8.5.2.6 Output Impedance

In testing a device, whether the terminal current or voltage is the dependent variable or the independent variable is subjective and conditional upon type of device (BJT or FET). However, pulsed measurement systems are usually implemented with sources of voltage pulses, for practical reasons. Thus, it is desirable to have negligible output impedance in the pulse generator or driving amplifier.

There exist, however, some situations where it is desirable to have significant output impedance in the pulse driver. For example, in testing FETs with very-fast pulses, it is usually necessary to use a 50 Ω output impedance with the gate-terminal pulser to prevent RF oscillations.

When current is the more convenient independent variable, a large driver output impedance can simulate a current source. With bipolar devices (BJTs and HBTs), it is desirable to perform measurements at particular values of base current. This is a very strong function of base-emitter voltage and hence difficult to control with a voltage source. With a large source resistance (e.g., 10 kΩ) in the base-voltage driver, a reasonable current source can be approximated and base current controlled. This will necessarily severely limit the rise time of a base-terminal pulse, so that typically this pulse would be first applied and allowed to stabilize before a fast pulse is applied to the collector terminal. This is fine for investigating isodynamic collector current in relation to dispersion effects due to collector voltage and power dissipation. However, the long base-current pulse implies that base voltage- and current-related dispersion effects are not isodynamic.

Output impedance is also used for current limiting and for safe exploration of the I/V-plane. The diode characteristic of the FET gate junction during forward conduction and breakdown means that gate current can become very large. A 50 Ω resistance in the gate-terminal pulser will limit this current to less than 20 mA V^{-1}. Similarly, 50 Ω in the drain-terminal pulser will limit drain current for a particular voltage pulse and constrain DUT output behavior to follow the load line determined by this 50 Ω load impedance and the applied voltage pulse. In this way, pulse voltage can be slowly increased to explore expanded regions of device operation safely. It will also curb transients.

8.5.2.7 Extending the Data Range

An important aspect of pulsed testing is that a wider range of data points can be tested. Beyond a certain range of terminal potentials or power, damage can be done to a device because of excessive temperature or current density. As the DUT temperature is a function of the time for which a given power level is applied, the shorter a pulse then the greater the voltage and/or instantaneous power that can be applied.

The conventional SOA of a device is that part of the I/V-plane for which the device can withstand static or continuous application of those voltage levels. Pulsed testing then extends this region, in particular to regions that are outside the static SOA, but are still encountered during normal RF operation of the device.

This gives an extended range of data for use in modeling device operation, not only for isodynamic I/V characteristics, but also for RF parameters for extraction of parasitic resistances and capacitances. With a pulsed S-parameter system coupled with a pulsed-I/V system, the voltage pulses can take the DUT to an isothermal point outside the static SOA, where S-parameters can then be measured during this pulse.

8.5.2.8 Repetition

The characteristics of a device can change due to the manner in which it is used. For example, an excursion into a breakdown region can alter although not damage a device, permanently modifying its characteristics. To investigate such phenomena, an I/V-grid can be measured before and after such an excursion. Changes in the device characteristics can then be observed in the difference between the measurements [12].

Of use in such investigations is the ability to specify an arbitrary list of pulse points. In this case, the list of points in the I/V-plane to be pulsed to would first list the regular grid, then the points of breakdown excursion, and then repeat of the same regular grid points. Additionally, scripting capabilities might be used to create a series of such measurements.

8.5.2.9 Onion-Ring Destructive Testing

Often it is desired to test a device until destruction. An example of this might be breakdown measurements. Sometimes it is difficult not to destroy a *fragile* device during testing—especially devices fabricated with an immature technology. In either case, it is desirable to structure the sequence of pulse points from safe voltage and power levels to increasing levels up to destruction. It is essential in this that all data up to the point of device destruction is preserved.

Here again, scripting capabilities and the use of a list of pulse points allow measurements to be structured as a series of layers of pulse points, increasing in power and/or voltage level. In this way, the characteristics of a device can be explored as an extension, in layers, of the safe device operation or constant power level. Inter-pulse averaging and a waiting period for device stabilization would not normally be used in this form of measurement.

8.5.2.9.1 Quiescent Measurement

It is important to measure the bias point representing the isodynamic conditions of the DUT. This is the terminal voltage and current before each pulse and as such gives the quiescent thermal and trapping state of the device. This needs to be measured as part of the pulse exercise if the pulse sequence used is such that the average device temperature is raised.

The time spent at the quiescent point is usually quite long, affording opportunity for considerable averaging. Additionally, when measuring many points on the I/V-plane, the quiescent point can be measured many times. Thus, a comparatively noise-free measurement can be obtained.

Sample points for quiescent data would usually be placed immediately before a pulse. A value for the quiescent condition can be obtained by averaging the samples taken prior to all pulses. It is assumed that the relaxation time at the quiescent condition, since the previous pulse, is very much greater than all relevant dispersion-effect time constants (unless these time constants are themselves being investigated). This is necessary if the samples are to be considered as representing a bias condition, rather than a transient condition.

Alternatively, or additionally, some samples might be taken immediately after a pulse. For these post-pulse samples to be considered to represent the bias condition, the pulse must be short enough for no significant dispersion effects to have occurred. Notwithstanding this, there may be useful information in observing relaxation after a pulse and in the change in device current immediately before and after a return from a pulse.

8.5.2.9.2 Other Techniques

As well as the various measurement techniques just discussed, there exist a range of practical issues. For example, with combined pulsed-I/V and pulsed-RF systems, the RF must be turned off while measuring DUT current. This means that experiment times are longer than might be expected, as the pulsed-I/V and pulsed-RF data are gathered separately.

Another consideration is that the applied voltage pulses are not square-shaped. Instrumentation and cable termination issues result in pulses having significant rise and fall times and in particular overshoot and settling. The devices being tested are generally fast enough to respond to the instantaneous voltages, rather than an averaged rectangular pulse. First, this means that sampling of both voltage and current must be performed, and that this must occur simultaneously. Second, as any pulse overshoot will be responded to, if this voltage represents a destructive level then damage may be done even when the target voltage settled to is safe. This particularly applies to gate voltage pulses approaching forward conduction or breakdown.

Arising from the fact that the DUT is far faster in response than the pulse instrumentation is the issue of pulsing trajectory. In pulsing from a bias point to the desired pulse point, the DUT will follow a path of voltage and current values across the I/V-plane, between the two points. Similarly, a path is followed in returning from the pulse point to the bias point. The actual trajectory followed between these two points will be determined by the pulse rise and fall times, overshoot and other transients, and by the relative inset of gate and drain pulses (Figure 8.9).

A problem can arise if, in moving between two safe points on the I/V-plane, the trajectory passes through a destructive point. An example is pulsing to a point of low drain voltage and high current from a bias point of high drain voltage and low current. Here drain voltage is pulsing to a lower voltage while gate voltage is pulsing to a higher value. If the gate pulse is applied first, then the DUT will move through a path of high voltage and high current. This is a problem if it represents destructive levels and is dependent upon trajectory time. A similar problem exists in returning from the pulse point to the bias point. In general, because gate/drain coincidence cannot be sufficiently well-controlled, consideration need be given to the trajectories that may be taken between two points on the I/V-plane and the suitability of these. With appropriate choice of leading and trailing overlaps between the gate and drain pulses, this trajectory can be controlled.

8.6 Data Processing

Having gathered data through pulsed measurements, various processing steps can follow. In this, reference need again be made to the pulse-domain paradigm. In the simplest case, the data consists of a grid of pulse points for a fixed bias point, sampled free of dispersion effects. To this could be added further data of grids for multiple bias points. Rate dependence can be included with data from pulse-profile measurements and grids with delayed sample times. In this way, the data can be considered as a sampling of a multidimensional space. The dimensions of this space are the terminal currents and voltages, both instantaneous and average, together with sample timing and temperature. Added to this can also be RF parameters at a range of frequencies.

Processing of this data can be done in two ways. First, the data can be considered as raw and processed to clean and improve it. Examples of this form of processing are interpolation and gridding. Second, data can be interpreted against specific models. Model parameter extraction is the usual objective here. However, to use fully the information available in the pulsed data, such models need to incorporate the dispersion effects within the pulse-domain paradigm.

8.6.1 Interpolation and Gridding

Data over the I/V-plane can be gathered rapidly about a grid of target pulse points. The grid of voltage values represents raw data points. Instrument output impedance and noise usually differentiate these from desired grid points. Interpolation and gridding can translate this data to the desired grid.

Data can be gathered rapidly if the precision of the target pulse-voltage values is relaxed. The data still represents accurate samples; however, the actual voltage values will vary considerably. This variation is not a problem in model extraction, but can be a problem in the comparison of different characteristic curves (for different quiescent conditions) and the display of a single characteristic curve for a specified terminal voltage.

Gridding is performed as the simple two-dimensional interpolation of current values as a function of input and output pulse-voltage values. A second- or third-order function is usually used. The interpolated voltage values represent a regular grid of desired values, whereas the raw data values are scattered. A least-squares fit can be used if a noise model is assumed, such as thermal noise. Nothing is assumed about the underlying data, except for the noise model and the assumption that the data local variation can be adequately covered by the interpolation function used.

8.6.1.1 Intrinsic Characteristics

The simplest of models for data interpretation all assume series access resistances at each terminal. Fixed resistances can be used to model probe and contact resistances, as connecting external terminals to an idealized internal nonlinear device. For measured terminal current and assumed values of resistances, the voltage across the terminal access resistances is calculated and subtracted to give intrinsic voltages. These voltages can then be used in model interpretation.

For example, consider a FET with gate, drain, and source access resistances of R_G, R_D, and R_S, respectively. If the measured terminal voltages and currents are v_{GS}, i_G, v_{DS}, and i_D respectively, then the intrinsic voltages can be obtained as

$$v'_{DS} = v_{DS} - i_D R_D - (i_D + i_G) R_S,$$

$$v'_{GS} = v_{GS} - i_G R_G - (i_D + i_G) R_S. \tag{8.2}$$

If v_{GS}, i_G, v_{DS}, and i_D are raw data, then a set of v'_{DS}, v'_{GS} values can be used to obtain a grid of intrinsic data. This is easy to do with copious amounts of data gathered over the I/V-plane.

8.6.1.2 Interpretation

The data, raw or gridded, can be used to extract information on specific effects under investigation. In the simplest case, small-signal transconductance and conductance can be obtained as gradients, such as di_D/dv_{GS} and di_D/dv_{DS} in the case of a FET. These could then be used in circuit design where the device is being operated at a specific bias point. A second example is in the extrapolation of plots of voltage and current ratios to give estimates of terminal resistances for use in determining intrinsic values. The advantage of pulsed testing here is that an extended range of data can be obtained, extending outside the static SOA.

Another example of data interpretation is the use of measured history dependence to give information on dispersion effects. If, in pulsed testing, the quiescent relaxation time is insufficient, then pulse samples will be affected by dispersion. The use of shuffling of the pulse sequence enhances sampling of dispersion. Models of dispersion can then be fitted to this data to extract parameters for dispersion, as a function of terminal voltages and of pulse timing.

8.6.1.3 Modeling

The paradigm of pulsed testing assumes that DUT terminal currents are functions of both instantaneous and of average terminal voltages. This means that device response to RF stimuli will be different for different average or bias conditions. Pulsed testing allows separation and measurement of these effects.

A model of device behavior, for use in simulation and design, must then either incorporate this bias dependence or be limited to use at one particular bias condition. The latter is the usual case, where behavior is measured for a particular bias condition, for modeling and use at that bias condition.

If a model incorporates the bias-dependent components of device behavior, the wider sample space of pulsed testing can be utilized in model parameter extraction. From I/V-grids sampled for multiple bias conditions, the bias dependency of terminal current can be extracted as a function of both instantaneous and bias terminal voltages. From pulse-profile measurements, dispersion effects can be modeled in terms of average terminal voltages, where this average moves from quiescent to pulse target voltage,

over the pulse period, according to a difference equation and exponential time constants. The actual parameter extraction consists of a least-squares fit of model equations to the range of data available, starting from an initial guess and iterating to final parameter values. The data used would be I/V-grids, pulse profiles, and RF measurements over a range of frequencies, at a range of bias points, depending on the scope of the model being used. Important in all this is a proper understanding of what the sampled DUT data represents, in the context of the pulse-domain paradigm, and of how the data are being utilized in modeling.

Empirical models that account for dispersion effects must calculate terminal currents in terms of the instantaneous and time-averaged potentials. In the case of a FET, the modeled drain current is a function of the instantaneous potentials v_{GS} and v_{DS}, the averaged potentials $<v_{GS}>$, $<v_{DS}>$ and average power $<I_{DS}v_{DS}>$. The time averages are calculated over the time constants of the relevant dispersion effects. A model of thermal dispersion is

$$I_{DS} = I_O(1 - \lambda R_T \langle I_{DS} v_{DS} \rangle), \tag{8.3}$$

where i_O includes other dispersion effects in a general form

$$I_O = I(v_{GS}, v_{DS}, \langle v_{GS} \rangle, \langle v_{DS} \rangle). \tag{8.4}$$

With a suitable value of λR_T, the thermal effects present in the characteristics of Figure 8.1 can be modeled and the other dispersion effects can be modeled with the correct function for i_O in Equation 8.4. The DC characteristics are given by the model when the instantaneous and time-averaged potentials track each other such that $<v_{GS}> = v_{GS}$, $<v_{DS}> = v_{DS}$, and $<I_{DS}v_{DS}> = I_{DS}v_{DS}$. In this case, the model parameters can be fitted to the measured DC characteristics and would be able to predict the apparently negative-drain conductance that they exhibit. In other words, the DC characteristics are implicitly described by

$$I_{DS} = I(V_{GS}, V_{DS}, V_{GS}, V_{DS})(1 - \lambda R_T I_{DS} V_{DS}). \tag{8.5}$$

Of course, this would be grossly inadequate for modeling RF behavior, unless the model correctly treats the time-averaged quantities as constants with respect to high-frequency signals.

For each quiescent point $(\langle v_{GS} \rangle, \langle v_{DS} \rangle)$, there is a unique set of isodynamic characteristics, which relate the drain current i_{DS} to the instantaneous terminal-potentials v_{GS} and v_{DS}. Models that do not provide time-averaged bias dependence must be fitted to the isodynamic characteristics of each quiescent condition individually. Models in the form of Equations 8.3 and 8.4 simultaneously determine the quiescent conditions and the appropriate isodynamic characteristics [13,14]. Pulsed measurements facilitate this characterization and modeling of device RF behavior with bias dependency.

8.7 Summary

Pulsed measurements yield an extended range of characteristic curves for a device that, at specific operating conditions, correspond to the high-frequency behavior of the device. Central to the pulse technique is the principle of the pulse-domain paradigm, which considers the characteristic curves to be a function of quiescent operating condition. Therefore, the basis for pulse techniques is the concept of measurements made in isodynamic conditions, which is effectively an invariable operating condition. The requirements for an isodynamic condition vary with the transistor type and technology and bias conditions. Pulsed-measurement equipment and specifications vary considerably in terms of cost and complexity. Pulse-measurement techniques exploit various timing schemes and ordering of pulse-measurements points to reveal insights into the dynamics of device operation.

Defining Terms

Characteristic curves: For FETs/HBTs, a graph showing the relationship between drain/collector current (or RF parameters) as a function of drain/collector potential for step values of gate/base potential.

Bias condition: For a device, the average values of terminal potential and currents when the device is operating with signals applied.

Bias-evolution characteristic: A set of of isodynamic characteristics repeated, so as to record their variation over time after a step change in bias condition.

Dispersion effects: Collective term for thermal, rate-dependent, electron trapping, and other anomalous effects that alter the characteristic curves with the bias condition changes.

DC characteristics: Characteristic curves relating quiescent currents to quiescent terminal potentials.

Isodynamic: Having the condition of invariant bias, quiescent, and thermal properties.

Isodynamic characteristic: Characteristic curves relating instantaneous terminal currents and voltages for isodynamic conditions.

Measurement cycle: A periodic repetition of a sequence of pulse events.

Pulsed bias: Pulsed stimulus that briefly biases a device during a pulsed-RF measurement.

Pulsed characteristics: Characteristic curves measured with pulsed-I/V or pulsed-RF measurements.

Pulsed-*I/V* measurement: Device terminal currents and voltages measured with pulse techniques.

Pulsed-RF measurement: Device RF parameters measured with pulse techniques.

Pulse event: A single pulse stimulus preceded by a period at the quiescent point.

Pulse repetition frequency (PRF): Rate of repetition of pulse events.

Quiescent condition: For a device, the value of terminal potential and currents when the device is operating without any signals applied.

Stabilizing period: Time required to establish a stable quiescent condition that is the steady-state component of the repetition of pulse events.

Soak time: Time required to establish a stable quiescent condition with a DC bias without pulse stimuli.

Time-evolution characteristic: A collection of time-domain pulsed measurements from a bias to points on a DC characteristic.

References

1. Parker, A.E. and Rathmell, J.G., Bias and frequency dependence of FET characteristics, *IEEE Trans. MTT*, vol. 51, no. 2, Feb. 2003, pp. 588–592.
2. Parker, A.E. and Rathmell, J.G., Broad-band characterization of FET self-heating, *IEEE Trans. MTT*, vol. 53, no. 7, July 2005, pp. 2424–2429.
3. Webster, R.T., Wu, S. and Anwar, A.F.M., Impact ionization in InAlAs/InGaAs/InAlAs HEMTs, *IEEE Elect. Device Lett.*, vol. 21, May 2000, pp. 193–195.
4. Rathmell, J.G. and Parker, A.E., Characterizing charge trapping in microwave transistors, *Microelectronics, MEMS and Nanotechnology* (John M. Bell and Vijay K. Varadan, Eds.), Brisbane Australia, 11–15 Dec. 2005, *SPIE The International Society for Optical Engineering.*
5. Teyssier, J.-P., Bouysse, P., Ouarch, Z., Baratand, D., Peyretaillade, T. and Raymond Quéré, R., 40-GHz/150-ns versatile pulsed measurement system for microwave transistor isothermal characterization, *IEEE Trans. MTT*, vol. 46, no. 12, pp. 2043–2052, Dec. 1998.
6. Scott, J., Sayed, M., Schmitz, M., Schmitz, P. and Parker, A.E., Pulsed-bias/pulsed-RF device measurement system requirements, *24th European Microwave Conference*, Cannes, France, 5–8 Sept. 1994, *European Microwave Association*, London.
7. Ernst, A.N., Somerville, M.H. and del Alamo, J.A., Dynamics of the kink effect in InAlAs/InGAs HEMT's, *IEEE Elect. Device Lett.*, vol. 18, no. 12, pp. 613–615, Dec. 1997.
8. Nanometrics Incorporated, "DIVA[Dynamics I-V Analyser]," 2007. [online]. Available: URL: http://www.nanometrics.com/products/Diva.html
9. Access Macquarie Ltd, "Pulsed-bias semiconductor parameter analyzer," 2004. [Online]. Available: URL: http://www.elec.mq.edu.au/cnerf/apspa.

10. Auriga Measurement Systems, "Auriga Measurement Systems haracterization Services," 2006. [Online]. Available: URL: http://www.auriga-ms.com.
11. Parker, A.E. and Scott, J.B., Method for determining correct timing for pulsed-I/V measurement of GaAs FETs, *IEE Elect. Lett.*, vol. 31, no. 19, pp. 1697–1698, 14 Sept. 1995.
12. Scott, J.B., Rathmell, J.G., Parker, A.E. and Sayed, M., Pulsed device measurements and applications, *IEEE Trans. MTT*, vol. 44, no. 12, pp. 2718–2723, Dec. 1996.
13. Parker, A.E. and Skellern, D.J., A realistic large-signal MESFET model for SPICE, *IEEE Trans. MTT*, vol. 45, no. 9, pp. 1563–1571, Sept. 1997.
14. Filicori, F., Vannini, G., Santarelli, A., Sanchez, A.M., Tazon, A. and Newport, Y., Empirical modeling of low frequency dispersive effects due to traps and thermal phenomena in III-V FET's, *IEEE Trans. MTT*, vol. 43, no. 12, pp. 2972–2981, Dec. 1995.

9

Microwave On-Wafer Test

Jean-Pierre Lanteri
Christopher Jones
John R. Mahon
M/A-COM TycoElectronics

9.1 On-Wafer Test Capabilities and Applications

9.1.1 Fixtured Test Limitations

Until 1985 the standard approach to characterize at microwave frequencies and qualify a semiconductor wafer before shipping was to dice it up, select a few devices, typically one in each quadrant, assemble them, and then test them in a fixture, recording s-parameters or power levels. Often, the parts were power transistors, the most common RF/microwave product then, and a part was used as a sample. For Gallium Arsenide (GaAs) monolithic microwave integrated circuits (MMICs), a transistor was similarly used for test coupon, or the MMIC itself. Typically, the parts were assembled in a leaded metal ceramic package, with epoxy or eutectic attach, and manually wedge bonded with gold wires for RF and bias connections. The package was then manually placed in a test fixture and held down by closing a clamp on the leads and body. The fixture was connected to the test equipment, typically a vector network analyzer (VNA) or a scalar power meter, by radio frequency (RF) coaxial cables to present a 50 Ω environment at the end of the coaxial cables. The sources of test uncertainty were numerous:

- Part placement in the package and bond wire loop profile, manually executed by an operator, lead to bond wire length differences and therefore matching variations for the device under test (DUT).
- Package model inaccuracy and variability from package to package.
- RF and ground contacts through physical pressure of the clamp, applying force to the body of the package and the leads, with variable results for effective lead inductance and resistance, and potential oscillations especially at microwave frequencies.
- Fixture de-embedding empirical model for the connectors and transmission lines used on the RF ports.
- Calibration of the test equipment at the connectorized interface between the RF cables and the test fixture, not at the part or package test planes.

Most of these technical uncertainties arise because the calibration plane is removed from the product plane and the intermediate connection is not well characterized or not reproducible.

The main drawbacks of fixtured tests from a customer and business perspective were:

- Inability to test the very product shipped, only a "representative" sample is used due to the destructive nature of the approach. Especially for MMICs where the yield loss can be significant, this can lead to the rejection of many defective modules and products after assembly, at a large loss to the user.
- Cost of fixtured test; sacrificing parts and packages used for the test.
- Long cycle time; typically a day or two are needed for the parts to make it through assembly.
- Low rate production test; part insertion in a fixture is practically limited to a part per minute.

A first step was to develop test fixtures for bare die that could be precisely characterized. One solution was a modular fixture, where the die is mounted on an insert of identical length, which is sandwiched between two end pieces with transmission line and connector. The two end pieces can be fully characterized with a VNA to the end point of the transmission lines by short-open-load-thru (SOLT) or thru-reflect-line (TRL) calibrations; wire bonding to preset inserts or between the two end pieces butted together. Then the die is attached to the insert, assembled in between the end pieces, and wire bonded to the transmission lines. This approach became the dominant one for precise characterization and model extraction. The main advances were removal of die placement, package, lead contact and fixture as sources of variability, at the expense of a complex assembly and calibration process. The remaining limitations are bond-loop variation, and destructiveness, and the length and cost of the approach, preventing its use in volume applications such as statistical model extraction or die acceptance tests.

9.1.2 On-Wafer Test Enabler: Coplanar Probes

The solution to accurate, high-volume microwave testing of MMICs came from Cascade Microtech, the first company to make RF and microwave probes commercially available, along with extensive application support; their history and many useful application notes are provided on their Website (*www.cascademicrotech.com*). On-wafer test was common place for DC and digital applications, with high-pin-count probe cards available, based upon needles mounted on metal or ceramic blades. Although a few companies had developed RF frequency probes for their internal use, they relied on shortened standard DC probes, not the coplanar ground-signal-ground (G-S-G) structure of Cascade Microtech's probes, and were difficult to characterize and use at microwave frequencies. The breakthrough idea to use a stable G-S-G configuration up to the probe tip enabled a reproducible 50 Ohms match to the DUT, leading to highly reproducible, nondestructive microwave measurements at the wafer level [1,2]. All intermediate interconnects were eliminated, along with their cost, delay, and uncertainty, provided that the DUT was laid out with the proper G-S-G inputs and outputs. Calibration patterns (Short, Open, Load, Thru, Line Stub) available on ceramic substrates or fabricated on the actual wafers provided standard calibration to the probe tips [3,4]. A few years later, PicoProbe (*www.picoprobe.com*) introduced a different mechanical embodiment of the same GSG concept.

About the same time, automatic probers with top plates fitted with probe manipulators for Cascade Microtech's probes became available. Agilent (then Hewlett Packard) introduced the 8510 VNA, a much faster and easier way to calibrate microwave test equipment, and 50 Ohms matched MMICs dominated microwave applications. These events combined to completely change the characterization and die selection process in the industry. By the late 1980s, many MMIC suppliers were offering wafer qualification based upon RF test results on standard transistor cells in a process control monitor (PCM) and providing RF-tested known good dies (KGD) to their customers.

9.1.3 On-Wafer Test Capabilities

At first, RF on-wafer testing was used only for the s-parameter test, for two port devices up to 18 GHz. Parameters of interest were gain, reflection coefficients, and isolation. Soon RF switching was introduced

to test complex MMICs in one pass, switching the two ANA ports between multiple DUT ports. Next came noise figure test on-wafer, using noise source and figure meter combined with ANA. Power test on-wafer required a new generation of equipment, pulsed vector analyzers, to become reliable, and provided pulsed power, power droop, and phase droop [5]. Soon many traditional forms of microwave test equipment were connected to the DUT through complex switching matrixes for stimuli and responses, such as multiple sources, amplifiers, spectrum analyzers, yielding intermodulation distortion products. Next came active source pull equipment, and later on active load pull [6], from companies such as ATN Microwave (*www.atnmicrowave.com*) and Cascade Microtech. The maximum s-parameter test frequency kept increasing, to 26 GHz, then 40 GHz, 50 GHz, and 75 GHz. In the late 1990s new parameters such as noise power ratio (NPR) and adjacent channel power ratio (ACPR) were required and could be accommodated by digitally modulated synthesizers and vector signal analyzers (Table 9.1). Today, virtually any microwave parameter can be measured on-wafer, including s-parameters up to 110 GHz.

9.1.4 On-Wafer RF Test Applications

On-wafer test ease of use, reasonable cost, and extensive parameter coverage has led to many applications in MMIC development and production, from device design and process development to high volume test for known good die (KGD). The main applications are summarized in Table 9.2. Of course, all of the devices to test need to have been designed with one of the standard probe pad layouts (S-G-S, G-S, or S-G) to allow for RF probing:

1. Model development and statistical model extraction is often performed on design libraries containing one type of element, generally field effect transistors (FET), but sometimes inductors or capacitors, implemented in many variations that are characterized to derive a parametric model of the element.[7] The parts must be laid out with G-S-G (or G-S only for low microwave frequencies) in a coplanar and/or microstrip configuration. This test task would have taken months ten years ago, and is now accomplished in a few days. The ability to automatically perform all these measurements on significant sample sizes has considerably increased the statistical relevance of the device models. They are stored in a statistical database automatically used by the design and yield simulation tools. This allows first pass design success for complex MMICs.
2. Process monitoring is systematically performed on production wafers, sample testing a standard transistor in a process control monitor (PCM) realized at a few places on each wafer. The layout is in a coplanar configuration that does not require back-side ground vias and therefore can be tested in process. Each time, a small signal model is extracted. Very good agreement between the tested s-parameters and the calculated ones from the extracted model can be seen in Figure 9.1. The results are used during fabrication for pass/fail screening of wafers on RF parameters, and supplement the statistical model data.

TABLE 9.1 On-Wafer RF Test Capabilities Evolution

Year	Product	Configuration	Test Capability	Equipment
1985	Amplifier	2-Port	18 GHz s-Parameters	ANA
1987	Amplifier	Switched Multi-Port	26 GHz s-Parameters	ANA + Switch Matrix
1989	LNA	2-Port	Noise Figure	ANA + Noise System
1990	HPA	2-Port	Pulsed Power	Pulsed Power ANA
1991	Amplifier	2-Port	Intermodulation	Spectrum Analyzer
1991	LNA	2-Port, Zin Variable	Noise Parameters	Active Source Pull, ANA
1992	Mixer	3-Port	Conversion Parameters	ANA, Spectrum Analyzer
1993	HPA	2-Port, Zout Variable	Load Power Contours	Active Load Pull, ANA
1995	T/R Module	Switched Multi-Port	40 GHz s-Parameter, NF, Power	ANA, Noise, Spectrum
1998	Transceiver	Multi-Port	Modulation Parameters	Vector Signal Analyzer
1999	Amplifier	2-Port	110 GHz s-Parameters	ANA

TABLE 9.2 On-Wafer RF Test Applications

Application	DUT	Technique	Test	Test time/DUT	Volume/year
FET Model Development	Standard transistor	Source or load pull	Noise parameters, load contours	10 min	100s
Statistical Model Extraction	Transistor library	S-Parameters, NF, PP, set load	Small and large signal models	1 min	1,000s
Process Monitoring	PCM transistor	S-Parameters, 50 ohms	Small signal model	10 s	10,000s
Know Good Die Test	MMIC or transistor	S-Parameters, NF, PP	Test specification	10–30 s	100,000s
Module or Carrier Test	Assembly or package	S-Parameters, NF, PP	Test specification	10-60 s	100,000s

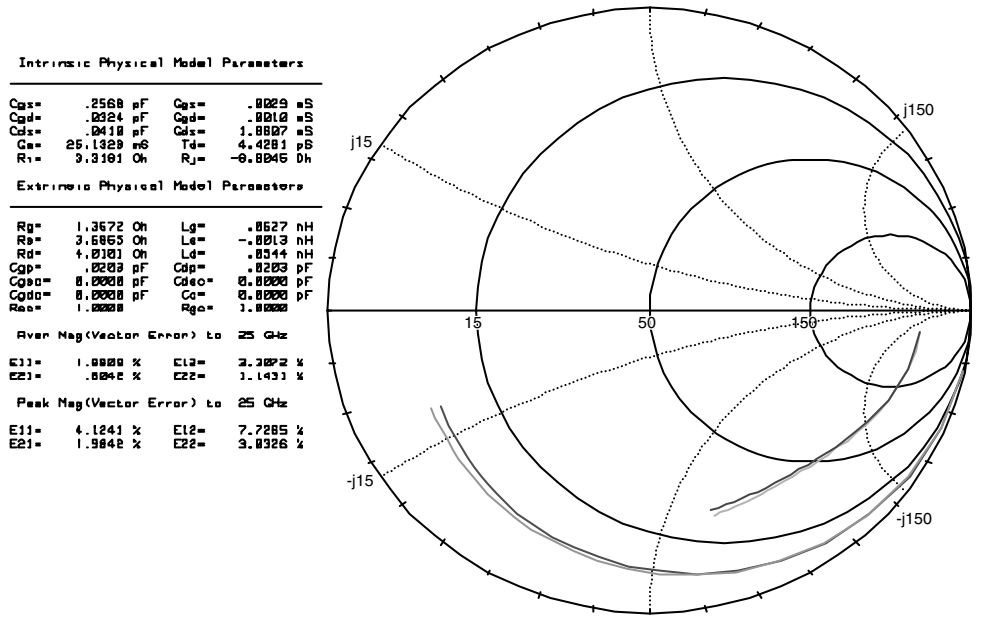


FIGURE 9.1 Equivalent circuit FET model extraction and fit with measurement.

3. On-wafer test is a production tool for dies, typically 100% RF tested when sold as is—as KGD—or used in expensive packages or modules. This is the norm for high power amplifiers in expensive metal ceramic packages, MMICs for transmit/receive (T/R) modules, bumped parts for flip-chip assembly, and military applications. The RF parameters of interest are measured at a few points across the DUT bandwidth, as seen in Figure 9.2, and used to make the pass/fail decision. The rejected dies on the wafer are either marked with an ink dot, or saved in an electronic wafer map, as seen in Figure 9.3, which is used by the pick-and-place equipment to pick the passing devices. Final RF test on-wafer is usually not performed on high-volume products. These achieve high yields and are all assembled in inexpensive packages, therefore it is easier and cheaper to plastic package all parts on the wafer to test them on automatic handlers and take the yield at this point.
4. The same "on-wafer" test application is used when testing packages, carriers, or modules manufactured in array form on ceramic or laminate substrates, or leadless packages held in an array format by a test fixture.

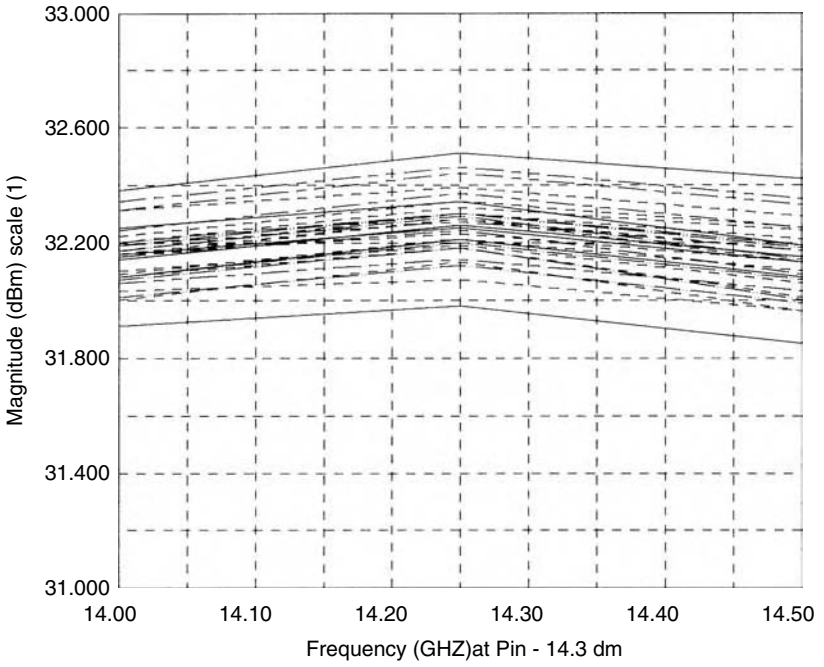


FIGURE 9.2 Pout response of Ku band PAs across a wafer.

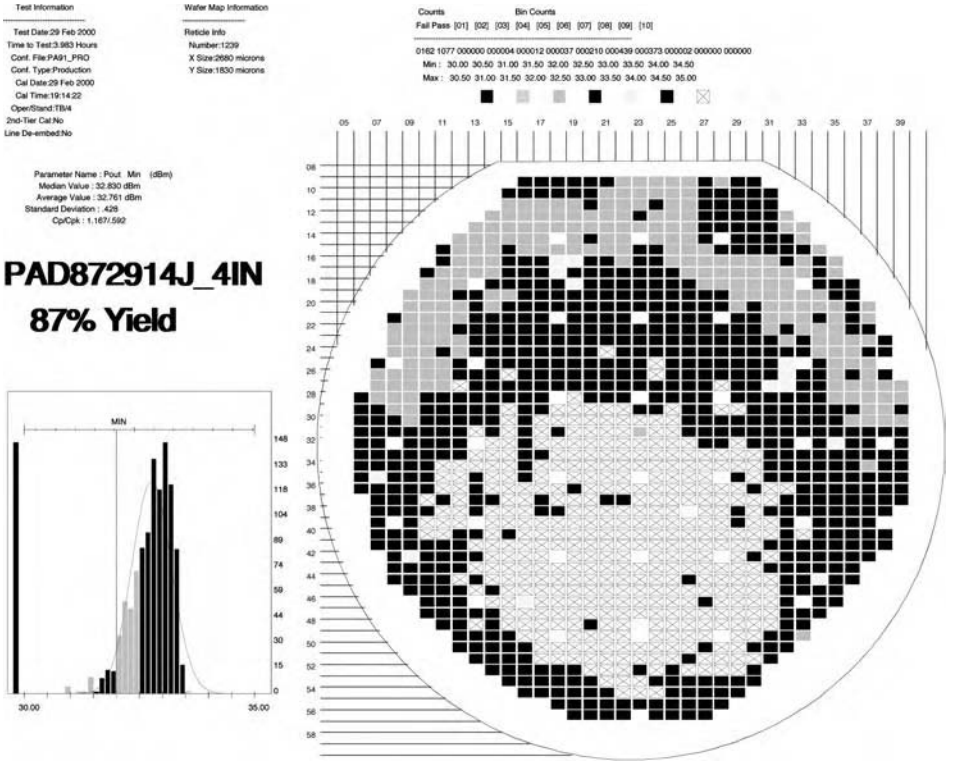


FIGURE 9.3 Wafter map of known good dies from on-wafer test.

9.2 Test Accuracy Considerations

In any test environment, three important variables to consider are accuracy, speed, and repeatability. The order of importance of these variables is based on price of the device, volume, and specification limits. High test speed is beneficial when it reduces the test cost-per-part and provides greater throughput without reaching an unacceptable level of accuracy and repeatability. Perfect accuracy would seem ideal, although in a high volume manufacturing environment "accuracy" is usually based on agreement between test results of two or more parties, primarily the vendor and end customer, for a specific product. The end customer, utilizing their available methods of measurement, usually defines most initial device specifications and sets the reference "accuracy," defining what parts work in the specific customer application. If due to methodology differences, a vendor's measurement is incompatible with that of a customer, yield and output can be affected without any benefit to the customer. It is not always beneficial, in this environment, to provide a more "accurate" method of measuring a product if the end customer is not testing it in the same fashion. Repeatability of the supplier measurement and correlation with the customer result are the more important criteria in that case.

Accuracy and repeatability considerations of any measurement system can be broken down into four primary parts, discussed in detail in the next sections.

9.2.1 Test Equipment Manufacturer

The manufacturer tolerances and supplied instrument error models are the first places to research when selecting the appropriate system. Most models will provide detailed-information on performance, dynamic range, and accuracy ratings of the individual instruments. Vendors like Agilent, Anritsu, Tektronix, and Boonton, to name a few, provide most hardware resources needed for automatic testing. There are many varieties of measurement instruments available on the market today. The largest single selection criterion of these is the frequency range. The options available diminish and the price increases dramatically as the upper frequency requirements increase. In the last decade many newer models with faster processors, countless menu levels, and more compact enclosures have come on the market making selections almost as difficult as buying a car. Most vendors will be competitive with each other in these matters. More important is support availability, access to resources when questions and problems arise, and software compatibility. Within the last decade many vendors have adopted a standard language structure for command programming of instruments known as SCPI (pronounced Skippy). This reduces software modification requirements when swapping instrumentation of one vendor with another. Some vendors have gone so far as to option the emulation of a more established competitor's model's instrument language to help inject their products into the market.

9.2.2 System Integration

Any system requiring full parametric measurement necessitates a complex RF matrix scheme to integrate all capabilities into a single function platform. Criteria such as frequency range, power levels, and device interface functionality drive the requirements of a RF matrix. Highly integrated matrices can easily exhibit high loss and poor matches that increase with frequency if care is not taken in the construction. These losses and mismatches can significantly degrade the accuracy of a system regardless of the calibration technique used. Assuming moderate power levels are to be used, frequency range is by far the most critical design consideration.

A system matrix must outperform the parts being tested on it. For complex systems requiring measurements such as intermodulation, harmonics, noise figure, or high port-to-port isolation, mechanical switches are the better alternative over solid state. Solid state switches would likely add their own performance limitations to the critical measurements being performed and cause erroneous results. Mechanical switches also have limitations to be considered. Although most mechanical switches have

excellent transfer, isolation, and return loss characteristics, there is one issue that is sometimes overlooked. The return loss contact repeatability can easily vary by ±5 milliunits and is additive based on the number of switches in series. To remove this error, directional couplers could be placed last in the matrix closest to the DUT and multiplexed to a common measurement channel within the network analyzer. This deviates from a conventional 2-port ANA configuration, but is worth consideration when measuring low VSWR devices.

9.2.3 Calibration Technique

Regardless of the environment, the level of system complexity and hardware resources can be minimized depending on the accuracy and speed requirements. Although the same criteria apply to both fixture and wafer environments, for optimum accuracy, errors can be minimized by focusing efforts on the physical limitations of the system integration, the most important being source and load matches presented to the DUT. By minimizing these parameter interactions, the accuracy of a scalar system can approach that of a full vector-corrected measurement system.

The level of integration and hardware availability dictates the calibration requirements and capabilities of any test system. Simple systems designed for only one or two functions may necessitate assumptions in calibration and measurement errors. As an example, performing noise figure measurements on wafer using only a scalar noise figure system required scalar offsets be applied to attribute the loss of the probe environment, which cannot be dynamically ascertained through an automated calibration sequence. The same can also apply to a simple power measurement system consisting of only a RF source and a conventional power meter and assuming symmetry of input and output probes. These methods can be and are used in many facilities, but can create large errors if care is not taken to minimize mismatch error terms that often come with contact degradation from repeated connections.

To obtain high accuracy up to the probe interface in a wafer environment requires a two-tier calibration method for certain measurements since it is usually difficult to provide a noise source or power sensor connection at the wafer plane. The most effective measurement tool for this second-tier calibration is a VNA. It not only provides full vector correction to the tips of the RF probes, but when the resulting vector measurements are used in conjunction with other measurement, such as noise figure and power, it can compensate for dynamic vector interactions between the measurement system and the device being tested. Equation 9.1, the vector relationship to the corrected input power (P_{A1}), and Equation 9.2, the scalar offset normally applied in a simpler system, illustrate the relationship that would not be taken into account during a scalar power measurement when trying to set a specific input power level to the DUT. Usually a simple offset, P_{offset}, is added to the raw power measured at port A_0 (P_{A0}) to correct for the incident power at the device input A_1 (P_{A1}) . This can create a large error when poor or moderate matches are present.

As an example, a device with a 10 dB return loss in a system with a 15 dB source match, not uncommon in a wafer environment, can create an error of close to ±0.5 dB in the input-power setting when system interactions are ignored.

$$P_{A1} = \left| \frac{P_{A0}}{\left(1 - E_{sf} S_{11a}\right)^2} \right| \left(P_{offset}\right) \tag{9.1}$$

$$P_{A1} = P_{A0}\left(P_{offset}\right) \tag{9.2}$$

A similar comparison can be shown for the noise figure. Equations 9.3 and 9.4 illustrate the difference between the vector and scalar correction of the raw noise figure (R_{NF}) as measured by a standard noise

figure meter. Depending on the system matches and the noise source gamma, the final corrected noise figure (C_{NF}) could vary considerably.

$$C_{NF} = R_{NF} + 10LOG\left(\frac{\left(\left|E_{10}^2\right|\right)\left(1-\left|G_{ns}^2\right|\right)}{\left(\left(1-\left|E_{sf}+\left(E_{10}^2 G_{ns}/\left(1-G_{ns}E_{df}\right)\right)^2\right|\right)\left(\left|\left(1-E_{df}G_{ns}\right)^2\right|\right)}\right) \tag{9.3}$$

$$C_{NF} = R_{NF} + 10LOG\left(\left|E_{10}^2\right|\right) \tag{9.4}$$

For small signal correction, the forward path of the standard 12 term, full 2-port error model as given in Figure 3.4 (network analyzer calibration) [8], is applied. Equation 9.5 gives the derivation of the actual forward transmission (S_{21a}) from these error terms combined with raw measured data. By minimizing the mismatched terms E_{sf}, E_{lf}, E_{sr}, E_{lr}, E_{xf}, and E_{df}, detailed in Section 4.2, Equation 9.5 simplifies to Equation 9.6. This simplified term is essentially the calculation used in standard scalar measurement systems and reflects an ideal environment. A further level of accuracy can be obtained when dealing with scalar systems that is very dependent on the type of device being tested. Looking at Equation 9.5 it can be seen that in deriving S_{21a} many relationships between the error terms and measured values provide products that can further minimize errors based on the return loss components of the DUT as well as isolation in the reverse path. This makes an active device with good return losses and high reverse isolation a good candidate for a scalar measurement system when only concerned with gain as the functional pass/fail criteria. On the other hand, a switch or other control product has a potential for being a problem due to the symmetrical nature of the device if care is not taken to minimize the match terms. An even poorer candidate for a scalar system would be discrete transistors, which normally have not been tuned for optimum matching in the measurement environment. Figure 9.4 is an on-wafer measurement comparison of a discrete FET measurement using

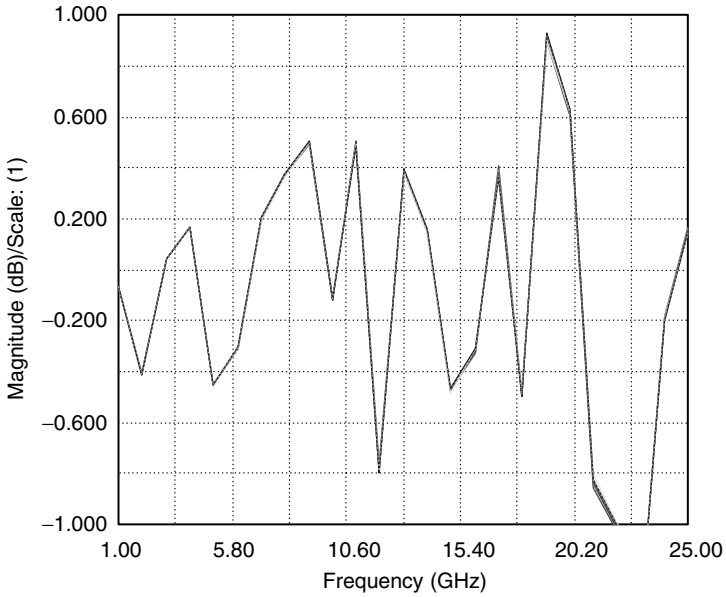


FIGURE 9.4 S_{21} vector to scalar measurement comparison of discrete FET (mismatched).

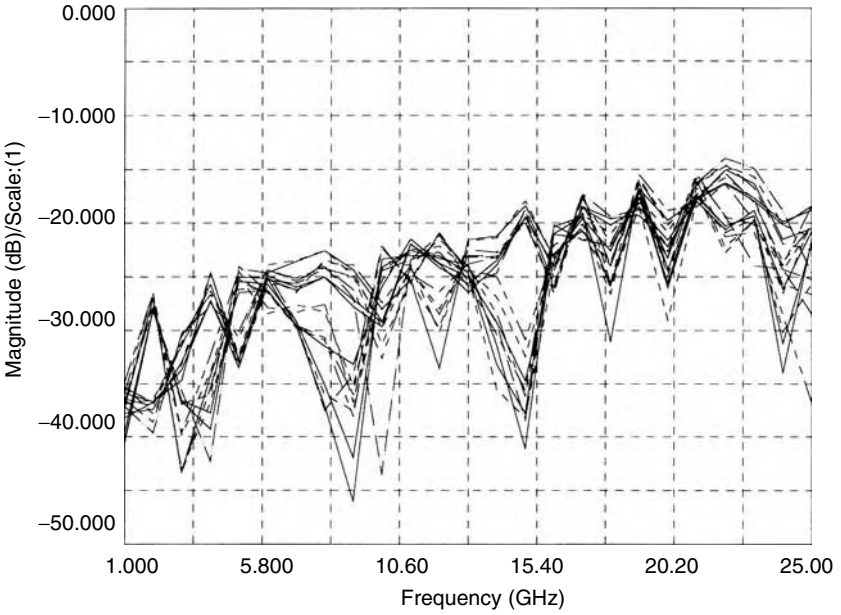


FIGURE 9.5 E_{lf} and E_{lr} error terms over a 5-month period.

both full 2-port error correction as in Equation 9.5 and the simplified scalar response Equation 9.6 from 1 to 25 GHz. The noticeable difference between these data sets is the "ripple" effect that is induced in the scalar corrected data, which stems from the vector sums of the error terms rotational relationship to the phase rotation of the measurement. Figure 9.5 shows the error terms E_{lf} and E_{lr} generated by multiple calibrations on the same vector test system used to measure the data in Figure 9.4. Although the values seem reasonable, the error induced in the final measurement is significant.

This error is largely based on the poor input and output match of the discrete FET, as shown in Figure 9.6, and their interaction with the system matches.

Figure 9.7, an example of better scalar-to-vector correlation, is an on-wafer measurement of a single pole double throw switch comparison using both full 2-port error correction as in Equation 9.5 and the simplified scalar response Equation 9.6 from 2 to 20 GHz. Although the system matches are comparable to the discrete FET measurement, the device input and output return losses are both below 15 dB (Figure 9.8). This product minimizes the errors induced by system to DUT interactions thus giving errors much smaller than that of the discrete FET measurement of Figure 9.4.

$$S_{21a} = \frac{\left(\left(S_{21m} - E_{xf}\right)/E_{tf}\right)\left(1+\left(S_{22m} - E_{dr}\right)\left(E_{sr} - E_{lf}\right)/E_{rr}\right)}{\left(1+\left(\left(S_{11m} - E_{df}\right)E_{sf}/E_{rf}\right)\right)\left(1+\left(\left(S_{22m} - E_{dr}\right)E_{sr}/E_{rr}\right)\right) - \left(\left(S_{21m} - E_{xf}\right)\left(S_{12m} - E_{xr}\right)E_{lf}E_{lr}/E_{tf}E_{tr}\right)} \tag{9.5}$$

$$S_{21a} = \frac{S_{21m}}{E_{tf}} \left| E_{lf}, E_{sf}, E_{sr}, E_{lr}, E_{df} \to 0 \right. \tag{9.6}$$

9.2.4 Dynamic Range

Dynamic range is the final major consideration for accuracy of a measurement system. Dynamic range of any measurement instrument can be enhanced with changes in bandwidth or averaging. This usually

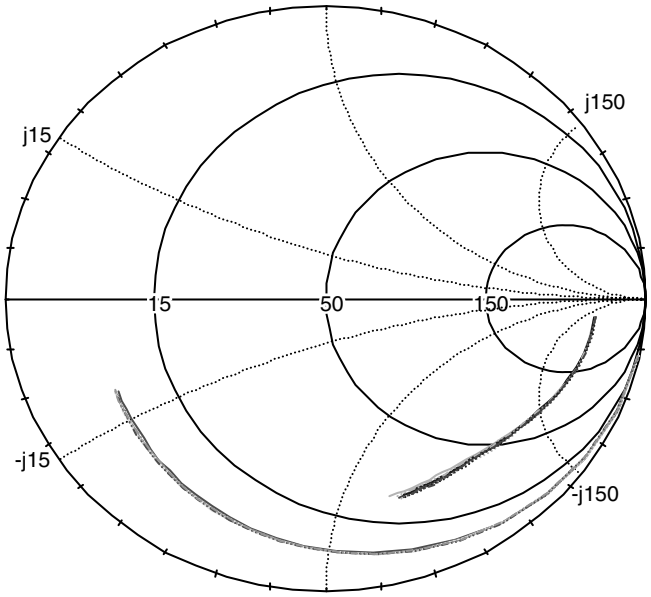


FIGURE 9.6 S_{11} and S_{22} of PCM FETs (mismatched) across a wafer.

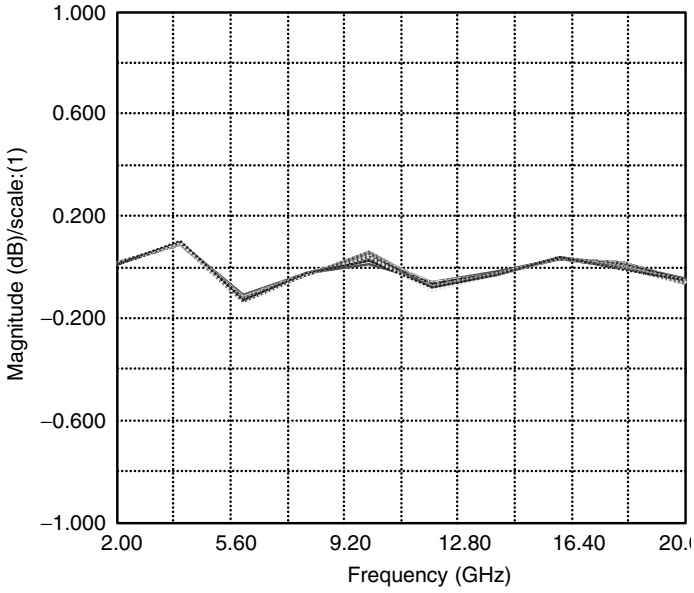


FIGURE 9.7 S_{21} vector to scalar measurement comparison for matched SPDT switch.

degrades the speed of the test. A perfect example of this is a standard noise figure measurement of a medium gain LNA using an HP 8970 noise figure meter. Noise figure was measured on a single device one hundred times using eight averages. The standard deviation is 0.02 dB, the cost for this is a 1.1-s measurement rate. By comparison, the same device measured with no averaging resulted in a standard deviation of 0.07 dB, but the measurement rate was less than 500 ms.

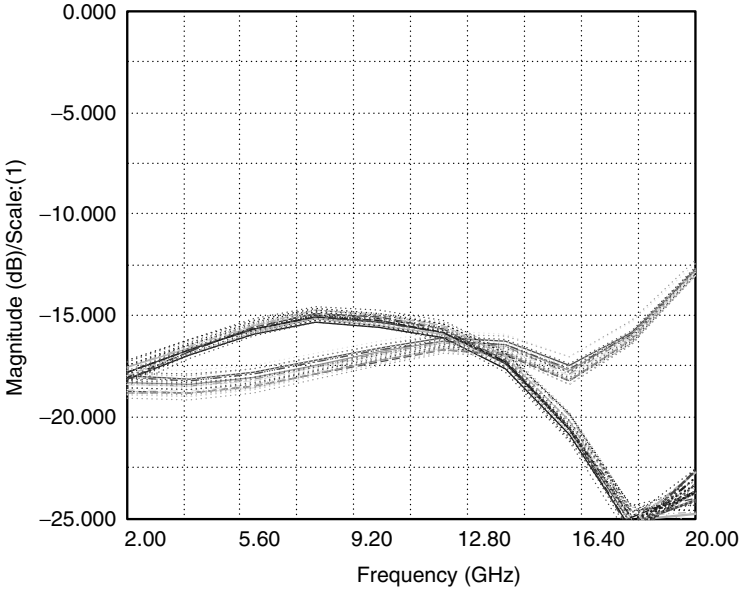


FIGURE 9.8 S_{11} and S_{22} of matched SPDT switch.

Other methods can be applied to enhance the accuracy of the measurement without losing the speed. Placing a high gain second stage LNA between the DUT and noise receiver will increase the dynamic range of the system and minimize the standard deviation obtained without losing the speed enhancement. These types of decisions should be made based on the parts performance and some experimentation.

Another obvious example is bandwidth and span setting on a spectrum analyzer. Sweep rates can vary from 50 milliseconds to seconds if optimization is not performed based on the requirements of the measurement. As in the noise measurement, this also should be evaluated based on the parts performance and some experimentation.

Highly customized systems that are optimized for one device type can overcome many dynamic range and mismatch error issues with additional components such as amplifiers, filters, and isolators. This can restrict or limit the capabilities of the system, but will provide speed enhancements and higher device output rates with minimal impact on accuracy.

9.3 On-Wafer Test Interface

On-wafer test of RF devices is almost an ideal measurement environment. Test interface technologies exist to support vector or scalar measurements. Common RF circuits requiring wafer test are: amplifiers, mixers, switches, attenuators, phase shifters, and coupling structures. The challenge is to select the interface technology or technologies that deliver the appropriate performance/cost relationship to support your product portfolio. Selection of test interface of wafer probes will be based on the measurements made and the desired product environment. It is common for high gain amplifiers to oscillate or for narrowband devices to shift frequency due to lack of bypass capacitors or other external components. It is recommended to consider wafer test during the circuit design stage to assure the circuit layout satisfies wafer test requirements.

A typical wafer probe system incorporates a test system, wafer prober, RF probes, and DC probes. Figure 9.9 shows a photograph of a typical production wafer prober. This prober has cassette feed, auto alignment, and is configured for a test system "test head." The test head connects to the test interface, which mounts in the hole on the left side of the machine. This prober uses a ring-type probe card as

FIGURE 9.9 Production wafer prober for RF test.

shown in Figure 9.10. Conventional RF probes are mounted to the prober top plate using micro-manipulators arranged in quadrants. This allows access to each of the four sides of the integrated circuit. Figure 9.11 shows a two-port high-frequency setup capable of vector measurements. Wafer prober manufacturers offer different top plates for different probe applications. Specification of top plate configuration is necessary for new equipment purchases.

Probe calibration standards are necessary to de-imbed the probe from the measurement. Calibrated open, short, and load standards are required for vector measurements. Probe suppliers offer calibrated standards designed specifically for their probes. For scalar measurements or when using complex probe assemblies, alternative calibration standards can be used, but with reduced measurement accuracy. Alternative calibration standards may be a custom test structure printed on a ceramic substrate or on a wafer test structure. Scalar offsets can be applied for probe loss if you have a method of probe qualification prior to use. In general you have to decide if you are performing characterization or just a functionality screen of the device. This is important to consider early since measurement accuracy defines the appropriate probe technology, which places physical restrictions on the circuit layout.

When selecting the probe technology for any application you should consider the calibration approach, the maximum-usable frequency, the number of RF and DC connections required, the ability to support off-chip matching components, the cost of probes, and the cost of the calibration circuits. By understanding the advantages and limitations of each probe approach, an optimum technology/cost decision can be made. Remember that the prober top plate can be specified for ring frames or micro-manipulator type probes. Machine definition often dictates the types of probes to be used.

Traditional RF probes convert a coax transmission line into coplanar signal and ground probe points. This allows a coplanar or microstrip circuit with ground vias to be measured. These probes are offered as ground-signal and ground-signal-ground. They have been widely used for accurate high frequency measurements for many years. The G-S-G probe offers improved performance above 12 GHz and can be used up to 100 GHz with proper construction. Probe spacing from signal to ground is referred to as the pitch. A common probe pitch is 0.006 in. Due to the small size, material selection significantly impacts

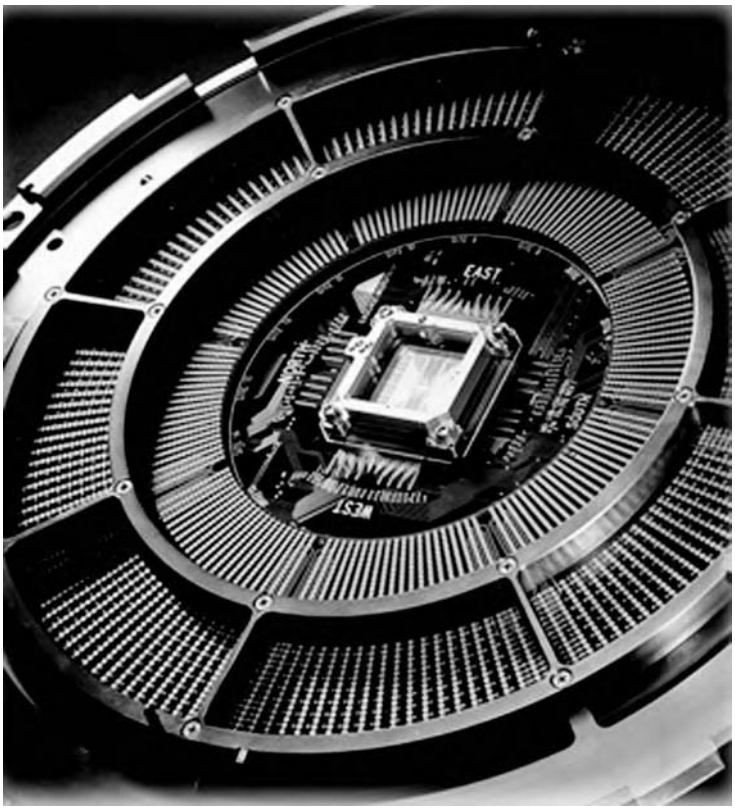


FIGURE 9.10 Ring-type RF probe card.

RF performance and physical robustness. Many companies including Cascade Microtech and PicoProbe specialize in RF probes.

Cost considerations of probes are important. RF probes or membranes can cost anywhere from $300 to $3,000 each. This adds up quickly when you need multiple probes per circuit, plus spares, plus calibration circuits. When possible it is recommended to standardize the RF probe pitch. This will minimize setup time and the amount of hardware that has to be purchased and maintained. When custom probes are to be used, be prepared to incur the cost of probe and the calibration circuit development.

Wafer level RF testing using coplanar probing techniques can easily be accomplished provided the constraints of the RF probe design are incorporated into the circuit layout. This usually requires more wafer area be used for the required probe patterns and ground vias. These are standard and preferred design criteria for high frequency devices requiring on-wafer test. Devices without ground vias may require alternative interface techniques such as custom probes or membrane probes.

Although typical RF circuits have two or three RF ports and several DC, there are many that require increased port counts. Advanced probing techniques have been developed to support the need for increased RF and DC ports as well as the need for near chip matching and bypass elements. Probe manufacturers have responded by producing custom RF/DC probe cards allowing multiple functions per circuit edge. Figure 9.12 is an example of a single side four-port RF probe connected to a calibration substrate. Probe manufacturers have also secured the ability to mount surface mount capacitors on the end of probe tips to provide close bypass elements.

Another approach is Cascade Microtech's Pyramid Probe. It is a patented membrane probe technology that offers impedance lines, high RF and DC port count, and close location of external components. Figure 9.13 shows the Pyramid Probe with an off-chip bypass capacitor. One important aspect of the

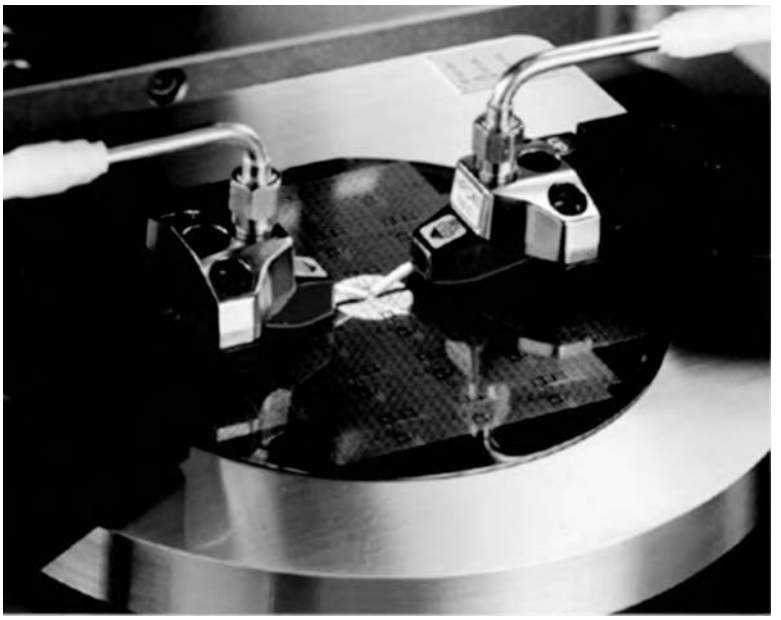


FIGURE 9.11 RF probes mounted on manipulators.

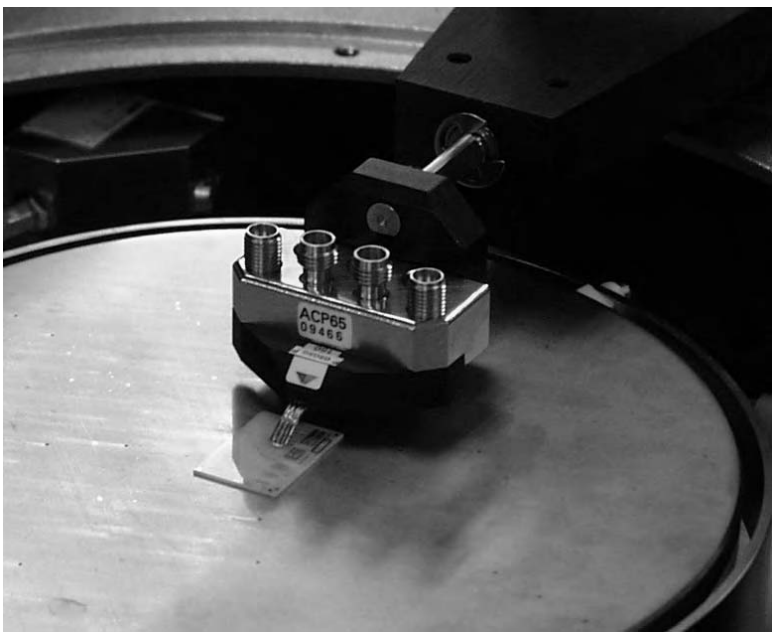


FIGURE 9.12 Four-part RF probe.

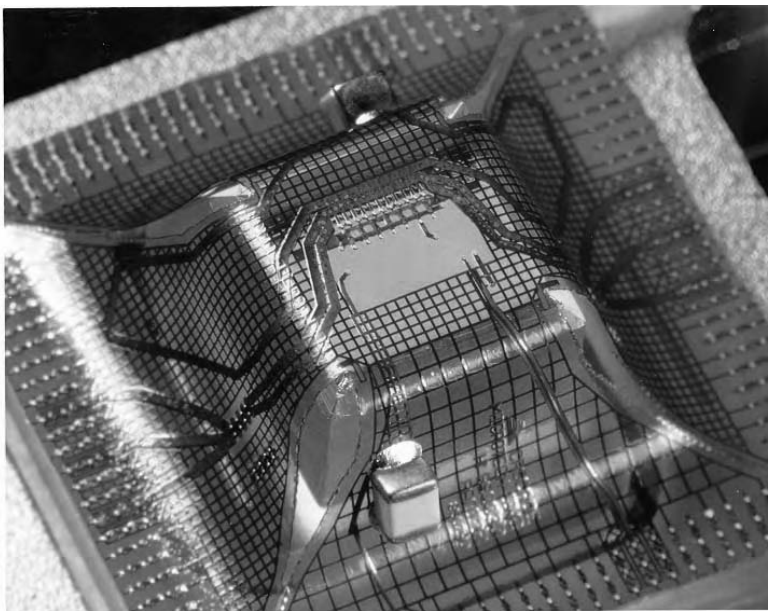


FIGURE 9.13 Cascade Microtech pyramid probe.

construction is that it incorporates an uninterrupted RF ground path throughout the membrane. This differs from the traditional coplanar probes that require the circuit to conduct the ground from one RF port to another. This allows for RF probing of lumped element circuits that do not utilize via holes and back side ground planes. This is becoming especially important to support developments such as chip scale packaging and multi-chip modules where the use of known good die is required for manufacturing.

For high volume devices where the circuit layout is optimized for the final package environment, considerations for on-wafer testing are secondary if not ignored. Products targeting the wireless market undergo aggressive die size reductions. Passive components such as capacitors, inductors, and resistors are often realized external to the integrated circuit. In this case the probes must be designed to simulate the packaged environment including the use of off chip components. Membrane technology is a good consideration for this. The membrane probe has the potential to emulate the package environment and external components that may be required at the final device level.

9.4 On-Wafer RF Test Benefits

The benefits of on-wafer RF testing are multiple and explain its success in the RF and microwave industry:

- Accuracy of RF test results with calibration performed at the probe tip, contact point to the DUT. The calibration techniques are now well established, supported by elaborate calibration standards, and easily implemented with software internally developed or purchased from the test equipment or probe vendors. This leads to accurate device models and higher first-pass design yields.
- Reproducibility of test results with stable impedance of the probe — be it 50 Ω or a custom impedance—and automatic probe-to-pad alignment performed by modern wafer probers. Set probe placement on the pads during test and calibration is critical, especially above 10 GHz and for DUTs presenting a narrowband match.
- Nondestructive test of the DUT, allowing shipment of RF KGD to the user. This ability is key for multi-chip module or flip-chip onboard applications. The correlation between on-wafer and

assembled device test results is excellent if the MMIC grounding is properly realized and the DC biasing networks are similar. For example, our experience producing 6 GHz power devices shows a maximum 0.2 dB difference in output power between wafer and module levels.

- Short cycle time for product test or statistical characterization and model extraction of library components, allowing for successful yield modeling and prediction.
- High throughput with complete automation of test and probing activities, and low cost, decreased by a factor of 10 in 10 years, to well below one dollar for a complex DUT today.

Wafer probing techniques are in fact gaining in importance today and are used for higher volume applications as chip size packages (CSP), and flip chip formats become more common, bypassing the traditional plastic packaging step and test handler. Another increasing usage of on-wafer test is for parts built in array formats such as multi-chip modules or ball grid arrays. For these applications, robust probes are needed to overcome the low planarity of laminate boards. Higher speed test equipment such as that used with automatic handlers is likely to become more prevalent in wafer level test to meet volume needs. The probing process must now be designed to form a continuous flow, including assembly, test, separation, sorting, and packaging.

References

1. Strid, E.W., 26 GHz Wafer Probing for MMIC Development and Manufacture, *Microwave Journal,* Aug. 1986.
2. Strid, E.W., On-Wafer Measurements with the HP 8510 Network Analyzer and Cascade Microtech Wafer Probes, RF & Microwave Measurement Symposium and Exhibition, 1987.
3. Cascade Microtech Application Note, On-Wafer Vector Network Analyzer Calibration and Measurements (www.cascademicrotech.com).
4. Cascade Microtech Technical Brief TECHBRIEF4-0694, A Guide to Better Network Analyzer Calibrations for Probe-Tip Measurements. (www.cascademicrotech.com)
5. Mahon, J.R. et al., On-Wafer Pulse Power Testing, ARFTG Conference, May 1990.
6. Poulin, D.D. et al., A High Power On-Wafer Pulsed Active Load Pull System, *IEEE Trans. Microwave Theory and Tech.,* MTT-40, 2412–2417, Dec. 1992.
7. Dambrine, G. et al., A New Method for Determining the FET Small Signal Equivalent Circuit, *IEEE Trans. Microwave Theory and Tech.,* MTT-36, 1151–1159, July 1988.
8. Staudinger, J., Network Analyzer Calibration, *CRC Modern Microwave and RF Handbook,* CRC Press, Boca Raton, FL, chap. 4.2, 2000.

10

High Volume Microwave Test

Jean-Pierre Lanteri
Christopher Jones
John R. Mahon
M/A COM TycoElectronics

10.1 High Volume Microwave Component Needs

10.1.1 Cellular Phone Market Impact

High volume microwave test has emerged in the early 1990s to support the growing demand for GaAs RFICs used in cellular phones. Prior to that date, most microwave and RF applications were military and only required 10,000s of pieces a year of a certain MMIC type, easily probed or tested by hand in mechanical fixtures. For most companies in this industry, the turning point for high volume was around 1995 when some RFIC parts for wireless telephony passed the million per year mark. Cellular phones have grown to over 300 million units shipped in 1999 and represent 80% of the volume of microwave and RF ICs manufactured, driving the industry and its technology.

The cellular phone needs in terms of volume, test cost, and acceptable defect rate demanded new test solutions (Table 10.1) be developed that relied on the following key elements:

1. "Low" frequency ICs, first around 900 MHz and later on around 1.8–2.4 GHz, with limited bandwidth, allowing simpler device interfaces and fewer test points over frequency. Previously, MMICs were mostly military T/R module functions with frequencies ranging from 2 to 18 GHz, with 30% or more bandwidths. They were tested at hundreds of frequencies, requiring specialized fast ramping Automatic Network Analyzers (ANA) such as Agilent's HP8510 or HP8530.

TABLE 10.1 Microwave and RF IC Test Needs Evolution

Year	Product	Application	Package	Price	Volume	Test Time	Test Cost	Escape Rate
1991	T/R Module	Radar	Carrier	$200	10K/Y	1 min	$30	1%
1993	T/R Switch	Radar/Com	Ceramic	$40	100K/Y	30 sec	$4	0.5%
1995	RF Switch	Com	Plastic	$10	Mil/Y	10 sec	$1	0.1%
1997	RF MMIC	Com	SOIC	$3	Mil/M	3 sec	$0.30	0.05%
1999	RF MMIC	Com	SOT	$1	Mil/W	1 sec	$0.10	0.01%

2. Standard plastic packages, based upon injection molding around a copper lead frame, to reach the low cost required in product assembly and test. Most early RFICs used large gull wing dual in-line packages (DIP), then small outline IC packages (SOIC), later small outline transistor packages (SOT), and today's micro leadframe flatpack (MLF).
3. Automatic handlers from the digital world, typically gravity fed, leveraging the plastic packages for full automation and avoiding human errors in bin selection. Previous metal or ceramic packages were mostly custom, bulky, and could only be handled automatically by pick-and-place type handlers, such as the one made by Intercontinental Devices in the early 1990s, barely reaching throughputs of a few hundred parts per hour.
4. Highly repeatable, accurate, and durable device contact interface and test board, creating the proper impedance environment for the device while allowing mechanized handling of the part. Most products before that were designed as matched to 50 Ohm impedance in and out, where cellular phone products will most often need to be matched in the user's system, and therefore on the test board. Adding to the difficulty, many handlers converted from digital applications hold the part in the test socket with a bulky mechanical clamp that creates ground discontinuities in the test board and spread the matching components further apart than designed in the part application.
5. Faster ANA test equipment through hardware and software advances, later supplanted by specialized RFIC testers. The very high volumes reached by some parts, over a million pieces a week, allow dedication of a customized system to their testing to reduce measurement time and cost. Therefore the optimum test equipment first evolved from a powerful ANA-based system (HP8510, for example) with noise figure meter, spectrum analyzer, and multiport RF switch matrix, to an ad hoc set of bench-top equipment around an integrated ANA or ANA/spectrum analyzer. Next appeared products inspired from the digital world concept of the "electronic pin" tester, with RF functionality at multiple ports, such as the HP84000, widely used today.
6. Large databases on networked workstations and PCs for test results collection and analysis. The value of the information does not reside in the pass or fail outcome of a specific part, but in the statistical trends and operational performance measures available to company management. They provide feedback on employee training, equipment and calibration reproducibility, equipment maintenance schedules, handler supplier selection, and packaging supplier tolerances to name a few.

Although the high volume techniques described in this chapter would apply to most microwave and RF components, they are best fitted for products that do not require a broadband matched environment and that are packaged in a form that can be automatically tested in high-speed handlers.

10.1.2 High Volume RF Component Functions and Test Specifications

We will focus in this section on the different functions in the RF front end of a wireless phone to illustrate the typical products tested, their function, specification, and performance. The generic building blocks of a RF front end (Figure 10.1) are switches (for antenna, transmit/receive (T/R), or band selection), input low noise amplifiers (LNA), output power amplifiers (PA), up- and downconverters (typically comprising a mixer), local oscillator amplifier (LOA), and intermediate frequency amplifier (IFA). In

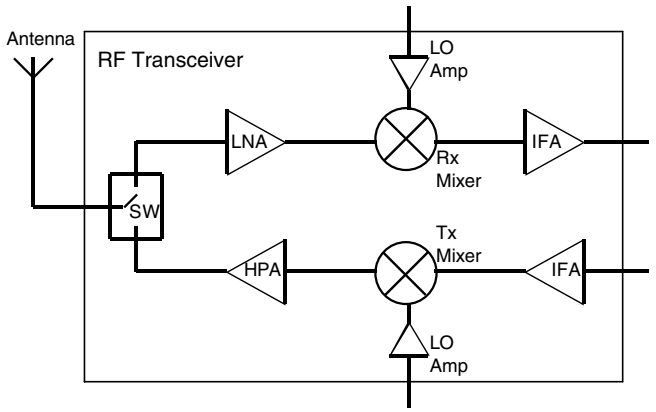


FIGURE 10.1 Typical RF transceiver building blocks.

TABLE 10.2 Typical Product Specifications for High Volume Test

Switch Parameters	Min	Max	LNA Parameters	Min	Max
Frequency Range	800 MHz	1000 MHz	Frequency Range	800 MHz	1000 MHz
Control Leakage	–10 μA	10 μA	Current Consumption	8 mA	12 mA
Insertion Loss		0.5 dB	Linear Gain	15 dB	18 dB
Isolation	25 dB		Noise Figure		2 dB
Input IP3	60 dBm		Input IP3	–4 dBm	
PA Parameters	**Min**	**Max**	**Mixer Parameters**	**Min**	**Max**
Frequency Range	800 MHz	1000 MHz	Frequency Range	800 MHz	1000 MHz
Linear Current	160 mA	200 mA	IF Frequency Range	DC	100 MHz
Linear Gain	27 dB	35 dB	Conversion Loss		7.5 dB
Pout @ Pin = –1 dBm	25 dBm	30 dBm	LO to RF Leakage	38 dB	
Current @ Pin = –1 dBm		300 mA	1 dB Compression	21 dBm	
1 dB Compression	22.5 dBm		IMD @ Pin = –10 dBm	65 dBc	

most cases, these products are single band, either cellular or PCs, although new dual band components are appearing, requiring two similar tests in sequence, one for each band.

The test equipment should therefore be capable of measuring DC parameters, network parameters such as gain or isolation, and spectral parameters such as IMD for most high volume products. Noise figure is required for LNAs and downconverters, and output power for HPAs. Typically, two types of RFIC testers will handle most parts, a general purpose RFIC for converters and eventually switches, and a specialized one for HPAs.

Typical specifications for the various parts are provided below. No specification is very demanding on the test instrument in absolute terms, but the narrow range of acceptance for each one requires outstanding reproducibility of the measurements, part after part. This will be the limiting factor in escape rate in our experience.

These specifications are dictated by the application and therefore largely independent of the technology used for fabrication of the RFIC. RFIC technology was predominantly GaAs metal semiconductor field effect transistor (MESFET) until 1997, when GaAs heterojunction bipolar transistor (HBT) appeared, soon followed by silicon products, in BiCMOS, SiGe BiCMOS, and CMOS technologies. The RF test is performed in a similar fashion for all implementation technologies of a given functionality.

10.1.3 High Volume Test Success Factors

The next sections will review in detail aspects of a successful back-end production of typical RF high volume parts; inexpensive, not too complex, packaged in plastic, produced at the rate of a million per week. The basic requirements addressed are:

- Test equipment selection, balancing highest test speed with lowest test cost for the product mix
- Automatic package handler keeping pace with the tester through parallel handling, and highly reliable
- Part contactor performing at the required frequency, lasting for many contacts
- Test software for fast set up of a new part with automatic revision control

Less obvious but key points for cost-effective high volume production are also discussed:

- Tester, contactor, and test board calibration approach for reproducible measurements
- Cost factors in a high volume test operation
- Data analysis capabilities for relating yield to design or process
- Test process monitoring tools, to ascertain the performance of the test operation itself

10.2 Test System Overview

10.2.1 Hardware: Rack and Stack versus High Speed IC Testers

Hardware considerations are based on the measurement requirements of your product set. To evaluate this, the necessary test dimensions should be determined. These dimensions can include but are not limited to swept frequency, swept spectrum, modulation schemes, swept power, and DC stimulus.

Commercially available hardware instruments can be combined to perform most RF/DC measurement requirements for manufacturing applications. These systems better known as "Rack and Stack" along with widely available third party instrument control software can provide a quick, coarse start-up for measurement and data collection, ideally suited for engineering evaluation. As the measurements become more integrated, the complexity required may exceed the generic capabilities of the third party software and may have to be supplemented with external software that can turn the original software into nothing more than a cosmetic interface.

To take the "Rack and Stack" system to a higher level requires a software expertise in test hardware communication and knowledge of the optimum sequencing of measurement events. Most hardware in a rack and stack system provides one dimension of competence, for example a network analyzer's optimum performance is achieved during a swept frequency measurement, a spectrum analyzer is optimized for frequency spectrum sweeps with fixed stimulus. Taking these instruments to a different dimension or repeating numerous cycles within their optimum dimension may not provide the speed required. Some instruments do provide multiple dimensions of measurement, but usually there is a setup or state change required that can add to the individual die test time. Another often-ignored aspect of these types of instruments is the overhead of the display processor, which is important in an engineering environment but an unnecessary time consumer in a manufacturing environment.

Commercially available high volume test systems usually provide equivalent speed in all dimensions by integrating one or two receivers with independently controlled stimulus hardware, unlike a network analyzer where the stimulus is usually linked to the receiver. These high-speed receivers combined with independently controlled downconverters, for IF processing, perform all the RF measurements that normally would take multiple instruments in a rack and stack system. Since these receivers are plug-in modules, whether for a PC back plane or a controlling chassis like a VXI card cage, they are also optimized for fast I/O performance and do not require a display processor, which can significantly impact the measurement speed. And since these receivers are usually based on DSP technology, complex modulation measurements such as ACPR can easily be made without additional hardware as would be required in most rack and stack systems.

TABLE 10.3 Speed Comparison of Rack and Stack and High Speed IC Tester

		Rack and Stack		High Speed IC	
Repeat Count	Measurement/Stimulus	Each (ms)	Total (ms)	Each (ms)	Total (ms)
3 Times	Set RF Source #1 Stimulus	100	300	50	150
3 Times	Set RF Source #2 Stimulus	100	300	50	150
12 Times	Set Analyzer to Span	250	3000	50	600
12 Times	Acquire Output Signal	50	600	40	480
	Total Time		4200		1380

In a normal measurement sequence of any complex device, the setting of individual stimulus far exceeds the time required to acquire the resulting output. A simple example of this would be a spectrum analyzer combined in a system with two synthesized sources to perform an intermodulation measurement at three RF frequencies. Accomplishing this requires extensive setting before any measurements can be made. Table 10.3 shows the measurement sequence and the corresponding times derived from a rack and stack system and a commercially available high-speed IC measurement system for comparison. The measurement repeatability of these systems is equivalent for this example, therefore the bandwidth of the instrument setting is comparable.

As shown in the table, the acquisition of the output signal shows relatively no speed improvement with a difference of only 120 mS total. The most significant improvement is the setting of the acquisition span on the high-speed IC tester. This speed is the same as the setting of a RF stimulus since the only overhead is the setting of the LO source required for the measurement downconversion. The only optimization that could be performed with the rack and stack system would be higher speed RF sources having internal frequency list and power leveling capability. The change in span setting on a standard spectrum analyzer will always be a speed inhibitor since it is not its optimum dimension of performance.

From this type of table a point can be determined where the cost of a high-speed IC tester outweighs the speed increase it will yield. This criteria is based on complex multifunction devices that require frequent dimension hopping as described above. Other component types, such as filters requiring only broadband frequency sweeps in a single dimension, would show less speed improvement with an increase in frequency points since network analyzers are optimized for this measurement type.

Various vendors for high-speed systems exist. Agilent Technologies (formerly Hewlett Packard), Roos Instruments, LTX, and Teradyne are just a few of the more well-known suppliers. The full system prices can range from a few hundred thousand dollars to well into the millions depending on the complexity/customization required.

A note of caution when purchasing a high speed IC tester: careful homework is warranted. Most IC testers are a three- to five-year commitment of capital dollars, and the one purchased should meet current and future product requirements. Close attention to measurement capabilities, hardware resources, available RF ports, DC pin count, and compatibility to existing test boards will avoid future upgrades, which are usually costly and delay time to market for new products if the required measurement capability is not immediately available.

10.2.2 System Software Integration

Software capabilities of third party systems require close examination, especially if it is necessary to integrate the outputs with existing resources on the manufacturing floor. Most high-speed IC testers focus on providing a test solution not a manufacturing solution. Network integration, software or test plan control, and data file organization is usually taken care of by the end customer. This software usually provides little operator input error checking or file name redundancy checking when dealing with multiple systems. The output file structure should have all the information required available in the file. Most third party systems provide an ASCII file output, which supports STDF (Standard Test Data Format), an industry standard data format invented by Teradyne. As with the hardware, the software is fixed at a revision level. It is important to suggest improvements to the vendors to make the system more

effective. Software revisions introduced by the vendor may not be available as fast as expected to correct observed deficiencies. It is still valuable to use the current revision level of the software to avoid known bugs and receive the best technical support.

10.2.3 RFIC Test Handlers

The primary function of the test handler is to move parts to the test site and then to sort them based on the test result. Package style and interface requirements will define what machines are available for consideration. The product will define the package and the handler is typically defined by the package. Common approaches include tube input—gravity handling, tray input—pick and place handling, and bulk input—turret handling. During the product design phase, selection of a package that works well with automation is highly recommended. The interface requirements are extremely critical for RF devices. Contact inductance, off chip matching components, and high frequency challenge our ability to realize true performance. The best approach is a vacuum pick up and plunge. This allows optimal RF circuit layout and continuous RF ground beneath the part.

Common test handler types and suppliers are listed in Table 10.4. Various options can be added to support production needs such as laser marking, vision inspection, and tape and reel. For specialized high volume applications, handlers are configured to accept lead frame input and tape and reel output providing complete reel-to-reel processing. When evaluating handlers for purchase, some extra time to identify process needs is very valuable. The machine should be configured for today's needs with the flexibility to address tomorrow's requirements. Base price, index time, jam rate, hard vs. soft tooling, conversion cost, tolerance to multiple package vendors, and vendor service should be considered. One additional quantitative rating is design elegance. An elegant design typically has the fewest transitions and fewest moving parts. Be cautious of machines that have afterthought solutions to hide their inherent limitations.

10.2.4 Contact Interface and Test Board

The test interface is comprised of a contactor and test board. The contactor provides compliance and surface penetration ensuring a low resistance connection is made to all device ports. Figure 10.2 shows a sectioned view of a pogo pin contactor. For RF applications the ideal contactor has zero electrical length and coupling capacitance. In the real-world contactors typically have 1–2 nH of series inductance and 0.2 to 0.4 pF of coupling capacitance. This can have significant impacts on electrical performance. Refer to Table 10.5 for a review of contactor manufacturers and parasitics. A more in-depth review of some available contactor approaches and suppliers is given in an article by Robert Crowley [1]. Parasitics of contactors can typically be compensated for in series ports using filter networks. Shunt ports however, such as an amplifier ground reference, challenge the use of contactors because the electrical length cannot be removed. The additional electrical length often shifts performance in magnitude or frequency beyond the range where scalar offsets can be used.

Fine pitch packaging has increased the challenges associated with contactor manufacturing and lifetime. Packages such as TSSOP, SOT, SC70, and the new micro leadframe flatpack (MLF) have pitches as

TABLE 10.4 Test Handler Manufacturers and Type

Manufacturer	Pick and Place	Gravity	Turret
Aetrium		X	X
Asseco	X	X	
Delta Daymark	X	X	
Exatron	X	X	
Intercontinental Microwave	X		
Ismeca	X		X
MultiTest	X	X	
Roos		X	

TABLE 10.5 Test Contactor Manufacturers and Type

Manufacturer	Approach	Self Inductance	Mutual Inductance	Capacitance
Agilent	"YieldPro"	0.3 nH		0.17 pF
Aries	Microstrip Contact	0.01 pF	0.05 nH	0.04 pF
Exatron	Particle Interconnect	0.26 nH		0.024 pF
Johnstech International	"S" Contact	1.0 nH	0.2 nH	0.07 pF
Oz Tek	Pogo Pin	2.4 nH	0.4 nH	0.09 pF
Prime Yield	"Surface Mount Matrix"			
Synergetix	Pogo Pin	1.3 nH	0.1 nH	0.1 pF
Tecknit	"Fuzz Button"	2.7 nH	0.3 pF	0.3 pF

Note: Values supplied are typical values from manufacturer's catalog. Refer to manufacturer for specific information to support your specific needs.

FIGURE 10.2 Pogo pin contactor.

small as 0.020 in. and may require a back-side ground connection. As contactor element size is reduced to support fine pitch packages, sacrifices are made in compliance and lifetime.

High-frequency contactors are typically custom machined and assembled making them expensive. Suppliers are quoting $1000 to $4000 for a single contactor. If this expense is amortized over 500,000 parts, the cost per insertion is about one-half cent. This may be acceptable for some high value added part, but certainly not for all RF parts in general. Add to this the need to support your product mix and the need for spares, and you will find that contactors can be more expensive than your capital test equipment. There is a true need for an industry solution to provide an affordable contactor with low parasitics, adequate compliance, and tolerance to tin-lead buildup.

The second half of the test interface is the test board, which interfaces the contactor to the test system. The test board can provide a simple circuit routing function or a matching circuit. It is common for RF circuits to utilize off-chip components for any non-active function. The production test board often requires tuning to compensate for contactor parasitics. This can result in a high Q-matching circuit that increases measurement variability due to the interaction between the part, the contactor, and the test board. It is recommended to consider the contactor and test board during the product design cycle allowing the configuration to be optimized for robust performance.

10.3 High Volume Test Challenges

10.3.1 Required Infrastructure

The recommended facility for test of RF semiconductor components is a class 100,000 clean room with full ESD protection. RF circuits, especially Gallium Arsenide, are ESD sensitive to as little as 100 volts. Although silicon tends to be more robust than Gallium Arsenide, the same precautions should be taken. The temperature and humidity control aids test equipment stable operation and helps prolong the life of other automated equipment. Facility requirements include HVAC, lights, pressurized air and nitrogen, vacuum, various electrical resources, and network lines.

As volume increases the information system becomes a critical part of running the operation. The ideal system aids the decision process, communicates instructions, monitors inventory, tracks work in process, and measures equipment and product performance. The importance of information automation and integration cannot be overemphasized. It takes vision, skill, and corporate support to integrate all technical, manufacturing, and business systems.

The human resources are the backbone of any high volume operation. Almost any piece of equipment or software solution can be purchased, but it takes a talented core team to assemble a competitive operation and keep it running. Strengths are required in operations, software, and test systems, products, data analysis, and automation.

10.3.2 Accuracy and Repeatability Challenges

Measurement accuracy and repeatability are significant challenges for most high volume RF measurements. All elements of the setup may contribute to measurement inaccuracies and variability. The primary considerations are test system, the test board, the contactor, and the test environment.

For this discussion we will assume that all production setups are qualified for accuracy. This allows us to focus this discussion on variability.

10.3.2.1 Measuring Variability

Gauge Repeatability and Reproducibility (Gauge R&R) measurements can be used to measure variability. In this context the measurement system is referred to as the gauge. The gauge measurement is a structured approach that measures "x" products, "y" times, on "z" machines allowing the calculation of "machine" variability. Variability is reported in terms of repeatability and reproducibility. Repeatability describes variability within a given setup such as variability of contact resistance in one test lot. Reproducibility describes the variability between setups such as between different test systems or on different days. An overview of gauge measurement theory and calculations can be found in any statistical textbook [2].

Figure 10.3 summarizes the sources of measurement variability within an automated test setup. The three locations are identified to allow easy gauge measurements.

Table 10.6 qualitatively rates the sources of measurement variability for repeatability and reproducibility. We can see that the system calibration and test board variations are large between setups while the contactor variations are large within a given setup. We will use these relationships in the case study to follow.

Variability is expressed in terms of standard deviation. This allows normalized calculations to be made. For example, the variability of any measurement is a combination of the variability of the product and the gauge. This can be expressed as:

$$\sigma^2 \text{ measured} = \sigma^2 \text{ product} + \sigma^2 \text{ gauge}$$

Based on Figure 10.3 the total variability of an automated test can be described as:

$$\sigma^2 \text{ total} = \sigma^2 \text{ product} + \sigma^2 \text{ system} + \sigma^2 \text{ board} + \sigma^2 \text{ contact}$$

And for any expression of variability we can distinguish between repeatability and reproducibility as:

$$\sigma^2 \text{ gauge} = \sigma^2 \text{ repeatability} + \sigma^2 \text{ reproducibility}$$

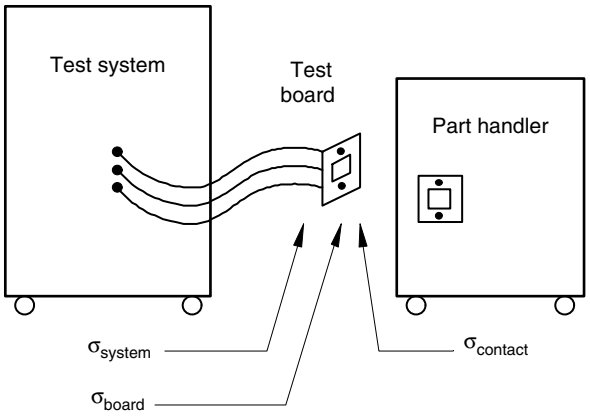


FIGURE 10.3 Sources of variability in an automated test setup.

TABLE 10.6 Repeatability and Reproducibility Comparison for the Complete Test Environment

Source	Description	Repeatability (within a setup)	Reproducibility (between setups)
Test System	Calibration	Low	High
Test Board	Matching Circuit	Low	High
Contactor	Contact Resistance	High	Low

TABLE 10.7 Gauge Test Design

"Machine"	# Machines	"Product"	# Products	# "Measurements"
Test System	4	Part soldered to test board	3	3
Test Board	4	Loose parts	3	3
Handler Contact	3	Loose parts	10	3

Table 10.7 recommends a gauge test design to characterize the components shown in Figure 10.3. In this design we are measuring the "Machine" variation using "Products" and repetitive "Measurements." In all cases, stable product fixturing techniques are required for measurement accuracy. For the handler contact measurement, a single test setup is recommended.

10.3.2.2 Case Study

A low yielding product has been identified. Feedback from the test floor suggests the yield depends on the machine used and the day it was tested. These are the signs that yield is largely affected by variability. The following presents an analytical process that identifies the variability that must be addressed to improve yields.

Step 1: Identify the Failure Mode—For this product we found one gain measurement to be more sensitive than others. In fact this single parameter was driving the final yield result. This simplifies the analysis allowing us to focus on one parameter.

Step 2: Quantify Measurement and Product Variability—A query of the test database showed 1086 production lots tested over a four-month span. For each production lot the average gain and standard deviation was reported. We define a typical gain value by taking the average of all production lot averages. Repeatability, or variability within any given test, was defined by finding the average of all production lot standard deviations. Reproducibility, or variability between tests, was found by taking the standard

deviation of the average gain values for all production lots. Gauge R&R testing was conducted to determine the repeatability and reproducibility of the "system," "board," and "contact" as described previously. This allows calculation of product variability as shown in Table 10.8.

Step 3: Relate Variability to Yield—Relating variability to yield will define the product's sensitivity to the measurement. This will allow us to focus our efforts efficiently to maximize yield while supporting the customers' requirements. We can calculate yield to each spec limit using Microsoft Excel's NORMDIST function as follows:

Percent below upper spec limit = Y(USL) = NORMDIST(USL, μ, σ, 1)
Percent above lower spec limit = Y(LSL) = 1 − NORMDIST(LSL, μ, σ, 1)

And we can calculate the final yield as follows:

Yield = Y(USL) − (1 − Y(LSL))

Prior to calculating yield we need to make some assumptions of how repeatability and reproducibility should be treated. For this analysis it is assumed that repeatability values will be applied to the standard deviation and reproducibility values will be used to shift the mean. Yield will be calculated assuming a worst case shift of the mean by one, two, and three standard deviations. The result will be plotted as Yield vs. Standard Deviation. The plot can be interpreted as the sensitivity of the parameter yield versus the measured variability of the test setup. This result is shown in Figure 10.4 using the data in Table 10.8, the USL = 26.5 dB, the LSL = 21.5 dB, and the Average Gain = 23.1 dB.

Figure 10.4 quickly communicates the severity of the situation and identifies the test board as the most significant contributor. Looking at the product by itself we see that its yield can vary between 90% and 43%. Adding the test system variability makes matters worse. Adding the test board shows the entire process is not capable of supporting the specification. There are three solutions to this problem. Change the specifications, reduce the variability, or control the variability. Changing the specification requires significant customer involvement and communication. From the customer's point of view, if a product is released to production, specification changes are risky and avoided unless threat of line shutdown is evident. Reducing variability is where your effort needs to be focused. This may require new techniques and technology to achieve. In the process of reducing variability lessons learned can be applied across all products resulting in increased general expertise that can support existing and future products. The last method that can be applied immediately is to control variability. This is a process of tightly measuring and approving your measurement hardware from test systems to surface mount components. Everything gets qualified prior to use. This may take significant logistics efforts to put in place, but the yield improvements can be substantial.

This study is of an extreme case. To communicate this issue in a generic sense we can compare the same product case for various Cpk values. Figure 10.5 displays total variability vs. Cpk values of 0.5, 1.0, and 1.5. We see that the case study shape is similar to the Cpk = 0.5 curve with a mean offset. It also shows that the process can be supported by a Cpk = 1.5 or greater. Anything less requires control of variability.

TABLE 10.8 Measurement and Product Variability

Data Source	Variability	Repeatability (one setup)	Reproducibility (across setups)	Total
Production Data	Total	1.6 dB	1.0 dB	1.89 dB
Gauge R&R	System	0.09 dB	0.23 dB	0.25 dB
	Board	0.13 dB	0.75 dB	0.76 dB
	Contact	0.54 dB	0.00 dB	0.54 dB
Calculation	Product	1.50 dB	0.62 dB	1.62 dB

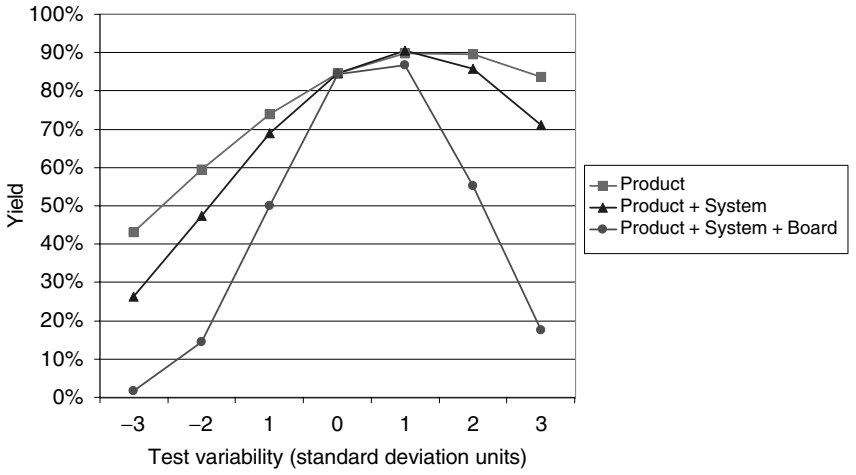


FIGURE 10.4 Yield versus variability for test system elements.

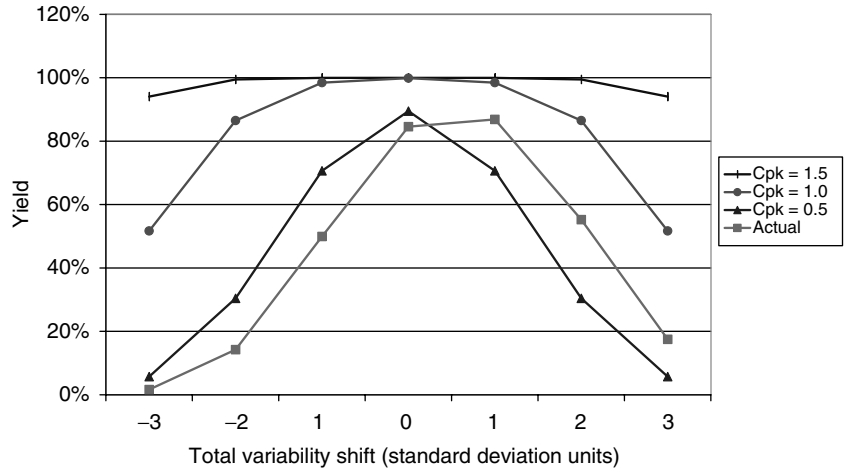


FIGURE 10.5 Yield versus variability as function of Cpk.

10.3.3 Volume and Cost Relationship

In general, cost of test reduces with increasing volume. Your ability to model available capacity will allow accurate estimation of cost. A generic capacity equation is:

$$Capacity = \frac{\left(Time\ Available\right)\left(Efficiency\right)}{Test\ Time + Handling\ Time} \tag{10.1}$$

Time available can be a day, month, or year as long as all time units are consistent. Efficiency is a measure of productive machine time. Efficiency accounts for all downtime within the time available due to equipment calibration, handler jams, material tracking operations, or anything else. For time intervals greater than a week you will find that efficiency converges. A typical range for initial estimates is 60% to 70%. Focus or lack of focus can swing the initial range by ±20%.

Cost of testing can be calculated using the estimated capacity and costs or with the actual cost and volume. The baseline result is shown below.

$$Unit\ Cost = \frac{Cost}{Volume} = \frac{Facility + Equipment + Labor + Materials}{(Capacity)(Yield)} \tag{10.2}$$

Example Cost of Test*:* A complex part enters production. A $650,000 test system and a $350,000 handler are required and have been purchased. The estimated test and handling times are both one-half second. Based on Equation 10.1 we can solve for the monthly capacity for varying efficiencies. This is shown in Table 10.9 for an average of 600 hours available per month.

TABLE 10.9 Efficiency versus Capacity

Efficiency	Capacity
40%	864,000
60%	1,296,000
80%	1,728,000
100%	2,160,000

We can see from Table 10.9 that there is a wide range of possible outcomes for capacity. In fact this is a very realistic result. If the objective was to install a monthly capacity of 1,600,000 parts, then the efficiency of operation defines if one or two systems are required. For this case an average of 74% efficiency will be required to support the job. Successful implementation requires consideration of machine design, vendor support, and operation skill sets to support 74% efficiency. If the efficiency cannot be met, then two systems need to be purchased.

Efficiency has little impact on the cost of test unless the volume is increased. This can be shown by expanding our example to calculate cost. We will assume fixed facilities and capital costs; variable labor and material costs; and 100% yield to calculate the cost per test insertion. The assumptions are summarized in Table 10.10.

Cost per insertion calculations are shown in Table 10.11 for varying volume and efficiencies.

Columns compare the cost per insertion to the volume of test. The improvements in cost are due to amortizing facility and capital costs across more parts. The impact is significant due to the high capital cost of the test system and handler. Rows compare the cost per insertion as compared to efficiency. The difference in cost is relatively low since the only savings are labor. For this dedicated equipment example, improving efficiency only has value if the capacity is needed. Given efficiency or capacity, the cost of test can be reduced by increasing volume through product mix.

10.3.4 Product Mix Impact

Product mix adds several challenges such as tooling costs and manufacturing setup time. Tooling costs include test boards, mounting hardware, product standards, documentation, and training. These costs can run as high as $10,000 or as low as the documentation depending on product similarity and your approach to standardization. Tooling complexity will ultimately govern your product mix through resource limitations. Production output, on the other hand, will be governed by setup time. Setup time is the time to break down a setup and configure for another part number. This can involve test system calibration, test board change and/or handler change. Typical setup time can take from ten minutes to four hours. The following example explores product mix, setup time, and volume.

Example: Setup Time—Assume that setup can vary between ten minutes and four hours, equal volumes of four products are needed, test plus handing time is 1.0 s, and the efficiency is 60%. Calculate the

TABLE 10.10 Cost Assumptions

Cost	Assumption	Fixed or Volume Dependent
Facility	$ per square foot of floor space	Fixed
Capital	3 year linear depreciation	Fixed
Labor	Labor and fringe	Volume Dependent
Materials	General Consumables	Volume Dependent
Yield	Not used	

TABLE 10.11 Cost vs. Volume vs. Efficiency

Efficiency/Volume	100%	90%	80%	70%	60%
400,000	$0.096	$0.097	$0.098	$0.100	$0.101
800,000	$0.053	$0.054	$0.055	$0.056	$0.058
1,200,000	$0.038	$0.039	$0.040	$0.042	$0.043
1,600,000	$0.031	$0.032	$0.033	N/A	N/A
2,000,000	$0.027	N/A	N/A	N/A	N/A

TABLE 10.12 Monthly Capacity of Four Products with Varying Setup Time and Delivery Intervals

Setup/Delivery	10 min.	30 min.	1 hour	2 hours	4 hours
Monthly	1,294,531	1,291,680	1,287,360	1,278,720	1,261,440
Weekly	1,290,125	1,278,720	1,261,440	1,226,880	1,157,760
Daily	1,251,936	1,166,400	1,036,800	777,600	259,200

optimum output assuring deliveries are required at monthly, weekly, or daily intervals. To do this we subtract four setup periods from the delivery interval, calculate the test capacity of the remainder of the interval, and then normalize to one-month output. Table 10.12 summarizes the results.

As you may have expected, long setup times and regular delivery schedules can significantly reduce capacity. When faced with a high-mix environment everything needs to be standardized from fixturing to calibration files to equipment types and operating procedures.

10.4 Data Analysis Overview

10.4.1 Product Data Requirements and Database

Tested parameters for average RF devices can range from as little as 3 to as many as 30 depending on the functional complexity. In a high volume environment, where output can reach over 500,000 devices daily with a moderate product mix, methods to monitor and evaluate performance criteria have to provide efficient access to data sets with minimal user interaction. Questions such as "How high is the yield?" and "What RF parameters are failing most?" are important in any test facility, but can be very difficult to monitor and answer as volumes grow.

Many arguments have been made concerning the necessity of collecting parameter information on high yielding devices. To answer the two questions asked above, only limited information need be gathered. Most testers are capable of creating bin summary reports that can assign a bin number to a failure mechanism and output final counts to summarize the results.

The "binning" method may yield enough information for many circuits, but will not give insight into questions about tightness of parameter distributions, insufficient (or over-sufficient) amount of testing, test limits to change to optimize the yield, or possible change in part performance. These can only be answered with full data analysis packages either supplied by third parties or developed in-house. Standard histogram (Figure 10.6) or wafer maps (Figure 10.7) can answer the first question by providing distributions, standard deviation, average values, and when supplied with limit specifications, CP and CPK values. XY or correlation plots (Figure 10.8) can answer the second question, but when dealing with 20 or so parameters, this can be very time consuming to monitor.

The last questions require tools focusing on multivariable correlation and historical analysis. Changing of limit specifications to optimize yield is a tricky process and should not be performed on a small sample base. Nor should the interdependency of multiple parameters be ignored. Control charts such as Box Plots (Figure 10.9) are ideal tools for monitoring performance variations over time.

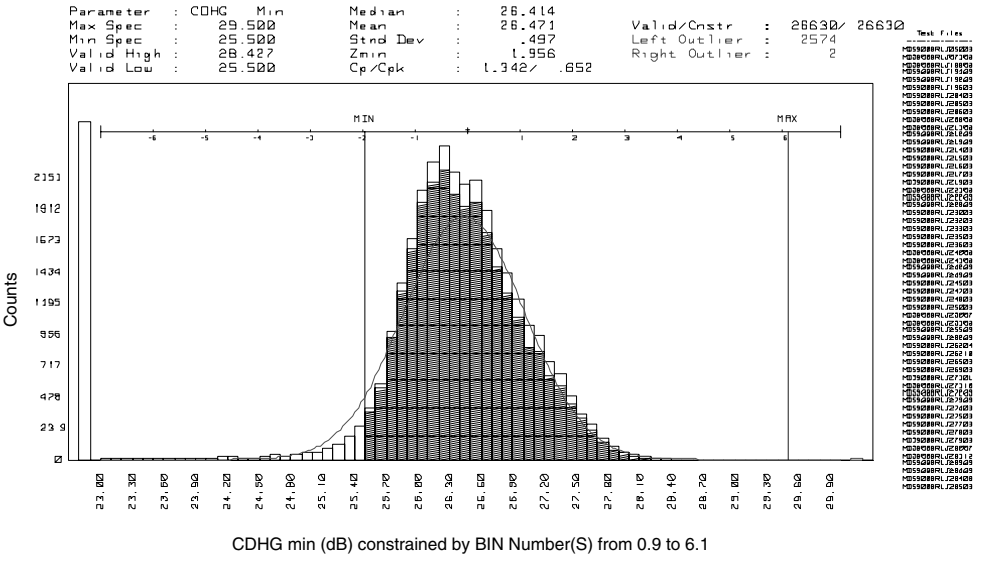


FIGURE 10.6 Histogram for distribution analysis.

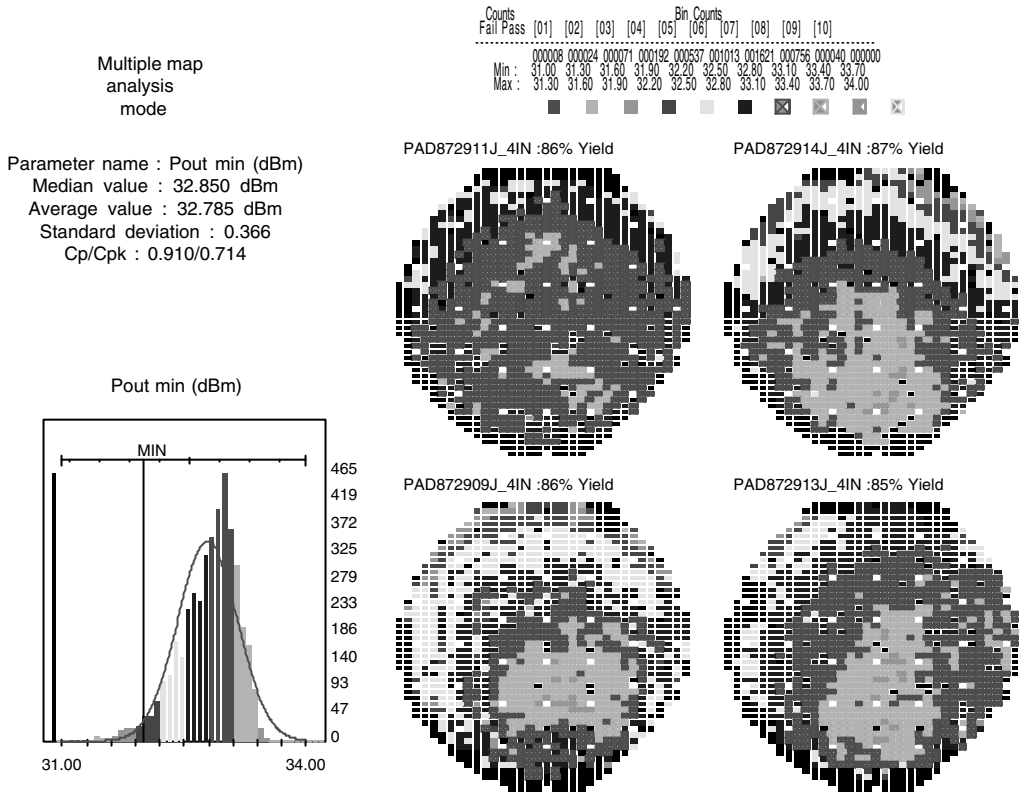


FIGURE 10.7 Wafer maps for yield pattern analysis.

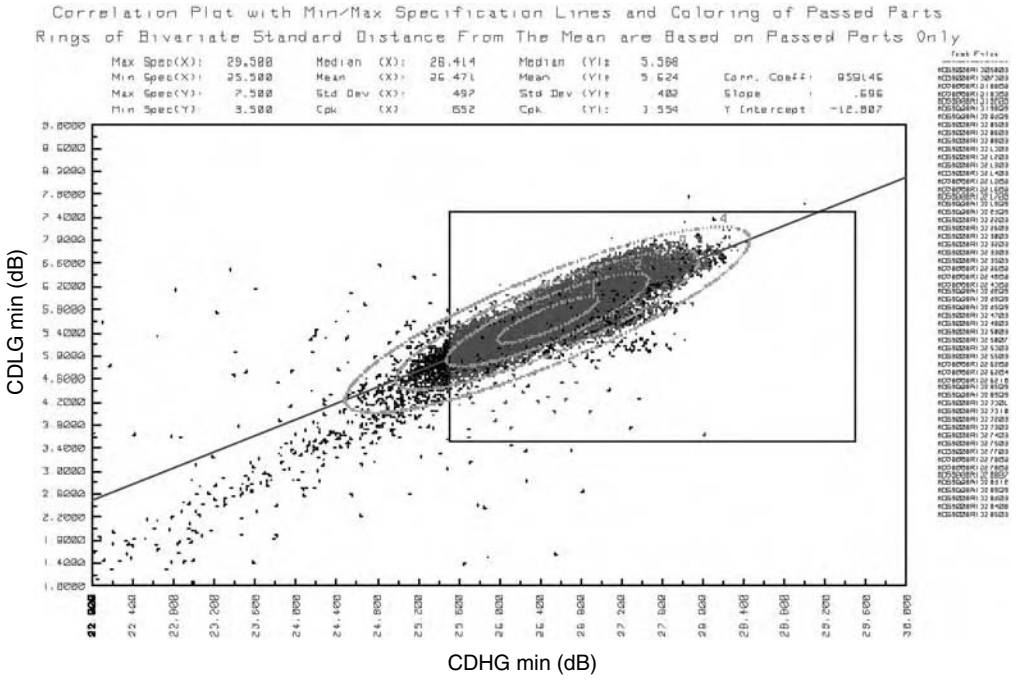


FIGURE 10.8 Scatter plot for parameter correlation analysis.

These same tools when applied in real time can usually highlight problem parameters to help drill down to the problem at hand. Yield analysis tools displaying low yielding test parameters or single failure mechanisms are critical for efficient feedback analysis to the test floor as well as the product lines.

10.4.2 Database Tools

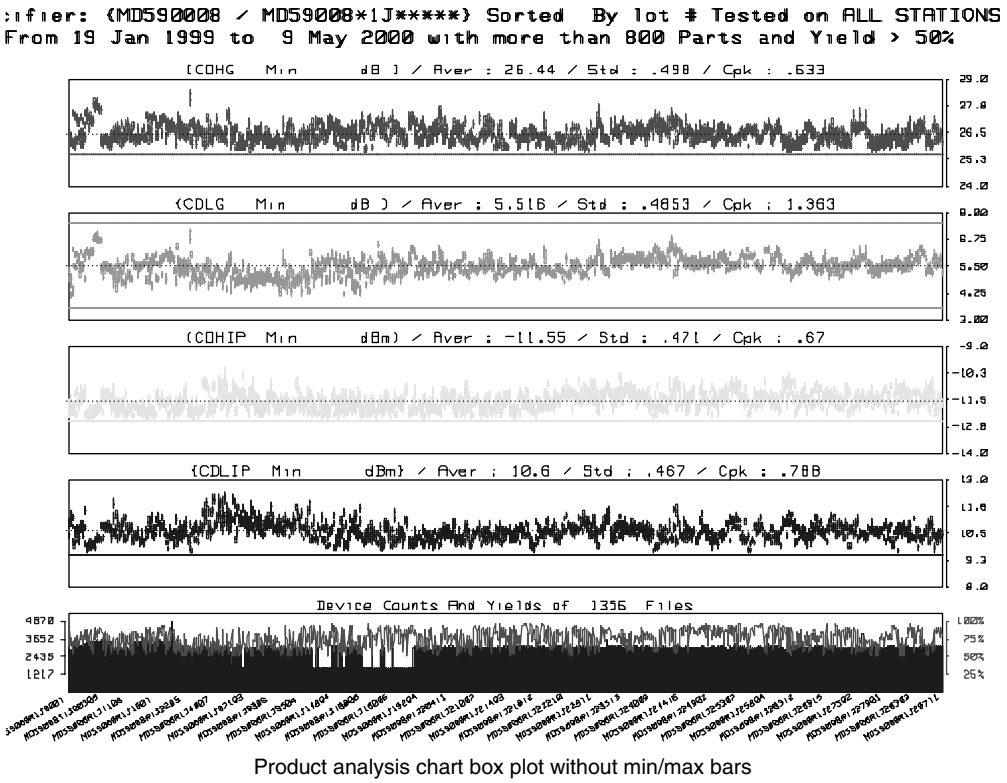
Analysis tools to quickly identify failure mechanisms are among the most important in high volume for quick feedback to the manufacturing floor. This requires that the database have full knowledge of not only the resulting data but also the high and low specifications placed on each individual parameter.

All databases, whether third party or custom, are depots for immense amounts of data with standard input and output utilities for organizing, feeding, and extracting information. The tools to display and report that information are usually independent of the database software.

Most third party database software packages can accommodate links to an exhaustive set of tools for extensive data analysis requirements. These external tools, again whether third party or custom, can be designed to provide fixed output reports for each device in question. But these databases usually require rigid data structures with fixed field assignments. Because of this, a high level of management for porting data, defining record structures, and organizing outputs is necessary when dealing with a continually changing product mix. Of course, if the application is needed for a few devices with compatible parameter tables, the management level will be minimal.

The alternative is creating a custom database structure to handle the dynamics of a high product mix for your specific needs. This is neither easy or recommended when starting fresh in today's market since it requires in-house expertise in selecting the appropriate platforms and data structures. But if the capability already exists and can handle the increased demand, it may be a more cost-effective path considering the available resources.

An important note on the consideration of third party vs. in-house is the ability to implement software changes as the need arises. With third party platforms these changes may not be instituted until the next



Product analysis chart box plot without min/max bars

FIGURE 10.9 Multiple parameter control charts for product performance analysis.

available revision or never if deemed highly custom. So be sure to select the appropriate mix to ensure this does not happen.

Regardless of the database option selected, data backups, network issues, and system integrity will still have to be maintained. Most systems today can use compression tools to maintain access to large amounts of data without the need to reload from externally archived tapes. Disc space is extremely cheap today. Even with high volume data collection requirements, information can be kept online for well over a year if necessary. More mature products can actually stop processing dense detailed information and only provide more condensed summary statistics used for tracking process uniformity.

10.4.3 Test Operation Data

To reduce the cost of testing and remain competitive in today's market, a constant monitoring of resource utilization is advantageous. A simple system utilization analysis can consist of a count test system, average cycle time of a device, and the quantity of parts in and parts out. This information is enough to get a rough idea of the average system utilization, but cannot give a complete picture when dealing with a large product base and package style mix. With detailed information of system throughput, pinpointing specific problem systems and focusing available resources to resolve the issues can be performed more efficiently. Output similar to the operational chart of Figure 10.10 can show information such as efficiency and system utilization within seconds to evaluate performance issues.

Another important aspect of monitoring is the availability of resources to floor personnel to help them react to issues as fast as possible. During the course of a measurement sequence, potential problems could arise that require immediate response. A continuous yield display will react slowly to a degradation in contact or measurement performance, especially after thousands of devices have been tested. For this

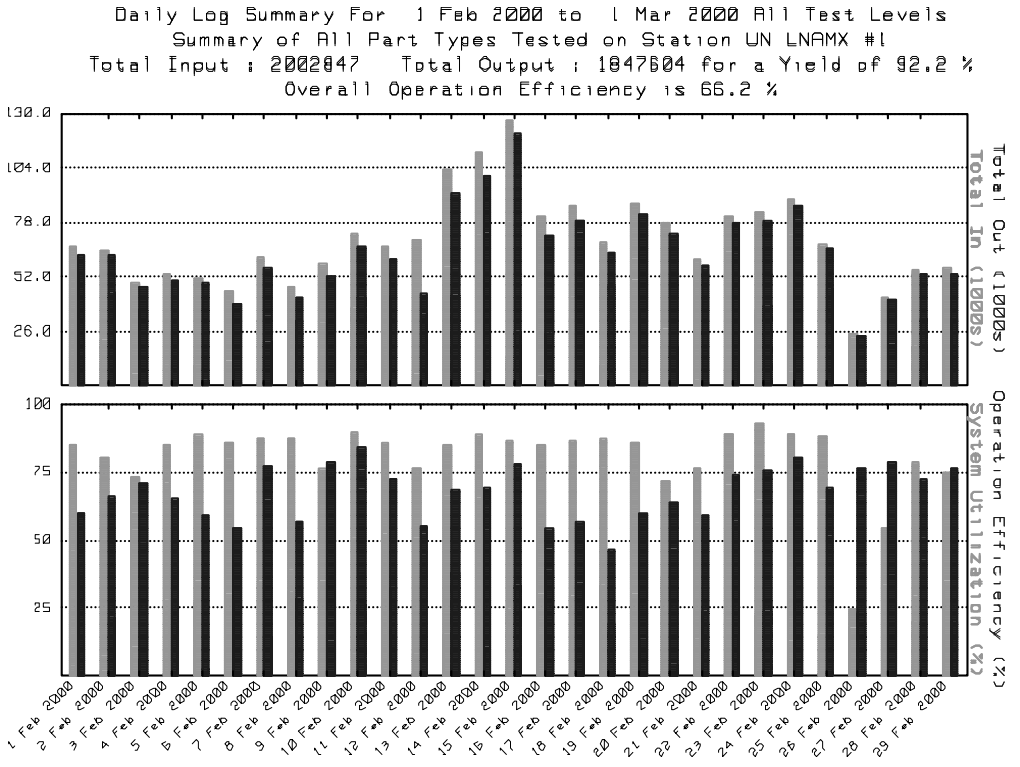


FIGURE 10.10 Yield and operation efficiency analysis tool.

reason it is beneficial to have a sample or instantaneous yield reported during the test cycle to alert operators for quick reaction.

10.5 Conclusion

High volume microwave testing has become an everyday activity for all RFIC suppliers. Microwave test equipment vendors have developed equipment with acceptable accuracy and reproducibility, and satisfactory speed. Actual test software is robust and allows automatic revision tracking. Package handlers are improving although they are the throughput bottleneck for most standard RFICs, and do not accept module packages easily. Test contactors remain a technical difficulty, especially for high frequency or high power applications. In general, "hardware" solutions for microwave high volume testing exist today.

The remaining challenge is to reduce the customer's cost of quality and the supplier test cost with existing equipment. The ability to understand the customer specifications, the test system limitations, the test information available, and their interaction is key to test effectiveness improvement today. Analysis tools and methods to exploit the vast amount of data generated are essential to pinpoint the areas of possible improvement. These tools can highlight the fabrication process, the calibration process, the specification versus process limits, the package supplier, or the handler as the first area to focus upon for cost and quality improvement. This "software" side of people with the appropriate knowledge and tools to translate data into actionable information is where we expect the most effort and the most progress.

References

1. Crowley, R., Socket Developments for CSP and FBGA Packages, *Chip Scale Review,* May 1998.
2. Montgomery, D.C., Introduction to Statistical Quality Control, chap. 9.6, 455–460.

11

Large Signal Network Analysis/Waveform Measurements

Joseph M. Gering
RF Micro Devices

11.1 Introduction

This chapter will review an area of microwave measurements concerned with measuring the instantaneous current and/or voltage waveforms at the terminals of a device-under-test (DUT) under realistic, often high power, operating conditions. (Note that "Device" in the acronym DUT generically means any device, circuit, or network.) While all engineers have measured voltage waveforms in their introductory, electronic circuits laboratories with an oscilloscope, this chapter will focus on signals at radio frequency (RF) and microwave frequencies and with techniques that provide calibrated accuracy akin to a network analyzer. The types of instruments that do these measurements go by many names, but the two most common names are large signal network analyzer (LSNA) and nonlinear vector network analyzer (NVNA). In this chapter, the acronym LSNA will be used to represent this entire class of instruments.

While not as well known as other measurement techniques presented in this book, the area is not new. Work with these techniques began in the late 1980s and has been steadily progressing ever since. Over its greater than 16-year history, the LSNA has mostly been used in academic and research laboratories, but recently it has been migrating into commercial, RF test laboratories. Applications for the LSNA are

primarily in the RF, semiconductor-related electronics area and include (but are certainly not limited to) the following:

1. *Semiconductor device development*: Observing the terminal voltages and currents of a device gives insight into the physical phenomena that limit the device's performance. These phenomena can be related to device breakdown, dispersion caused by trapping, or thermal effects, to list a few.
2. *Semiconductor device model extraction and verification*: With well-established measurement reference planes, accurate measurements can be obtained that can be directly compared to a device, RF simulation. Also, because of the instrument's ability to accommodate a variety of RF signals/stimuli, the device and model can be extensively exercised. This includes small- and large-signal, single- and multitone, continuous and pulsed, and matched and mismatched operation.
3. *Amplifier circuit analysis*: Because the LSNA provides both scalar and phase information, it is a powerful tool to analyze amplifier circuits. It can measure quantities such as AM/AM distortion, AM/PM distortion, and vector IM distortion. This can be done in both a nominal 50 Ω environment or in a mismatched environment with the additional hardware of a load-pull system. Also, the LSNA simplifies testing with active injection, where an RF signal is applied simultaneously at the input and output of the DUT.
4. *Device/circuit behavioral modeling*: Because of the wide variety of signals that can be applied and measured with a LSNA, data sets can be easily collected to fit or train phenomenological or behavioral models. These models are often used where the circuit is sufficiently complex that a full, device-level simulation is too slow, ill-conditioned, or awkward. By fitting a functional form to a circuit's response to realistic signals at realistic operating conditions, a more straight forward model can be created and used in higher level system simulations.

In this chapter, some basic mathematical background to the LSNA will be reviewed. Next, typical measurement systems and their calibration techniques will be presented followed by some extensions to basic LSNA measurements. Lastly, this chapter will conclude with a brief summary and references. For convenience, this chapter will only consider a two-port DUT, since typical LSNAs only measure a two-port. This is not a fundamental limitation, and these systems and techniques can be extended to multiport networks.

11.2 Mathematical Background

Unlike a conventional vector network analyzer (VNA), a LSNA does not measure any two-port parameters of the DUT. Instead, it measures the waveform of the voltage and current or incident and emanating traveling waves at the ports of the DUT. All two-port parameter representations of networks are linear representations, and, as such, are not strictly applicable to the large-signal, nonlinear operation of a network. The waveform can be represented as a time varying signal, as a complex, frequency spectrum, or as a complex modulation envelope [1–3]. The majority of LSNAs actually "measure" in the frequency domain. As such it is useful to consider the classic representation of a Fourier series [4]:

$$f(t) = \sum_{n=-\infty}^{\infty} C_n \mathrm{e}^{jn\omega t}, \tag{11.1}$$

where

$$C_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) \mathrm{e}^{-jn\theta} \mathrm{d}\theta \tag{11.2}$$

with $\theta = \omega t$. Since the time-domain waveforms are real quantities, $C_{-n} = C_n^*$, where the asterisk means complex conjugate, and C_0 is real. Given this, Equation 11.1 can be rearranged to be

$$f(t) = C_0 + \sum_{n=1}^{\infty} \mathrm{Re}\left\{2C_n e^{jn\omega t}\right\}. \tag{11.3}$$

If one redefines the series coefficient to be

$$K_n = \begin{cases} C_0 & \text{for } n = 0, \\ 2C_n & \text{for } n > 0. \end{cases} \tag{11.4}$$

Equation 11.3 can be written as

$$f(t) = \sum_{n=0}^{\infty} \mathrm{Re}\left\{K_n e^{jn\omega t}\right\} = \mathrm{Re}\left\{\sum_{n=0}^{\infty} K_n e^{jn\omega t}\right\}. \tag{11.5}$$

Because the LSNA is measuring in the frequency domain and is thus not restricted to an even frequency grid, Equation 11.5 can be extended to a more general case,

$$f(t) = \mathrm{Re}\left\{\sum_{n=0}^{N} K_n e^{j\omega_n t}\right\}, \tag{11.6}$$

where the harmonically related frequency $n\omega$ has been replaced with a frequency list ω_n. In addition, the waveform is assumed to have finite bandwidth, so the series is only summed from $n = 0$ to N.

The extension to Equation 11.6 allows the use of a non-evenly-spaced, frequency list. This is useful when examining a signal which has a fundamental carrier frequency with harmonics of that carrier as well as a modulation on the carrier with a finite bandwidth around the carrier. For example, if one were stimulating a DUT with a two-tone signal with frequencies $\omega_c \pm \omega_m$, one could envision the frequency list given in Table 11.1. With this list the input tones would be at indices 10 and 12. The third-order intermodulation products would be at indices 8 and 14, and the fifth-order intermodulation products would be at indices 6 and 16.

TABLE 11.1 Simple Frequency List for a Two-Tone Signal

n	ω_n	n	ω_n
0	0 (DC)	14	$\omega_c + 3\,\omega_m$
1	ω_m	15	$\omega_c + 4\,\omega_m$
2	$2\,\omega_m$	16	$\omega_c + 5\,\omega_m$
3	$3\,\omega_m$	17	$2\,\omega_c - 5\,\omega_m$
4	$4\,\omega_m$	18	$2\,\omega_c - 4\,\omega_m$
5	$5\,\omega_m$	19	$2\,\omega_c - 3\,\omega_m$
6	$\omega_c - 5\,\omega_m$	20	$2\,\omega_c - 2\,\omega_m$
7	$\omega_c - 4\,\omega_m$	21	$2\,\omega_c - \omega_m$
8	$\omega_c - 3\,\omega_m$	22	$2\,\omega_c$
9	$\omega_c - 2\,\omega_m$	23	$2\,\omega_c + \omega_m$
10	$\omega_c - \omega_m$	24	$2\,\omega_c + 2\,\omega_m$
11	ω_c	25	$2\,\omega_c + 3\,\omega_m$
12	$\omega_c + \omega_m$	26	$2\,\omega_c + 4\,\omega_m$
13	$\omega_c + 2\,\omega_m$	27	$2\,\omega_c + 5\,\omega_m$

Before proceeding further, it is worth taking a moment to clarify a convention concerning voltages, currents, and traveling waves. If one has a periodic, band-limited, time-varying voltage at the port of a DUT, $v(t)$, one can express it in a Fourier series in the form of Equation 11.6 as

$$v(t) = \mathrm{Re}\left\{\sum_{n=0}^{N} V_n e^{j\omega_n t}\right\}. \tag{11.7}$$

The voltage coefficients, V_n, are the peak voltages at the frequency ω_n. This is more clearly seen for a simple sinusoidal voltage, where, for example, V_n is a real number, V_1, for $n = 1$ and 0 otherwise. This yields $v(t) = V_1 \cos(\omega_1 t)$. Similarly, the Fourier coefficients for the current, I_n, are also peak values. Therefore, at any frequency index, n, the power delivered to the DUT is given by

$$P_n = \tfrac{1}{2}\mathrm{Re}\{V_n I_n^*\}. \tag{11.8}$$

This convention for the Fourier coefficients is consistent with standard phasor analysis, as well as with microwave simulators such as Agilent's Advanced Design System, and with commercial LSNAs such as the Maury Microwave MT4463.

Conversely, the classic traveling waves, a and b, are the "square-root-of-power" waves defined with s-parameters (see [5–7]). These traveling waves are inherently rms values, so that the square of their magnitude is power. For convenience, instead of using these classic traveling waves when dealing with waveform measurements, traveling voltage-waves are used whose Fourier coefficients are also peak values. These are typically denoted by the capital letters A and B. This simplifies the equations translating between a voltage/current representation and a traveling wave representation:

$$\begin{aligned} V &= A + B, \qquad I = \frac{A - B}{Z_0}, \\ A &= \frac{V + Z_0 I}{2}, \quad B = \frac{V - Z_0 I}{2}. \end{aligned} \tag{11.9}$$

Z_0 is the characteristic impedance of the port and is considered to be real. This gives the following for the incident power, P_{inc}, and emanating (i.e., reflected) power, P_{em}, from a port:

$$\begin{aligned} P_{\mathrm{inc}} &= \frac{|A|^2}{2Z_0}, \\ P_{\mathrm{em}} &= \frac{|B|^2}{2Z_0}. \end{aligned} \tag{11.10}$$

The power delivered to the port is simply the difference of the incident and emanating powers.

With the above analysis, one can take waveform data collected as complex Fourier coefficients, view it directly as a frequency spectrum, or convert it to a time-domain waveform using Equation 11.6. In addition to examining these waveforms as functions of frequency or time, there are numerous other graphs possible. One example is a dynamic load-line plot, which plots the output current, $i(t)$, on the y-axis versus the output voltage, $v(t)$, on the x-axis.

There is another representation or a way to view the waveform data, and that is in the envelope domain. If the RF signal is periodic at a carrier frequency, ω_{c}, and a modulation frequency, ω_{m}, where $\omega_{\mathrm{c}} \gg \omega_{\mathrm{m}}$, then, reverting back to harmonically related frequencies for the moment, Equation 11.6 can be rewritten as

$$f(t) = \mathrm{Re}\left\{\sum_{n=0}^{N}\sum_{l=-L}^{L} K_{nl} e^{j(n\omega_{\mathrm{c}} + l\omega_{\mathrm{m}})t}\right\} = \mathrm{Re}\left\{\sum_{n=0}^{N}\left[\sum_{l=-L}^{L} K_{nl} e^{jl\omega_{\mathrm{m}} t}\right] e^{jn\omega_{\mathrm{c}} t}\right\}, \tag{11.11}$$

where the portion of Equation 11.11 in square brackets is a slowly time-varying complex function that represents the complex modulation of the RF carrier and its harmonics. (Note that $K_{nl} = 0$ for $n = 0$ and $l < 0$.) There are times in practice when the base-band terms ($n = 0$) as well as the higher harmonics ($n \geq 2$) of the carrier can be ignored. This is typically the case with power amplifiers, where a DC block and output low-pass filter combine to allow only the modulated carrier to be transmitted. Considering for the moment just the modulation at the fundamental of the carrier, the modulation function can be written as

$$\hat{f}(t) = \sum_{l=-L}^{L} K_{1l} e^{jl\omega_m t}. \tag{11.12}$$

When this is applied to the traveling wave representation, graphs of $|\hat{B}_2(t)|$ versus $|\hat{A}_1(t)|$ will show AM/AM behavior, and graphs of $\angle\hat{B}_2(t)$ versus $|\hat{A}_1(t)|$ will show AM/PM behavior, where the subscripts are the port indices. Also, these plots will highlight memory effects as loops or hysteresis-like characteristics.

This mathematical background has relied on generalized equations. To sum up this discussion, an example might be in order. Consider a two-tone voltage given by

$$v(t) = (1\,\text{V})\cos[2\pi(1\,\text{GHz} - 20\,\text{MHz})t] + (1\,\text{V})\cos[2\pi(1\,\text{GHz} + 20\,\text{MHz})t]. \tag{11.13}$$

Figure 11.1 gives a plot of this voltage versus time. Equation 11.13 can be rewritten in the form of Equation 11.6 as

$$v(t) = \text{Re}\left\{(1\,\text{V})e^{j2\pi(1\,\text{GHz}-20\,\text{MHz})t} + (1\,\text{V})e^{j2\pi(1\,\text{GHz}+20\,\text{MHz})t}\right\}, \tag{11.14}$$

which easily produces the plot of Figure 11.2 for the voltage versus frequency. Lastly, this can be further manipulated to give

$$v(t) = \text{Re}\left\{\left[(1\,\text{V})e^{j2\pi(-20\,\text{MHz})t} + (1\,\text{V})e^{j2\pi(+20\,\text{MHz})t}\right]e^{j2\pi(1\,\text{GHz})t}\right\}, \tag{11.15}$$

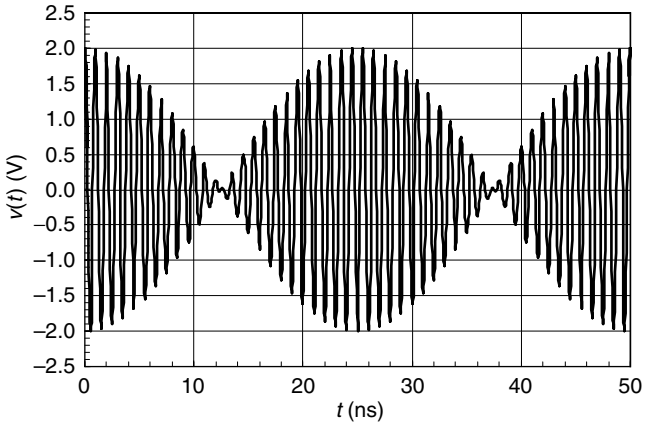


FIGURE 11.1 Voltage waveform as a function of time for a two-tone signal.

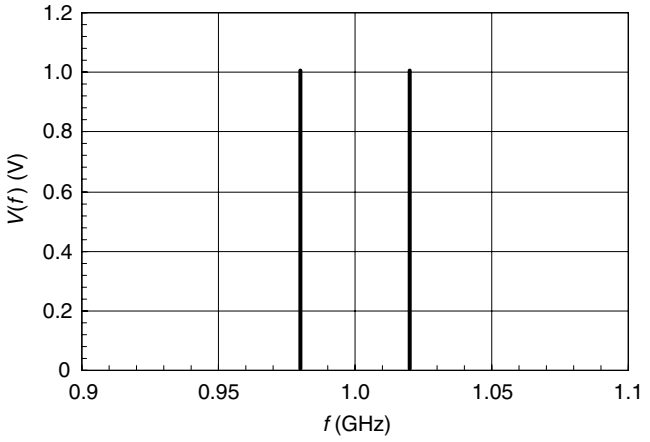


FIGURE 11.2 Voltage Fourier coefficients as a function of frequency for a two-tone signal.

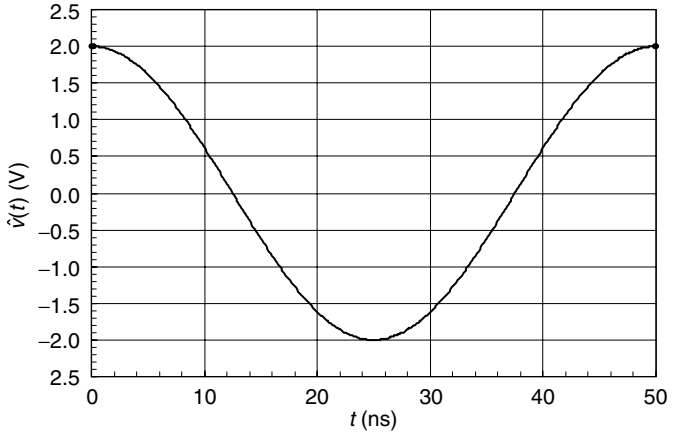


FIGURE 11.3 Voltage envelope as a function of time for a two-tone signal.

which gives a complex envelope function of

$$\hat{v}(t) = (1\,\mathrm{V})e^{j2\pi(-20\ \mathrm{MHz})t} + (1\,\mathrm{V})e^{j2\pi(+20\ \mathrm{MHz})t} = (2\,\mathrm{V})\cos[2\pi(20\ \mathrm{MHz})t], \tag{11.16}$$

which is plotted in Figure 11.3.

11.3 Measurement Systems

As mentioned previously, most engineers have performed low-frequency waveform measurements with an oscilloscope. In fact, some of the early work with waveform measurements used sampling oscilloscopes [8,9]. However, the dynamic range/resolution of oscilloscopes proved to be limiting when compared to the alternatives that will be discussed shortly. Also, in the past, oscilloscope typically only returned a time-domain waveform and were less programmable than microwave instruments, which made oscilloscopes less convenient. The two approaches that have gained acceptance are based on down-converting the RF signal using a sampling down-converter (such as in the Hewlett–Packard

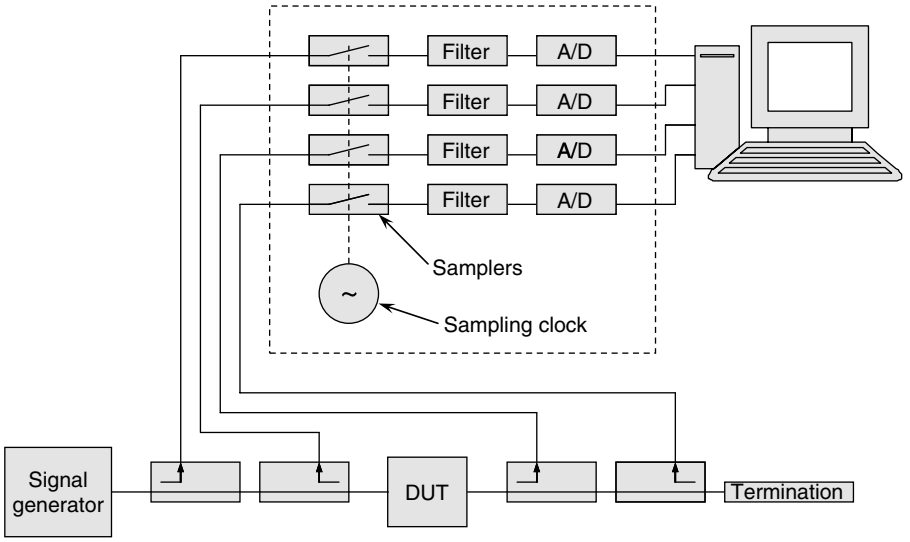


FIGURE 11.4 Block diagram of a generic sampling down-converter based LSNA.

Microwave Transition Analyzer or Maury Microwave LSNA) or a mixer down-converter (commonly used in a VNA).

11.3.1 Sampling Down-Converter Based LSNA

Figure 11.4 gives a generic block diagram for a simple LSNA based on sampling down-converters. The DUT is driven with an RF signal from the signal generator, and the output power of the DUT is dissipated in the termination. (For simplicity, the DUT bias is not drawn in this diagram.) The directional couplers sample the incident and emanating RF signals on both ports of the DUT. These are fed to the samplers that sample the RF signal at a certain rate determined by the sampling clock. This operation intentionally uses aliasing and converts the RF signal down to a lower frequency. One requirement in this system is that all the signals (including their harmonics and modulation tones) be periodic and be on an established frequency grid. All the various frequencies in the original RF signals are converted down to a collection of frequencies on a lower frequency grid at base band [10]. This down-converted signal is then filtered to remove any residual RF and then converted by an analog-to-digital converter for post-processing in a computer. Because the down-conversion process is performed on the entire signal spectrum at once, the down-converted signal preserves the relative magnitudes and phases of the various frequency components in the original signal in each of the new, down-converted components. As such, when the down-converted signal is brought into the computer for processing, no other reference is needed to establish the phase (unlike the mixer-based approach to be discussed shortly). Finally, because the entire signal is down-converted at once, the sampler approach tends to be a faster measurement.

In a simplified fashion, this is the method employed by a presently available, commercial LSNA, the Maury Microwave MT4463 [11]. The instrument, a joint effort with NMDG Engineering, has been a Maury product since 2003; however, it was originally developed by Hewlett–Packard/Agilent for many years prior to 2003. The MT4463 is a combination of a custom-designed-and-built test set, down-converter, and phase calibrator, along with other commercially available microwave equipment. While this virtually turn-key system has been around for a while and is being used by numerous research and academic laboratories, its roots are based in another piece of equipment: the Hewlett–Packard Microwave Transition Analyzer (MTA). The MTA, which came out in the early 1990s, used sampling down-converters to down convert, measure, and display (or transfer to a computer) the Fourier coefficients or waveform

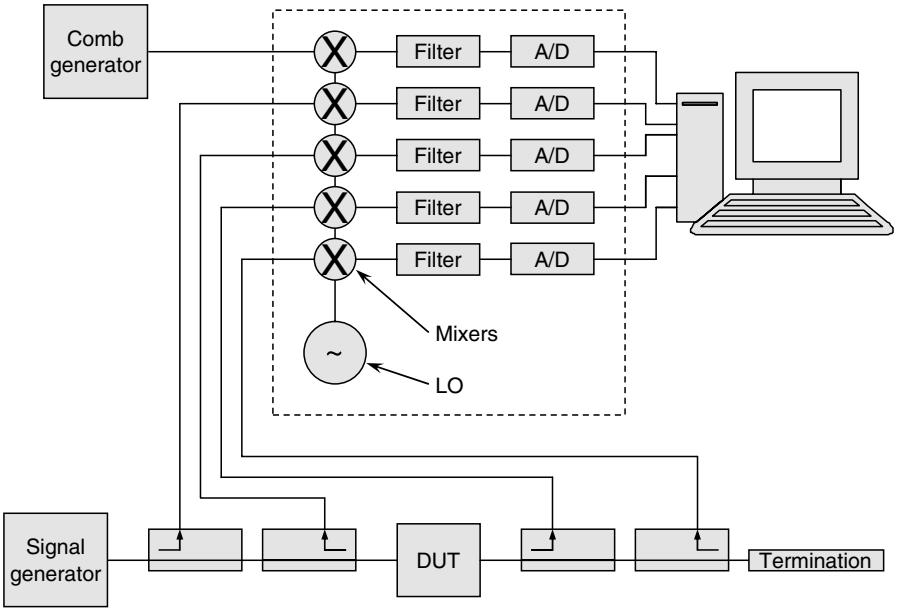


FIGURE 11.5 Block diagram of a generic mixer down-converter based LSNA.

of two RF channels. Many research groups used the MTA to build their own LSNAs [12–20]. This work helped to promote the sampling down-converter approach as well as helped establish uses for the LSNA.

11.3.2 Mixer Down-Converter Based LSNA

Figure 11.5 gives a generic block diagram for a simple LSNA based on mixer down-converters. This diagram is quite similar to the sampling down-converter diagram in Figure 11.4 with two notable exceptions. First, the down-converting elements are mixers driven by a local oscillator (LO). Second, there is a fifth mixer channel receiving a signal from a reference comb generator. As before, the RF signal starts with the signal generator, ends with the termination, and is sampled by the directional couplers. This system still has the restriction that the RF signal must be on a known frequency grid. However, this system has the further requirement that the frequency grid be equal to the output of the RF comb generator, which means that the frequency grid must be evenly spaced. Also, while not shown in Figure 11.5, the signal generator and comb generator must be frequency locked to each other. At each frequency to be measured, the LO is tuned to down-convert that frequency to an intermediate frequency (IF), which is filtered, digitized, and read into the computer.

With a mixer system, the relative magnitudes and phases of the down-converted signals depend on the magnitude and phase of the LO. The magnitude of the LO can be controlled; however, each time the LO is tuned to a new frequency, its phase is randomly changed. So each time a frequency is measured, its phase has a random offset from all previous measurements, even those previous measurements at the same frequency. A comb generator is measured on a fifth channel to get around this. The comb generator is designed to produce a full spectrum signal with a component at each of the frequencies to be measured with a very stable phase offset between these different frequency components. (A driven step recovery diode is one such comb generator.) Every time a frequency is measured, the phases of the four DUT channels are referenced to the phase of the fifth reference channel. Thus a stable phase relationship between frequencies is achieved. Lastly, because this approach measures one frequency at a time, it tends to be slower. However, heterodyne mixers are capable of very high dynamic ranges, so this technique is similarly capable of much better accuracy than other approaches.

Mixer-like down-conversion is the technique used in a VNA. While there are currently no commercial LSNAs that use the mixer approach, there has been work on this approach based on commercial VNAs [21–25]. Despite the slightly more cumbersome requirement of a phase reference, this approach has one other practical benefit. If implemented with a conventional VNA, this approach is potentially more cost effective, since the VNA required for this LSNA could easily be used for other duties such as conventional *s*-parameter measurements.

11.3.3 Comparison of Sampling and Mixer-Based LSNAs

It is always risky quoting performance numbers because they are always subject to improvement or change, but they are nonetheless worth a brief mention. One of the big advantages of the mixer-based approach is dynamic range. The Maury MT4463 can approach dynamic ranges of 90 dBc. By comparison, modern VNAs have dynamic ranges approaching 140 dBc. Whether this is a prohibitive difference depends on the application. One of the biggest inconveniences of the mixer-based LSNA is that the frequency grid is set by the reference comb generator. The minimum frequency step size from a commercial comb generator known to this author is 80 MHz. Blockley et al. [24] have demonstrated a custom circuit with a 300 kHz minimum spacing. Unfortunately, this spacing would be too large if one were attempting to use a multitone signal that would cover the same modulation bandwidth as some modern communication signals. By comparison, the MT4463 has an arbitrary, software limit of a 1 kHz minimum spacing. Neither system can measure base-band signals, but the system by Blockley et al. [24] comes closest since it will measure down to 300 kHz, while the MT4463 is presently limited to a 600 MHz minimum frequency. There are obviously other differences between these two approaches; however, the ones mentioned tend to be the most significant practical differences.

11.4 Calibration

When implemented, both of the LSNAs described above will measure the raw, uncorrected traveling waves incident upon and emanating from a DUT across a frequency grid. The next obvious requirement is to error correct those raw measurements to calibrated measurements at the reference plane of the DUT [18,26–32]. Figure 11.6 shows a drawing of the error network for the LSNA. As indicated with the blocks labeled A and B in Figure 11.6, it is assumed that there is no coupling or leakage between the input and output sides of the DUT. The test set functioning, the down-conversion, and the digitizing are required to be linear processes so that the raw traveling waves are true representations of the traveling waves at the DUT. As such, one could envision representing the A and B blocks as *s*-parameter matrices with the appropriate, additional, frequency indexing. This is essentially what is done, but the formulation is written

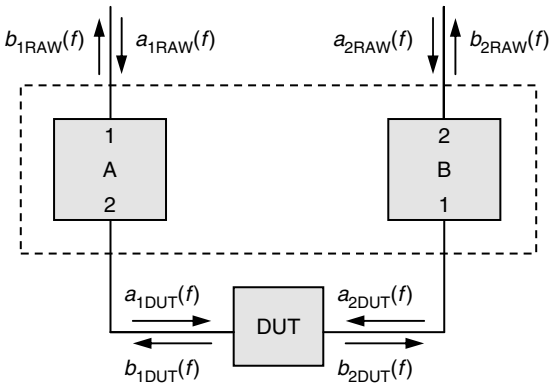


FIGURE 11.6 Drawing of the error network for a LSNA.

as follows to help in the calibration procedure [11]:

$$\begin{bmatrix} a_{1\mathrm{DUT}}^{(n)} \\ b_{1\mathrm{DUT}}^{(n)} \\ a_{2\mathrm{DUT}}^{(n)} \\ b_{2\mathrm{DUT}}^{(n)} \end{bmatrix} = K^{(n)} \begin{bmatrix} 1 & \beta_1^{(n)} & 0 & 0 \\ \gamma_1^{(n)} & \delta_1^{(n)} & 0 & 0 \\ 0 & 0 & \alpha_2^{(n)} & \beta_2^{(n)} \\ 0 & 0 & \gamma_2^{(n)} & \delta_2^{(n)} \end{bmatrix} \begin{bmatrix} a_{1\mathrm{RAW}}^{(n)} \\ b_{1\mathrm{RAW}}^{(n)} \\ a_{2\mathrm{RAW}}^{(n)} \\ b_{2\mathrm{RAW}}^{(n)} \end{bmatrix} \tag{11.17}$$

The superscript (n) denotes the frequency index. The 4×4 matrix is the equivalents of the *s*-parameters of the A and B blocks. These relate the DUT quantities to the raw quantities within one frequency and are normalized so that the upper left matrix element is one. The complex coefficient, $K^{(n)}$ provides the link between frequencies in both magnitude and phase. It also sets the overall magnitude and phase reference. With a conventional VNA calibration, the end goal is to compute corrected *s*-parameters, which are ratioed quantities. As such, the VNA error correction is rarely set up to compute the absolute magnitude of the traveling waves. However, with the LSNA, one of the outputs of the measurement is the waveform at the DUT terminals. To calculate this correctly, an absolute reference must be established, which is one of the functions of $K^{(n)}$. To find the values of Equation 11.17, the calibration procedure is broken into three parts: an *s*-parameter calibration, a power calibration, and a phase calibration. These will be discussed next.

11.4.1 *s*-Parameter Calibration

To generate the terms in the 4×4 matrix of Equation 11.17, it is sufficient to perform a standard *s*-parameter calibration at each frequency. Any technique such as short-open-load-thru (SOLT), thru-reflect-line (TRL), line-reflect-match (LRM), line-reflect-reflect-match (LRRM), and so on, can be used. (Since *s*-parameter calibration is a fairly well-established and well-published subject, the reader is directed to References [33–50] for further information.) However, it should be noted that the block diagrams of Figures 11.4 and 11.5 show forward-only measurements. This is sufficient, and in fact is often necessary, for large-signal measurements, but all *s*-parameter calibration procedures require both forward and reverse measurements. Therefore, during the *s*-parameter calibration, it will be necessary to swap the signal generator and termination in Figures 11.4 and 11.5. Typically, a more elaborate test set is used, which employs a high power RF switch to automate the toggling between the forward and reverse measurements. If the *s*-parameters of the A and B blocks of Figure 11.6 are denoted as A_{ij} and B_{ij}, respectively, then the 4×4 matrix of Equation 11.17 can be written as

$$\begin{bmatrix} 1 & -\dfrac{A_{22}^{(n)}}{\Delta_A^{(n)}} & 0 & 0 \\ \dfrac{A_{11}^{(n)}}{\Delta_A^{(n)}} & -\dfrac{1}{\Delta_A^{(n)}} & 0 & 0 \\ 0 & 0 & \dfrac{\Delta_B^{(n)}}{\Delta_A^{(n)}} \dfrac{A_{12}^{(n)}}{B_{21}^{(n)}} & -\dfrac{B_{11}^{(n)}}{\Delta_A^{(n)}} \dfrac{A_{12}^{(n)}}{B_{21}^{(n)}} \\ 0 & 0 & \dfrac{B_{22}^{(n)}}{\Delta_A^{(n)}} \dfrac{A_{12}^{(n)}}{B_{21}^{(n)}} & \dfrac{-1}{\Delta_A^{(n)}} \dfrac{A_{12}^{(n)}}{B_{21}^{(n)}} \end{bmatrix}, \tag{11.18}$$

where $\Delta_A^{(n)} = A_{11}^{(n)} A_{22}^{(n)} - A_{21}^{(n)} A_{12}^{(n)}$ and $\Delta_B^{(n)} = B_{11}^{(n)} B_{22}^{(n)} - B_{21}^{(n)} B_{12}^{(n)}$. For those familiar with the standard, VNA, 12-term, error model, Figure 11.7 shows the flow diagrams comparing the 12-term error model and the A/B-block model used here. The two models can be equated with the following set of

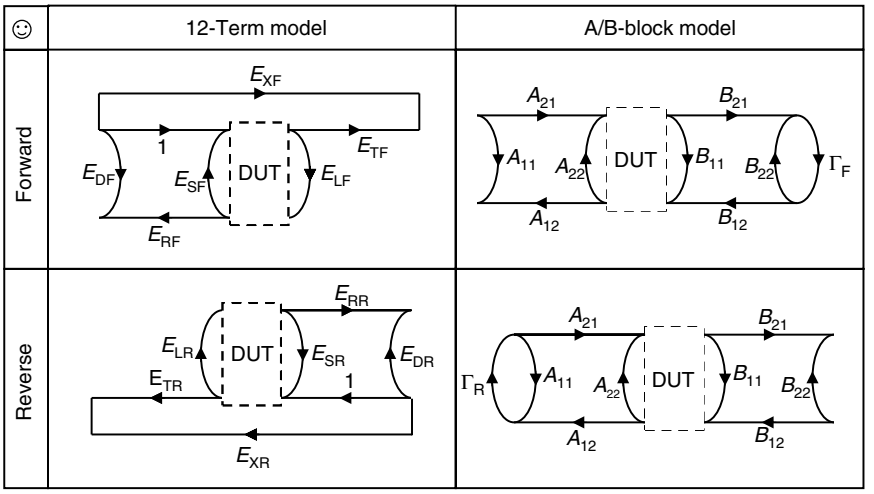


FIGURE 11.7 Flow diagrams of the 12-term error model and A/B-block model.

equations:

$$
\begin{aligned}
E_{DF} &= A_{11} & E_{DR} &= B_{22} \\
E_{SF} &= A_{22} & E_{SR} &= B_{11} \\
E_{RF} &= A_{12}A_{21} & E_{RR} &= B_{12}B_{21} \\
E_{XF} &= 0 & E_{XR} &= 0 \\
E_{LF} &= B_{11} + \frac{B_{12}B_{21}\Gamma_F}{1 - B_{22}\Gamma_F} & E_{LR} &= A_{22} + \frac{A_{12}A_{21}\Gamma_R}{1 - A_{11}\Gamma_R} \\
E_{TF} &= \frac{A_{21}B_{21}}{1 - B_{22}\Gamma_F} & E_{TR} &= \frac{A_{12}B_{12}}{1 - A_{11}\Gamma_R}
\end{aligned}
\tag{11.19}
$$

Γ_F and Γ_R are the "switch" reflection coefficients in a standard VNA calibration. In the case of a LSNA *s*-parameter calibration, these account for the external termination in the forward and reverse measurements. It is interesting to note that while the reflection coefficient of the termination in Figures 11.4 and 11.5 will affect the traveling waves and thus the waveforms at the port of the DUT, it is not necessary to know what that reflection coefficient is to measure error corrected waveforms. This is useful since once the LSNA is calibrated, it allows the user to vary the termination or inject other signals from the "outside" of the couplers without altering the calibration.

11.4.2 Power Calibration

As mentioned previously, the *s*-parameter calibration is a relative calibration that relates the various traveling wave quantities to each other at a given frequency. To bridge between frequencies, additional calibration steps are needed. This frequency-to-frequency connection is provided by the $K^{(n)}$ factor in Equation 11.17. The magnitude of the $K^{(n)}$ factor is established through a power calibration. If a power meter is connected in place of port 1 of the DUT, it will naturally measure the power delivered to the power meter from the LSNA. Taking the traveling waves, *a* and *b*, as traveling voltage-waves, and stepping the signal generator through frequency, one obtains

$$
P_{\text{meter}}^{(n)} = \frac{\left|a_{1\text{DUT}}^{(n)}\right|^2 - \left|b_{1\text{DUT}}^{(n)}\right|^2}{2Z_0} = \left|K^{(n)}\right|^2 \frac{\left|a_{1\text{RAW}}^{(n)} + \beta_1^{(n)} b_{1\text{RAW}}^{(n)}\right|^2 - \left|\gamma_1^{(n)} a_{1\text{RAW}}^{(n)} + \delta_1^{(n)} b_{1\text{RAW}}^{(n)}\right|^2}{2Z_0}, \tag{11.20}
$$

which can be solved for $|K^{(n)}|$.

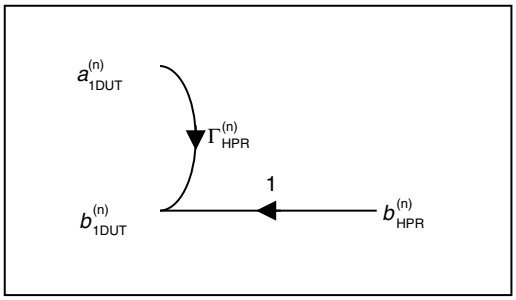


FIGURE 11.8 Flow diagram of the HPR connected to port 1.

11.4.3 Phase Calibration

The last step in the LSNA calibration is to find the phase for $K^{(n)}$ of Equation 11.17. While the *s*-parameter calibration is obviously done with a VNA and the power calibration is partially done with newer VNAs, the phase calibration is unique to the LSNA. This step is performed by measuring a standard with a known, stable phase relationship between frequencies. The "standard" that is commonly used is a comb generator for the same reason that it was used as the phase reference with the mixer-based LSNA. Incidentally, the comb generator can also be thought of as a pulse generator, since a square wave signal will have a rich spectrum at the pulse frequency and its harmonics. Strictly speaking, the comb generator, which is often called a harmonic phase reference (HPR), is a transfer standard. It is a stable device that is measured with a technique that is traceable to the National Institute of Standards and Technology (NIST). The original procedure for this was to use a nose-to-nose calibration to calibrate a sampling oscilloscope, which in turn was used to characterize the pulse response of the HPR [32,51,52]. This technique works well despite limitation at very high frequencies and is still used to calibrate the HPR with the Maury MT4463. However, in 2001 NIST reported a new method for calibrating the HPR based on electro-optic sampling [53]. Because of its improved accuracy, this method is taking over the role as the primary standard in calibrating the HPR, while the nose-to-nose technique is still used but used as a transfer standard off the electro-optic sampling method.

For the phase calibration, to determine the phase of $K^{(n)}$, the HPR is connected at the port 1 reference plane in place of the DUT. The signal generator is not connected to the LSNA for this procedure, since the HPR is its own source. Figure 11.8 shows the signal flow diagram for the HPR connected to port 1. The HPR signal ($b_{HPR}^{(n)}$) and reflection coefficient ($\Gamma_{HPR}^{(n)}$) are provided by the HPR characterization. From the flow diagram, one can write

$$b_{1DUT}^{(n)} = b_{HPR}^{(n)} + \Gamma_{HPR}^{(n)} a_{1DUT}^{(n)}. \tag{11.21}$$

From Equation 11.17, one has

$$a_{1DUT}^{(n)} = K^{(n)} \left(a_{1RAW}^{(n)} + \beta_1^{(n)} b_{1RAW}^{(n)} \right) \tag{11.22}$$

and

$$b_{1DUT}^{(n)} = K^{(n)} \left(\gamma_1^{(n)} a_{1RAW}^{(n)} + \delta_1^{(n)} b_{1RAW}^{(n)} \right). \tag{11.23}$$

Combining Equations 11.21 through 11.23 yields

$$K^{(n)} = \frac{b_{HPR}^{(n)}}{\left(\gamma_1^{(n)} a_{1RAW}^{(n)} + \delta_1^{(n)} b_{1RAW}^{(n)} \right) - \Gamma_{HPR}^{(n)} \left(a_{1RAW}^{(n)} + \beta_1^{(n)} b_{1RAW}^{(n)} \right)}. \tag{11.24}$$

From Equation 11.24 one might be inclined to obtain both the magnitude and phase of $K^{(n)}$, but this typically is not done. Equation 11.24 is only used to obtain the phase of $K^{(n)}$ because slight power variations such as from unwanted loss between the HPR and LSNA will have a direct impact on the magnitude of $b_{\mathrm{HPR}}^{(n)}$, but will have much less influence on the phase of $b_{\mathrm{HPR}}^{(n)}$, since the phase depends on the shape of the HPR pulse.

11.4.3.1 Issues with the LSNA Calibration

The calibration approach described is the fundamental method used to calibrate most LSNAs. However, there are several things that complicate the calibration. First, the situation is slightly more complicated in a calibration where the power meter and HPR cannot be placed directly at the port 1 reference plane of the DUT (e.g., when the DUT is to be RF probed). In this case, an additional reference plane is introduced that is compatible with the power meter and HPR, which is characterized and linked to the DUT and raw measurement planes through additional *s*-parameter networks [3]. This is necessary for certain measurements; however, it understandably adds to the complexity of the arithmetic and subtracts from the overall system accuracy.

The other major complication concerns the modulation tones around the carrier frequencies and its harmonics. It is implied in the above calibration discussion that the calibration is performed at every frequency point. This is often not the case. Instead, the calibration described above is only performed at the fundamental carrier and its harmonics. The calibration is then extended to handle modulation frequencies around the calibrated carrier-frequencies [30]. One drawback to this approach is that some users have only accounted for the differences in the IF path after the down converters. For the RF path, they make a "narrow band" assumption and use the error terms derived with the above procedure across the modulation frequency band. While this might be sufficient in some cases, there is much work aimed at measuring signals with wider modulation bandwidths [54,55]. The main limitation to applying the full calibration procedure to all frequencies (apart from the increased calibration time) is the lack of a phase standard (HPR/comb generator) with fine enough frequency steps. Fortunately, there is work in this area, which may allow the calibration procedure to be generalized [56].

11.5 Measurement Extensions

With the systems and calibration procedures described so far, one can measure the full, large-signal waveforms at the terminals of a DUT. Perhaps the most straightforward measurement is to excite the DUT with a single-tone input signal and examine the reflected and transmitted outputs at the fundamental and harmonics. In one sense, this is where the LSNA resembles a nonlinear-VNA. While these measurements may be the most prevalent, the LSNA is by no means limited to just this basic case. The following material will present some of the measurements that take a LSNA beyond a VNA.

11.5.1 Source-Pull, Load-Pull, and Active Injection

A VNA strives to present a matched (usually 50 Ω) termination to the DUT. In contrast, the LSNA measurement is independent of the input signals (as long as they are on the frequency grid) and terminations. A common large-signal measurement is a source-pull/load-pull measurement where the source impedance and load terminations are varied using tuners. The tuners of a source-pull/load-pull measurement can be easily incorporated with a LSNA [57]. To make the tuners noninvasive to the LSNA calibration, one would place them "outside" the couplers in the block diagrams of Figures 11.4 and 11.5. However, this is often not practical, since the LSNA test set will add loss that will limit the tuning range of the tuners. Instead, tuners are placed next to the DUT, and the *s*-parameters of the tuners are mathematically added to the error correction matrix of Equation 11.17. This is a useful approach to present the optimal impedances to a DUT and simultaneously observe the waveforms at the DUT's terminals.

The function of the load tuner is to reflect a portion of the DUT's output signal back into the DUT. As an alternative to a load tuner (or even in conjunction with one), a second source can be used to create a signal that is injected back into the DUT's output. The second source can be put in place of the termination in Figures 11.4 and 11.5 (with an isolator to shunt the DUT output to a load and protect the source). In this fashion, this active injection does not alter the LSNA calibration and can synthesize the same conditions as a load tuner. In fact, with active injection, one can achieve effective load terminations that a tuner cannot realize. One can even offset the frequency of the injected signal creating a two-tone condition. Since a two-tone can be viewed as a single-tone with modulation, one is in effect varying the effective load in real time [58,59].

11.5.2 Multisine (Multitone) Signals

As has been discussed, the LSNA is set up to measure signals with a periodic modulation. Hence, an obvious input signal is a multisine or multitone signal. Many RF signal generators make this easy by providing the user a means to enter a complex, periodic modulation spectrum to be applied to a carrier signal. As long as the frequencies of the multisine are matched to those of the LSNA, the LSNA can measure the input signal from the generator as well as the outputs produced by the DUT. One of the uses of the LSNA is to study devices and circuits used in modern telecommunications. These signals often have complex modulations. One area of study is how to represent these communications signals with multitone signals that achieve the same probability distribution as the original signal [60–64]. These constructed signals can then be used in place of the original signal with instruments like the LSNA.

11.5.3 Hot *s*-Parameters

One last topic that deserves mention is "Hot *s*-parameters." As mentioned earlier, *s*-parameters, like all two-port parameters, do not apply to large-signal measurements with nonlinearities. So, strictly speaking, a hot *s*-parameter or large-signal *s*-parameter is a contradiction in terms. Nonetheless, they can be properly accounted for as a linearization of the large-signal behavior of a DUT [65]. With a proper treatment (and in some cases without), hot *s*-parameters started out as a means to study device instabilities and have been used to examine a device's basic properties as well [66–70]. As this is an ancillary use of a LSNA, a more complete discussion will be avoided, and the reader is directed to References 65–70.

11.6 Summary

This chapter has covered large-signal network analysis through waveform measurements. It has focused on the LSNA by mainly presenting the system architecture and calibration procedure. This is a unique instrument that extends large-signal, nonlinear measurement capability beyond a conventional VNA or conventional load-pull. Lastly, there are a few review papers that discuss the LSNA or the LSNA in context with other microwave measurements that might be useful to the reader [71–73].

References

1. E. Van Damme, J. Verspecht, F. Verbeyst, M. Vanden Bossche, "Large-signal network analysis—a measurement concept to characterize nonlinear devices and systems," Agilent Technologies, Belgium, June 2002. [Online]. Available: http://www.janverspecht.com/
2. *Large Signal Network Analyzer User's Guide.* Ontario, CA: Maury Microwave Corp., 2005.
3. J. Verspecht, "Large-signal network analysis 'Going beyond S-parameters'," presented at ARFTG Conference Short Course, Boulder, CO, Dec. 2003. [Online]. Available: http://www.janverspecht.com/
4. M. R. Speigel, *Mathematical Handbook.* New York: McGraw-Hill, 1968, pp. 131–135.
5. K. Kurokawa, "Power waves and the scattering matrix," *IEEE Trans. Microw. Theory Tech.*, vol. 13, Mar. 1965, pp. 194–202.

6. G. Gonzalez, *Microwave Transistor Amplifier: Analysis and Design*. Prentice Hall, Englewood Cliffs, NJ, 1990.
7. R. B. Marks and D. F. Williams, "A general waveguide circuit theory," *J. Res. Natl. Inst. Stand. Tech.*, vol. 97, Sep.–Oct. 1992, pp. 533–562.
8. M. Sipila, K. Lehtinen, and V. Porra, "High-frequency periodic time-domain waveform measurement system," *IEEE Trans. Microw. Theory Tech.*, vol. 36, Oct. 1988, pp. 1397–1405.
9. G. Kompa and F. van Raay, "Error-corrected large-signal waveform measurement system combining network analyzer and sampling oscilloscope capabilities," *IEEE Trans. Microw. Theory Tech.*, vol. 38, Apr. 1990, pp. 358–365.
10. J. Verspecht, "The return of the sampling frequency convertor," in *62nd ARFTG Conf. Dig.*, Dec. 2003, pp. 155–164.
11. *Large Signal Network Analyzer System Manual*. Ontario, CA: Maury Microwave Corp., 2005.
12. F. van Raay and G. Kompa, "A new on-wafer large-signal waveform measurement system with 40 GHz harmonic bandwidth," in *IEEE MTT-S Int. Microw. Symp. Dig.*, vol. 3, June 1992, pp. 1435–1438.
13. C. J. Wei, Y. A. Tkachenko, and J. C. M. Hwang, "Non-invasive waveform probing for nonlinear network analysis," in *IEEE MTT-S Int. Microw. Symp. Dig.*, vol. 3, June 1993, pp. 1347–1350.
14. M. Demmler, P. J. Tasker, and M. Schlechtweg, "On-wafer large signal power, S-parameter and waveform measurement system," in *Proc. Third International Workshop on Integrated Nonlinear Microwave and Millimeterwave Circuits*, 1994, pp. 153–158.
15. C.-J. Wei, Y. E. Lan, J. C. M. Hwang, W.-J. Ho, and J. A. Higgins, "Waveform-based modeling and characterization of microwave power heterojunction transistors," *IEEE Trans. Microw. Theory Tech.*, vol. 43, Dec. 1995, pp. 2899–2903.
16. T. Van den Broeck and J. Verspecht, "Calibrated vectorial nonlinear-network analyzers," in *IEEE MTT-S Int. Microw. Symp. Dig.*, vol. 2, June 1994, pp. 1069–1072.
17. J. G. Leckey, A. D. Patterson, and J. A. C. Stewart, "A vector corrected waveform and load line measurement system for large signal transistor characterization," in *IEEE MTT-S Int. Microw. Symp. Dig.*, vol. 3, June 1995, pp. 1243–1246.
18. J. Verspecht, P. Debie, A. Barel, and L. Martens, "Accurate on wafer measurement of phase and amplitude of the spectral components of incident and scattered voltage waves at the signal ports of a nonlinear microwave device," in *IEEE MTT-S Int. Microw. Symp. Dig.*, vol. 3, June 1995, pp. 1029–1032.
19. D. Barataud, A. Mallet, J. P. Fraysse, F. Blache, M. Campovecchio, J. M. Nebus, J. P. Villoite, and J. Verspecht, "A novel time domain characterization technique of intermodulation in microwave transistors: application to the visualization of the distortion of high efficiency power amplifiers," in *IEEE MTT-S Int. Microw. Symp. Dig.*, vol. 3, June 1997, pp. 1687–1690.
20. D. Barataud, F. Blache, A. Mallet, P. P. Bouysse, J.-M. Nebus, J. P. Villotte, J. Obregon, J. Verspecht, and P. Auxemery, "Measurement and control of current/voltage waveforms of microwave transistors using a harmonic load-pull system for the optimum design of high efficiency power amplifiers," *IEEE Trans. Instrum. Meas.*, vol. 48, Aug. 1999, pp. 835–842.
21. U. Lott, "A method for measuring magnitude and phase of harmonics generated in nonlinear microwave two-ports," in *IEEE MTT-S Int. Microw. Symp. Dig.*, vol. 1, May 1988, pp. 225–228.
22. U. Lott, "Measurement of magnitude and phase of harmonics generated in nonlinear microwave two-ports," *IEEE Trans. Microw. Theory Tech.*, vol. 37, Oct. 1989, pp. 1506–1511.
23. D. Barataud, C. Arnaud, B. Thibaud, M. Campovecchio, J.-M. Nebus, and J. P. Villotte, "Measurements of time-domain voltage/current waveform at RF and microwave frequencies based on the use of a vector network analyzer for the characterization of nonlinear devices—applications to high-efficiency power amplifiers and frequency-multipliers optimization," *IEEE Trans. Instrum. Meas.*, vol. 47, Oct. 1998, pp. 1259–1264.
24. P. Blockley, D. Gunyan, and J. B. Scott, "Mixer-based, vector-corrected, vector signal/network analyzer offering 300 kHz–20 GHz bandwidth and traceable phase response," in *IEEE MTT-S Int. Microw. Symp. Dig.*, vol. 3, June 2005, pp. 1497–1500.
25. P. Blockley, D. Gunyan, and J. B. Scott, "Wide-bandwidth, vector-corrected measurement with high spurious-free dynamic range," in *65th ARFTG Conf. Dig.*, June 2005.
26. J. Verspecht, *Calibration of a Measurement System for High Frequency Nonlinear Devices*. Belgium: Doctoral Dissertation—Vrije Universiteit Brussel, Nov. 1995. [Online]. Available: http://www.janverspecht.com/

27. J. Verspecht, F. Verbeyst, and M. Vanden Bossche, "Network analysis beyond S-parameters: characterizing and modeling component behavior under modulated large-signal operating conditions," in *56th ARFTG Conf. Dig.*, Dec. 2000.
28. J. Verspecht. "Large-signal network analysis: going beyond S-parameters," Agilent Technologies, Belgium, 2001. [Online]. Available: http://www.janverspecht.com/
29. J. Scott, B. Behnia, M. Vanden Bossche, A. Cognata, J. Verspecht, F. Verbeyst, M. Thorn, and D. R. Scherrer, "Removal of cable and connector dispersion in time-domain waveform measurements on 40 Gb integrated circuits," in *IEEE MTT-S Int. Microw. Symp. Dig.*, vol. 3, June 2002, pp. 1669–1672.
30. S. Vandenplas, J. Verspecht, F. Verbeyst, E. Vandamme, M. Vanden Bossche, "Calibration issues for the large signal network analyzer (LSNA)," in *60th ARFTG Conf. Dig.*, Dec. 2002, pp. 99–106.
31. "Large-signal network analysis technology for HF analogue and fast switching components—theoretical considerations," NMDG Engineering, Belgium, 2004. [Online]. Available: http://www.nmdg.be/
32. F. Verbeyst, *Contributions to Large-Signal Network Analysis.* Belgium: Doctoral Dissertation—Vrije Universiteit Brussel, Sep. 2006. [Online]. Available: http://users.skynet.be/verbeyst.herremans
33. S. Rehnmark, "On the calibration process of automatic network analyzer systems," *IEEE Trans. Microw. Theory Tech.*, vol. 22, Apr. 1974, pp. 457–458.
34. R. A. Speciale, "A generalization of the TSD network-analyzer calibration procedure, covering n-port scattering-parameter measurements, affected by leakage terms," *IEEE Trans. Microw. Theory Tech.*, vol. MTT-25, Dec. 1977, pp. 1100–1115.
35. J. Fitzpatrick, "Error models for systems measurements," *Microwave J.*, vol. 21, May 1978, pp. 63–66.
36. A. A. M. Saleh, "Explicit formulas for error correction in microwave measuring sets with switching-dependent port mismatches," *IEEE Trans. Instrum. Meas.*, vol. IM-28, Mar. 1979, pp. 67–71.
37. G. F. Engen and C. A. Hoer, "'Thru-reflect-line': an improved technique for calibrating the dual six-port automatic network analyzer," *IEEE Trans. Microw. Theory Tech.*, vol. MTT-27, Dec. 1979, pp. 987–993.
38. G. H. Bryant, *Principles of Microwave Measurements.* London: Peter Peregrinus Ltd., 1988.
39. S. Lautzenhiser, A. Davidson, and K. Jones, "Improve accuracy of on-wafer tests via LRM calibration," *Microwaves & RF*, vol. 29, Jan. 1990, pp. 105–109.
40. R. B. Marks, "A multiline method of network analyzer calibration," *IEEE Trans. Microw. Theory Tech.*, vol. 39, July 1991, pp. 1205–1215.
41. J. V. Butler, D. K. Rytting, M. F. Iskander, R. D. Pollard, and M. Vanden Bossche, "16-term error model and calibration procedure for on-wafer network analysis measurements," *IEEE Trans. Microw. Theory Tech.*, vol. 39, Dec. 1991, pp. 2211–2217.
42. H. Heuermann and B. Schiek, "Robust algorithms for Txx network analyzer self-calibration procedures," *IEEE Trans. Instrum. Meas.*, vol. 43, Feb. 1994, pp. 18–23.
43. H. Van Hamme and M. Vanden Bossche, "Flexible vector network analyzer calibration with accuracy bounds using an 8-term or a 16-term error correction model," *IEEE Trans. Microw. Theory Tech.*, vol. 42, June 1994, pp. 976–987.
44. K. Silvonen, "LMR 16—a self-calibration procedure for a leaky network analyzer," *IEEE Trans. Microw. Theory Tech.*, vol 45, July 1997, pp. 1041–1049.
45. G. Vandersteen, Y. Rolain, J. Schoukens, and A. Verschueren, "An improved sliding-load calibration procedure using a semiparametric circle-fitting procedure," *IEEE Trans. Microw. Theory Tech.*, vol 45, July 1997, pp.1027–1033.
46. H. Heuermann and B. Schiek, "15-term self-calibration methods for the error-correction of on-wafer measurements," *IEEE Trans. Instrum. Meas.*, vol. 46, Oct. 1997, pp. 1105–1110.
47. F. Purroy and L. Pradell, "New theoretical analysis of the LRRM calibration technique for vector network analyzers," *IEEE Trans. Instrum. Meas.*, vol. 50, Oct. 2001, pp. 1307–1314.
48. "Applying error correction to network analyzer measurements," Agilent Technologies, Application Note 1287-3, Mar. 2002. [Online]. Available: http://www.agilent.com
49. D. F. Williams, J. C. M. Wang, and U. Arz, "An optimal vector-network-analyzer calibration algorithm," *IEEE Trans. Microw. Theory Tech.*, vol. 51, Dec. 2003, pp. 2391–2401.
50. A. Davidson, K. Jones, and E. Strid, "LRM and LRRM calibrations with automatic determination of load inductance," Cascade Microtech, Beaverton, OR, AR114. [Online]. Available: http://www.cascademicrotech.com/

51. J. Verspecht, "Broadband sampling oscilloscope characterization with the 'Nose-to-Nose' calibration procedure: a theoretical and practical analysis," *IEEE Trans. Instrum. Meas.*, vol. 44, Dec. 1995, pp. 991–997.
52. K. A. Remley, "Nose-to-nose oscilloscope calibration phase error inherent in the sampling circuitry," in *60th ARFTG Conf. Dig.*, Dec. 2002, pp. 85–97.
53. D. F. Williams, P. D. Hale, T. S. Clement, and J. M. Morgan, "Calibrating electro-optic sampling systems," in *IEEE MTT-S Int. Microw. Symp. Dig.*, vol. 3, May 2001, pp. 1527–1530.
54. K. A. Remley, D. M. M. P. Schreurs, D. F. Wiliams, and J. Wood, "Extended NVNA bandwidth for long-term memory measurements," in *IEEE MTT-S Int. Microw. Symp. Dig.*, vol. 3, June 2004, pp. 1739–1742.
55. W. Van Moer and Y. Rolain, "An improved broadband conversion scheme for the large signal network analyzer," in *IEEE MTT-S Int. Microw. Symp. Dig.*, vol. 3, June 2005, pp. 1501–1504.
56. L. Gommé, A. Barel, Y. Rolain, and F. Verbeyst, "Fine frequency grid phase calibration setup for the large signal network analyzer," in *IEEE MTT-S Int. Microw. Symp. Dig.*, vol. 3, June 2006, pp. 1444–1447.
57. G. Simpson, "Using the LSNA in device characterization for PA design," presented at IEEE Topical Workshop on Power Amplifiers for Wireless Communications, San Diego, CA, Sep. 2004.
58. F. Verbeyst and M. Vanden Bossche, "Real-time and optimal PA characterization speeds up PA design," in *34th European Microw. Conf. Dig.*, vol. 1, Oct. 2004, pp. 431–434.
59. F. Verbeyst and M. Vanden Bossche, "Measurement-based behavioral model under mismatched conditions: a new and easy approach for an accurate model," in *2005 European Microw. Conf. Dig.*, vol. 1, Oct. 2005, pp. 605–608.
60. K. A. Remley, D. F. Williams, D. M. M.-P. Schreurs, G. Loglio, and A. Cidronali, "Phase detrending for measured multisine signals," in *61st ARFTG Conf. Dig.*, June 2003, pp. 73–83.
61. M. Myslinski, D. Schreurs, K. A. Remley, M. D. McKinley, and B. Nauwelaers, "Large-signal behavioral model of a packaged RF amplifier based on QPSK-like multisine measurements," in *EGAAS Dig.*, vol. 1, Oct. 2005, pp. 185–188.
62. M. Li, K. M. Gharaibeh, K. G. Gard, and M. B. Steer, "Accurate multisine representation of digital communication signals for characterization of nonlinear circuits," in *Proc. 2006 Radio and Wireless Symp.*, Jan. 2006, pp. 527–530.
63. M. Myslinski, K. A. Remley, M. D. Mckinley, D. Schreurs, and B. Nauwelaers, "A measurement-based multisine design procedure," presented at INMMiC Integrated Non-linear Microwave and Millimetre-wave Circuits Workshop, Aveiro, Portugal, Jan. 2006.
64. K. M. Gharaibeh, K. G. Gard, and M. B. Steer, "In-band distortion of multisines," *IEEE Trans. Microw. Theory Tech.*, vol. 54, Aug. 2006, pp. 3227–3236.
65. J. Verspecht, D. F. Williams, D. Schreurs, K. Remley, and M. D. McKinley, "Linearization of large-signal scattering functions," *IEEE Trans. Microw. Theory Tech.*, vol. 53, Apr. 2005, pp. 1369–1376.
66. "Hot S22 and hot K-factor measurements," Anritsu, Morgan Hill, CA, 11410-00295, Aug. 2002. [Online]. Available: http://www.anritsu.com
67. T. Gasseling, D. Barataud, S. Mons, J.M. Nebus, J.P. Villotte, and R. Quere, "A new characterization technique of 'four hot S parameters' for the study of nonlinear parametric behaviors of microwave devices," in *IEEE MTT-S Int. Microw. Symp. Dig.*, vol. 3, June 2003, pp. 1663–1666.
68. T. Gasseling, D. Barataud, S. Mons, J.-M. Nebus, J. P. Villotte, J. J. Obregon, and R. Quere, "Hot small-signal S-parameter measurements of power transistors operating under large-signal conditions in a load–pull environment for the study of nonlinear parametric interactions," *IEEE Trans. Microw. Theory Tech.*, vol. 52, Mar. 2004, pp. 805–812.
69. J. Verspecht, D. Barataud, J.-P. Teyssier, and J.-M. Nébus, "Hot S-parameter techniques: $6 = 4 + 2$," in *66th ARFTG Conf. Dig.*, Dec. 2005.
70. M. Odyniec, "Large signal S-parameters," in *67th ARFTG Conf. Dig.*, June 2006.
71. "Large-signal network analysis technology for HF analogue and fast switching components—applications," NMDG Engineering, Belgium, 2004. [Online]. Available: http://www.nmdg.be/
72. J. Verspecht, "Large-signal network analysis," *IEEE Microw. Mag.*, vol. 6, Dec. 2005, pp. 82–92.
73. D. Rytting, "Network analyzers from small signal to large signal measurements," in *67th ARFTG Conf. Dig.*, June 2006.

II

Circuits

12

Receivers

Warren L. Seely
Ubidyne, Inc.

An electromagnetic signal picked up by an antenna is fed into a receiver. The ideal receiver rejects all unwanted noise including other signals. It does not add any noise or interference to the desired signal. The signal is converted, regardless of form or format, to fit the characteristics required by the detection scheme in the signal processor, which in turn feeds an intelligible user interface (Figure 12.1). The unit must require no new processes, materials, or devices not readily available. This ideal receiver adds no weight, size, or cost to the overall system. In addition, it requires no power source and generates no heat. It has an infinite operating lifetime in any environment, and will never be obsolete. It will be flexible, fitting all past, present, and future requirements. It will not require any maintenance, and will be transparent to the user, who will not need to know anything about it in order to use it. It will be fabricated in an "environmentally friendly" manner, visually pleasing to all who see it, and when the user is finally finished with this ideal receiver, he will be able to recycle it in such a way that the environment is improved rather than harmed. Above all else, this ideal receiver must be wanted by consumers in very large quantities, and it must be extremely profitable to produce. Fortunately, nobody really expects to achieve all of these "ideal" characteristics, at least not yet! However, each of these characteristics must be addressed by the engineering design team in order to produce the best product for the application at hand.

12.1 Frequency

Receivers represent a technology with tremendous variety. They include AM, FM, analog, digital, direct conversion, single and multiple conversions, channelized, frequency agile, spread spectrum, chirp, frequency hopping, and others. The applications are left to the imaginations of the people who create them. Radio, telephones, data links, radar, sonar, TV, astronomical, telemetry, and remote control are just a few of those applications. Regardless of the application, the selection of the operating frequencies is fundamental to obtaining the desired performance.

The actual receive frequencies are generally beyond the control of the design team, being dictated, controlled, and even licensed by various domestic or foreign government agencies, or by the customer. When a product is targeted for international markets, the allocated frequencies can take on nightmare

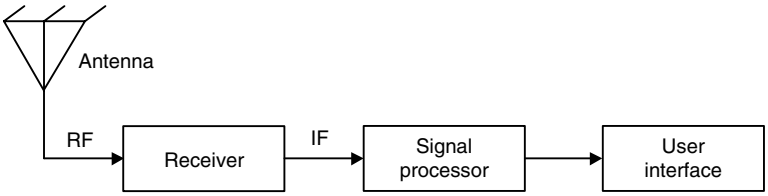


FIGURE 12.1 The receiver.

qualities due to differing allocations, adjacent interfering bands, and neighboring country restrictions or allocations. It will usually prove impossible to get the ideal frequency for any given application, and often the allocated spectrum will be shared with other users and multiple applications. Often the spectrum is available for a price, usually to the highest bidder. Failure to utilize the purchased spectrum within a specified time frame may result in forfeiture of what is now an asset; an expensive mistake. This has opened up the opportunity to speculate and make (or lose) large sums of money by purchasing spectrum to either control a market or resell to other users. For some applications where frequency allocation is up to the user, atmospheric or media absorption, multipathing, and background noise are important factors that must be considered. These effects can be detrimental or used to advantage. An example includes cross links for use with communications satellites, where the cross link is unaffected by absorption since it is above the atmosphere. However, the frequency can be selected to use atmospheric absorption to provide isolation between ground signals and the satellite cross links. Sorting out these problems is time consuming and expensive, but represents a fundamental first step in receiver design.

12.2 Dynamic Range

The receiver should match the dynamic range of the desired signal at the receiver input to the dynamic range of the signal processor. Dynamic range is defined as the range of desirable signal power levels over which the hardware will operate successfully. It is limited by noise, signal compression, and interfering signals and their power levels.

12.2.1 Power and Gain

The power in any signal(s), whether noise, interference or the desired signal, can be measured and expressed in Watts (W), decibels referenced to 1 Watt (dBW), milliwatts (mW) or decibels referenced to one milliwatt (dBm). The power decibel is 10 times the LOG of the dimensionless power ratio. The power gain of a system is the ratio output signal power to the input signal power expressed in decibels (dB). The gain is positive for components in which the output signal is larger than the input, negative if the output signal is smaller. Negative gain is loss, expressed as attenuation (dB). The power gain of a series component chain is found by simple multiplication of the gain ratios, or by summing the decibel gains of the individual components in the chain. All of these relationships are summarized in Figure 12.2.

12.2.2 Noise

Thermal noise arises from the random movement of charge carriers. The thermal noise power (n_T) is usually expressed in dBm (N_T), and is the product of Boltzman's constant (k), system temperature in degrees Kelvin (T), and a system noise bandwidth in Hertz (b_n). The system noise bandwidth (b_n) is defined slightly different from system bandwidth. It is determined by measuring or calculating the total system thermal average noise power ($n_{tot\text{-}ave}$) over the entire spectrum and dividing it by the system peak average noise power ($n_{pk\text{-}ave}$) in a 1 Hz bandwidth. This has the effect of creating a system noise bandwidth in which the noise is all at one level, that of the peak average noise power. For a 1 Hz system noise bandwidth at the input to a system at room temperature (25°C), the thermal noise power is about –174 dBm. These relationships are summarized in Figure 12.3.

$$Decibel = 10\ log\left[\frac{p}{p_{ref}}\right]$$

$$P(dBW) = 10\ log\left[\frac{p(W)}{1W}\right]$$

$$P(dBm) = 10\ log\left[\frac{p(mW)}{1mW}\right]$$

$0\ dBW = 1\ Watt$
$0\ dBm = 1\ mW$
$1000\ mW = 1\ W$
$30\ dBm = 0\ dBW$

$$g(-) = \frac{p_{out}}{p_{in}}$$

$$G(dB) = 10\ log(g)$$

$$g_{\text{total}}(-) = g_1 * g_2 * \ldots * g_N = \frac{p_{Nout}}{p_{1in}}$$

$$G_{\text{total}}(dB) = G_1(dB) + G_2(dB) + \cdots + G_N(dB) = \Sigma G_i$$

$Loss(dB) = -G(dB)$
$Loss(dB) = Attenuation\ (dB)$

FIGURE 12.2 Power and gain relationships.

$$n_T = kTb_n$$

$$b_n(Hz) = \frac{n_{tot\text{-}ave}(W)}{n_{pk\text{-}ave}(W/Hz)}$$

$$k = 1.38 * 10^{-23}\ \frac{W_{\text{sec}}}{K}$$

$$T(K) = T(^\circ C) + 273.15$$

$$n_T = 1.38 * 10^{-23}\ \frac{W_{\text{sec}}}{K} * (25^\circ C + 273.15)K * 1Hz = 4.46 * 10^{-21}\ W = 4.46 * 10^{-18}\ mW$$

$$N_T = 10\ log(n_T) = -204\ dBW = -174\ dBm$$

FIGURE 12.3 Noise power relationships.

12.2.3 Receiver Noise

The bottom end of the dynamic range is set by the lowest signal level that can reasonably be expected at the receiver input and by the power level of the smallest acceptably discernable signal as determined at the input to the signal processor. This bottom end is limited by thermal noise at the input, and by the gain distribution and addition of noise as the signal progresses through the receiver. Once a signal is below the minimum discernable signal (MDS) level, it will be lost entirely (except for specialized spread spectrum receivers). The driving requirement is determined by the signal clarity needed at the signal processor. For analog systems, the signal starts to get fuzzy or objectionably noisy at about 10 dB above the noise floor. For digital systems, the allowable bit error rate determines the acceptable margin above the noise floor. Thus the signal with margin sets the threshold minimum desirable signal level.

Noise power at the input to the receiver will be amplified and attenuated like any other signal. Each component in the receiver chain will also add noise. Passive devices such as filters, cables, and attenuators will cause a drop in both signal and noise power alike. These passive devices also contribute a small amount of internally generated thermal noise. Thus the actual noise figure of a passive device is slightly higher than the attenuation of that component. This slight difference is ignored in receiver design since the actual noise figures and losses vary by significantly larger amounts. Passive mixers will generally have a noise figure about 1 dB greater than the conversion loss. Active devices can exhibit loss or gain, and signal and noise power at the input will experience the same effect when transferred to the output.

$$f_n = \frac{s_i / n_i}{s_o / n_o} = \frac{n_o}{g n_i}$$

$$NF = 10\ log(f_n)$$

$$T_n = T(f_n - 1)\ \textit{where T is in kelvin}$$

$$N_o = NF + G + N_i$$

$$f_t = f_1 + \frac{f_2 - 1}{g_1} + \frac{f_3 - 1}{g_1 * g_2} + \dot{E} + \frac{f_n - 1}{\Pi g_n}$$

$$\Delta f_{n\text{-}bandwidth} = \frac{g_1 f_1 + (f_2 - 1)^{b_{n2} / b_{n1}}}{g_1 f_1 + f_2 - 1} \quad b_{n2} > b_{n1}$$

$$\Delta f_{n\text{-}image} = 1 + \frac{l_{ar}}{f_x}$$

$$f_{total} = f_{cascade} * \Delta f_{n\text{-}bandwidth} * \Delta f_{n\text{-}image}$$

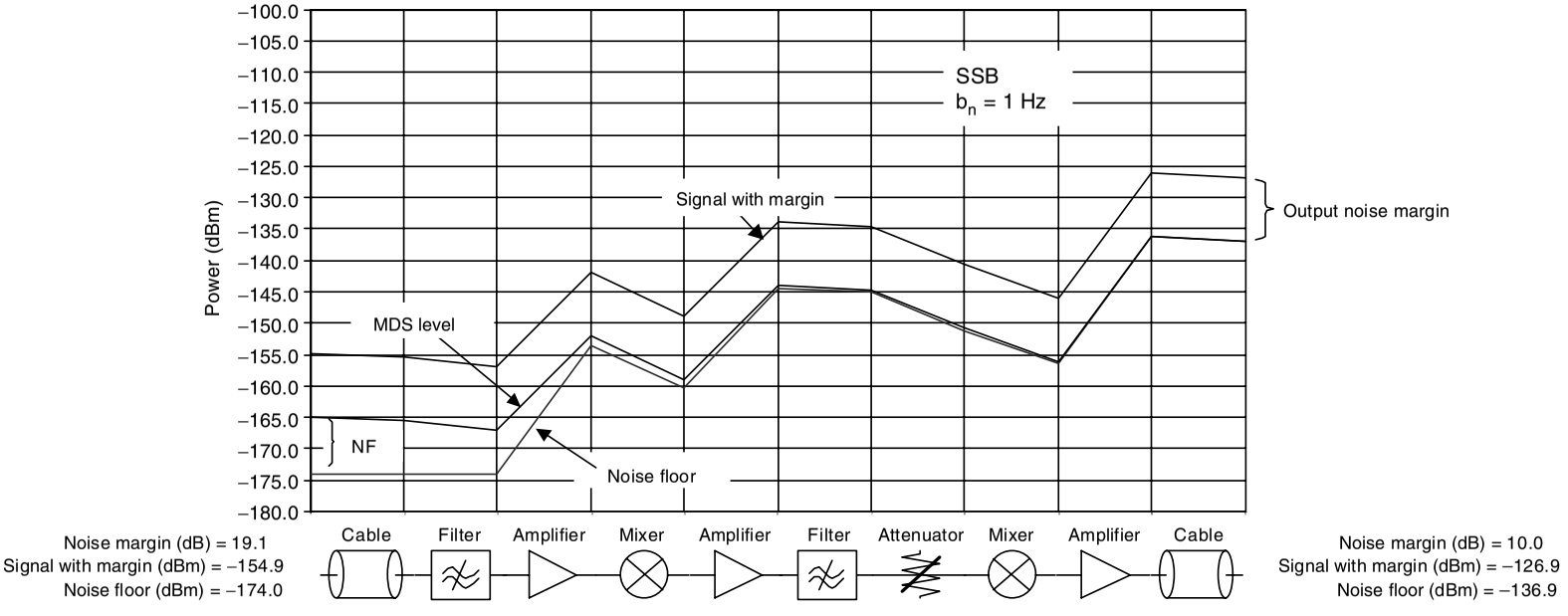
FIGURE 12.4 Receiver noise relationships.

However, the internally generated noise of an active device will be substantial and must be accounted for, requiring reasonably accurate noise figures and gain data on each active component.

The bottom end dynamic range of a receiver component cascade is easily described by the noise equations shown in Figure 12.4. The first three equations for noise factor (f_n), noise figure (NF), and noise temperature (T_n) are equivalent expressions to quantify noise. The noise factor is a dimensionless ratio of the input signal-to-noise ratio and the output signal-to-noise ratio. Replacing the signal ratio with gain results in the final form shown. Noise figure is the decibel form of noise factor, in units of dB. Noise temperature is the conversion of noise factor to an equivalent input temperature that will produce the output noise power, expressed in Kelvin. Convention dictates using noise temperature when discussing antennas and noise figure for receivers and associated electronics. By taking the decibel equivalent of the noise factor, the expression for noise out (N_o) is obtained, where noise in (N_i) is in dBm and noise figure (NF) and gain (G) are in dB. The cascaded noise factor (f_t) is found from the sum of the added noise due to each cascaded component divided by the total gain preceding that element. Use the cascaded noise factor (f_t) followed by the noise out (N_o) equation to determine the noise level at each point in the receiver.

Noise factor is generally computed for a 1 Hz bandwidth and then adjusted for the narrowest filter in the system, which is usually downstream in the signal processor. Occasionally, it will be necessary to account for noise power added to a cascade when components following the narrowest filter have a relatively broad noise bandwidth. The filter will eliminate noise outside its band up to that filter. Broader band components after the filter will add noise back into the system depending on their noise bandwidth. This additional noise can be accounted for using the equation for $\Delta f_{\text{n-bandwidth}}$, where subscript 1 indicates the narrowband component followed by the wideband component (subscript 2). Repeated application of this equation may be necessary if several wideband components are present following the filter. Image noise can be accounted for using the relationship for $\Delta f_{\text{n-image}}$ where l_{ar} is the dimensionless attenuation ratio between the image band and desired signal band and f_x is the noise factor of the system up to the image generator (usually a mixer). Not using an image filter in the system will result in a $\Delta f_{\text{n-image}} = 2$ resulting in a 3 dB increase in noise power. If a filter is used to reject the image by 20 dB, then a substantial reduction in image noise will be achieved. Finally, the corrections for bandwidth and image are easily incorporated using the relationship for the cascaded total noise factor, f_{total}.

A simple single sideband (SSB) receiver example, normalized to a 1 Hz noise bandwidth, is shown in Figure 12.5. It demonstrates the importance of minimizing the use of lossy components near the receiver



Gain (dB)	–0.5	–1.5	15.0	–7.0	15.0	–0.7	–6.0	–5.5	20.0	–0.8
NF (dB)	0.5	1.5	5.5	8.0	8.0	0.7	6.0	6.5	8.0	0.8
Cumilative gain (dB)	–0.5	–2.0	13.0	6.0	21.0	20.3	14.3	8.8	28.8	28.0
Cumilative NF (dB)	0.5	2.0	7.5	7.7	8.6	8.6	8.6	8.7	9.1	9.1
Output signal with margin (dBm)	–155.4	–156.9	–141.9	–148.9	–133.9	–134.6	–140.6	–146.1	–126.1	–126.9
Output noise floor (dBm)	–174.0	–174.0	–153.5	–160.3	–144.4	–145.1	–151.1	–156.5	–136.1	–136.9
Output MDS level (dBm)	–165.4	–166.9	– 151.9	–158.9	–143.9	–144.6	–150.6	–156.1	–136.1	–136.9

FIGURE 12.5 Example SSB receiver noise and signal cascade normalized to $b_n = 1$.

front end, as well as the importance of a good LNA. A 10 dB output signal-to-noise margin has been established as part of the design. Using the –174 dBm input thermal noise level and the individual component gains and noise figures, the normalized noise level can be traced through the receiver, resulting in an output noise power of –136.9 dBm. Utilizing each component gain and working backwards from this point with a signal results in the MDS power level in the receiver. Adding the 10 dB signal-to-noise margin to the MDS level results in the signal with margin power level as it progresses through the receiver. The signal and noise levels at the receiver input and output are indicated. The design should minimize the gap between the noise floor and the MDS level. Progressing from the input toward the output, it is readily apparent that the noise floor gets closer to and rapidly converges with the MDS level due to the addition of noise from each component, and that lossy elements near the input hurt system performance. The use of the low noise amplifier as close to the front end of the cascade as possible is critical in order to mask the noise of following components in the cascade and achieve minimum noise figure. The overall cascaded receiver gain is easily determined by the difference in the signal levels from input to output. The noise floor margins at both the input and output to the receiver are also easily observed, along with the receiver noise figure. Note also that the actual noise power does not drop below the thermal noise floor, which is always the bottom limit. Finally, the actual normalized signal with margin level of –154.9 dBm at the input to the receiver is easily determined.

12.2.4 Intermodulation

Referring to Figure 12.6, the upper end of the dynamic range is limited by nonlinearity, compression, and the resulting intermodulation in the presence of interfering signals. The in-band two-tone output third-order intercept point (3OIP) is a measure of the nonlinearity of a component. This particular product is important because it can easily result in an undesired signal that is close to the desired signal, which would be impossible to eliminate by filtering. By definition, the 3OIP is found by injecting 2 equal amplitude signals (F_1 and F_2) that are not only close to each other in frequency, but are also both within the passband of the component or system. The third-order intermodulation products are then given by $\pm nF_1 \pm mF_2$ where $n + m = 3$. For third-order products, n and m must be 1 or 2. Since negative frequencies are not real, this will result in two different third-order products which are near each other and within the passband. The power in the intermodulation products is then plotted, and both it and Pout are projected until they

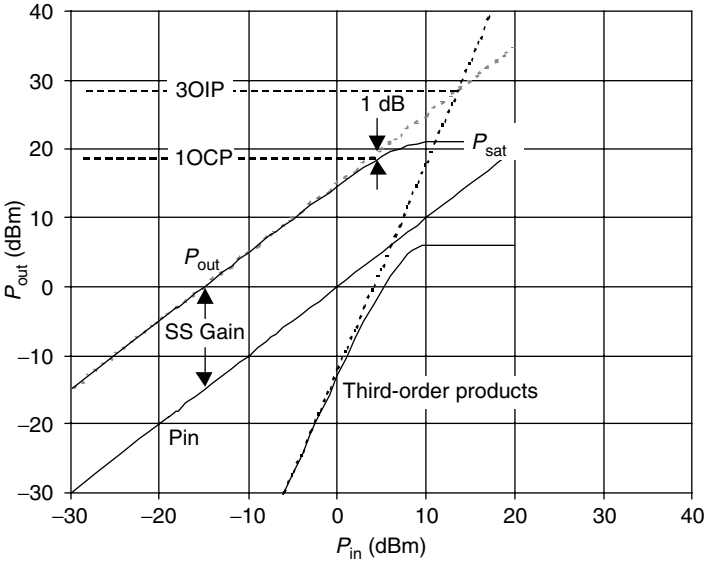


FIGURE 12.6 3OIP, P_{sat}, and 1 dB OCP.

$$\frac{1}{p_{3iip,tot}} = \frac{1}{p_{3iip,1}} + \frac{g_1}{p_{3iip,2}} + \cdots + \frac{\Pi g_n}{p_{3iip,n}}$$

$$P_{3OIP} = P_{3IIP} + G_{ss}$$

$$SFDP = \frac{2}{3}(3IIP - MDS_{input}) = \frac{2}{3}(3OIP - MDS_{output})$$

$$MSI_{out} = NOISEFLOOR_{out} + SPDR$$

FIGURE 12.7 Receiver 3OIP, P_{sat}, and 1 dB OCP cascade.

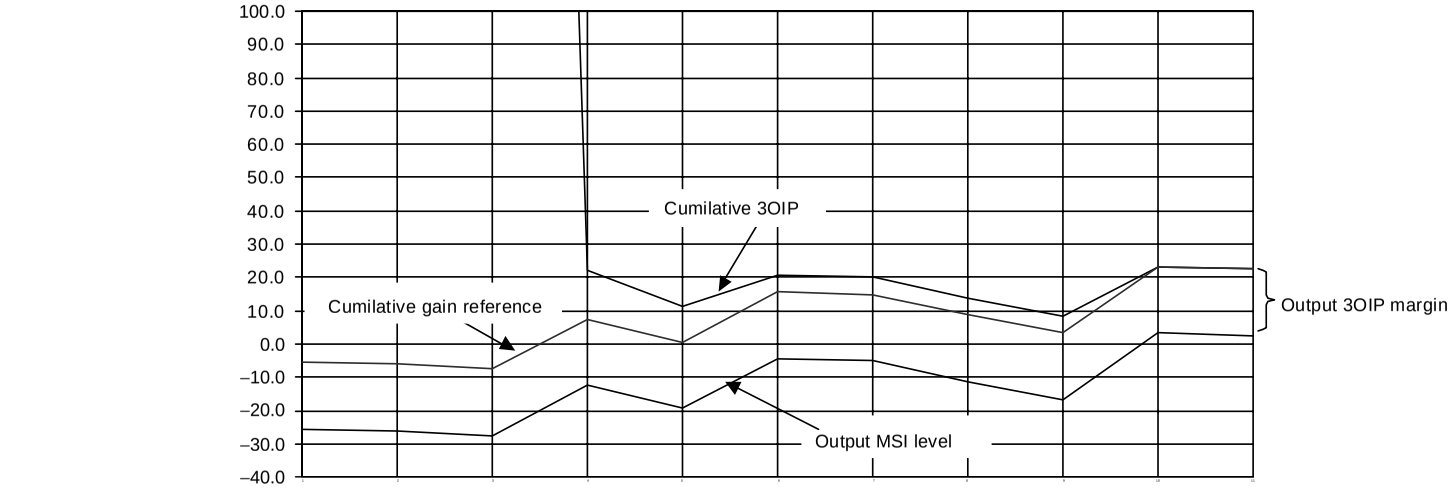
intersect, establishing the 3OIP. The desired signal is projected using a 1:1 slope, while the third-order products are projected using a 3:1 slope. The output saturation power (P_{sat}) is the maximum power a device will produce. The output 1 dB compression point (1 dB OCP) is the point at which the gain is lowered by 1 dB from small signal conditions due to signal compression on its way to complete saturation. In general, higher values mean better linearity and thus better performance. However, component costs rapidly increase along with these numbers, especially above saturated levels of about +15 dBm, limiting what can be achieved within project constraints. Thus, one generally wants to minimize these parameters in order to produce an affordable receiver. For most components, a beginning assumption of square law operation is reasonable. Under these conditions, the 1 dB OCP is about 3 dB below P_{sat}, and the 3OIP is about 10 dB above the 1 dB OCP. When the input signal is very small (i.e., small signal conditions), Pout increases on a 1:1 slope, and third-order products increase on a 3:1 slope. These numbers can vary significantly based on actual component performance and specific loading conditions. This whole process can be reversed, which is where the value of the concept lies. By knowing the small signal gain, Pin or Pout, and the 3OIP, all the remaining parameters, including 1OCP, P_{sat}, and third-order IM levels can be estimated. As components are chosen for specific applications, real data should be utilized where possible. Higher order products may also cause problems, and should be considered also. Finally, any signal can be jammed if the interfering signal is large enough and within the receiver band. The object is to limit the receiver's susceptibility to interference under reasonable conditions.

12.2.5 Receiver Intermodulation

Analog receiver performance will start to suffer when in-band third-order products are within 15 dB of the desired signal at the detector. This level determines the maximum signal of interest (MSI). The margin for digital systems will be determined by acceptable bit error rates. The largest signal that the receiver will pass is determined by the saturated power level of the receiver. Saturating the receiver will result in severe performance problems, and will require a finite time period to recover and return to normal performance. Limiting compression to 1 dB will alleviate recovery.

Analyzing the receiver component cascade for 3OIP, P_{sat}, 1 dB OCP, and MSI will provide insight into the upper limits of the receiver dynamic range, allowing the designer to select components that will perform together at minimum cost and meet the desired performance (Figure 12.7). The first equation handles the cascading of known components to determine the cumulated input 3rd order input intermod point. After utilizing this equation to determine the cascaded 3OIP up to the component being considered, the second equation can be utilized to determine the associated P_{3OIP} . Successive application will result in completely determining the cascaded performance. The last two equations determine the third-order IM spur free dynamic range (SPDR) and the maximum spur-free level or maximum signal of interest (MSI) at the output of the receiver.

An example receiver 3OIP and signal cascade is shown in Figure 12.8. The results of the cumulative 3OIP are plotted. A gain reference is established by setting and determining the cumulative gain and matching it to the cumulative 3OIP at the output. A design margin of 20 dB is added to set the MSI power level for the cascade.



Gain (dB)	–0.5	–1.5	15.0	–7.0	15.0	–0.7	–6.0	–5.5	20.0	–0.8
3IIP (dBm)	999.5	1000.5	7.0	21.0	7.0	999.7	1005.0	28.5	5.0	999.8
3OIP (dBm)	999.0	999.0	22.0	14.0	22.0	999.0	999.0	23.0	25.0	999.0
Cumilative gain (dB)	–0.5	–2.0	13.0	6.0	21.0	20.3	14.3	8.8	28.8	28.0
Cumilative 3IIP (dBm)	999.5	997.2	9.0	5.5	–0.3	–0.3	–0.3	–0.5	–5.5	–5.5
Cumilative 3OIP (dBm)	999.0	995.2	22.0	11.5	20.7	20.0	14.0	8.3	23.3	22.5
Cumilative gain reference (dBm)	–6.0	–7.5	7.5	0.5	15.5	14.8	8.8	3.3	23.3	22.5
Output MSI level (dBm)	–26.0	–27.5	–12.5	–19.5	–4.5	–5.2	–11.2	–16.7	3.3	2.5

FIGURE 12.8 Example receiver 3OIP and signal cascade.

12.2.5 Receiver Dynamic Range

Combining the results for the noise and intermodulation from the above discussion into one graph results in a graphical representation of the dynamic range of the receiver (Figure 12.9). Adjusting for the 6 MHz bandwidth moves the noise plots up by 10 log(6e6) = 67.8 dB. The cumulative 3OIP and gain reference plot remain at the previously determined levels. The SPDR = 2(–12.5 dBm – (–97.1 dBm))/3 = 61.1 dB is calculated and then used to determine the MSI level = –69.1 dBm + 61.1 dB = –8 dBm at the output. The MSI level on the graph is set to this value, backing off to the input by the gain of each component.

The receiver gain, input, and output dynamic ranges, signal levels which can be easily handled, and the appropriate matching signal processing operating range are readily apparent, being between the MSI level and the signal with noise margin. The receiver NF, 3OIP, gain, and SPDR are easily determined from the plot. Weaknesses and choke points, as well as expensive parts are also apparent, and can now be attacked and fixed or improved. In general, components at or near the input to the receiver dominate the noise performance, and thus the lower bounds on dynamic range. Components at or near the output dominate the nonlinear performance, and thus the upper bounds on dynamic range.

The use of 3OIP and noise floors is just one way commonly used to characterize the dynamic range of a receiver. Other methods include determining compression and saturation curves, desensitization, noise power ratios, and intercept analysis for other IM products such as 2OIP up to as high as possibly 15OIP. Specific applications will determine the appropriate analysis required in addition to the main SFDR analysis described above.

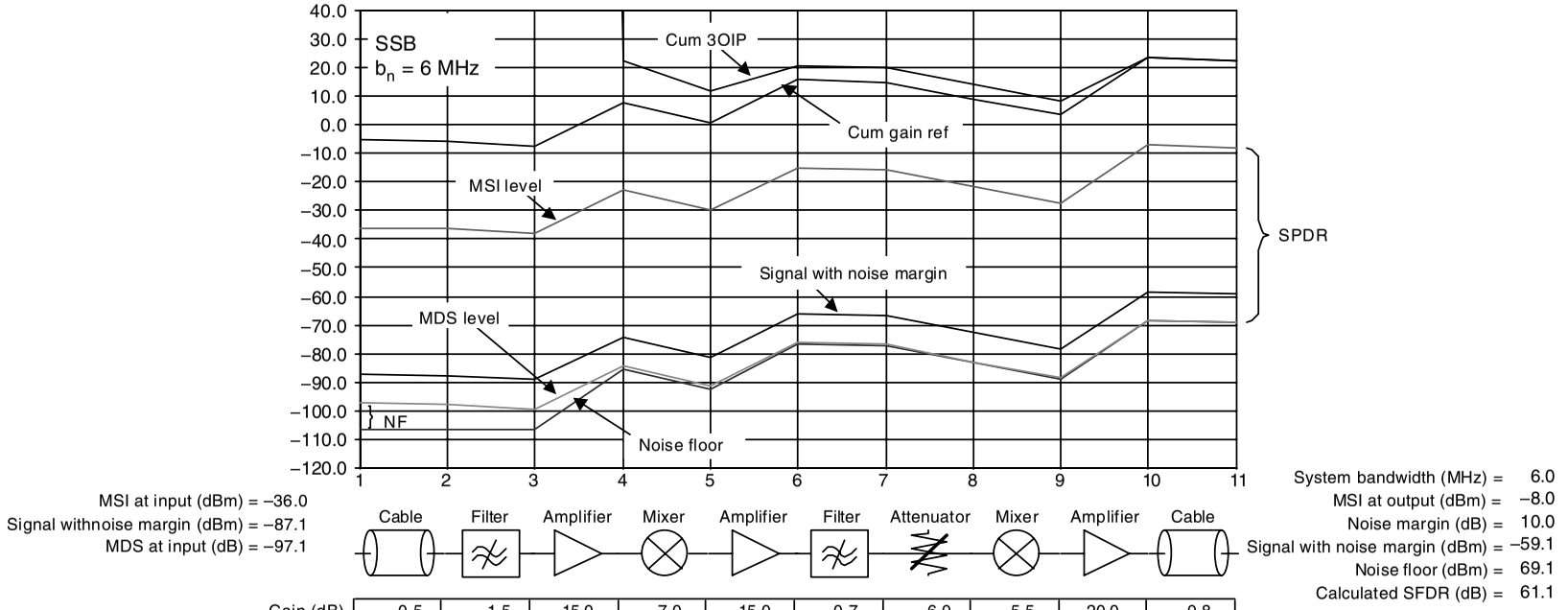
12.3 The Local Oscillator Chain

A reference signal or local oscillator (LO) is generally required in order to up- or down-convert the desired signal for further processing. The design of the LO chain is tied to the receiver components by frequency, power level, and phase noise. The LO signal will have both amplitude and phase noise components, both of which will degrade the desired signal. Often, a frequency agile LO is required. The LO can be generated directly by an oscillator, multiplied or divided from another frequency source, created by mixing several signals, injection or phase locked to a reference source, digitally synthesized, or any combination thereof.

12.3.1 Amplitude and Phase Noise

A pure tone can be represented as a vector of a given amplitude (α) rotating with a fixed angular velocity (ω) as shown in Figure 12.10-1. A random noise source can be viewed similarly, but has random phase equally distributed over time about 360°, and random amplitude based on a probability distribution. At any given instant in time the vector will change amplitude and angle. A plot of the noise vector positions for a relatively long period of time would appear solid near the origin and slowly fade away at larger radii from the origin (Figure 12.10-2). A plot of the number of hits versus distance from the origin would result in the probability distribution. This random noise is the same noise present in all electronic systems. Combining the pure tone with the random noise results in the vector addition of both signals (Figure 12.10-3). At any given instance in time the combined angular velocity will change by $\Delta\omega$, and the amplitude will change by $\Delta\alpha$. The frequency jittering about ω and the amplitude wavering about α result in AM and PM noise components, which distort the signal of interest. Thus, phase noise is a measure of signal stability.

The design of the LO chain usually includes at least one final amplifier stage that is completely saturated. This sets the LO amplitude to a fixed level, minimizes temperature variations, and minimizes or eliminates the AM noise component. Saturation results in a gain reduction of several dB and simultaneously limits the maximum amplitude that can be achieved. Thus as the random noise vector changes the LO amplitude, the saturated amplifier acts to eliminate the output amplitude change. The AM contribution to noise is cleaned up.



Gain (dB)	−0.5	−1.5	15.0	-7.0	15.0	−0.7	−6.0	−5.5	20.0	−0.8
NF (dB)	0.5	1.5	5.5	8.0	8.0	0.7	6.0	6.5	8.0	0.8
3IIP (dBm)	999.5	1000.5	7.0	21.0	7.0	999.7	1005.0	28.5	5.0	999.8
3OIP (dBm)	999.0	999.0	22.0	14.0	22.0	999.0	999.0	23.0	25.0	999.0
Cumilative gain (dB)	−0.5	−2.0	13.0	6.0	21.0	20.3	14.3	8.8	28.8	28.0
Cumilative NF (dB)	0.5	2.0	7.5	7.7	8.6	8.6	8.6	8.7	9.1	9.1
Output signal with margin (dBm)	−87.6	−89.1	−74.1	−81.1	−66.1	−66.8	−72.8	−78.3	−58.3	−59.1
Output noise floor (dBm)	−106.2	−106.2	−85.7	−92.5	−76.6	−77.3	−83.3	-88.7	−68.3	−69.1
Output MDS level (dBm)	−97.6	−99.1	−84.1	−91.1	−76.1	−76.8	−82.8	−88.3	−68.3	−69.1
Cumilative 3IIP (dBm)	999.5	997.2	9.0	5.5	−0.3	−0.3	−0.3	−0.5	−5.5	-5.5
Cumilative 3OIP (dBm)	999.0	995.2	22.0	11.5	20.7	20.0	14.0	8.3	23.3	22.5
Cumilative gain reference (dBm)	−6.0	−7.5	7.5	0.5	15.5	14.8	8.8	3.3	23.3	22.5
Output MSI level (dBm)	−36.5	−38.0	−23.0	−30.0	−15.0	−15.7	−21.7	−27.2	−7.2	−8.0

FIGURE 12.9 Example SSB receiver spur free dynamic range normalized to b_n = 6 MHz.

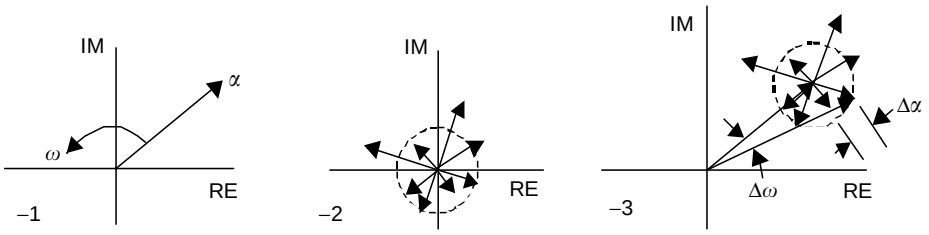


FIGURE 12.10 Phase noise, AM, and PM noise.

The phase noise in the LO chain must be attacked directly at the source. Clean, low phase noise signal generation in oscillators is achieved by the use of very high Q filter components, incorporating bipolar devices as the active oscillator element, maximizing the source power generation, and careful design of the conduction cycle within the oscillator itself. Once a clean signal is created, it must be kept clean.

Frequency multiplying or dividing the signal will also multiply or divide the phase noise by a factor of $20*\log(N)$ at any given offset from the base signal. Conversely, if the signal is multiplied or divided, then the spectrum is respectively stretched or contracted by the factor N. The mixing process also mixes the phase noise, but the net result will depend on the mixer types utilized. Injection locking will replicate the injection source modified by the multiplication or division factor N. A phase lock loop exhibits close in noise dependent on the reference source and loop circuitry, but the far out phase noise is set by the source used in the loop itself. Finally, the LO is utilized in the receiver chain to perform frequency conversion. The resulting converted signal will have components of phase noise from the original signal, the LO signal, and from noise in the mixing component.

12.4 The Potential for Trouble

Receivers are designed to work with very small signals, transforming these to much larger signals for handling in the signal processor. This inherent process is open to many pitfalls resulting in the vast majority of problems encountered in creating a viable product. It will only take a very small interfering or spurious signal to wreak havoc. Interfering signals generally can be categorized as externally generated, internally generated, conducted, electromagnetically coupled, piezoelectrically induced, electromechanically induced, and optically coupled or injected. Some will be fixed, others may be intermittent, even environmentally dependent. Most of these problem areas can be directly addressed by simple techniques, precluding their appearance altogether. However, ignoring these potential problem areas usually results in disaster, primarily because they are difficult to pinpoint as to cause and effect, and because eliminating them may be difficult or impossible without making major design changes and fabricating new hardware in order to verify the solution. This can easily turn into a long-term iterative nightmare. Additionally, if multiple problems are present, whether or not they are perceived as multiple problems or as a single problem, the amount of actual time involved in solving them will go up exponentially! Oscillator circuits are generally very susceptible to any and all problems, so special consideration should be given in their design and use. Finally, although the various cause, effect, and insight into curing problems are broken down into component parts in the following discussion, it is often the case that several concepts must be combined to correctly interpret and solve any particular problem at hand.

12.4.1 Electromechanical

Vibrations and mechanical shocks will result in physical relative movement of hardware. Printed circuit boards (PCBs), walls, and lids may bow or flutter. Cables and wire can vibrate. Connectors can move. Solder joints can fracture. PCBs, walls, and lids capacitively load the receiver circuitry, interconnects, and cabling. Movement, even very small deflections, will change this parasitic loading, resulting in small

changes in circuit performance. In sensitive areas, such as near oscillators and filters, this movement will induce modulation onto the signals present. In phase-dependent systems, the minute changes in physical makeup and hence phase length of coaxial cable will appear as phase modulation. Connector pins sliding around during vibration can introduce both phase and amplitude noise. These problems are generally addressed by proper mechanical design methods, investigating and eliminating mechanical resonance, and minimizing shock susceptibility. Don't forget that temperature changes will cause expansion and contraction, with similar but slower effects.

12.4.2 Optical Injection

Semiconductor devices are easily affected by electromagnetic energy in the optical region. Photons impinging on the surface of an active semiconductor create extra carriers, which appear as noise. A common occurrence of this happens under fluorescent lighting common in many offices and houses. The 60 Hz "hum" is present in the light given off by these fixtures. The light impinges on the surface of a semiconductor in the receiver, and 60 Hz modulation is introduced into the system. This is easily countered by proper packaging to prevent light from hitting optically sensitive components.

12.4.3 Piezoelectric Effects

Piezoelectric materials are reciprocal, meaning that the application of electric fields or mechanical force changes the electromechanical properties, making devices incorporating these materials highly susceptible to introducing interference. Even properly mounted crystals or SAW devices, such as those utilized in oscillators, will move in frequency or generate modulation sidebands when subjected to mechanical vibration and shock. Special care should therefore be given to any application of these materials in order to minimize these effects. This usually includes working closely with the original equipment manufacturer (OEM) vendors to ensure proper mounting and packaging, followed by extensive testing and evaluation before final part selection and qualification.

12.4.4 Electromagnetic Coupling

Proper design, spacing, shielding, and grounding is essential to eliminate coupled energy between circuits. Improper handling of each can actually be detrimental to achieving performance, adding cost without benefit, or delaying introduction of a product while problems are solved. Proper design techniques will prevent inadvertent detrimental E–M coupling. A simple example is a reject filter intended to minimize LO signal leakage into the receiver, where the filter is capable of the required performance, but the packaging and placement of the filter allow the unwanted LO to bypass the filter and get into the receiver anyway.

It is physically impossible to eliminate all E–M resonant or coupled structures in hardware. A transmission line is created by two or more conductors separated by a dielectric material. A waveguide is created by one or more conductive materials in which a dielectric channel is present, or by two or more nonconductive materials with a large difference in relative dielectric constant. Waveguides do not have to be fully enclosed in order to propagate E–M waves. In order to affect the hardware, the transmission line or waveguide coupling must occur at frequencies that will interfere with operation of the circuits, and a launch into the structure must be provided. Properly sizing the package (a resonant cavity) is only one consideration. Breaking up long, straight edges and introducing interconnecting ground vias on multilayer PCBs can be very effective. Eliminating loops, opens and shorts, sharp bends, and any other "antenna like" structures will help.

E-field coupling usually is associated with high impedance circuits, which allow relatively high E-fields to exist. E-field or capacitive coupling can be eliminated or minimized by any grounded metal shielding. M-field coupling is associated with low impedance circuits in which relatively high currents and the associated magnetic fields are present. M-field or magnetic coupling requires a magnetic shielding material. In either case, the objective is to provide a completely shielded enclosure. Shielding metals must be thick enough to attenuate the interfering signals. This can be determined by E or M skin effect

calculations. Alternatively, absorbing materials can also be used. These materials do not eliminate the basic problem, but attempt to mask it, often being very effective, but usually relatively expensive for production environments. Increased spacing of affected circuitry, traces, and wires will reduce coupling. Keeping the E–M fields of necessary but interfering signals orthogonal to each other will add about 20 dB or more to the achieved isolation.

Grounding is a problem that could be considered the "plague" of electronic circuits. Grounding and signal return paths are not always the same, and must be treated accordingly. The subject rates detailed instruction, and indeed entire college level courses are available and recommended for the serious designer. Basically, grounding provides a reference potential, and also prevents an unwanted differential voltage from occurring across either equipment or personnel. In order to accomplish this objective, little or no current must be present. Returns, on the other hand, carry the same current as the circuit, and experience voltage drops accordingly. A return, in order to be effective, must provide the lowest impedance path possible. One way to view this is by considering the area of the circuit loop, and making sure that it is minimized. In addition, the return conductor size should be maximized.

12.5 Summary

A good receiver design will match the maximum dynamic range possible to the signal processor. In order to accomplish this goal, careful attention must be given to the front end noise performance of the receiver and the selection of the low noise amplifier. Equally important in achieving this goal is the linearity of the back end receiver components, which will maximize the SFDR. The basic receiver calculations discussed above can be utilized to estimate the attainable performance. Other methods and parameters may be equally important and should be considered in receiver design. These include phase noise, noise power ratio, higher order intercepts, internal spurious, and desensitization.

Further Reading

Sklar, Bernard, *Digital Communications Fundamentals and Applications*, Prentice-Hall, Englewood Cliffs, NJ, 1988.

Tsui, Dr. James B., *Microwave Receivers and Related Components,* Avionics Laboratory, Air Force Write Aeronautical Laboratories, 1983.

Watkins-Johnson Company Tech-notes, Receiver Dynamic Range: Part 1, Vol. 14, No. 1.

Watkins-Johnson Company Tech-notes, Receiver Dynamic Range: Part 2, Vol. 14, No. 2.

Steinbrecher, D., Achieving Maximum Dynamic Range in a Modern Receiver, *Microwave Journal,* Sept. 1985.

13

Transmitters

Warren L. Seely
Ubidyne, Inc.

A signal is generated by the frequency synthesizer and amplified by the transmitter (Figure 13.1), after which it is fed to the antenna for transmission. Modulation and linearization may be included as part of the synthesized signal, or may be added at some point in the transmitter. The transmitter may include frequency conversion or multiplication to the actual transmit band. An ideal transmitter exhibits many of the traits of the ideal receiver described in Chapter 12. Just as with the receiver, the task of creating a radio transmitter begins with defining the critical requirements, including frequencies, modulations, average and peak powers, efficiencies, adjacent channel power (ACP)or spillover, and phase noise. Additional transmitter parameters that should be considered include harmonic levels, noise powers, spurious levels, linearity, DC power allocations, and thermal dissipations. Nonelectrical, but equally important considerations include reliability, environmentals such as temperature, humidity, and vibration, mechanicals such as size, weight, and packaging or mounting, interfaces, and even appearance, surface textures, and colors. As with most applications today, cost is becoming a primary driver in the design and production of the finished product. For most RF/microwave transmitters, power (PA) amplifier considerations dominate the cost and design concerns.

Safety must also be considered, especially when high voltages or high power levels are involved. To paraphrase a popular educational TV show; "be sure to read and understand all safety related materials that are applicable to your design before you begin. And remember, there is nothing more important than shielding yourself and your coworkers (assuming you like them) from high voltages and high levels of RF/microwave power." Another safety issue for portable products concerns the use of multiple lithium

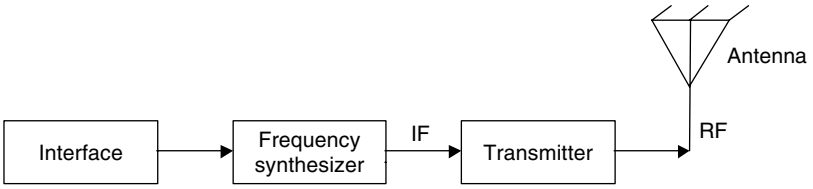


FIGURE 13.1 The transmitter.

batteries connected in parallel. Special care must be taken to insure that the batteries charge and discharge independent of each other in order to prevent excessive I–R heating, which can cause the batteries to explode. It is your life—spend a little time to become familiar with these important issues.

13.1 ACP, Modulation, Linearity, and Power

The transmitter average and peak output powers are usually determined from link/margin analysis for the overall system. The transmitter linearity requirements are determined from the transmit power levels, phase noise, modulation type, filtering, and allowed ACP spillover. Linearity is intimately tied to the transmitter saturated power, which in turn is tied to the 1 dB compression point. The need for high linearity is dependent on the maximum acceptable ACP in adjacent channels or bands. This spillover of power will cause interference in adjacent channels or bands, making it difficult or impossible to use that band. The maximum ACP may be regulated by government agencies, or may be left up to the user. Often the actual transmitted power requirement is less stringent than the linearity requirement in determining the necessary power handling capability or saturated power of the transmitter. In order to achieve required linearity, the transmitter may operate significantly backed off from the saturated power capability, even under peak power operation. Since the cost of a transmitter rapidly increases with its power handling capability, and the linearity requirements are translated into additional power requirements, a great deal of the cost of a transmitter may actually be associated with linearity rather than transmit power level. If the added cost to achieve necessary linearity through additional power capability is significant, linearizing the transmitter can be cost effective.

13.2 Power

The single most important specification affecting the final system cost is often the transmitter saturated power, which is intimately linked to the transmitter linearity requirements. This parameter drives the PA device size, packaging, thermal paths and related cooling methods, power supply, and DC interconnect cable sizes, weight, and safety, each of which can rapidly drive costs upward. The power level analysis may include losses in the cables and antenna, transmitter and receiver antenna gains, link conditions such as distance, rain, ice, snow, trees, buildings, walls, windows, atmospherics, mountains, waves, water towers, and other issues that might be pertinent to the specific application. The receiver capabilities are crucial in determining the transmitter power requirements. Once a system analysis has been completed indicating satisfactory performance, then the actual PA requirements are known. The key parameters to the PA design are frequency, bandwidth, peak and average output power, duty cycle, linearity, gain, bias voltage and current, dissipated power, and reliability mean time to failure (MTBF, usually given as maximum junction temperature). Other factors may also be important, such as power added efficiency (PAE), return losses, isolations, stability, load variations, cost, size, weight, serviceability, manufacturability, etc.

13.3 Linearization

Linearity, as previously indicated, is intimately tied to the transmitter power. The need for high linearity is dependent on the maximum acceptable ACP in adjacent channels or bands. This spillover of power

will cause interference in those bands, making it difficult or impossible to use that band. The ACP spillover is due to several factors, such as phase noise, modulation type, filtering, and transmit linearity. The basic methods used for linearization include the class A amplifier in back-off, feedforward, Cartesian and polar loops, adaptive predistortion, envelope elimination and recovery (EER), linear amplification using nonlinear components (LINC), combined analog locked-loop universal modulation (CALLUM), *I–V* trajectory modification, device tailoring, and Doherty amplification. Each of these methods strives to improve the system linearity while minimizing the overall cost. The methods may be combined for further improvements. Economical use of the methods may require the development of application-specific integrated circuits (ASICs). As demand increases these specialized ICs should become available as building blocks, greatly reducing learning curves, design time, and cost.

13.4 Efficiency

Power added efficiency (η_a or PAE) is the dimensionless ratio of RF power delivered from a device to the load (p_{out}) minus the input incident RF power (p_{incident}) versus the total DC power dissipated in the device (p_{DC}). It is the most commonly used efficiency rating for amplifiers and accounts for both the switching and power gain capabilities of the overall amplifier being considered. High PAE is essential to reducing the overall cost of high power transmitter systems, as previously discussed in the power section above. As with power, *PAE* affects the PA device size, packaging, thermal paths and related cooling methods, power supply, and DC interconnect cable sizes, weight, and safety, each of which can rapidly drive up cost.

$$PAE = \eta_a = \frac{p_{load} - p_{incident}}{p_{DC}}$$

13.5 The *I–Q* Modulator

The *I–Q* modulator is a basic building block used in numerous applications, and is an essential element of many linearization methods. The basic block diagram is shown in Figure 13.2, along with the associated symbol that will be used in the following discussions. The modulator consists of two separate mixers that are driven 90° out of phase with each other to generate quadrature (I and Q) signals. The 90° port is usually driven with the high level bias signal, allowing the mixer compression characteristic to minimize amplitude variation from the 90° hybrid. The configuration is reciprocal, allowing either up- or downconversion.

13.6 Class A Amplifier in Back Off

An amplifier is usually required near the output of any transmitter. The linearity of the amplifier is dependent on the saturated power that the amplifier can produce, the amplifier bias and design, and the characteristics of the active device itself. An estimate of the DC power requirements and dissipation for each stage in the PA chain can be made based on the peak or saturated power, duty cycle, and linearity requirements. The maximum or saturated power (P_{sat} in dBW or dBm, depending on whether power is in W or mW) can be estimated (Figure 13.3) from the product of the RMS voltage and current swings across the RF load $(V_{\max} - V_{\text{on}})/2\sqrt{2}$ and $I_{\text{on}}/2\sqrt{2}$, and from the loss in the output matching circuits (L_{out} in dB). As previously discussed in the receiver section, the 3OIP is about 6 dB above the saturated power for a square law device, but can vary by as much as 4 dB lower to as much as 10 dB higher for a given actual device. Thus it is very important to determine the actual 3OIP for a given device, using vendor data, simulation, or measurement. One must take into account the effects of transmitter components both prior to and after the PA, utilizing the same analysis technique used for receiver intermodulation. The ACP output intercept point (AOIP) will be closely correlated to the 3OIP, and will act in much the same way, except for the actual value of the AOIP. The delta between the AOIP and 3OIP will be modulation dependent, and must be determined through simulation or measurement at this time. Once

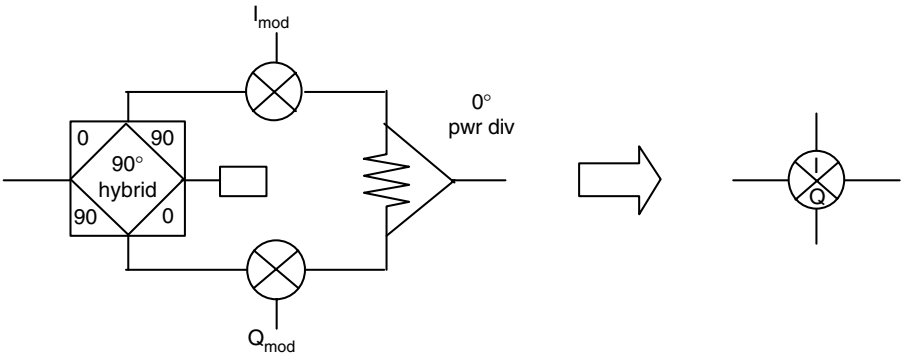


FIGURE 13.2 *I–Q* modulator block diagram.

$$R_{on} = \frac{V_{on}}{I_{on}}$$

$$P_{sat} = 10Log\left[\frac{(V_{max} - V_{on})I_{on}}{8}\right] - L_{out}$$

$$P_{dis} = P_{DC} + P_{in} - P_{out}$$

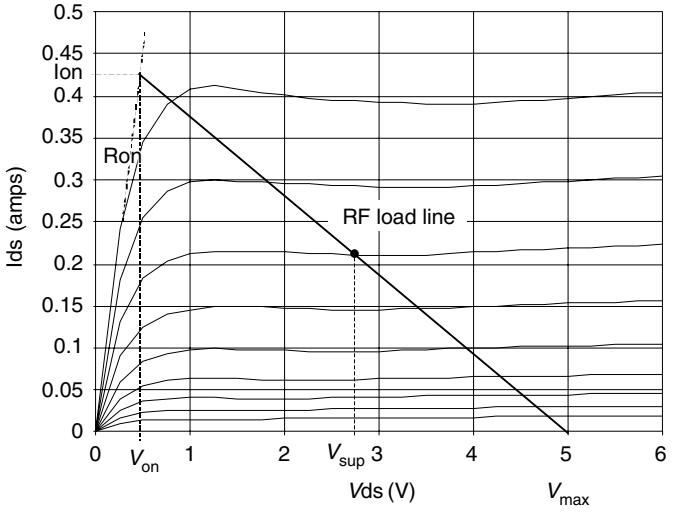


FIGURE 13.3 Class A amplifier saturated power estimation.

this has been determined, it is a relatively easy matter to determine the required back off from P_{sat} for the output amplifier, often as much as 10 to 15 dB. Under these conditions, amplifiers that require a high intercept but in which the required peak power is much lower that the peak power that is available, linearization can be employed to lower the cost.

13.7 Feed Forward

Although the feedforward amplifier (Figure 13.4) is a simple concept, it is relatively difficult to implement, especially over temperature and time. The applied signal is amplified to the desired power level by the PA, whose output is sampled. The PA introduces distortion to the system. A portion of the input signal is delayed and then subtracted from the sampled PA signal, nulling out the original signal, leaving only the unwanted distortion created by the PA. This error signal is then adjusted for amplitude and recombined with the distorted signal in the output coupler, canceling out the distortion created by the PA. The resulting linearity improvement is a function of the phase and amplitude balances maintained, especially over temperature and time. The process of generating an error signal will also create nonlinearities, which will limit the ultimate improvements that are attainable, and thus are a critical part of the design.

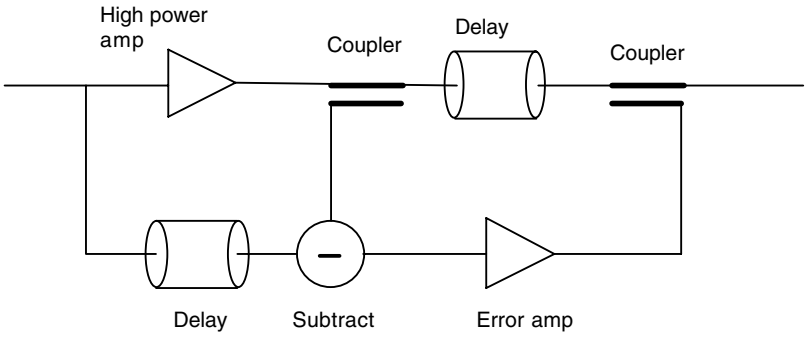


FIGURE 13.4 Feedforward amplifier.

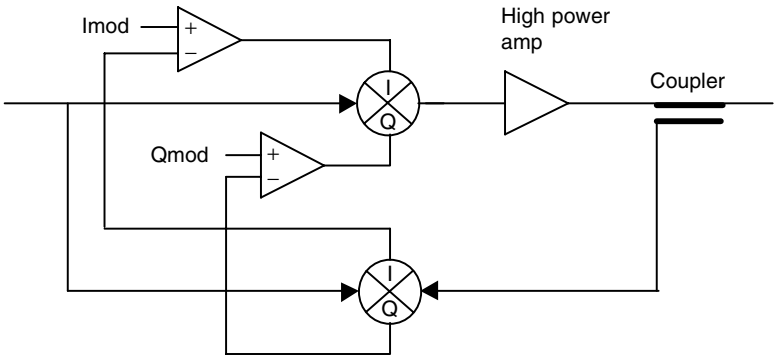


FIGURE 13.5 Cartesian loop.

13.8 Cartesian and Polar Loops

The Cartesian loop (CL) (Figure 13.5) is capable of both good efficiency and high linearity. The efficiency is primarily determined by the amplifier efficiency. The loop action determines the linearity achieved. A carrier signal is generated and applied to the input of the CL, where it is power divided and applied to two separate *I*–*Q* mixers. The high power carrier path is quadrature modulated and then amplified by the output PA, after which the distorted modulated signal is sampled by a coupler. The sample distorted signal is then demodulated by mixing it with the original unmodulated carrier, resulting in distorted modulation in quadrature or *I*–*Q* form. These distorted *I* and *Q* modulations are then subtracted from the original *I* and *Q* modulation to generate error *I* and *Q* modulation (hence the name Cartesian), which will continuously correct the nonlinearity of both the PA and the *I*–*Q* modulator. Loop gain and phase relationships are critical to the design, and as with any feedback scheme, care must be taken to prevent loop oscillation. The *I*–*Q* modulators and the sampling coupler utilize 90-degree power dividers with limited bandwidth over which satisfactory performance can be attained. The loop delay ultimately will limit the attainable bandwidth to about 10%. Even with these difficulties, the CL is a popular choice. Much of the circuitry is required anyway, and can be easily integrated into ASICs, resulting in low production costs.

Whereas the Cartesian loop depends on quadrature I and Q signals, the related polar loop uses amplitude and phase to achieve higher linearity. The method is much more complex since the modulation correction depends on both frequency modulating the carrier as well as amplitude modulating it. Ultimately the performance will be worse than that of the Cartesian loop, as well as being more costly by a considerable margin. For these reasons, it is not used.

13.9 Fixed Predistortion

Fixed predistortion methods are conceptually the simplest form of linearization. A PA will have nonlinearities that distort the original signal. By providing complimentary distortion prior to the PA, the predistorted signal is linearized by the PA. The basic concept can be divided into digital and transfer characteristic methods, both with the same objective. In the digital method (Figure 13.6), digital signal processing (DSP) is used to provide the required predistortion to the signal. This can be applied at any point in the system, but is usually provided at baseband where it can be cheaply accomplished. The information required for predistortion must be determined and then stored in memory. The DSP then utilizes this information and associated algorithms to predistort the signal, allowing the PA to correct the predistortion, resulting in high linearity. When hardware is used to generate the predistortion, the predistorting transfer characteristic must be determined, and appropriate hardware must be developed. There are no algorithms or methods to accomplish this, so it can be a formidable task. In either case, the improvements in linearity are limited by the lack of any feedback to allow for deviations from the intended operation, and by the ability to actually determine and create the required predistortion. In short, it is cheap, but don't expect dramatic results!

13.10 Adaptive Predistortion

Linearization by adaptive predistortion (Figure 13.7) is very similar to fixed methods, with the introduction of feedback in the form of an error function that can be actively minimized on a continuous basis. The ability to change under operational conditions requires some form of DSP. The error signal is generated in the same way used for Cartesian loop systems, but is then processed by the DSP, allowing the DSP to

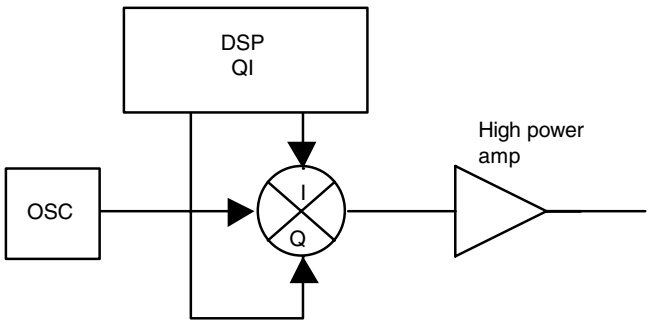


FIGURE 13.6 Fixed digital predistortion.

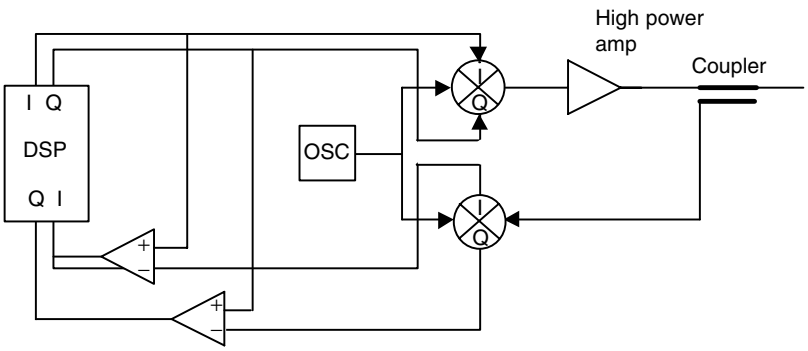


FIGURE 13.7 Adaptive predistortion.

minimize the error by modifying or adapting the applied predistortion. The disadvantages in this method center on the speed of the DSP and the inability of the system to react due to loop delay. It must see the error before it can correct for it.

13.11 Envelope Elimination and Recovery

The highly efficient EER amplifier (Figure 13.8) accepts a fully modulated signal at its input and power divides the signal. One portion of the signal is amplitude detected and filtered to create the low-frequency AM component of the original signal. The other portion of the signal is amplitude limited to strip off or eliminate all of the AM envelope, leaving only the FM component or carrier. Each of these components is then separately amplified using high-efficiency techniques. The amplified AM component is then utilized to control the FM amplifier bias, modulating the amplified FM carrier. Thus the original signal is recovered, only amplified. While this process works very well, an alternative is available that utilizes the DSP capabilities to simplify the whole process, cut costs, and improve performance. In a system, the input half of the EER amplifier can be eliminated and the carrier FM modulated directly by DSP-generated tuning of a voltage-controlled oscillator (VCO). The DSP-generated AM is amplified and used to control the FM amplifier bias. The result is the desired modulated carrier.

13.12 Linear Amplification Using Nonlinear Components

The LINC transmitter (Figure 13.9) concept is quite simple. The DSP creates two separate amplitude and phase-modulated signals, each in quadrature (*I*–*Q*) format. These signals are upconverted by *I*–*Q* modulators to create two separate phase-modulated signals that are separately applied to high-efficiency output power amplifiers. The amplified FM signals are then combined at the output, the signals being such that all of the unwanted distortion is cancelled by combining 180° out of phase, and all of the

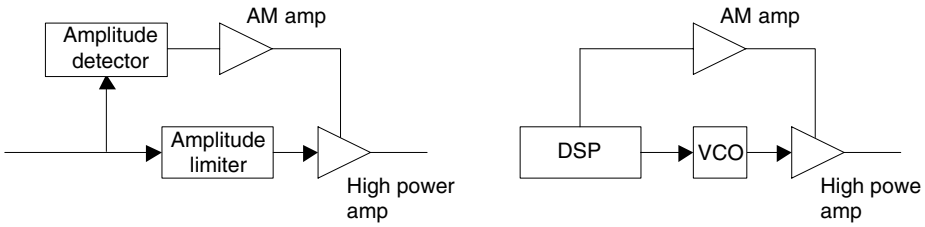


FIGURE 13.8 Envelope elimination and recovery.

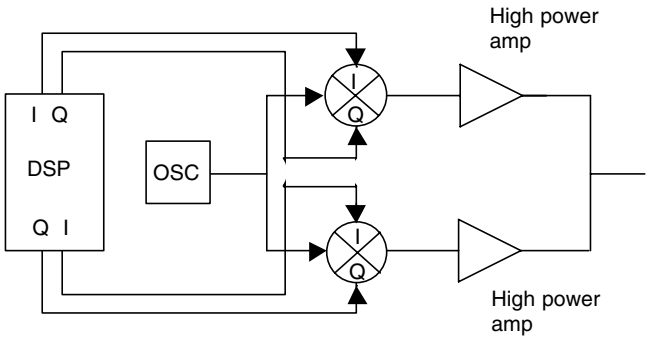


FIGURE 13.9 LINC transmitter.

desired signal components are added by combining in phase. The challenge in this method is in the DSP generation of the original pair of quadrature signals required for the desired cancellation and combination at the output of the transmitter. Another area of concern with this linearization method is the requirement for amplitude and phase matching of the two channels, which must be tightly controlled in order to achieve optimum performance.

13.13 Combined Analog Locked-Loop Universal Modulation

The CALLUM linearization method is much simpler than it looks at first glance (Figure 13.10). Basically, the top portion of the transmitter is the LINC transmitter discussed above. An output coupler has been added to allow sampling of the output signal, and the bottom half of the diagram delineates the feedback method that generates the two quadrature pairs of error signals in the same way as used in the Cartesian loop or EER methods. This feedback corrects for channel differences in the basic LINC transmitter, substantially improving performance. Since most of the signal processing is performed at the modulation frequencies, the majority of the circuit is available for ASIC implementation.

13.14 *I–V* Trajectory Modification

In *I–V* trajectory or cyclic modification, the idea is to create an active *I–V* characteristic that changes with applied signal level throughout each signal cycle, resulting in improved linear operation (see device tailoring below). A small portion of the signal is tapped off or sampled at the input or output of the amplifier and, based on the continuously sampled signal amplitude, the device bias is continuously modified at each point in the signal cycle. The power range over which high PAE is achieved will be compressed. This method requires a good understanding of the PA device, and excellent modeling. Also, the sampling and bias modification circuitry must be able to react at the same rate or frequencies as the PA itself while providing the required voltage or current to control the PA device. Delay of the sampled

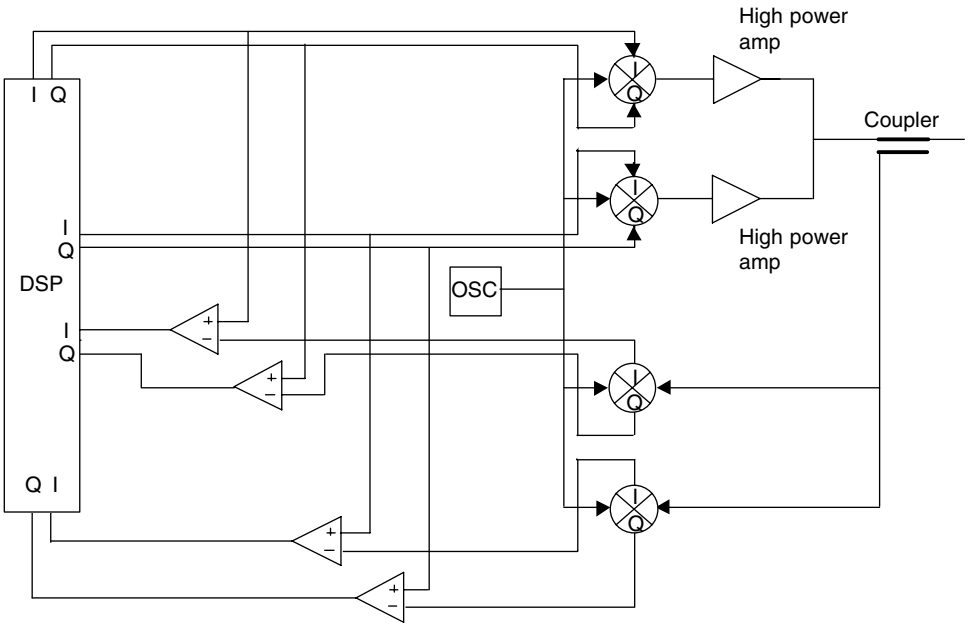


FIGURE 13.10 CALLUM.

signal to the time the bias is modified is critical in obtaining performance. This method is relatively cheap to implement, and can be very effective in improving linearity.

13.15 Doherty Amplification

The simplified form of the Doherty amplifier, which maintains high PAE over a much wider power range than a single amplifier, is shown in Figure 13.11. In the low power path, a 90° phase shifter is used to compensate for the 90° phase shifter/impedance inverter required in the high power path. The low PA is designed to operate efficiently at a given signal level. The class C high power under low power conditions does not turn on, and thus represents high impedance at the input. The high power 90° phase shifter/impedance inverter provides partial matching for the low PA under these conditions. As the signal level increases, the low PA saturates, the class C high PA turns on, and the power of both amplifiers sum at the output. Under these conditions the high power 90° phase shifter/impedance inverter matches the high PA to the load impedance. Although the modulation bandwidths are not a factor in this technique, the bandwidth is limited by the phase and amplitude transfer characteristics of the 90° elements. This concept can be extended by adding more branches, or by replacing the low PA with a complete Doherty amplifier in itself.

13.16 Device Tailoring

For designers with access to a flexible semiconductor foundry service, linearity can be improved directly at the device level. The most obvious way of accomplishing this is by modifying the semiconductor doping to achieve the desired linearity while maintaining other performance parameters. In the ideal device, both the real and reactive device impedance would remain constant and linear (constant derivatives) as the *I–V* load line or trajectory is traversed for increasing amplitude signals. This linear operation would continue up to the signal amplitude at which both clipping and pinch-off simultaneously occur (ideal biasing). Thus the ideal device would be perfectly linear for any signal below P_{sat}. The objective should be to come as close to this ideal as possible in order to maximize linearity. At best this is a difficult task involving a great deal of device and process engineering. Another strategy might involve trying to minimize the amplitude-dependent parasitic effects such as leakage currents and capacitive or charge-related problems. A third strategy would be to modify the linearity by paralleling two or more devices of varying design together resulting in the desired performance. This is relatively easy to implement through device layout, with best results achieved when this is accomplished at the lowest level of integration (i.e., multiple or tapered gate lengths within a single gate finger, or a stepped or tapered base layer). The results can be quite dramatic with respect to linearity, and best of all the added recurring cost is minimal. As with trajectory modification, the power range for high efficiency operation is compressed.

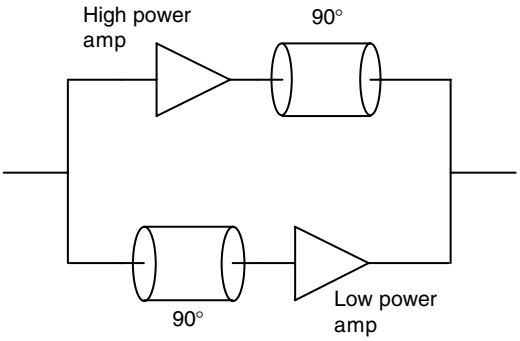


FIGURE 13.11 Doherty amplifier.

13.17 Summary

The key transmitter parameters of ACP, modulation, linearity, and power are all tightly correlated. These parameters must be determined early in transmitter design so that individual component parameters can be determined and flow-down specifications can be made available to component designers. Linear operation is essential to controlling the power spill-over into adjacent channels (ACP). The basic linearization methods commonly used have been described. These include the class A amplifier in back-off, feedforward, Cartesian and polar loops, adaptive predistortion EER, LINC, CALLUM, *I–V* trajectory modification, device tailoring, and Doherty amplification. Combining methods in such a way as to take advantage of multiple aspects of the nonlinear problem can result in very good performance. An example might be the combination of a Cartesian loop with device tailoring. Unfortunately, it is not yet possible to use simple relationships to calculate ACP directly from linearity requirements, or conversely, required linearity given the ACP. The determination of these requirements is highly dependent on the modulation being used. However, simulators are available that have the capability to design and determine the performance that can be expected.

Further Reading

Bateman, A., Haines, D., and Wilkinson, R., Linear Transceiver Architectures, Communications Research Group, University of Bristol, England.

Boloorian, M. and McGeeham, J., The Frequency-Hopped Cartesian Feedback Linear Transmitter, *IEEE Transactions on Vehicular Technology,* 45, 4, 1996.

Casadevall, F., The LINC Transmitter, *RF Design,* Feb. 1990.

Correlation Between P1db and ACP in TDMA Power Amplifiers, *Applied Microwave & Wireless,* Mar. 1999.

Kenington, P., Methods Linearize RF Transmitters and Power Amps, Part 1, *Microwaves & RF,* Dec. 1998.

Kenington, P., Methods Linearize RF Transmitters and Power Amps, Part 2, *Microwaves & RF,* Jan. 1999.

Sundstrom, L. and Faulkner, M., Quantization Analysis and Design of a Digital Predistortion Linearizer for RF Power Amplifiers, *IEEE Transactions on Vehicular Technology,* 45, 4, 1996.

Zavosh, F., Runton, D., and Thron, C., Digital Predistortion Linearizes CDMA LDMOS Amps, *Microwaves & RF,* Mar. 2000.

14

Low Noise Amplifier Design

Jakub Kucera
Urs Lott
AnaPico AG

Signal amplification is a fundamental function in all wireless communication systems. Amplifiers in the receiving chain that are closest to the antenna receive a weak electric signal. Simultaneously, strong interfering signals may be present. Hence, these low noise amplifiers mainly determine the system noise figure and intermodulation behavior of the overall receiver. The common goals are therefore to minimize the system noise figure, provide enough gain with sufficient linearity, and assure a stable 50 Ω input impedance at a low power consumption.

14.1 Definitions

This section introduces some important definitions used in the design theory of linear RF and microwave amplifiers. Further, it develops some basic principles used in the analysis and design of such amplifiers.

14.1.1 Gain Definitions

Several gain definitions are used in the literature for high-frequency amplifier designs.

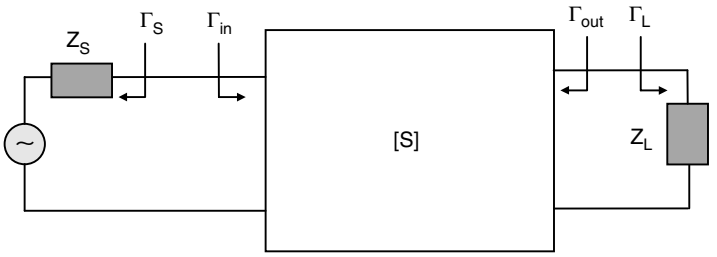


FIGURE 14.1 Amplifier block diagram. Z_S: source impedance, Z_L: load impedance, Γ_S: source reflection coefficient, Γ_{in}: input reflection coefficient, Γ_{out}: output reflection coefficient, Γ_L: load reflection coefficient.

The transducer gain G_T is defined as the ratio between the effectively delivered power to the load and the power available from the source. The reflection coefficients are shown in Figure 14.1.

$$G_T = \frac{1-|\Gamma_S|^2}{|1-\Gamma_S \cdot S_{11}|^2} \cdot |S_{21}|^2 \cdot \frac{1-|\Gamma_L|^2}{|1-\Gamma_L \cdot \Gamma_{OUT}|^2}$$

The available gain G_{AV} of a two-port is defined as the ratio of the power available from the output of the two-port and the power available from the source.

$$G_{AV} = \frac{1-|\Gamma_S|^2}{|1-\Gamma_S \cdot S_{11}|^2} \cdot |S_{21}|^2 \cdot \frac{1}{|1-\Gamma_{OUT}|^2} \quad \text{with } \Gamma_{OUT} = S_{22} + \frac{S_{12} \cdot S_{21} \cdot \Gamma_S}{1-\Gamma_S \cdot S_{11}}$$

The entire available power at one port can be transferred to the load, if the output is terminated with the complex conjugate load.

The available gain G_{AV} is a function of the two-port scattering parameters and of the source reflection coefficient, but independent of the load reflection coefficient Γ_L. The available gain gives a measure for the maximum gain into a conjugately matched load at a given source admittance.

The associated gain G_{ASS} is defined as the available gain under noise matching conditions.

$$G_{ASS} = \frac{1-|\Gamma_{opt}|^2}{|1-\Gamma_{opt} \cdot S_{11}|^2} \cdot |S_{21}|^2 \cdot \frac{1-|\Gamma_L|^2}{|1-\Gamma_L \cdot \Gamma_{OUT}|^2}$$

14.1.2 Stability and Stability Circles

The stability of an amplifier is a very important consideration in the amplifier design and can be determined from the scattering parameters of the active device, the matching circuits, and the load terminations (see Figure 14.1). Two stability conditions can be distinguished: unconditional and conditional stability.

Unconditional stability of a two-port means that the two-port remains stable (i.e., does not start to oscillate) for any passive load at the ports. In terms of the reflection coefficients, the conditions for

unconditional stability at a given frequency are given by the following equations:

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

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

$$\left|\Gamma_S\right| < 1 \text{ and } \left|\Gamma_L\right| < 1$$

In terms of the scattering parameters of the two-port, unconditional stability is given, when

$$K = \frac{1-\left|S_{11}\right|^2 - \left|S_{22}\right|^2 + \left|\Delta\right|^2}{2\left|S_{12}\cdot S_{21}\right|} > 1$$

and

$$\left|\Delta\right| < 1$$

with $\Delta = S_{11}\cdot S_{22} - S_{12}\cdot S_{21}$. K is called the stability factor [1].

If either $|S_{11}| > 1$ or $|S_{22}| > 1$, the network cannot be unconditionally stable because the termination $\Gamma_L = 0$ or $\Gamma_S = 0$ will produce or $|\Gamma_{IN}| > 1$ or $|\Gamma_{OUT}| > 1$.

The maximum transducer gain is obtained under simultaneous conjugate match conditions $\Gamma_{IN} = \Gamma_S^*$ and $\Gamma_{OUT} = \Gamma_L^*$. Using

$$\Gamma_{IN} = S_{11} + \frac{S_{12}\cdot S_{21}\cdot \Gamma_L}{1-\Gamma_L\cdot S_{22}} \text{ and } \Gamma_{OUT} = S_{22} + \frac{S_{12}\cdot S_{21}\cdot \Gamma_S}{1-\Gamma_S\cdot S_{11}}$$

a closed-form solution for the source and load reflection coefficients Γ_S and Γ_L can be found. However, a simultaneous conjugate match having unconditional stability is not always possible if $K < 1$ [2].

Conditional stability of a two-port means that for certain passive loads (represented as $\Gamma_L < 1$ or $\Gamma_S < 1$) oscillation may occur. These values of Γ_L and Γ_S can be determined by drawing the stability circles in a Smith chart. The source and load stability circles are defined as

$$\left|\Gamma_{IN}\right| = 1 \text{ and } \left|\Gamma_{OUT}\right| = 1$$

On one side of the stability circle boundary, in the Γ_L plane, $|\Gamma_{IN}| > 1$ and on the other side $|\Gamma_{IN}| < 1$. Similarly, in the Γ_S plane, $|\Gamma_{OUT}| > 1$ and on the other side $|\Gamma_{OUT}| < 1$. The center of the Smith chart ($\Gamma_L = 0$) represents a stable operating point, if $|S_{11}| < 1$, and an unstable operating point, if $|S_{11}| > 1$ (see Figure 14.2). Based on these observations, the source and load reflection coefficient region for stable operation can be determined.

With unconditional stability, a complex conjugate match of the two-port is possible. The resulting gain then is called the maximum available gain (MAG) and is expressed as

$$MAG = \left|\frac{S_{21}}{S_{12}}\right| \cdot \left(K - \sqrt{K^2 - 1}\right)$$

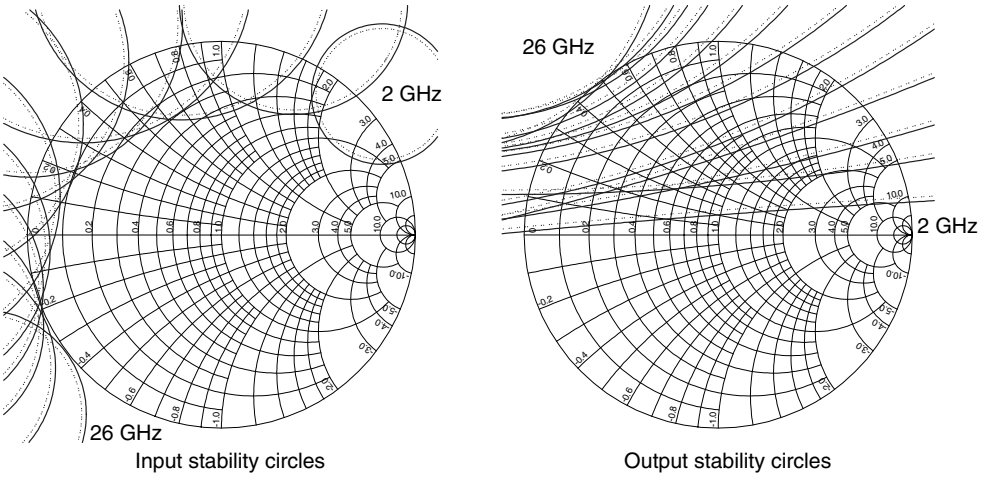


FIGURE 14.2 Source (input) and load (output) stability circles in the Smith chart for the MESFET NE710 over the frequency range from 2 to 26 GHz. Unstable region is indicated by the dotted line.

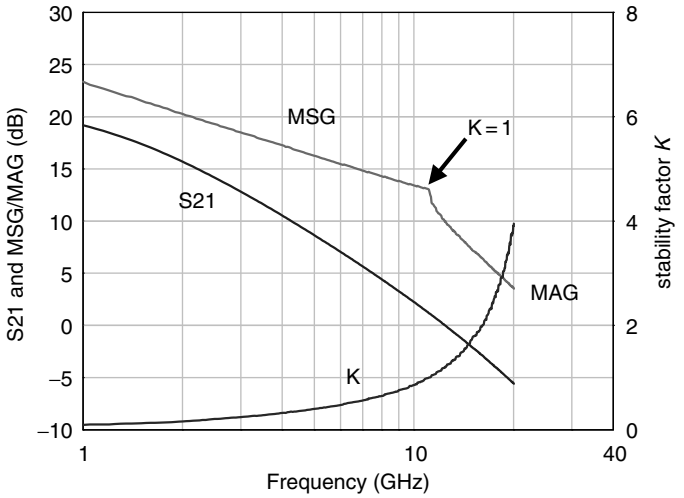


FIGURE 14.3 Maximum stable gain (MSG), maximum available gain (MAG), S_{21}, and stability factor K for a typical MESFET device with 0.6 μm gate length.

The maximum stable gain MSG is defined as the maximum transducer gain for which $K = 1$ holds, namely

$$MSG = \left|\frac{S_{21}}{S_{12}}\right|$$

MSG is often used as a figure of merit for potentially unstable devices (see Figure 14.3).

It must be mentioned, however, that the stability analysis as presented here in its classical form, is applicable only to a single-stage amplifier. In a multistage environment, the above stability conditions are insufficient, because the input or output planes of an intermediate stage may be terminated with

active networks. Thus, taking a multistage amplifier as a single two-port and analyzing its *K*-factor is helpful, but does not guarantee overall stability. Literature on multistage stability analysis is available [3].

14.1.3 Representation of Noise in Two-Ports

The LNA can be represented as a noise-free two-port and two partly correlated input noise sources i_n and v_n as shown in Figure 14.4. The partial correlation between the noise sources i_n and v_n can be described by splitting i_n into a fully correlated part i_c and a noncorrelated part i_u as

$$i_n = i_c + i_u$$

The fully correlated part i_c is defined by the correlation admittance [4] Y_{cor}

$$i_c = Y_{cor} v_n$$

The source impedance $Z_s = R_s + jB_s$ shows thermal noise i_s which depends on the bandwidth as

$$\overline{i_s^2} = 4kTG_s\Delta f$$

Finally, the noise factor F can be expressed in terms of these equivalent input noise generators as

$$F = 1 + \left|Y_s + Y_{cor}\right|^2 \frac{\overline{v_n^2}}{\overline{i_s^2}} + \frac{\overline{i_u^2}}{\overline{i_s^2}}$$

Details on the use of the correlation matrix and the derivation of the noise factor from the noise sources of the two-port can be found in References 4 and 5.

14.1.4 Noise Parameters

The noise factor F of a noisy two-port is defined as the ratio between the available signal-to-noise power ratio at the input to the available signal-to-noise ratio at the output.

$$F = \frac{S_{in}}{N_{in}} \Big/ \frac{S_{out}}{N_{out}}$$

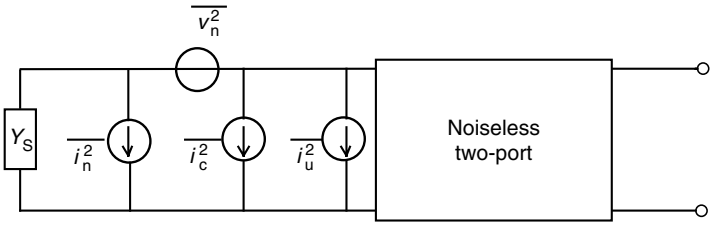


FIGURE 14.4 Representation of noisy two-port as noiseless two-port and two partly correlated noise sources (v_n and $i_c + i_u$) at the input.

The noise factor of the two-port can also be expressed in terms of the source admittance $Y_s = G_s + jB_s$ as

$$F = F_{min} + \frac{R_n}{G_s}\left|Y_s - Y_{opt}\right|^2$$

where F_{min} is the minimum achievable noise factor when the optimum source admittance $Y_{opt} = G_{opt} + jB_{opt}$ is presented to the input of the two-port, and R_n is the equivalent noise resistance of the two-port. Sometimes the values Y_s, Y_{opt}, and R_n are given relative to the reference admittance Y_0.

The noise performance of a two-port is fully characterized at a given frequency by the four noise parameters F_{min}, R_n, and real and imaginary parts of Y_{opt}.

Several other equivalent forms of the above equation exist, one of them describing F as a function of the source reflection coefficient Γ_s.

$$F = F_{min} + \frac{4R_n}{Z_0}\frac{\left|\Gamma_s - \Gamma_{opt}\right|^2}{\left|1+\Gamma_{opt}\right|^2 \cdot \left(1-\left|\Gamma_s\right|^2\right)}$$

When measuring noise, the noise factor is often represented in its logarithmic form as the noise figure NF

$$NF = 10\log F$$

Care must be taken not to mix up the linear noise factor and the logarithmic noise figure in noise calculations.

14.1.5 Noise Circles

Noise circles refer to the contours of constant noise figure for a two-port when plotted in the complex plane of the input admittance of the two-port. The minimum noise figure is presented by a dot, while for any given noise figure higher than the minimum, a circle can be drawn. This procedure is adaptable in the source admittance notation as well as in the source reflection coefficient notation. Figure 14.5 shows the noise circles in the source reflection plane.

Noise circles in combination with gain circles are efficient aids for circuit designers when optimizing the overall amplifier circuit network for low noise with high associated gain.

14.1.6 Friis Formula: Cascading Noisy Two-Ports

When several noisy two-ports are connected in cascade, the overall noise characteristics are described by [5]

$$F_{tot} = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 \cdot G_2} + \ldots + \frac{F_i - 1}{G_1 \cdot G_2 \ldots \cdot G_{i-1}}$$

where F_i and G_i are noise factor and available gain of the ith two-port. The available gain depends on the output admittance of the previous stage.

14.1.7 Noise Measure *M*

The overall noise factor of an infinite number of identical cascaded amplifiers is $F = 1 + M$ with

$$M = \frac{F-1}{1-\frac{1}{G}}$$

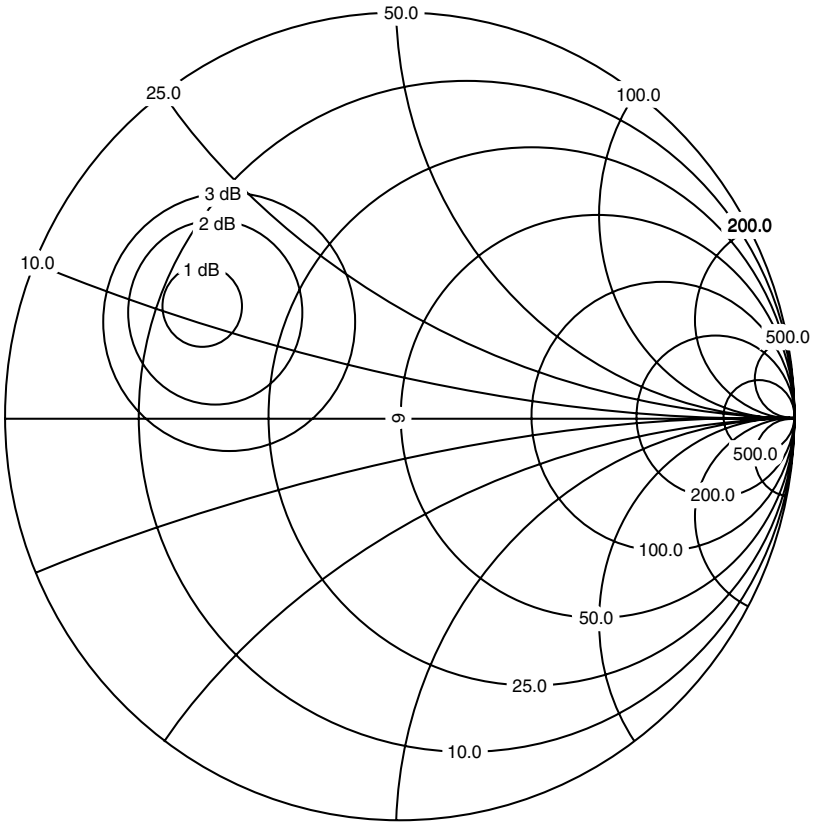


FIGURE 14.5 Noise circles in the input reflection coefficient plane.

M is here called the noise measure. The noise measure is useful for comparing the noise performance of devices or LNAs with different power gains.

14.2 Design Theory

The apparent structural simplicity of an LNA with its relatively few components is misleading. The design should be easy, but the trade-offs complicate the design. A simultaneous noise and power matching involves a more complicated matching network and the achievable dynamic range is often limited by the given low supply voltage and the maximum allowed current consumption. The LNA must provide enough gain so that the noise contributions from the following components become small. But the maximum tolerable gain is limited by the linearity requirements of the following receiver chain.

The most important design considerations in a high-frequency amplifier are stability, power gain, bandwidth, noise, and DC power consumption requirements.

A systematic mathematical solution, aided by graphical methods, is developed to determine the input and output matching network for a particular noise, gain, stability, and gain criteria. Unconditionally stable designs will not oscillate with any passive termination, while designs with conditional stability require careful analysis of the loading to assure stable operation.

14.2.1 Linear Design Procedure for Single-Stage Amplifiers

1. *Selection of device and circuit topology:* Select the appropriate device based on required gain and noise figure. Also decide on the circuit topology (common-base/gate or common-emitter/source).

The most popular circuit topology in the first stage of the LNA is the common-emitter (source) configuration. It is preferred over a common-base (-gate) stage because of its higher power gain and lower noise figure. The common-base (-gate) configuration is a wideband unity-current amplifier with low input impedance ($\approx 1/g_m$) and high predominantly capacitive output impedance. Wideband LNAs requiring good input matching use common-base input stages. At high frequencies the input impedance becomes inductive and can be easily matched.

2. *Sizing and operating point of the active device:* Select a low noise DC operating point and determine scattering and noise parameters of the device. Typically, larger input transistors biased at low current densities are used in low noise designs. At RF frequencies and at a given bias current, unipolar devices such as MOSFET, MESFET, and HEMT are easier to match to 50 Ω when the device width is larger. Both, (hetero-) BJTs and FETs show their lowest intrinsic noise figure when biased at approximately one tenth of the specified maximum current density. Further decreasing the current density will increase the noise figure and reduce the available gain.
3. *Stability and RF feedback:* Evaluate stability of the transistor. If only conditionally stable, either introduce negative feedback (high-resistive DC parallel or inductive series feedback) or draw stability circles to determine loads with stable operation.
4. *Select the source and load impedance:* Based on the available power gain and noise figure circles in the Smith chart, select the load reflection coefficient Γ_L that provides maximum gain, the lowest noise figure (with $\Gamma_S = \Gamma_{opt}$), and good VSWR. In unconditionally stable designs, Γ_L is

$$\Gamma_L = \left(S_{22} + \frac{S_{12} \cdot S_{21} \cdot \Gamma_{opt}}{1 - \Gamma_{opt} \cdot S_{11}} \right)^*$$

In conditionally stable designs, the optimum reflection coefficient Γ_S may fall into an unstable region in the source reflection coefficient plane. Once Γ_S is selected, Γ_L is selected for the maximum gain $\Gamma_L = \Gamma_{OUT}$, and Γ_L must again be checked to be in the stable region of the load reflection coefficient plane.

5. *Determine the matching circuit:* Based on the required source and load reflection coefficients, the required ideal matching network can be determined. Depending on the center frequency, lumped elements or transmission lines will be applied. In general, there are several different matching circuits available. Based on considerations about insertion loss, simplicity, and reproducibility of each matching circuit, the best selection can be made.
6. *Design the DC bias network:* A suitable DC bias network is crucial for an LNA, which should operate over a wide temperature and supply voltage range and compensate parameter variations of the active device. Further, care must be given that no excessive additional high-frequency noise is injected from the bias network into the signal path, which would degrade the noise figure of the amplifier. High-frequency characteristics including gain, noise figure, and impedance matching are correlated to the device's quiescent current. A resistor bias network is generally avoided because of its poor supply rejection. Active bias networks are capable of compensating temperature effects and rejecting supply voltage variations and are therefore preferred.

For bipolar circuits, a simple grounded emitter DC bias network is shown in Figure 14.6a. The high-resistive bias network uses series feedback to stabilize the current of the active device against device parameter variations. However, the supply rejection of this network is very poor, which limits its applicability. A bypassed emitter resistor is often used at low frequencies to stabilize the DC bias point (Figure 14.6b). At RF and microwave frequencies, the bypass capacitor can cause unwanted high-frequency instability and must be applied with care. Furthermore, an emitter resistor will degrade the noise figure performance of the amplifier, if the resistor is not fully bypassed at the signal frequency.

More advanced active bias circuits use (temperature compensated) bandgap references and generate a reference current, which is mirrored to the amplifying device through a high value

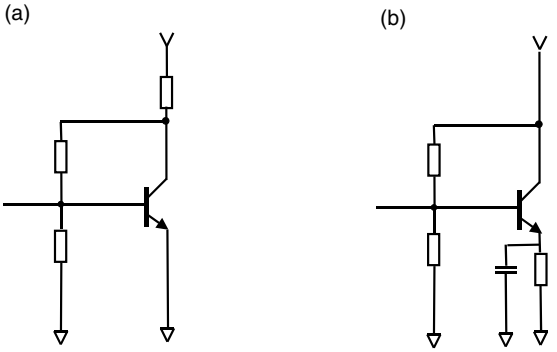


FIGURE 14.6 Passive bias network for bipolar amplifiers.

resistor or an RF choke to minimize noise injection. Another popular method is to generate a proportional to absolute temperature (PTAT) current source. The amplifier gain is proportional to the transconductance which itself is proportional to the collector current and inversely proportional to the temperature ($g_m = qI_c/kT$). With the transistor biased with a current proportional to temperature, the gain remains roughly constant over temperature. Combining bandgap circuits with PTAT sources leads to excellent supply and temperature variation suppression [7].

The implementation of appropriate bias methods for FET amplifiers is generally more involved. The most critical parameter affecting the bias point is the threshold voltage. Stable voltage reference and PTAT current sources are typically based on the Schottky diode barrier height and involve rather sophisticated circuitry [8].

7. *Optimize entire circuit with lossy matching elements:* The final design optimization of the LNA circuit must include the nonidealities of the matching elements, parasitic components such as bond wire inductance, as well as fabrication tolerance aspects. This last design phase today is usually performed on a computer-aided design (CAD) system.

   The dominant features of an LNA (gain, noise, matching properties) can be simulated with excellent accuracy on a linear CAD tool. The active device is characterized by its scattering parameters in the selected bias point, and the four noise parameters. The passive components are described by empirical or equivalent circuit models built into the linear simulation tool. If good models for elements like millimeter-wave transmission lines are not available, these elements must be described by their measured scattering parameters, too.

   Alternatively, a nonlinear simulator with a full nonlinear device model allows direct performance analysis over varying bias points. The use of nonlinear CAD is mandatory for compression and intermodulation analysis.

   Advanced CAD tools allow for direct numerical optimization of the circuit elements toward user-specified performance goals. However, these optimizers should be used carefully, because it can be very difficult to transform the conflicting design specifications into optimization goals. In many cases, an experienced designer can optimize an LNA faster by using the "tune" tools of a CAD package.

14.3 Practical Design of a Low Noise Amplifier

The last section presented a design procedure to design a stable low noise amplifier based on linear design techniques. In practice, there are nonidealities and constraints on component sizing that typically degrade the amplifier performance and complicate the design. In fact, the presented linear design method does not take power consumption versus linearity explicitly into account. Some guidelines are provided in this section that may facilitate the design.

14.3.1 Hybrid versus Monolithic Integrated LNA

With the current trend to miniaturized wireless devices, LNAs are often fabricated as monolithic integrated circuits, usually referred to as monolithic microwave integrated circuit) (MMIC). High volume applications such as cell phones call for even higher integration in the RF front end. Thus the LNA is integrated together with the mixer, local oscillator, and sometimes even parts of the transmitter or the antenna [9].

Depending on the IC technology, monolithic integration places several additional constraints on the LNA design. The available range of component values may be limited, in particular, the maximum inductance and capacitance values are often smaller than required. Integrated passive components in general have lower quality factors Q because of their small size. In some cases, the first inductor of the matching circuit must be realized as an external component.

The electromagnetic and galvanic coupling between adjacent stages is often high due to the close proximity of the components. Furthermore, the lossy and conducting substrate used in many silicon-based technologies increases coupling. At frequencies below about 10 GHz, transmission lines cannot be used for matching because the required chip area would make the IC too expensive, at least for commercial applications.

Finally, monolithic circuits cannot be tuned in production. On the other hand, monolithic integration also has its advantages. The placement of the components is well controlled and repeatable, and the wiring length between components is short. The number of active devices is almost unlimited and adds very little to the cost of the LNA. Each active device can be sized individually.

For applications with low volume where monolithic integration is not cost effective, LNAs can be built as hybrid circuits, sometimes called MIC (microwave integrated circuit). A packaged transistor is mounted on a ceramic or organic substrates. The matching circuit is realized with transmission lines or lumped elements. Substrates such as alumina allow very high quality transmission line structures to be fabricated. Therefore, in LNAs requiring ultimate performance, for example, for satellite ground stations, hybrid circuit technology is sometimes used even if monolithic circuits are available.

14.3.2 Multistage Designs

Sometimes a single amplifier stage cannot provide the required gain and multiple gain stages must be provided. Multiple gain stages complicate the design considerably. In particular, the interstage matching must be designed carefully (in particular in narrowband designs) to minimize frequency shifts and ensure stability. The ground lines of the different gain stages must often be isolated from each other to avoid positive feedback, which may cause parasitic oscillations. Moreover, some gain stages may need some resistive feedback to enhance stability.

Probably the most widely used multistage topology is the cascode configuration. A low noise amplifier design that uses a bipolar cascode arrangement as shown in Figure 14.7 offers performance advantages in wireless applications over other configurations. It consists of a common-emitter stage driving a common-base stage. The cascode derives its excellent high-frequency properties from the fact that the collector load of the common-emitter stage is the very low input impedance of the common-base stage. Consequently, the Miller effect is minimal even for higher load impedances and an excellent reverse isolation is achieved. The cascode has high output impedance, which may become difficult to match to 50 Ω. A careful layout of the cascode amplifier is required to avoid instabilities. They mainly arise from parasitic inductive feedback between the emitter of the lower and the base of the upper transistor. Separating the two ground lines will enhance the high-frequency stability considerably.

14.3.3 Stability Considerations

Figure 14.8 shows a possible strategy for a stable design of a two-stage amplifier. Separated ground and supply lines of the two gain stages minimize positive feedback. RC parallel feedback further enhances in-band stability. Low frequency oscillations caused through unstable bias lines can be attenuated by adding small resistors and blocking capacitors into the supply line.

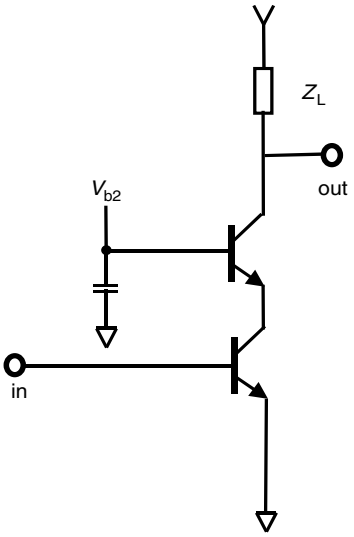


FIGURE 14.7 Cascode amplifier.

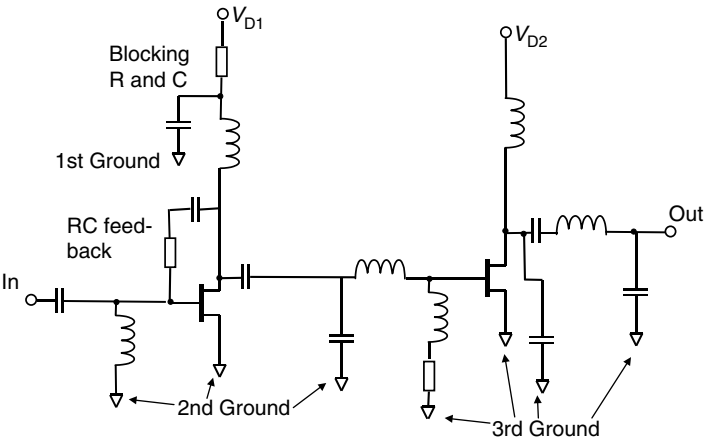


FIGURE 14.8 Design for stability.

14.3.4 Feedback

Negative feedback is widely used in amplifier design to stabilize gain against parameter changes in the active device due to supply voltage variations and temperature changes. RF feedback is used in many LNAs to ensure high-frequency stability and make noise and power match coincident. A well-known technique is adding inductance at the emitter (source) of the active device. The inductance L interacts with the base-emitter (gate-source) capacitance C_{IN} and device transconductance g_m to produce a resistive component to the input impedance $g_m(L/C_{IN})$, while no additional noise source is introduced (except for the parasitic series resistance of the inductor). Neglecting the Miller capacitance, the input impedance of an inductively degenerated FET stage is

$$Z_{IN} = \frac{1}{j\omega C_{IN}} + j\omega L + g_m \frac{L}{C_{IN}}$$

This method of generating a real term to the input impedance is preferable to resistive methods as only negligible additional noise is introduced. Moreover, the inductance has the helpful side effect of shifting the optimum noise match closer to the complex conjugate power match and reducing the signal distortion. However, the benefits are accompanied by a gain reduction.

14.3.5 Impedance Matching

Following the design procedure in the last section, the conditions for a conjugate match at the input and output ports are satisfied at one frequency. Hence, reactive matching inherently leads to a narrowband design. The input bandwidth is given by

$$BW = \frac{f_0}{Q_{IN}}$$

where f_0 is the center frequency and Q_{IN} is the quality factor of the input matching network. The bandwidth can be increased by increasing the capacitance or decreasing the inductance of the matching network.

Using multistage impedance transformators (lumped element filters or tapers) can broaden the bandwidth, but there is a given limit for the reflection coefficient-bandwidth product using reactive elements [10]. In reality, each matching element will contribute some losses, which directly add to the noise figure.

Select an appropriate matching network based on physical size and quality factor (transmission line length, inductance value): long and high-impedance transmission lines show higher insertion loss. Thus, simple matching typically leads to a lower noise figure.

At higher microwave and millimeter-wave frequencies, balanced amplifiers are sometimes used to provide an appropriate noise and power match over a large bandwidth.

14.3.6 Temperature Effects

Typically, LNAs must operate over a wide temperature range. As transistor transconductance is inversely proportional to the absolute temperature, the gain and amplifier stability may change considerably. When designing LNAs with S-parameters at room temperature, a stability margin should be included to avoid unwanted oscillations at low temperatures, as the stability tends to decrease.

14.3.7 Parasitics

Parasitic capacitance, resistance, or inductance can lead to unwanted frequency shifts, instabilities, or degradation in noise figure and gain and rarely can be neglected. Hence, accurate worst-case simulations with the determined parasitics must be made. Depending on the frequency, the parasitics can be estimated based on simple analytical formulas, or must be determined using suitable electromagnetic field-simulators.

14.4 Design Examples

In this section a few design examples of recently implemented low noise amplifiers for frequencies up to 5.8 GHz are presented. They all were manufactured in commercial IC processes.

14.4.1 A Fully Integrated Low Voltage, Low Power LNA at 1.9 GHz [11]

Lowest noise figure can only be achieved when minimizing the number of components contributing to the noise while simultaneously maximizing the gain of the first amplifier stage. Any resistive matching and loading will degrade the noise figure and dynamic behavior and increase power consumption.

In GaAs MESFET processes, the semi-insulating substrate and thick metallization layers allow passive matching components such as spiral inductors and metal-insulator-metal (MIM) capacitors with high quality factors. These lumped passive components are ideally suited for integrated impedance matching

at low GHz frequencies. A fully integrated matching network improves the reproducibility and saves board space while it increases expensive chip area.

It is generally known that GaAs MESFETs have excellent minimum noise figures in the lower GHz frequency range. Still few designs achieve noise figures close to the transistor F_{min}. In fact, several factors prevent F_{min} being attained in practice. If a small input device is employed, a large input impedance transformation involving large inductance values is required. Larger MMIC inductors have higher series resistance and consequently introduce more noise. Further a simultaneous noise and power match often needs additional inductive source degeneration, again introducing noise and reducing gain. For a given maximum power dissipation, very large devices, in contrast, must be biased at very low current densities at which the F_{min} and the gain are degraded. Consequently a trade-off must be made for an optimum design.

The employed GaAs technology features three types of active devices: an enhancement and two depletion MESFETs with different threshold voltages. The enhancement device has a higher maximum available gain, a slightly lower minimum noise figure, but somewhat higher distortion compared to the depletion type. Another advantage of the enhancement FET is that a positive gate bias voltage can be used, which greatly simplifies single-supply operation.

Preliminary simulations are performed using a linear simulator based on measured S- and noise parameters of measured active devices at various bias points and using a scalable large signal model within a harmonic balance simulator in order to investigate the influence of the transistor gate width on the RF performance. The current consumption of the transistor is set at 5.5 mA independent of the gate width. The simulations indicate a good compromise between gain, NF, and intermodulation performance at a gate width of 300 μm (Figure 14.9).

The LNA schematic is shown in Figure 14.10. The amplifier consists of a single common-source stage, which uses a weak inductive degeneration at the source (the approximately 0.3 nH are realized with several parallel bondwires to ground). The designed amplifier IC is fabricated in a standard 0.6 μm E/D MESFET foundry process. The matching is done on chip using spiral inductors and MIM capacitors. The complete LNA achieves a measured 50 Ω noise figure of 1.1 dB at 1.9 GHz with an associated gain of 16 dB at a very low supply voltage of $V_{dd} = 1\ V$ and a total current drain of $I_{dd} = 6$ mA. Figure 14.11 depicts the measured gain and 50 Ω noise figure versus frequency.

The low voltage design with acceptable distortion performance and reasonable power gain can only be achieved using a reactive load with almost no voltage drop.

Figure 14.12 shows, respectively, the supply voltage and supply current dependence of the gain and noise figure. As can be seen, the amplifier still achieves 10 dB gain and a 1.35 dB noise figure at a supply voltage of only 0.3 V and a total current consumption of 2.3 mA. Sweeping the supply voltage from 1 to 5 volts, the gain varies less than 0.5 dB and the noise figure less than 0.15 dB, respectively. IIP_3 and –1 dB compression point are also insensitive to supply voltage variations as shown in Figure 14.13.

Below 1 V, however, the active device enters the linear region resulting in a much higher distortion.

Finally, the input and output matchings are measured for the LNA. At the nominal 1 V supply, the input and output return loss are –8 dB and –7 dB, respectively.

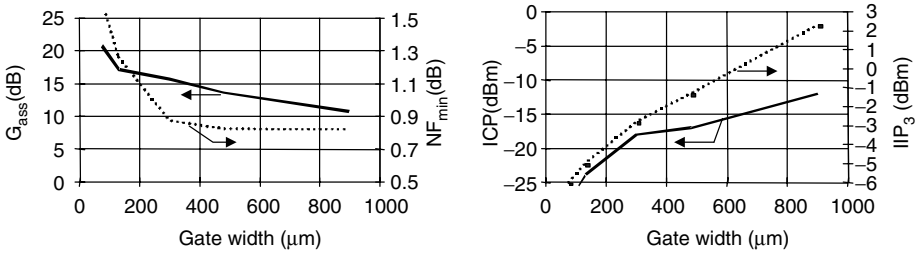


FIGURE 14.9 Simulated performance of an enhancement FET versus device width at a constant power dissipation.

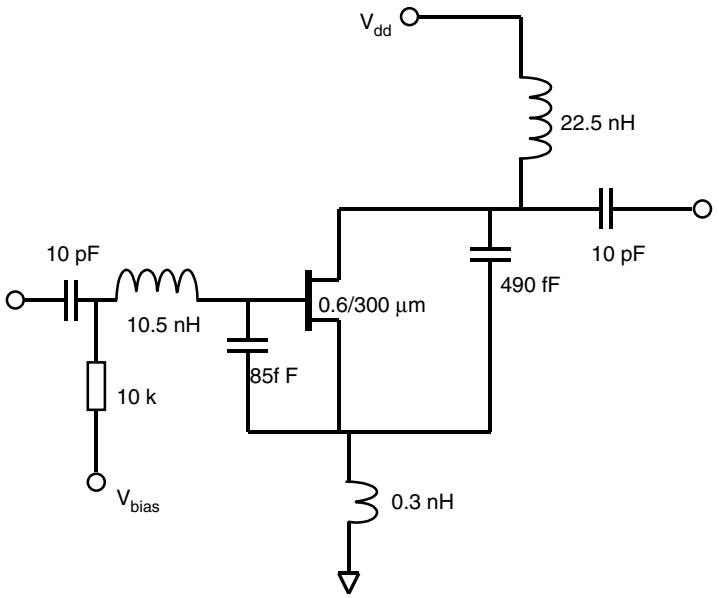


FIGURE 14.10 Schematic diagram of the low voltage GaAs MESFET LNA.

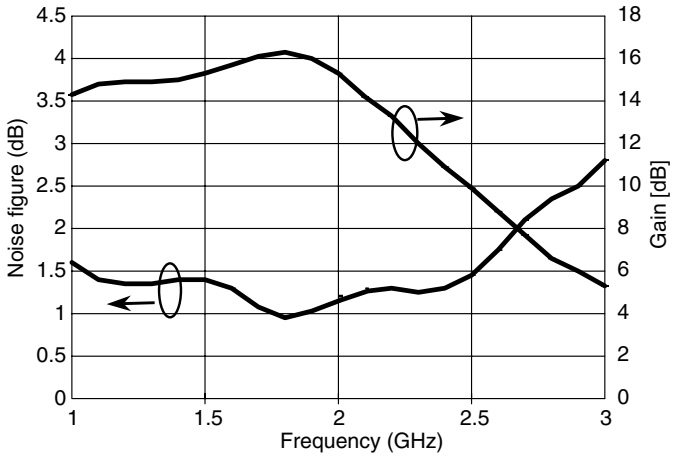


FIGURE 14.11 Measured gain and 50 Ω noise figure versus frequency.

14.4.2 A Fully Matched 800 MHz to 5.2 GHz LNA in SiGe HBT Technology

Bipolar technology is particularly well suited for broadband amplifiers because BJTs typically show low input impedances in the vicinity of 50 Ω and hence can be easily matched. A simplified schematic diagram of the monolithic amplifier is shown in Figure 14.14. For the active devices of the cascode LNA large emitter areas (47 μm^2), biased at low current densities are employed to simplify the simultaneous noise and power match. Input and output matching is consequently obtained simply by the aid of the bondwire inductance at the input and output ports and the chip ground.

The LNA was fabricated with MAXIM's GST-3 SiGe process and subsequently was mounted on a ceramic test package for testing. No additional external components are required for this single-supply LNA.

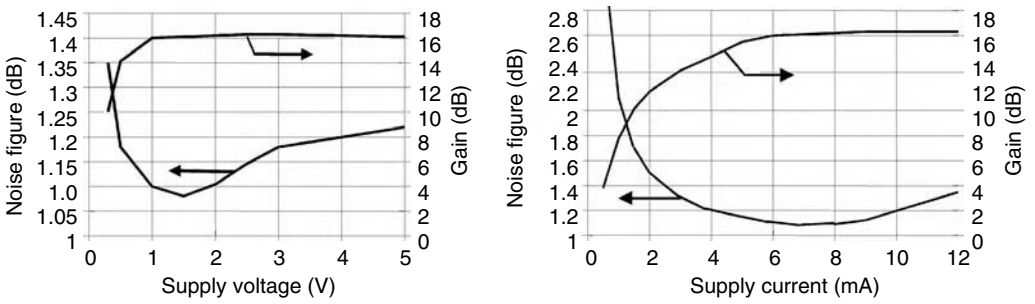


FIGURE 14.12 Measured gain and noise figure versus supply voltage (I_{dd} = 6 mA) and noise figure versus supply current (V_{dd} = 1 V).

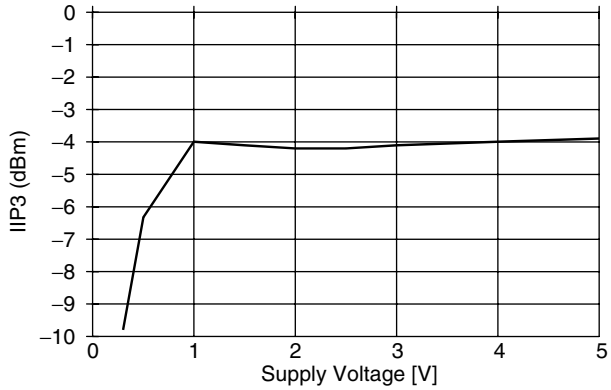


FIGURE 14.13 Measured input IP3 versus supply voltage.

Figure 14.15 shows the 50 Ω noise figure and associated gain over the frequency range of interest. A relatively flat gain curve is measured from 500 MHz up to 3 GHz. Beyond 3 GHz the gain starts to roll off. The circuit features 14.5 dB of gain along with a 2 dB noise figure at 2 GHz. At 5.2 GHz, the gain is still 10 dB and the noise figure is below 4 dB. The input return loss is less than –10 dB between 2.5 and 6.5 GHz. At 1 GHz it increases to –6 dB.

The distortion performance of the amplifier was measured at the nominal 3 V supply for two frequencies, 2.0 and 5.2 GHz, respectively. At 2 GHz, the –1 dB compression point is +2 dBm at the output. At 5.2 GHz the value degrades to 0 dBm.

The LNA is comprised of two sections: the amplifier core and a PTAT reference. The core is biased with the PTAT to compensate for the gain reduction with increasing temperature. The gain is proportional to the transconductance of the transistor, which itself is proportional to collector current and inversely proportional to temperature. The PTAT biasing increases the collector current with temperature to keep the gain roughly constant over temperature. Simultaneously, the biasing shows a good supply rejection as shown in Figure 14.16.

A chip photograph of the 0.5 × 0.6 mm^2 large LNA is shown in Figure 14.17.

14.4.3 A Fully Matched Two-Stage Low Power 5.8 GHz LNA [12]

A fully monolithic LNA achieves a noise figure below 2 dB between 4.3 GHz and 5.8 GHz with a gain larger than 15 dB at a DC power consumption of only 6 mW using the enhancement device of a standard 17 GHz f_T 0.6 μm E/D-MESFET process.

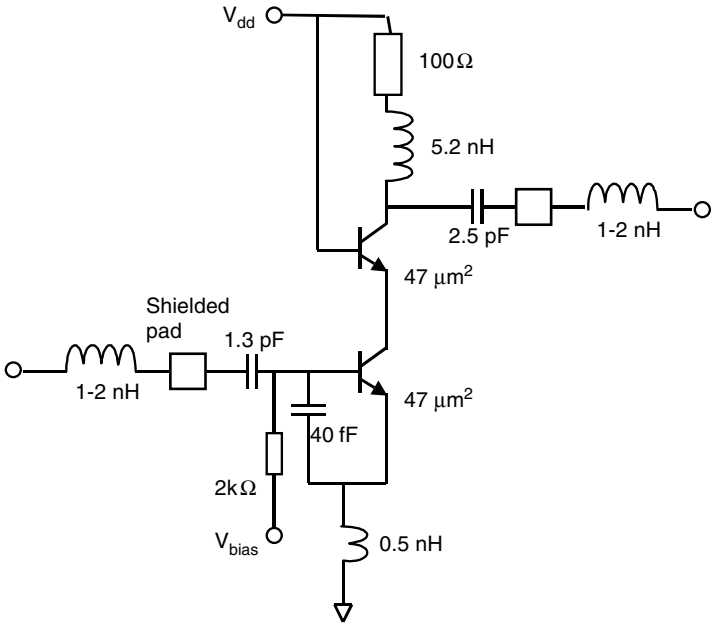


FIGURE 14.14 Schematic diagram of the SiGe HBT LNA.

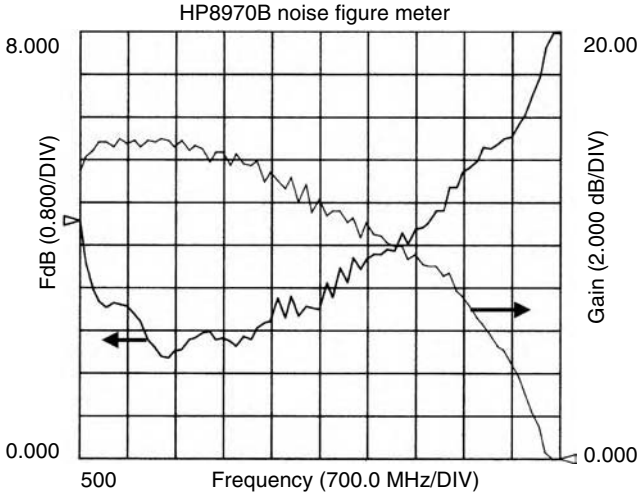


FIGURE 14.15 Measured LNA gain and noise figure versus frequency (V_{dd} = 3 V, I_{dd} = 8.8 mA).

A schematic diagram of the integrated LNA core is shown in Figure 14.18. The circuit consists of two common-source gain stages to provide enough power gain. The first stage uses an on-chip inductive degeneration of the source to achieve a simultaneous noise and power match, and to improve RF stability. Both amplifier stages are biased at the same current. The noise contributions of the biasing resistors are negligible.

The output of each stage is loaded with a band pass LC section to increase the gain at the desired frequency. The load of the first stage, together with the DC block between the stages, is also used for inter-stage matching.

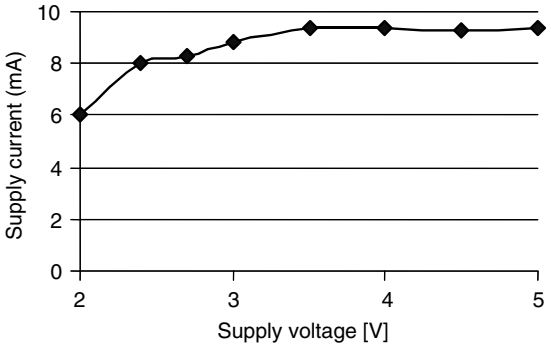


FIGURE 14.16 Supply current versus supply voltage.

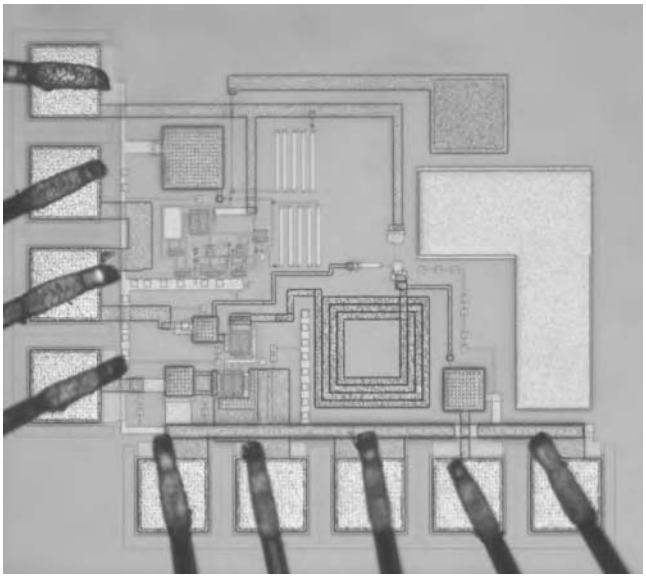


FIGURE 14.17 Chip photograph of the SiGe HBT LNA (0.5×0.6 mm^2).

The DC biasing is done on-chip with a combination of E/D MESFETs (Figure 14.19). The bias circuit is able to effectively stabilize the bias point for voltages from 1 V to beyond 4 V without any feedback network within the amplifier. It also can accurately compensate for threshold voltage variations.

The correlation of the threshold voltages of enhancement and depletion devices due to simultaneous gate recess etch of both types is used in the bias circuit to reduce the bias current variations over process parameter changes. Figure 14.20 shows the simulated deviation from the nominal current as a function of threshold voltage variations. The device current remains very constant even for extreme threshold voltage shifts.

If the RF input device is small, a large input impedance transformation is required. The third-order intercept point can be degraded and larger inductor values are needed sacrificing chip area and noise figures, due to the additional series resistance of the inductor. If instead a very large device is used, the current consumption is increased, unless the current density is lowered. Below a certain current density the device gain will decrease, the minimum noise figure will increase, and a reliable and reproducible biasing of the device becomes difficult as the device is biased close to the pinch-off voltage. To achieve high quality factors, all inductors are implemented using the two top wiring levels with a total metal thickness of 6 μm. The spiral inductors were analyzed using a 2.5D field simulator in order to accurately determine their equivalent circuit.

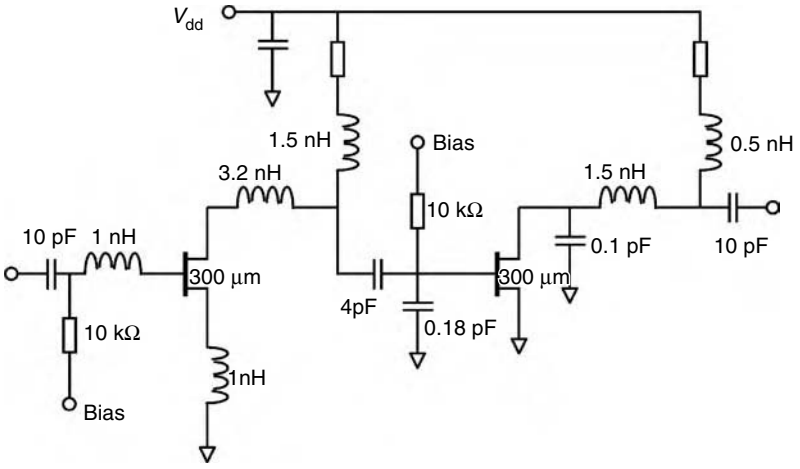


FIGURE 14.18 Schematic diagram of the low noise amplifier.

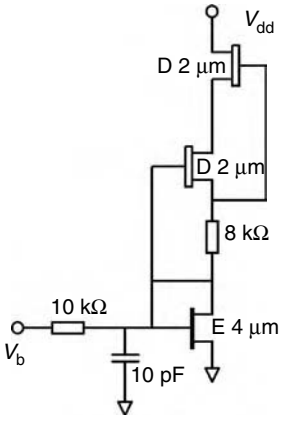


FIGURE 14.19 Schematic diagram of the employed bias circuit.

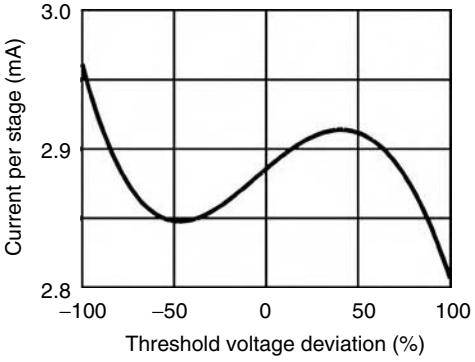


FIGURE 14.20 Simulated current dependence on threshold voltage variations.

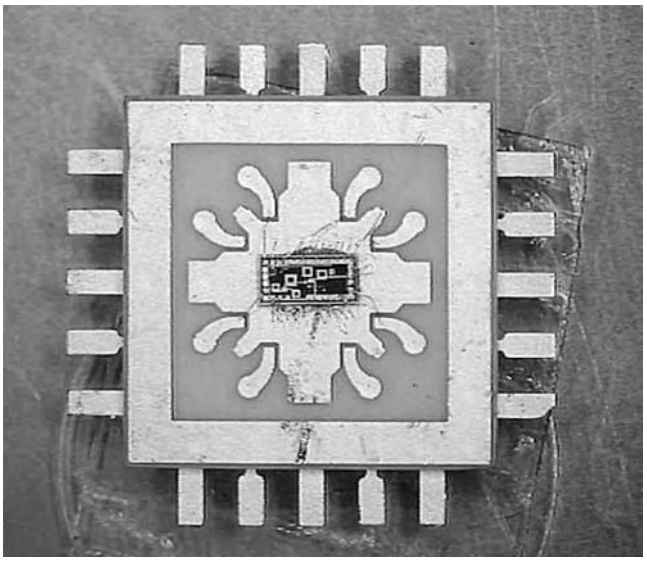


FIGURE 14.21 Photograph of the chip mounted in the test package.

Sample test chips were mounted in a ceramic test package (Figure 14.21) to investigate the influence of the bonding wires and the package parasitics.

In Figures 14.22 and 14.23 the influence of the bond wires on the input and return loss, gain, and noise figure, respectively, is shown. The optimum input matching is shifted from 5.2 to 5.8 GHz with the bond wire included. In an amplifier stage with moderate feedback one would expect the bond wire to shift the match toward lower frequencies. However, due to the source inductor the inter-stage matching circuit strongly interacts with the input port, causing a frequency shift in the opposite direction.

As expected, the gain curve of the packaged LNA (Figure 14.23) is flatter and the gain is slightly reduced because of the additional ground inductance arising from the ground bond wires (approx. 40 pH).

At the nominal supply current of 6 mA the measured 50 Ω noise figure is 1.8 dB along with more than 15 dB gain from 5.2 to 5.8 GHz as given in Figure 14.23. For the packaged LNA the noise figure is slightly degraded due to losses associated with the package and connectors.

At 5.5 GHz the minimum noise figure of the device including the source inductor at the operating bias point is 1.0 dB and the associated gain is 8.5 dB. The minimum noise figure of an amplifier with two identical stages is therefore 1.6 dB. Thus, only a small degradation of the noise figure by the on-chip matching inductor is introduced at the input.

At 5.2 GHz a measured –1 dB compression point of 0 dBm at the output confirms the excellent distortion characteristics of GaAs MESFET devices at very low power consumption. The measured input referenced third order intercept point (IIP3) is –6 dBm.

14.4.4 0.25 µm CMOS LNAs for 900 MHz and 1.9 GHz [13,14]

CMOS technology starts to play a significant role in integrated RF transceivers for the low GHz range with effective gate lengths reaching the deep submicron regions. Competitive circuit performance at low power dissipation is becoming possible even for critical building blocks such as the LNA. In fact, quarter-micron CMOS seems to be the threshold where robust designs can be realized with current consumption competitive to BJT implementations. Further downscaling calls for a reduction in supply voltage which will ultimately limit the distortion performance of CMOS-based designs.

Designing a low noise amplifier in CMOS is complicated by the lossy substrate, which requires a careful layout to avoid noise injection from the substrate. The schematic diagrams of two demonstrated 0.25 µm

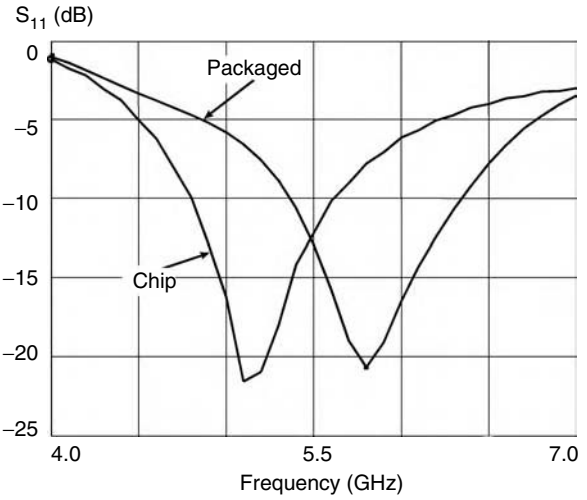


FIGURE 14.22 Input return loss versus frequency of chip and packaged LNA.

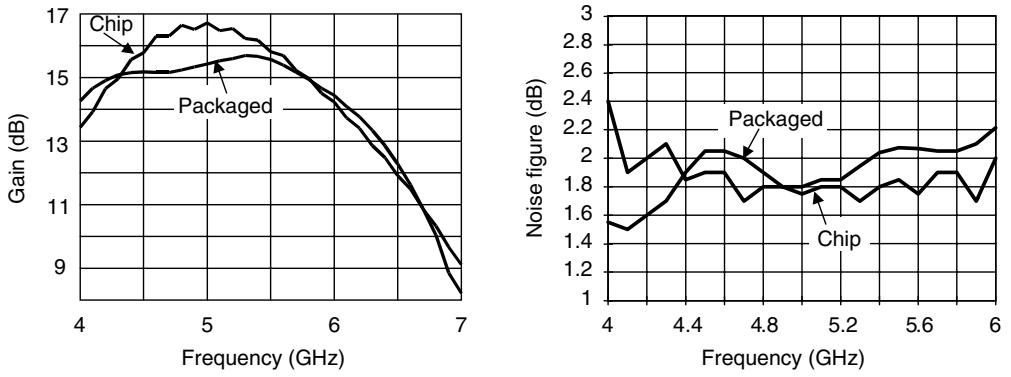


FIGURE 14.23 Gain and noise figure versus frequency (V_{dd} = 1 V and I_{dd} = 6 mA).

CMOS LNAs for 900 MHz and 1.9 GHz are shown in Figure 14.24a,b, respectively. Both circuits use two stages to realize the desired gain.

The first amplifier consisting of an externally matched cascode input stage and a transimpedance output stage consumes 10.8 mA from a 2.5 V supply. The cascode is formed using two 600-μm wide NMOS devices loaded by a 400 Ω resistor. The inductance of approximately 1.2 nH formed by the bondwire at the source of the first stage is used to simplify the matching of an otherwise purely capacitive input impedance. The directly coupled transimpedance output stage isolates the high-gain cascode and provides a good 50 Ω output matching. A simple biasing is included on the chip. At the nominal power dissipation and 900 MHz, the LNA achieves 16 dB gain and a noise figure of below 2 dB. The input and output return losses are –8 dB and –12 dB, respectively. The distortion performance of the LNA can well be estimated by measuring the input referred third order intercept point and the –1 dB compression point. They are –7 dBm and –20 dBm, respectively.

The 1.9 GHz LNA shown in Figure 14.24b employs a resistively loaded common-source stage followed by a reactively loaded cascode stage. To use inductors to tune out the output capacitance and to realize the 50 Ω output impedance is a viable alternative to using the transimpedance output stage. The circuit

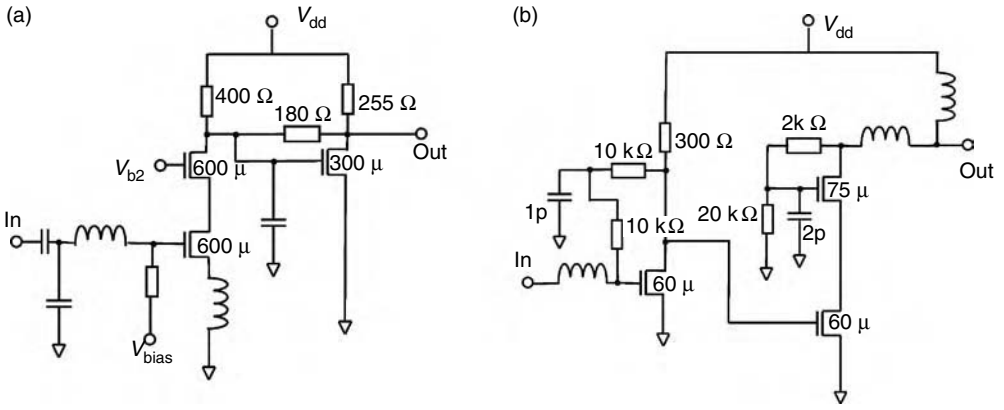


FIGURE 14.24 Schematic diagram of two 0.25 µm CMOS LNAs for 900 MHz (a) and 1900 MHz (b).

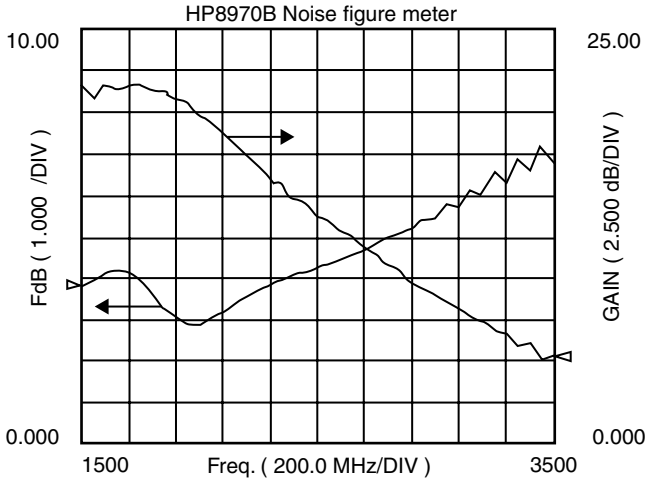


FIGURE 14.25 Measured gain and noise figure of the 1900 MHz CMOS LNA of Figure 14.3b.

employs a self-biasing method by feeding the DC drain voltage of the first stage to the gates. The supply rejection is consequently poor.

The LNA achieves 21 dB gain and a 3 dB noise figure while drawing 10.8 mA from a 2.7 V supply. In Figure 14.25 the measured gain and noise figure versus frequency are plotted. At the nominal bias the input referred –1 dB compression point is –25 dBm, which corresponds to –4 dBm at the output. The input and output return loss are –5 dB and –13 dB, respectively

A comparison between the two amplifiers presented reveals some interesting points:

- The 900 MHz LNA explicitly makes use of the bondwire inductance to reduce the (otherwise purely capacitive) input impedance while the fist stage of the 1.9 GHz amplifier is connected to the chip ground. Both amplifiers use an external inductor for the input matching and both achieve a relatively poor input match.
- No explicit inter-stage matching is employed in either of the amplifiers. The 900 MHz amplifier uses the second stage as an impedance transformer.

- The 900 MHz amplifier employs ten times wider devices biased at lower current densities compared to its 1.9 GHz counterpart. As a consequence, the bias current becomes more sensitive to threshold voltage variations due to fabrication.
- At comparable power consumption the two amplifiers show roughly same distortion performance.

14.4.5 A Highly Selective LNA with Electrically Tunable Image Reject Filter for 2 GHz [15]

LNA designs with purely reactive passive components are inherently narrowband. IC technologies on high resistivity substrates allow reproducible passive components (inductors, capacitors, varactors, transmission lines) with excellent quality factors. They are well suited for designs to include a frequency selectivity which goes beyond a simple matching. In particular, amplifiers with adjustable image rejection can be realized. To show the potential of highly frequency selective LNAs as viable alternative to image reject mixers, an LNA for 1.9 GHz is demonstrated, which allows a tunable suppression of the image frequency. The schematic diagram of the circuit is shown in Figure 14.26. The amplifier consists of two cascaded common-source stages loaded with LC resonant circuits. Undesired frequencies are suppressed using series notch filters as additional loads. Each of the two notch filters is formed by a series connection of a spiral inductor and a varactor diode. The two notches resonate at the same frequencies and must be isolated by the amplifier stages.

A careful design must be done to avoid unwanted resonances and oscillations. In particular, immunity against variations in the ground inductance and appropriate isolation between the supply lines of the two stages must be included. Only the availability of IC technologies with reproducible high-Q, low-tolerance passive components enables the realization of such highly frequency-selective amplifiers.

The LNA draws 9.5 mA from a 3 V supply. At this power dissipation, the input referred –1 dB compression point is measured at –24 dBm.

The measured input and output reflection is plotted in Figure 14.27. The tuning voltage is set to 0 V. The excellent input match changes only negligibly with varying tuning voltage. The input matching shows a high-pass characteristic formed by the series C–L combination instead to the commonly used low-pass. So, the inductor can also act as a bias choke and the input matching can contribute to the suppression of lower frequency interferer. Moreover, the employed matching achieves better noise performance than the high-pass matching network.

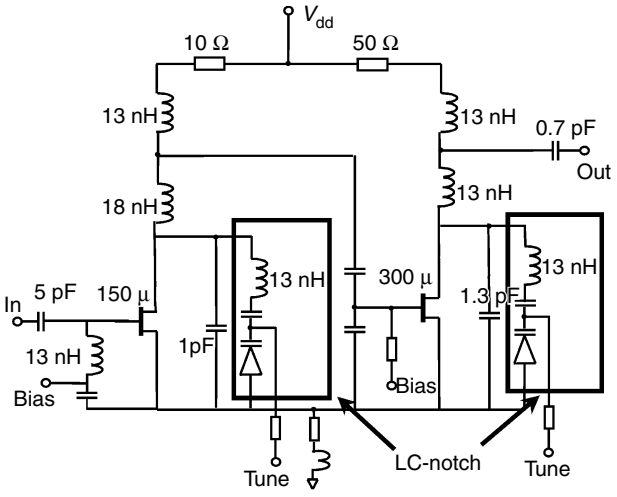


FIGURE 14.26 Schematic diagram of the selective frequency LNA at 2 GHz.

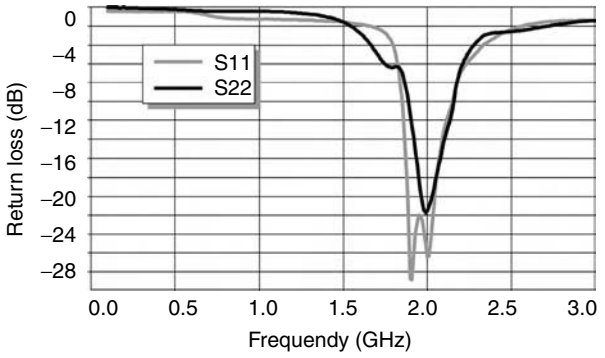


FIGURE 14.27 Measured input and output return loss of the 2 GHz selective LNA.

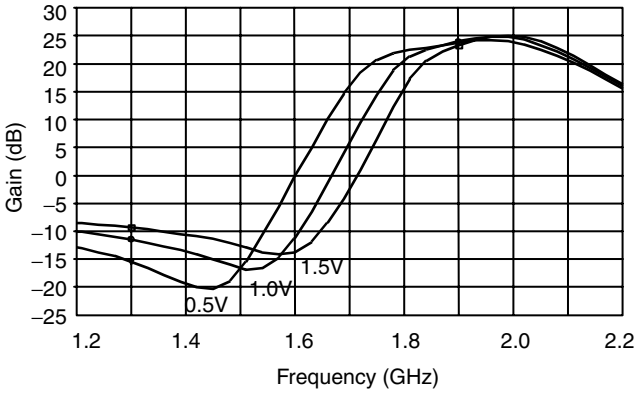


FIGURE 14.28 Selective amplifier gain versus frequency for different notch filter control voltages.

The power gain versus frequency for different notch tuning voltages is shown in Figure 14.28. By varying the tuning voltage from 0.5 to 1.5 V, the filter center frequency can be adjusted from 1.44 to 1.6 GHz. At all tuning voltages the unwanted signal is suppressed by at least 35 dB.

The temperature dependence of gain and noise figure was measured. The temperature coefficients of the gain and noise figure are –0.03 dB/°C and +0.008 dB/°C, respectively. The noise figure of the LNA at different temperatures is plotted in Figure 14.29.

A chip photograph of the fabricated 1.6 × 1.0 mm^2 LNA is depicted in Figure 14.30. More than 50% of the chip area is occupied by the numerous spiral inductors.

14.5 Future Trends

RF and microwave functions are increasingly often realized as integrated circuits (ICs) to reduce size and power consumption, enhance reproducibility, minimize costs, and enable mass production.

14.5.1 Design Approach

The classical noise optimization is based on linear methods and does not take power consumption and linearity requirements explicitly into account. Further, these methods offer only little guidance about how to select the active device dimensions. However, LNA circuit design practices are increasingly influenced by the improvements in the device models in terms of accuracy. Powerful optimization tools

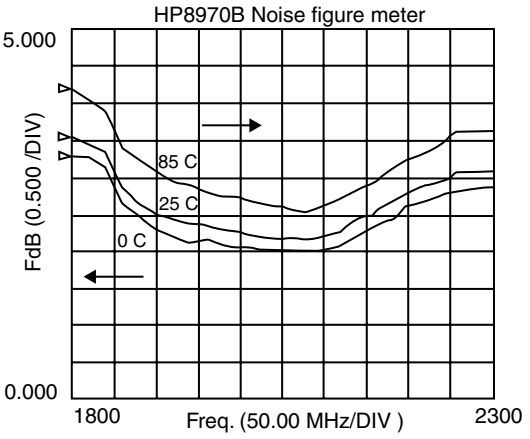


FIGURE 14.29 Amplifier noise figure at various temperatures.

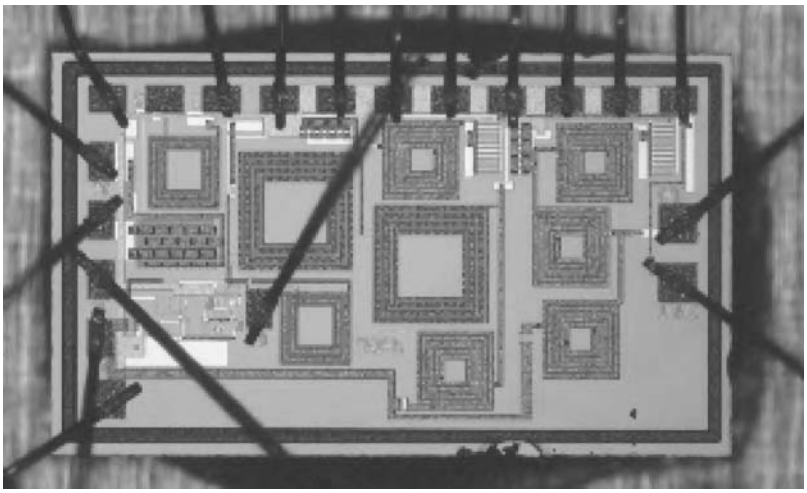


FIGURE 14.30 Chip photograph of the frequency selective LNA.

become available and eases the design procedure. However, a detailed understanding of the basic material will remain necessary for an efficient and robust LNA circuit design.

14.5.2 Device Models

The plurality of bias conditions applied to integrated circuits requires the flexibility of bias-dependent device models. State-of-the-art BJT models (such as Gummel-Poon) already work very well in RF simulations. More recently, sophisticated, semiempirical MOSFET models (such as BSIM3, MM9, or EKV) became suitable for RF simulations. Using accurate models, designs do not need to rely on sample scattering parameters of test devices and tolerance simulations can be implemented.

14.5.3 Circuit Environment

New RF design practices away from the 50 Ω impedance culture will affect the selection of the device size and operation point, but will leave the design procedure basically unchanged. The obstacles in the

quest for higher integrated RF radios are the requirements on system noise figure, substrate crosstalk, and parasitic coupling. Trends to alleviate the unwanted coupling involve using fully differential circuit design, which in turn increases the power consumption.

14.5.4 IC Technologies

In recent years, the advances in device shrinking have made silicon devices (BJTs and more recently MOSFETs) become competitive with III–V semiconductors in terms of gain and minimum noise figure at a given power dissipation in the low GHz range.

The introduction of SiGe and SiC layers further enhance the cutoff frequencies and reduce power dissipation of silicon-based transistors. Furthermore, the use of thick (copper) metallization layers allow relatively low-loss passive components such as MIM capacitors and spiral inductors. Silicon-on-insulator (SOI) technologies will further cut substrate losses and parasitic capacitance and reduce bulk crosstalk.

With the scaling toward minimum gate length of below 0.25 μm, the use of CMOS has become a serious option in low-noise amplifier design. In fact, minimum noise figures of 0.5 dB at 2 GHz and cutoff frequencies of above 100 GHz for 0.12 μm devices[16] can easily compete with any other circuit technology. While intrinsic CMOS device F_{min} is becoming excellent for very short gate lengths, there remains the question of how closely amplifier noise figures can approach F_{min} in practice, particularly if there is a constraint on the allowable power consumption.

References

1. J. M. Rollett, Stability and power-gain invariance of linear two ports, *IEEE Trans. on Circuit Theory,* CT-9, 1, 29–32, Mar. 1962, with corrections, CT-10, 1, 107, Mar. 1963.
2. G. Gonzales, *Microwave Transistor Amplifiers Analysis and Design,* 2nd edition, Prentice Hall, Englewood Cliffs, NJ, 1997.
3. G. Macciarella, et al., Design criteria for multistage microwave amplifiers with match requirements at input and output, *IEEE Trans. Microwave Theory and Techniques,* MTT-41, 1294–98, Aug. 1993.
4. G. D. Vendelin, A. M. Pavio, and U. L. Rohde, *Microwave Circuit Design using Linear and Nonlinear Techniques,* 1st edition, John Wiley & Sons, New York, 1990.
5. H. Hillbrand and P. H. Russer, An efficient method for computer aided noise analysis of linear amplifier networks, *IEEE Transactions on Circuit and Systems,* CAS-23, 4, 235–38, Apr. 1976.
6. H. T. Friis, Noise figure for radio receivers, *Proc. of the IRE,* 419–422, Jul. 1944.
7. H. A. Ainspan, et al., A 5.5 GHz low noise amplifier in SIGe BiCMOS, in *ESSCIRC98 Digest,* 80–83.
8. S. S. Taylor, A GaAs MESFET Schottky diode barrier height reference circuit, *IEEE Journal of Solid-State Circuits,* 32, 12, 2023–29, Dec. 1997.
9. J. J. Kucera, U. Lott, and W. Bächtold, A new antenna switching architecture for mobile handsets, *2000 IEEE Int'l Microwave Symposium Digest,* 3, 1281–1284.
10. R. M. Fano, Theoretical limitations on the broad-band matching of arbitrary impedances, *Journal of the Franklin Institute,* 249, 57–83, Jan. 1960, and 139–155, Feb. 1960.
11. J. J. Kucera and W. Bächtold, A 1.9 GHz monolithic 1.1 dB noise figure low power LNA using a standard GaAs MESFET foundry process, *1998 Asia-Pacific Microwave Conference Digest,* 383–386.
12. J. J. Kucera and U. Lott, A 1.8 dB noise figure low DC power MMIC LNA for C-band, *1998 IEEE GaAs IC Symposium Digest,* 221–224.
13. Q. Huang, P. Orsatti and F. Piazza, Broadband, 0.25 μm CMOS LNAs with sub-2dB NF for GSM applications, *IEEE Custom Integrated Circuits Conference,* 67–70, 1998.
14. Ch. Biber, Microwave modeling and circuit design with sub-micron CMOS technologies, PhD thesis, Diss. ETH No. 12505, Zurich, 1998.
15. J. J. Kucera, Highly integrated RF transceivers, PhD thesis, Diss. ETH No. 13361, Zurich, 1999.
16. R. R. J. Vanoppen, et al., RF noise modeling of 0.25 μm CMOS and low power LNAs, *1997 IEDM Technical Digest,* 317–320.

15

Microwave Mixer Design

Anthony M. Pavio
Microwave Specialties

At the beginning of the 20th century, RF detectors were crude, consisting of a semiconductor crystal contacted by a fine wire ("whisker"), which had to be adjusted periodically so that the detector would keep functioning. With the advent of the triode, a significant improvement in receiver sensitivity was obtained by adding amplification in front of and after the detector. A real advance in performance came with the invention by Edwin Armstrong of the super regenerative receiver. Armstrong was also the first to use a vacuum tube as a frequency converter (mixer) to shift the frequency of an incoming signal to an intermediate frequency (IF), where it could be amplified and detected with good selectivity. The superheterodyne receiver, which is the major advance in receiver architecture to date, is still employed in virtually every receiving system.

The mixer, which can consist of any device capable of exhibiting nonlinear performance, is essentially a multiplier or a chopper. That is, if at least two signals are present, their product will be produced at the output of the mixer. This concept is illustrated in Figure 15.1. The RF signal applied has a carrier frequency of w_s with modulation $M(t)$, and the local oscillator signal (LO or pump) applied has a pure sinusoidal frequency of w_p. From basic trigonometry we know that the product of two sinusoids produces a sum and difference frequency.

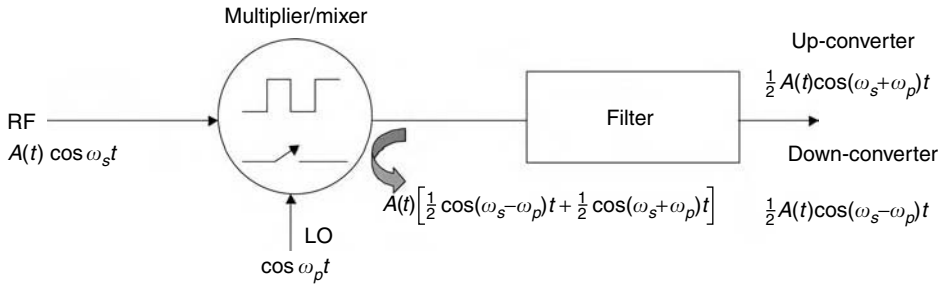


FIGURE 15.1 Ideal mixer model.

The voltage-current relationship for a diode can be described as an infinite power series, where V is the sum of both input signals and I is the total signal current. If the RF signal is substantially smaller than the LO signal and modulation is ignored, the frequency components of the signal are:

$$w_d = nw_p \pm w_s \tag{15.1}$$

As mentioned above, the desired component is usually the difference frequency ($|w_{p+} w_s|$ or $|f_p - f_s|$), but sometimes the sum frequency ($f_s + f_p$) is desired when building an up-converter, or a product related to a harmonic of the LO can be selected.

A mixer can also be analyzed as a switch that is commutated at a frequency equal to the pump frequency w_p. This is a good first-order approximation of the mixing process for a diode since it is driven from the low-resistance state (forward bias) to the high-resistance state (reverse bias) by a high-level LO signal.

The concept of the switching mixer model can also be applied to field-effect transistors used as voltage-controlled resistors. In this mode, the drain-to-source resistance can be changed from a few ohms to many thousands of ohms simply by changing the gate-to-source potential. At frequencies below 1 GHz, virtually no pump power is required to switch the FET, and since no DC drain bias is required, the resulting FET mixer is passive. However, as the operating frequency is raised above 1 GHz, passive FET mixers require LO drive powers comparable to diode or active FET designs.

Mixers can be divided into several classes: (1) single ended, (2) single balanced, or (3) double balanced. Depending on the application and fabrication constraints, one topology can exhibit advantages over the other types. The simplest topology (Figure 15.2a) consists of a single diode and filter networks. Although there is no isolation inherent in the structure (balance), if the RF, LO, and IF frequencies are sufficiently separated, the filter (or diplexer) networks can provide the necessary isolation. In addition to simplicity, single diode mixers have several advantages over other configurations. Typically, the best conversion loss is possible with a single device, especially at frequencies where balun or transformer construction is difficult or impractical. Local oscillation requirements are also minimal since only a single diode is employed and DC biasing can easily be accomplished to reduce drive requirements. The disadvantages of the topology are: (1) sensitivity to terminations; (2) no spurious response suppression; (3) minimal tolerance to large signals; and (4) narrow bandwidth due to spacing between the RF filter and mixer diode.

The next topology commonly used is the single balanced structure shown in Figure 15.2b. These structures tend to exhibit slightly higher conversion loss than that of a single-ended design, but since the RF signal is divided between two diodes, the signal power-handling ability is better. More LO power is required, but the structure does provide balance. The double-balanced mixer (Figure 15.2c) exhibits the best large signal-handling capability, port-to-port isolation, and spurious rejection. Some high-level mixer designs can employ multiple-diode rings with several diodes per leg in order to achieve the ultimate in large-signal performance. Such designs can easily require hundreds of milliwatts of pump power.

15.1 Single-Diode Mixers

The single-diode mixer, although fondly remembered for its use as an AM "crystal" radio or radar detector during World War II, has become less popular due to demanding broadband and high dynamic range requirements encountered at frequencies below 30 GHz. However, there are still many applications at millimeter wave frequencies, as well as consumer applications in the microwave portion of the spectrum, which are adequately served by single-ended designs. The design of single-diode mixers can be approached in the same manner as multi-port network design. The multi-port network contains all mixing product frequencies regardless of whether they are ported to external terminations or terminated internally. With simple mixers, the network's main function is frequency component separation; impedance matching requirements are secondary (Figure 15.3). Hence, in the simplest approach, the network must be capable of selecting the LO, RF, and IF frequencies (Figure 15.4).

However, before a network can be designed, the impedance presented to the network by the diode at various frequencies must be determined. Unfortunately, the diode is a nonlinear device; hence,

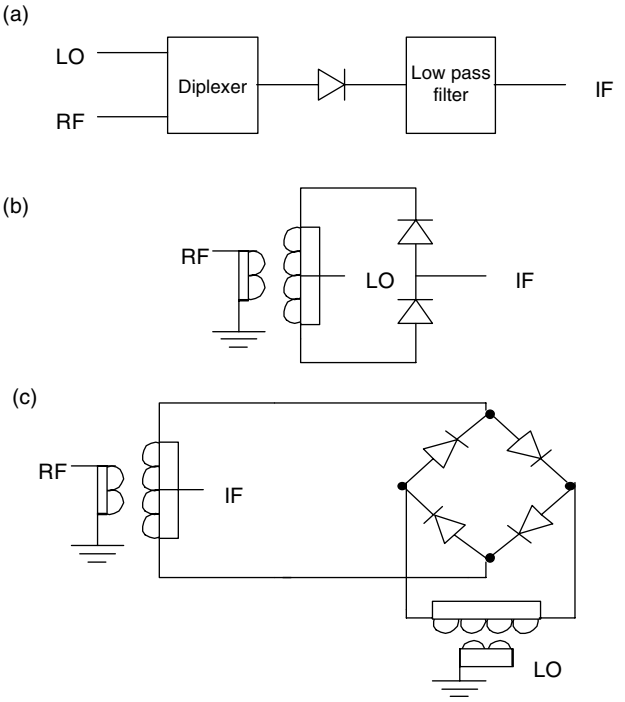


FIGURE 15.2 Typical mixer configurations. (a) Single ended; (b) single balanced; (c) double balanced.

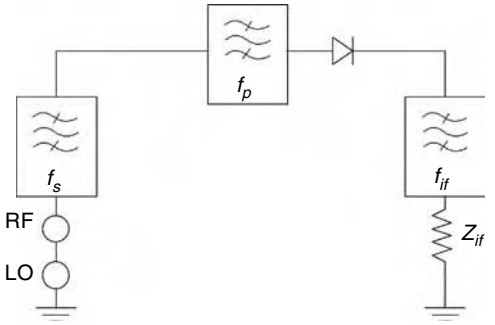


FIGURE 15.3 Filtering requirements for single-diode mixer.

determining its characteristics is more involved than determining an "unknown" impedance with a network analyzer. Since the diode impedance is time varying, it is not readily apparent that a stationary impedance can be found. Stationary impedance values for the RF, LO, and IF frequencies can be measured or determined if sufficient care in analysis or evaluation is taken.

15.2 Single-Balanced Mixers

Balanced mixers offer some unique advantages over single-ended designs such as LO noise suppression and rejection of some spurious products. The dynamic range can also be greater because the input RF signal is divided between several diodes, but this advantage is at the expense of increased pump power.

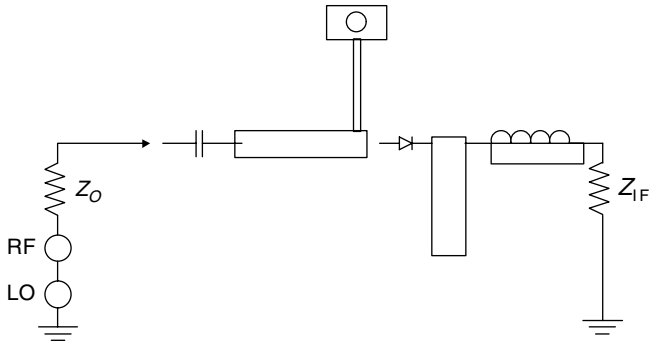


FIGURE 15.4 Typical single-ended mixer.

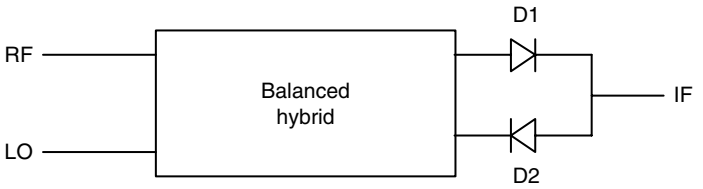


FIGURE 15.5 Single-balanced mixer topology.

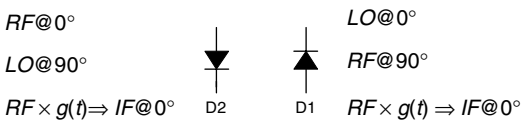


FIGURE 15.6 Signal phase relationships in quadrature coupled hybrid single-balanced mixer.

Both the increase in complexity and conversion loss can be attributed to the hybrid or balun, and to the fact that perfect balance and lossless operation cannot be achieved.

There are essentially only two design approaches for single-balanced mixers; one employs a 180° hybrid, while the other employs some form of quadrature structure (Figure 15.5). The numerous variations found in the industry are related to the transmission-line media employed and the ingenuity involved in the design of the hybrid structure. The most common designs for the microwave frequency range employ either a branch-line, Lange, or "rat-race" hybrid structure (Figure 15.6). At frequencies below about 5 GHz, broadband transformers are very common, while at frequencies above 40 GHz, waveguide and MMIC structures become prevalent.

15.3 Double-Balanced Mixers

The most commonly used mixer today is the double-balanced mixer. It usually consists of four diodes and two baluns or hybrids, although a double-ring or double-star design requires eight diodes and three hybrids. The double-balanced mixer has better isolation and spurious performance than the single-balanced designs described previously, but usually requires greater amounts of LO drive power, are more difficult to assemble, and exhibit somewhat higher conversion loss. However, they are usually the mixer of choice because of their spurious performance and isolation characteristics.

A typical single-ring mixer with transformer hybrids is shown in Figure 15.7. With this configuration the LO voltage is applied across the ring at terminals LO^- and LO^+, and the RF voltage is applied across terminals RF^- and RF^+. As can be seen, if the diodes are identical (matched), nodes RF^- and RF^+ are virtual grounds; thus no LO voltage appears across the secondary of the RF transformer. Similarly, no RF voltage appears across the secondary of the LO balun. Because of the excellent diode matching that can be obtained with diode rings fabricated on a single chip, the L-to-R isolation of microwave mixers can be quite good, typically 30 to 40 dB.

Transmission-line structures which are naturally balanced, such as slotline and finline, can also be used as balanced feed in mixer design. However, all of the structures above, and the more complex transmission-line structures to follow, exhibit one major drawback compared to a transformer hybrid: There is no true RF center tap. As will be seen, this deficiency in transmission-line structures, extensively complicates the design of microwave-balanced mixers.

The lack of a balun center tap does indeed complicate the extraction of IF energy from the structure, but if the IF frequency is low, diplexing can be employed to ease performance degradation. This concept is illustrated in the following example of the center section of a double-balanced 2 to 12 GHz mixer (Figure 15.8). It will be assumed that because of the soft-substrate transmission-line media and frequency range, a packaged diode ring with known impedances can be used. For Si diodes in this frequency range,

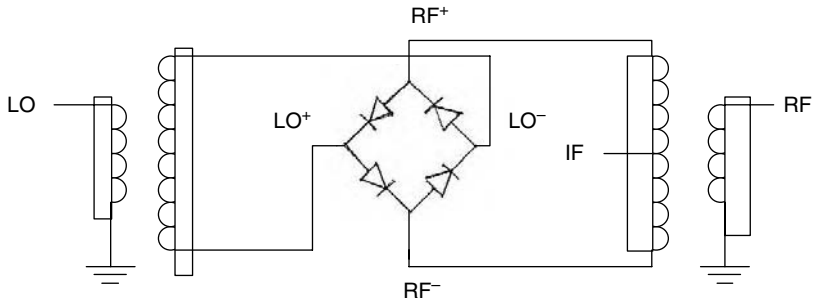


FIGURE 15.7 Transformer coupled double-balanced mixer.

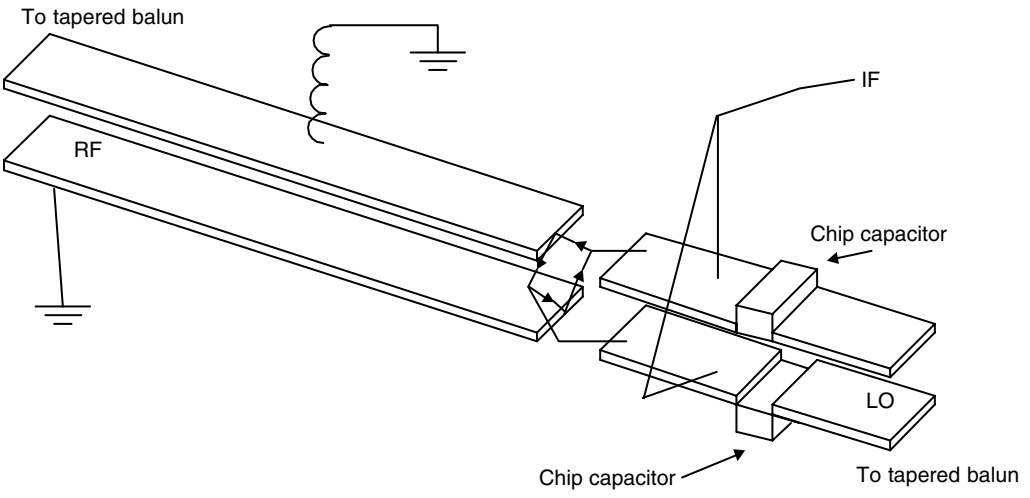


FIGURE 15.8 Double-balanced mixer center section.

the typical LO impedance range (magnitude) is on the order of 75, while the RF impedance is approximately 50. With these values in mind, microstrip-to-parallel plate transmission-line baluns can be fabricated on soft-substrate material.

As can be seen, both the RF and LO baluns terminate at the diode ring and provide the proper phase excitation. But since there is no center tap, the IF must be summed from the top and bottom of either balun. This summing is accomplished with bond wires that have high reactances at microwave frequencies but negligible inductances in the IF passband. Blocking capacitors form the second element in a high-pass filter, preventing the IF energy to be dissipated externally. An IF return path must also be provided at the terminals of the opposite balun. The top conductor side of the balun is grounded with a bond wire, providing a low-impedance path for the IF return and a sufficiently large impedance in shunt with the RF path. The ground-plane side of the balun provides a sufficiently low impedance for the IF return from the bottom side of the diode ring. The balun inductance and blocking capacitor also form a series resonant circuit shunting the IF output; therefore, this resonant frequency must be kept out of the IF passband.

The upper-frequency limit of mixers fabricated using tapered baluns and low parasitic diode packages, along with a lot of care during assembly, can be extended to 40 GHz. Improved "high-end" performance can be obtained by using beam-lead diodes. Although this design technique is very simple, there is little flexibility in obtaining an optimum port VSWR since the baluns are designed to match the magnitude of the diode impedance. The IF frequency response of using this approach is also limited, due to the lack of a balun center tap, to a frequency range below the RF and IF ports.

15.4 FET Mixer Theory

Interest in FET mixers has been very strong due to their excellent conversion gain and intermodulation characteristics. Numerous commercial products employ JFET mixers, but as the frequency of operation approaches 1 GHz, they begin to disappear. At these frequencies and above, the MESFET can easily accomplish the conversion functions that the JFET performs at low frequencies. However, the performance of active FET mixers reported to date by numerous authors has been somewhat disappointing. In short, they have not lived up to expectations, especially concerning noise-figure performance, conversion gain, and circuit-to-circuit repeatability. However, they are simple and low cost, so these sins can be forgiven.

Recently, growing interest is GaAs monolithic circuits is again beginning to heighten interest in active MESFET mixers. This is indeed fortunate, since properly designed FET mixers offer distinct advantages over their passive counterparts. This is especially true in the case of the dual-gate FET mixer; since the additional port allows for some inherent LO-to-RF isolation, it can at times replace single balanced passive approaches. The possibility of conversion gain rather than loss is also an advantage, since the added gain may eliminate the need for excess amplification, thus reducing system complexity.

Unfortunately, there are some drawbacks when designing active mixers. With diode mixers, the design engineer can make excellent first-order performance approximations with linear analysis; also, there is the practical reality that a diode always mixes reasonably well almost independent of the circuit. In active mixer design, these two conditions do not hold. Simulating performance, especially with a dual-gate device, requires some form of nonlinear analysis tool if any circuit information other than small-signal impedance is desired. An analysis of the noise performance is even more difficult.

As we have learned, the dominant nonlinearity of the FET is its transconductance, which is typically (especially with JFETs) a squarelaw function. Hence it makes a very efficient multiplier.

The small-signal circuit [1] shown in Figure 15.9 denotes the principal elements of the FET that must be considered in the model. The parasitic resistances R_g, R_d, and R_s are small compared to R_{ds} and can be considered constant, but they are important in determining the noise performance of the circuit. The mixing products produced by parametric pumping of the capacitances C_{gs}, C_{dg}, and C_{ds} are typically small and add only second-order effects to the total circuit performance. Time-averaged values of these capacitances can be used in circuit simulation with good results.

This leaves the FET transconductance g_m, which exhibits an extremely strong nonlinear dependence as a function of gate bias. The greatest change is transconductance occurs near pinch off, with the most

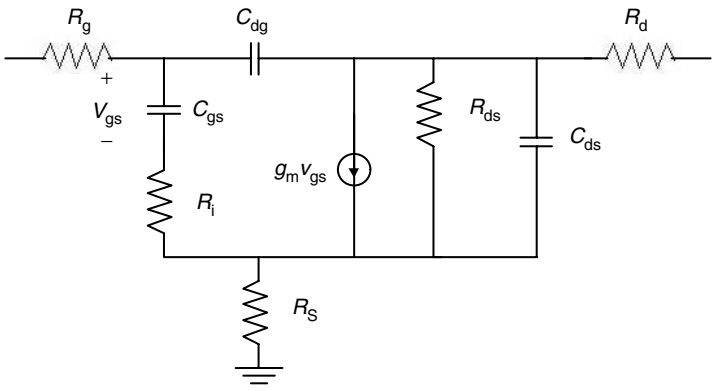


FIGURE 15.9 Typical MESFET model.

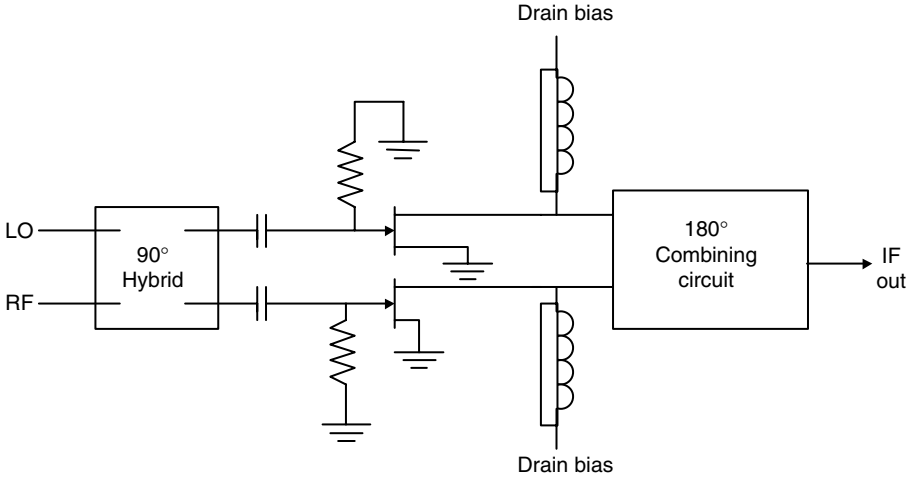


FIGURE 15.10 Typical FET single-balanced mixer.

linear change with respect to gate voltage occurring in the center of the bias range. As the FET is biased toward I_{dss}, the transconductance function again becomes nonlinear. It is in these most nonlinear regions that the FET is most efficient as a mixer.

If we now introduce a second signal, V_c, such that it is substantially smaller than the pump voltage, across the gate-to-source capacitance C_{gs}, the nonlinear action of the transconductance will cause mixing action within the FET producing frequencies $|nw_p \pm w_1|$, where n can be any positive or negative integer. Any practical analysis must include mixing products at both the gate and drain terminal, and at a minimum, allow frequency components in the signal, image, LO, and IF to exist.

Double-balanced FET mixers can also be designed using transformer hybrids [1]. Figure 15.10 shows a typical balanced FET mixer, which can be designed to operate from VHF to SHF. An additional balun is again required because of the phase relationships of the IF signal. This structure is completely balanced and exhibits spurious rejection performance, similar to diode mixers constructed for the same frequency range. However, the intermodulation and noise-figure performance of such structures is superior to those of simple four-diode designs. For example, third-order intercept points in excess of 33 dBm, with

associated gains of 6 dB, are common in such structures. High-level multiple-diode ring mixers, which would require substantially more LO power, would exhibit comparable intermodualtion characteristics, but would never exhibit any gain.

There are a variety of interesting mixer topologies in widespread use that perform vital system functions that cannot be simply classified as balanced mixers. Probably the most popular configuration is the image rejection or single-sideband mixer. However, a variety of subharmonically pumped and self-oscillating mixers are in limited use [1].

Reference

1. G. D. Vendelin, A. M. Pavio, and U. L. Rohde, *The Design of Amplifiers Mixers and Oscillators Using the S-Parameter Method*, John Wiley and Son, New York, 1990.

16

Modulation and Demodulation Circuitry

Charles Nelson
California State University

16.1 Some Fundamentals: Why Modulate?

Because this chapter uses a building block approach, it may seem to be a long succession of setting up straw men and demolishing them. To some extent, this imitates the development of radio and TV, which has been going on for most of the century just ended. A large number of concepts were developed as the technology advanced; each advance made new demands upon the hardware. At first, many of these advances were made by enthusiastic amateurs who had no fear of failure and viewed radio communication the way Hillary viewed Everest—something to be surmounted "because it was there." Since about World War II, there have been increasing numbers of engineers who understood these principles and could propose problem solutions that might have worked the first or second time they were tried. The author fondly hopes this book will help to grow a new cadre of problem solvers for the 21st century.

What probably first motivated the inventors of radio was the need for ships at sea to make distress calls. It may be interesting to note that the signal to be transmitted was a digital kind of thing called Morse Code. Later, the medium became able to transmit equally crucial analog signals, such as a soldier warning, "Watch out!! The woods to your left are full of the abominable enemy!" Eventually, during a period without

widespread military conflict, radio became an entertainment medium, with music, comedy, and news, all made possible by businessmen who were convinced you could be persuaded, by a live voice, to buy soap, and later, detergents, cars, cereals not needing cooking, and so on. The essential low and high frequency content of the signal to be transmitted has been very productive for problems to be solved by radio engineers.

The man on radio, urging you to buy a "pre-owned" Cadillac, puts out most of his sound energy below 1000 Hz. A microphone observes pressure fluctuations corresponding to the sound and generates a corresponding voltage. Knowing that all radio broadcasting is done by feeding a voltage to an antenna, the beginning engineer might be tempted to try sending out the microphone signal directly. A big problem with directly broadcasting such a signal is that an antenna miles long would be required to transmit it efficiently. However, if the frequency of the signal is shifted a good deal higher, effective antennas become much shorter and more feasible to fabricate. This upward translation of the original message spectrum is perhaps the most crucial part of what we have come to call "modulation." However, the necessities of retrieving the original message from the modulated signal may dictate other inclusions in the broadcast signal, such as a small or large voltage at the center, or "carrier" frequency of the modulated signal. The need for a carrier signal is dictated by what scheme is used to transmit the modulated signal, which determines important facts of how the signal can be demodulated.

More perspective on the general problem of modulation is often available by looking at the general form of a modulated signal,

$$f(t) = A(t)\cos\theta(t).$$

If the process of modulation causes the multiplier A(t) out front to vary, it is considered to be some type of "amplitude" modulation. If one is causing the angle to vary, it is said to be "angle" modulation, but there are two basic types of angle modulation. We may write

$$\theta(t) = \omega_c t + \phi(t).$$

If then our modulation process works directly upon $\omega_c = 2\pi f_c$, we say we have performed "frequency" modulation. If, instead, we directly vary the phase factor $\phi(t)$, we say we have performed "phase" modulation. The two kinds of angle modulation are closely related, so that we may do one kind of operation to get the other result, by proper preprocessing of the modulation signal. Specifically, if we put the modulating signal through an integrating circuit before we feed it to a phase modulator, we come out with frequency modulation. This is, in fact, often done. The dual of this operation is possible but is seldom done in practice. Thus, if the modulating signal is fed through a differentiating circuit before it is fed to a frequency modulator, the result will be phase modulation. However, this process offers no advantages to motivate such efforts.

16.2 How to Shift Frequency

Our technique, especially in this chapter, will be to make our proofs as simple as possible; specifically, if trigonometry proves our point, it will be used instead of the convolution theorem of circuit theory. Yet, use of some of the aspects of convolution theory can be enormously enlightening to those who understand. Sometimes, as it will in this first proof, it may also indicate the kind of circuit that will accomplish the task. We will also take liberties with the form of our modulating signal. Sometimes we can be very general, in which case it may be identified as a function m(t). At other times, it may greatly simplify things if we write it very explicitly as a sinusoidal function of time

$$m(t) = \cos\omega_m t.$$

Sometimes, in the theory, this latter option is called "tone modulation," because, if one listened to the modulating signal through a loudspeaker, it could certainly be heard to have a very well-defined "tone" or pitch. We might justify ourselves by saying that theory certainly allows this, because any particular signal we must deal with could, according the theories of Fourier, be represented as a collection, perhaps infinite, of cosine waves of various phases. We might then assess the maximum capabilities of a communication system by choosing the highest value that the modulating signal might have. In AM radio, the highest modulating frequency is typically about f_m = 5000 Hz. For FM radio, the highest modulation frequency might be f_m = 19 kHz, the frequency of the so-called FM stereo "pilot tone."

In principle, the shifting of a frequency is very simple. This is fairly obvious to those understanding convolution. One theorem of system theory says that multiplication of time functions leads to convolution of the spectra. Let us just multiply the modulating signal by a so-called "carrier" signal. One is allowed to have the mental picture of the carrier signal "carrying" the modulating signal, in the same way that homing pigeons have been used in past wars to carry a light packet containing a message from behind enemy lines to the pigeon's home in friendly territory. So, electronically, for "tone modulation," we need only to accomplish the product

$$\phi(t) = A\cos\omega_m t\cos\omega_c t.$$

Now, we may enjoy the consequences of our assumption of tone modulation by employing trigonometric identities for the sum or difference of two angles:

$$\cos(A+B) = \cos A\cos B - \sin A\sin B \quad \text{and} \quad \cos(A-B) = \cos A\cos B + \sin A\sin B$$

If we add these two expressions and divide by two, we get the identity we need:

$$\cos A\cos B = 0.5\left[\cos(A+B) + \cos(A-B)\right].$$

Stated in words, we might say we got "sum and difference frequencies," but neither of the original frequencies. Let's be just a little more specific and say we started with f_m = 5000 Hz and f_c = 1 MHz, as would happen if a radio station whose assigned carrier frequency was 1 MHz were simply transmitting a single tone at 5000 Hz. In "real life," this would not be done very often, but the example serves well to illustrate some definitions and principles. The consequence of the mathematical multiplication is that the new signal has two new frequencies at 995 kHz and 1005 kHz. Let's now just add one modulating tone at 3333 Hz. We would have added two frequencies at 9666.667 kHz and 1003.333 kHz. However, if this multiplication was done purely, *there is no carrier frequency term present.* For this reason, we say we have done a type of "suppressed carrier" modulation. Also, furthermore, we have two *new* frequencies for *each* modulating frequency. We define all of those frequencies above the carrier as the "upper sideband" and all the frequencies below the carrier as the "lower sideband." The whole process we have done here is named "double sideband suppressed carrier" modulation, often known by its initials DSB–SC. Communication theory would tell us that the signal spectrum, before and after modulation with a single tone at a frequency f_m, would appear as in Figure 16.1. Please note that the theory predicts equal positive and negative frequency components. There is no deep philosophical significance to negative frequencies. They simply make the theory symmetrical and a bit more intuitive.

16.3 Analog Multipliers, or "Mixers"

First, there is an unfortunate quirk of terminology; the circuit that multiplies signals together is in communication theory usually called a "mixer." What is unfortunate is that the engineer or technician

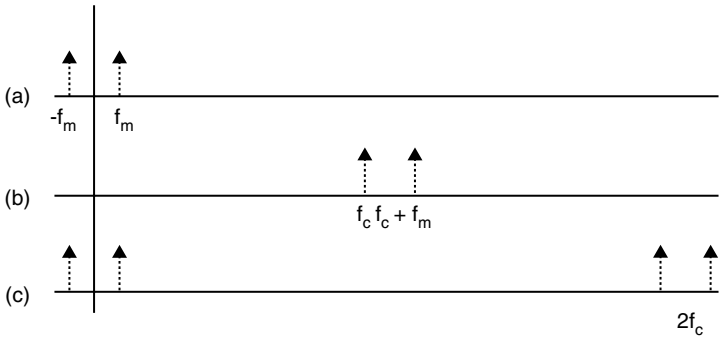


FIGURE 16.1 Unmodulated, modulated, and synchronously demodulated signal spectra. (a) Spectrum of tone-modulating signal. (b) Spectrum (positive part only) of double sideband suppressed carrier signal. (c) Spectrum of synchronously detected DSB–SC signal (except for part near $-2f_c$).

who produces sound recordings is very apt to feed the outputs of many microphones into potentiometers, the outputs of which are sent in varying amounts to the output of a piece of gear, and *that* component is called a "mixer." Thus, the communication engineer's mixer multiplies and the other adds. Luckily, it will usually be obvious which device one is speaking of.

There are available a number of chips (integrated circuits) designed to serve as analog multipliers. The principle is surprisingly simple, although the chip designers have added circuitry which no doubt optimizes the operation and perhaps makes external factors less influential. The reader might remember that the transconductance g_m for a bipolar transistor is proportional to the collector current; its output is proportional to the g_m *and* the input voltage, so in principle one can replace an emitter resistor with the first transistor, which then controls the collector current of the second transistor. If one seeks to fabricate such a circuit out of discrete transistors, one would do well to expect a need to tweak operating conditions considerably before some approximation of analog multiplication occurs. Recommendation: buy the chip. Best satisfaction will probably occur with a "four-quadrant multiplier." The alternative is a "two-quadrant multiplier," which might embarrass one by being easily driven into cut-off.

Another effective analog multiplier is alleged to be the dual-gate FET. The width of the channel in which current flows depends upon the voltage on each of two gates which are insulated from each other. Hence, if different voltages are connected to the two gates, the current that flows is the product of the two voltages. Both devices we have discussed so far have the advantage of having some amplification, so the desired resulting signal has a healthy amplitude. A possible disadvantage may be that spurious signals one does *not* need may also have strong amplitudes.

Actually, the process of multiplication may be the byproduct of any distorting amplifier. One can show this by expressing the output of a distorting amplifier as a Taylor series representing output in terms of input. In principle, such an output would be written

$$V_o = a_0 + a_1\left(v_1 + v_2\right) + a_3\left(v_1 + v_2\right)^2 + \text{ smaller terms.}$$

One can expand $(v_1 + v_2)^2$ as $v_1^2 + 2v_1v_2 + v_2^2$, so this term yields second harmonic terms of each input plus the product of inputs one was seeking. However, the term $a_1(v_1 + v_2)$ also yielded each input, so the carrier here would not be suppressed. If it is fondly desired to suppress the carrier, one must resort to some sort of "balanced modulator." An "active" (meaning there is amplification provided) form of a balanced modulator may be seen in Figure 16.2; failure to bias the bases of the transistors should assure that the voltage squared term is large.

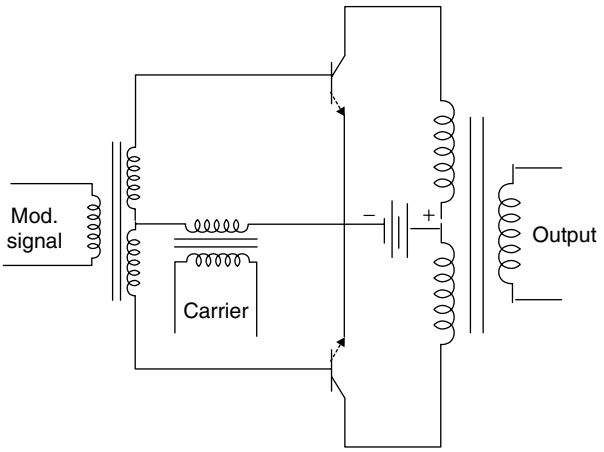


FIGURE 16.2 Balanced modulator.

One will also find purely passive mixers with diodes connected in the shape of a baseball diamond with one signal fed between first and third base, the other from second to home plate. Such an arrangement has the great advantage of not requiring a power supply; the disadvantage is that the amplitude of the sum or difference frequency may be small.

16.4 Synchronous Detection of Suppressed Carrier Signals

At this point, the reader without experience in radio may be appreciating the mathematical tricks but wondering, if one can accomplish this multiplication, can it be broadcast and the original signal retrieved by a receiver? A straightforward answer might be that multiplying the received signal by another carrier frequency signal such as $\cos\omega_c t$ will shift the signal back exactly to where it started and also up to a center frequency of twice the original carrier. This is depicted in part c of Figure 16.1. The name of this process is "synchronous detection." (In the days when it was apparently felt that communications enjoyed a touch of class if one used words having Greek roots, they called it "homodyne detection." If the reader reads a wide variety of journals, he/she may still encounter the word.) The good/bad news about synchronous detection is that the signal being used in the detector multiplication must have the *exact frequency and phase* of the original carrier, and such a signal is not easy to supply. One method is to send a "pilot" carrier, which is a small amount of the correct signal. The pilot tone is amplified until it is strong enough to accomplish the detection.

Suppose the pilot signal reaches high enough amplitude but is phase-shifted an amount θ with respect to the original carrier. We would then in our synchronous detector be performing the multiplication:

$$m(t)\cos\omega_c t\cos(\omega_c t+\theta).$$

To understand what we get, let us expand the second cosine using the identity for the sum of two angles,

$$\cos(\omega_c t+\theta)=\cos\omega_c t\cos\theta-\sin\omega_c t\sin\theta.$$

Hence, the output of the synchronous detector may be written as

$$m(t)\cos^2\omega_c t\cos\theta - m(t)\cos\omega_c t\sin\omega_c t\sin\theta =$$
$$(0.5)\left[m(t)\cos\theta(1-\cos 2\omega_c t) - m(t)\sin\theta\sin 2\omega_c t\right].$$

The latter two terms can be eliminated using a low-pass filter, and one is left with the original modulating signal, m(t), attenuated proportionally to the factor cos θ, so major attenuation does not appear until the phase shift approaches 90°, when the signal would vanish completely. Even this is not totally bad news, as it opens up a new technique called "quadrature amplitude modulation."

The principle of QAM, as it is abbreviated, is that entirely different modulating signals are fed to carrier signals that are 90° out of phase; we could call the carrier signals cos $\omega_c t$ and sin $\omega_c t$. The two modulating signals stay perfectly separated if there is no phase shift to the carrier signals fed to the synchronous detectors. The color signals in a color TV system are QAM'ed onto a 3.58 MHz subcarrier to be combined with the black-and-white signals, after they have been demodulated using a carrier generated in synchronism with the "color burst" (several periods of a 3.58 MHz signal), which is cleverly "piggy-backed" onto all the other signals required for driving and synchronizing a color TV receiver.

16.5 Single Sideband Suppressed Carrier

The alert engineering student may have heard the words "single sideband" and be led to wonder if we are proposing sending one more sideband than necessary. Of course it is true, and SSB–SC, as it is abbreviated, is the method of choice for "hams," the amateur radio enthusiasts who love to see night fall, when their low wattage signals can bounce between the earth and a layer of ionized atmospheric gasses 100 or so miles up until they have reached halfway around the world. It turns out that a little phase shift is not a really drastic flaw for voice communications, so the "ham" just adjusts the variable frequency oscillator being used to synchronously demodulate incoming signals until the whistles and squeals become coherent, and then he/she listens.

How can one produce single sideband? For many years it was pretty naïve to say, "Well, let's just filter one sideband out!" This would have been very naïve because, of course, one does not have textbook filters with perfectly sharp cut-offs. Recently, however, technology has apparently provided rather good "crystal lattice filters" which are able fairly cleanly to filter the extra sideband. In general, though, the single sideband problem is simplified if the modulating signal does not go to really deep low frequencies; a microphone that does not put out much below 300 Hz might have advantages, as it would leave a transition region of 600 Hz between upper and lower sidebands in which the sideband filter could have its amplitude response "roll off" without letting through much of the sideband to be discarded. Observe Figure 16.3, showing both sidebands for a baseband signal extending only from 300 Hz to 3.0 kHz.

Another method of producing single sideband, called the "phase-shift method," is suggested if one looks at the mathematical form of just one of the sidebands resulting from tone modulation. Let us just look at a lower sideband. The mathematical form would be

$$v(t) = A\cos(\omega_c - \omega_m)t = A\cos\omega_c t\cos\omega_m t + A\sin\omega_c t\sin\omega_m t$$

Mathematically, one needs to perform DSB–SC with the original carrier and modulating signals (the cosine terms) and also with the two signals each phase shifted 90°; the resulting two signals are then added to obtain the lower sideband. Obtaining a 90° phase shift is not difficult with the carrier, of which

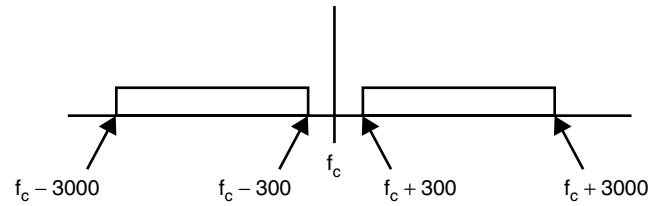


FIGURE 16.3 Double sideband spectrum for modulating signal 300–3000 Hz.

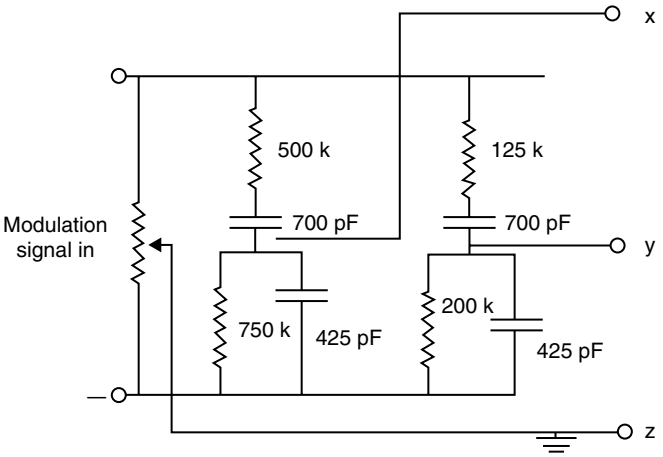


FIGURE 16.4 Audio network for single sideband modulator.

there is only one, but we must be prepared to handle a band of modulating signals, and it is not an elementary task to build a circuit that will produce 90° phase shifts over a range of frequency. However, a reasonable job will be done by the circuit of Figure 16.4 when the frequency range is limited (e.g., from 300 to 3000 Hz). Note that one does *not* modulate directly with the original modulation signal, but that the network uses each input frequency to generate two signals which are attenuated equal amounts and 90° away from each other. These voltages would be designated in the drawing as V_{xz} and V_{yz}. In calculating such voltages, the reader should note that there are two voltage dividers connected across the modulating voltage, determining V_x and V_y, and that from both of these voltages is subtracted the voltage from the center-tap to the bottom of the potentiometer. Note also that the resistance of the potentiometer is not relevant as long as it does not load down the source of modulating voltage, and that a good result has been found if the setting of the potentiometer is for 0.224 of the input voltage.

16.6 Amplitude Modulation as Double Sideband with Carrier

The budding engineer must understand that synchronous detectors are more expensive than many people can afford, and that a less expensive detection method is needed. What fills this bill much of the time is called the "envelope detector." Let us examine some waveforms, first for DSB–SC and then for a signal having a large carrier component. Figure 16.5 shows a waveform in which not very different carrier and modulating frequencies were chosen so that a spreadsheet plot would show a few details.

An ideal circuit we call an envelope detector would follow the topmost excursion of the waveform sketched here. Now, the original modulating signal was a sine wave, but the topmost excursion would

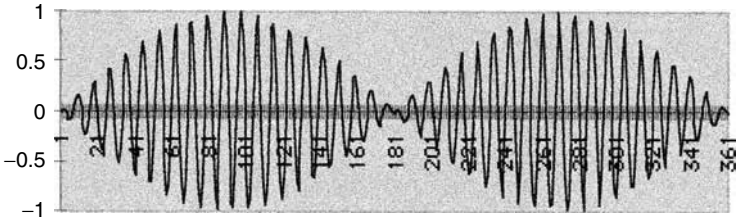


FIGURE 16.5 Double sideband suppressed carrier signal.

be a *rectified* sinusoid, thus containing large amounts of harmonic distortion. How can one get a waveform that will be detected without distortion by an envelope detector? What was plotted was 1.0 cos ω_{ct}t cos ω_mt. We suspect we must add some amount of carrier B cos ω_ct. The sum will be

$$\phi_{AM}(t) = B\cos\omega_c t + 1.0\cos\omega_{ct}t\cos\omega_m t = \cos\omega_c t\left[B + 1.0\cos\omega_m t\right].$$

This result is what is commonly called "amplitude modulation." Perhaps the most useful way of writing the time function for an amplitude modulation signal having tone modulation at a frequency f_m follows:

$$\phi_{AM}(t) = A\cos\omega_c t\left[1 + a\cos\omega_m t\right].$$

In this expression, we can say that A is the peak amplitude of the carrier signal that would be present if there were no modulation. The total expression inside the [ ] brackets can be called the "envelope" and the factor "a" can be called the "index of modulation." As we have written it, if the index of modulation were >1, the envelope would attempt to go negative; this would make it necessary, for distortion-free detection, to use synchronous detection. "a" is often expressed as a percentage, and when the index of modulation is less than 100%, it is possible to use the simplest of detectors, the envelope detector. We will look at the envelope detector in more detail a bit later.

16.7 Modulation Efficiency

It is good news that sending a carrier along with two sidebands makes inexpensive detection using an envelope detector possible. The accompanying bad news is that the presence of carrier does not contribute *at all* to useful signal output; the presence of a carrier only leads after detection to DC, which may be filtered out at the earliest opportunity. Sometimes, as in video, the DC is needed to set the brightness level, in which case DC may need to be added back in at an appropriate level.

To express the effectiveness of a communication system in establishing an output signal-to-noise ratio, it is necessary to define a "modulation efficiency," which, in words, is simply the fraction of output power that is put into sidebands. It is easily figured if the modulation is simply one or two purely sinusoidal tones; for real-life modulation signals, one may have to express it in quantities that are less easy to visualize.

For tone modulation, we can calculate modulation efficiency by simply evaluating the carrier power and the power of all sidebands. For tone modulation, we can write:

$$\phi_{AM}(t) = A\cos\omega_c t\left[1 + a\cos\omega_m t\right] =$$
$$A\cos\omega_c t + (aA)/2\left[\cos(\omega_c + \omega_m)t + \cos(\omega_c - \omega_m t)\right].$$

Now, we have all sinusoids, the carrier, and two sidebands of equal amplitudes, so we can write the average power in terms of peak amplitudes as:

$$P = 0.5\left[A^2 + 2 \times \left(aA/2\right)^2\right] = 0.5A^2\left[1 + a^2/2\right].$$

Then modulation efficiency is the ratio of sideband power to total power, for modulation by a single tone with modulation index "a," is:

$$\eta = \frac{\left(aA/2\right)^2}{0.5A^2\left(1 + a^2/2\right)} = \frac{a^2}{2 + a^2}.$$

Of course, most practical modulation signals are not so simple as sinusoids. It may be necessary to state how close one is to overmodulating, which is to say, how close to negative modulating signals come to driving the envelope negative. Besides this, what is valuable is a quantity we shall just call "m," which is the ratio of average power to peak power for the modulation function. For some familiar waveforms, if the modulation is sinusoidal, m = 1/2. If modulation were a symmetrical square wave, m = 1.0; any kind of symmetrical triangle wave has m = 1/3. In terms of m, the modulation efficiency is

$$\eta = \frac{ma^2}{1 + ma^2}$$

16.8 The Envelope Detector

Much of the detection of modulated signals, whether the signals began life as AM or FM broadcast signals or the sound or the video of TV, is done using envelope detectors. Figure 16.7 shows the basic circuit configuration.

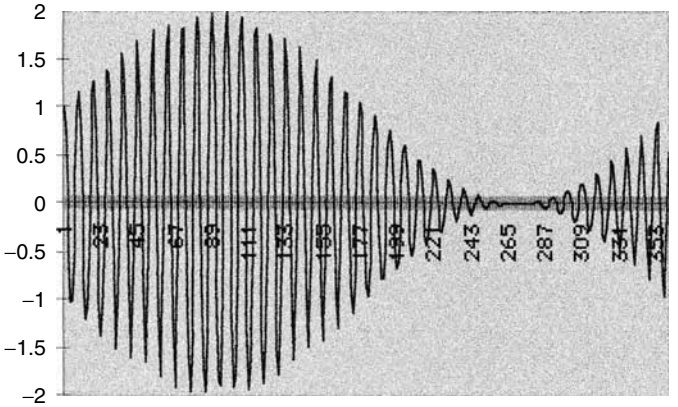


FIGURE 16.6 Amplitude-modulated signal.

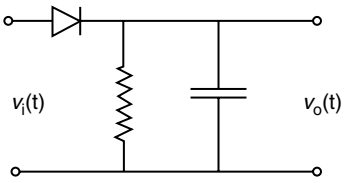


FIGURE 16.7 Simple envelope detector schematic.

The input signal is of course as shown in Figure 16.6. It is assumed that the forward resistance of the diode is 100 ohms or less. Thus, the capacitor is small enough that it gets charged up to the peak values of the high frequency signal, but then when input drops from the peak, the diode is reverse-biased so the capacitor can only discharge through R. This discharge voltage is of course given by

$$V(0)\exp(-t/RC).$$

Now the problem in AM detection is that we must have the minimum rate of decay of the voltage be at least the maximum decay of the envelope of the modulated wave. We might write the envelope as a function of time:

$$E(t) = A(1 + a\cos\omega_{mt}t),$$

where A is the amplitude of the carrier before modulation and "a" is the index of modulation, which must be less than one for accurate results with the envelope detector. Then, when we differentiate, we get

$$\frac{dE}{dt} = -\omega_m Aa\sin(\omega_m t).$$

We want this *magnitude* to be less than or equal to the maximum magnitude of the rate of decay of a discharging capacitor, which is E(0)/RC. For what is written as E(0), we will write the instantaneous value of the envelope, and the expression becomes

$$A(1 + a\cos\omega_{mt}t) \geq RC(\omega_m aA\sin(\omega_m t)).$$

The As cancel, and we have

$$RC \leq \frac{1 + a\cos(\omega_m t)}{\omega_m a\sin(\omega_m t)};$$

our major difficulty occurs when the right-hand side has its minimum value.

If we differentiate with respect to $\omega_m t$, we get

$$\frac{\omega_m a \sin\left(\omega_m t \times \left(-\omega_m a \sin(\omega_m t) - \left(1 + a\cos(\omega_m t)\right)(\omega_m)^2 a \cos\omega_m t\right)\right)}{\left(\omega_m a \sin(\omega_m t)\right)^2}.$$

We set the numerator equal to zero to find its maximum. We find we have

$$-(\omega_m a)^2\left[\sin^2(\omega_m t) + \cos^2(\omega_m t)\right] - a(\omega_m)^2 \cos\omega_m t$$
$$= -(\omega_m a)^2 - a(\omega_m)^2 \cos\omega_m t = 0.$$

Hence, the maximum occurs when $\cos \omega_m t = -a$, and of course by identity, at that time, $\sin \omega_m t = \sqrt{1-a^2}$.

Inserting these results into our inequality for the time constant RC, we have

$$RC \le \frac{1-a^2}{\omega_m a\sqrt{1-a^2}} = \frac{\sqrt{1-a^2}}{\omega_m a}.$$

Example 16.1

Suppose we say 2000 Hz is the main problem in our modulation scheme, our modulation index is 0.5, and we choose R = 10 k to make it large compared to the diode forward resistance, but not *too* large. What should be the capacitor C?

Solution We use the equality now and get

$$C = \frac{\sqrt{1-0.5^2}}{0.5 \times 4000\pi \times 10{,}000} = 13.8 \text{ nF}.$$

16.9 Envelope Detection of SSB Using Injected Carrier

Single sideband, it might be said, is a very forgiving medium. Suppose that one were attempting synchronous detection using a carrier that was off by a Hertz or so, compared to the original carrier. Because synchronous detection works by producing sum and difference frequencies, 1 Hz error in carrier frequency would produce 1 Hz error in the detected frequency. Because SSB is mainly used for speech, it would be challenging indeed to find anything wrong with the reception of a voice one has only ever heard over a static-ridden channel. Similar results would also be felt in the following, where we add a carrier to the sideband and find that we have AM, albeit with a small amount of harmonic distortion.

Example 16.2

Starting with just an upper sideband $B \cos(\omega_c + \omega_m)t$, let us add a carrier term $A \cos \omega_c t$, manipulate the total, and prove that we have an envelope to detect. First we expand the sideband term as

$$\Phi_{SSB}(t) = B\left[\cos\omega_c t \cos\omega_m t - \sin\omega_c t \sin\omega_m t\right].$$

Adding the carrier term A cos $\omega_c t$ and combining like terms, we have

$$\phi(t) = \cos\omega_c t\left(A + B\cos\omega_m t\right) - B\sin\omega_c t\sin\omega_m t.$$

In the first circuits class, we see that if we want to write a function of one frequency in the form E(t) cos ($\omega_c t$ + phase angle), the amplitude of the multiplier E is the square root of the squares of the coefficients of cos $\omega_c t$ and sin $\omega_c t$. Thus,

$$\begin{aligned} E(t) &= \sqrt{\left(A + B\cos\omega_m t\right)^2 + \left(B\sin\omega_m t\right)^2} \\ &= \sqrt{A^2 + 2AB\cos\omega_m t + B^2\left(\cos^2\omega_m t + \sin^2\omega_m t\right)}. \end{aligned}$$

Now, of course, the coefficient of B^2 is unity for all values of $\omega_m t$. We find that best performance occurs if $B \ll A$. Then we would have our expression for the envelope (and thus it is detectable using an envelope detector):

$$E(t) = \sqrt{A^2 + B^2 + 2AB\cos\omega_m t} = \sqrt{A^2 + B^2}\sqrt{1 + \frac{2AB}{A^2 + B^2}\cos\omega_m t}.$$

Our condition that $B \ll A$ allows us to say the coefficient of cos $\omega_m t$ is really small compared to unity. We use the binomial theorem to approximate the second square root: $(1 + x)^n \approx 1 + x/2 + (1/2)^2 (-1/2)x^2$ when $x \ll 1$. Using our approximation, $x \approx (2B/A) \cos \omega_m t$. In our expansion, the x term is the modulation term we were seeking, the x^2 term contributes second harmonic distortion. Using the various approximations, and stopping after we find the second harmonic (other harmonics *will* be present, of course, but in decreasing amplitudes), we have

$$\text{Detected } f(t) = B\cos\omega_m t - (1/2)\left(B^2/A\right)\cos^2(\omega_m t).$$

When we use trig identities to get the second harmonic, we get another factor of one half; the ratio of detected second harmonic to fundamental is thus (1/4)(B/A). Thus, for example, if B is just 10% of A, second harmonic is only 2.5% of fundamental.

16.10 Direct versus Indirect Means of Generating FM

Let us first remind ourselves of basics regarding FM. We can write the time function in its simplest form as

$$\phi_{FM}(t) = A\cos\left(\omega_c t + \beta\sin\omega_m t\right).$$

Now, the alert reader might be saying, "Hold on! That looks a lot like phase modulation. If $\beta = 0$, the phase would increase linearly in time, as an unmodulated signal, but gets advanced or retarded a maximum of β." One needs to remember the definition of instantaneous frequency, which is

$$f_i = \frac{1}{2\pi}\frac{d}{dt}\left(\omega_c t + \beta\sin\omega_m t\right) = \frac{1}{2\pi}\left(2\pi f_c + \beta 2\pi f_m\cos\omega_m t\right) = f_c + \beta f_m\cos\omega_m t.$$

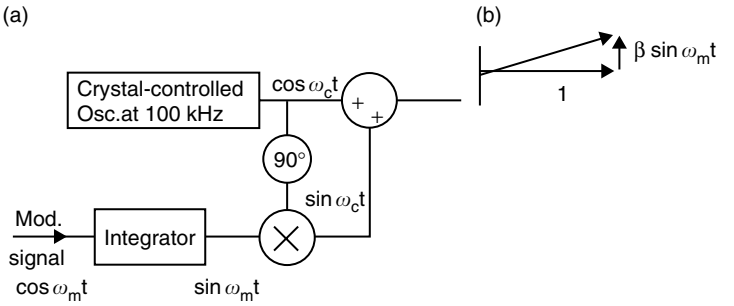


FIGURE 16.8 (a) Crystal-stabilized phase modulator; (b) phasor diagram.

Thus, we can say that instantaneous frequency departs from the carrier frequency by a maximum amount βf_m, which is the so-called "frequency deviation." This has been specified as a maximum of 75 kHz for commercial FM radio but 25 kHz for the sound of TV signals.

Now, certainly, the concept of directly generating FM has an intellectual appeal to it. The problems of direct FM are mainly practical; if the very means of putting information onto a high frequency "carrier" is in varying the frequency, it perhaps stands to reason the center value of the operating frequency will not be well nailed down. Direct FM could be accomplished as in Figure 16.8a, but Murphy's law would be very dominant and one might expect center frequency to drift continually in one direction all morning and the other way all afternoon, or the like. This system is sometimes stabilized by having an FM detector called a discriminator tuned to the desired center frequency, so that output would be positive if frequency got high and negative for frequency low. Thus, instantaneous output could be used as an error voltage with a long time constant to push the intended center frequency toward the center, whether the average value is above or below.

The best known method of indirect FM gives credit to the man who, more than any other, saw the possibilities of FM and that its apparent defects could be exploited for superior performance, Edwin Armstrong. He started with a crystal-stabilized oscillator around 100 kHz, from which he obtained also a 90° phase-shifted version. A block diagram of just this early part of the Armstrong modulator is shown in Figure 16.8a.

The modulating signal is passed through an integrator before it goes into an analog multiplier, to which is also fed the *phase-shifted* version of the crystal-stabilized signal. Thus, we feed $\cos \omega_c t$ and $\sin \omega_c t \sin \omega_m t$ into a summing amplifier. The phasor diagram shows the two signals with $\cos \omega_c t$ as the reference. There is a small phase shift given by $\tan^{-1}(\beta \sin \omega_m t)$ where β here gives the maximum amount of phase shift as a function of time. To see how good a job we have done, we need to expand $\tan^{-1}(x)$ in a Taylor series. We find that

$$\tan^{-1}(x) \approx x - (x)^3/3 + (x)^5/5 .$$

We see that we have a term proportional to the modulating signal (x) and others that must represent odd-order harmonic distortion, if one accounts for the fact that we have resorted to a subterfuge, using a phase modulator to produce frequency modulation. Assuming that our signal finally goes through a frequency detector, we find that the amount of third harmonic as a fraction of the signal output is $\beta^2/4$. Now, in frequency modulation, the maximum amount of modulation which is permitted is in terms of frequency deviation, an amount of 75 kHz. The relation between frequency deviation and maximum phase shift is

$$\Delta f = \beta f_m ,$$

where Δf is the frequency deviation, β is maximum phase shift, and f_m is modulation frequency. Since maximum modulation is defined in terms of Δf, the maximum value of β permitted will correspond to *minimum* modulation frequency. Let us do some numbers to illustrate this problem.

Example 16.3

Suppose we have a high fidelity broadcaster wishing to transmit bass down to 50 Hz with maximum third harmonic distortion of 1%. Find the maximum values of β and Δf.

Solution We have $\beta^2/4 = 0.01$. Solving for β, we get $\beta = 0.2$.

Then, $\Delta f = 0.2 \times 50$ Hz = 10 Hz.

One can recall that the maximum value of frequency deviation allowed in the FM broadcast band is 75 kHz. Thus, use of the indirect modulator has given us much lower frequency deviation than is allowed, and clearly some kind of desperate measures are required. Such are available, but do complicate the process greatly. Suppose we feed the modulated signal into an amplifier which is not biased for low distortion, that is, its Taylor series looks like

$$a_1x + a_2x^2 + a_3x^3, \text{ etc.}$$

Now the squared term leads to second harmonic, the cubed one gives third harmonic, and so on. The phase-modulated signal looks like $A \cos(\omega_c t + \beta \sin \omega_m t)$ and the term a^2x^2 *not only doubles the carrier frequency, but also the maximum phase shift* β. Thus, starting with the rather low frequency of 100 kHz, we have a fair amount of multiplying room before we arrive in the FM broadcast band 88 to 108 MHz. Unfortunately, we may need different amounts of multiplication for the carrier frequency than we need for the depth of modulation. Let's carry on our example and see the problems that arise. First, if we wish to go from

$$\Delta f = 10 \text{ Hz to } 75{,}000 \text{ Hz},$$

that leads to a total multiplication of 75,000/10 = 7,500.

The author likes to say we are limited to frequency doublers and triplers. Let's use as many triplers as possible; we divide the 7500 by 3 until we get close to an even power of 2:

$$7500/3 = 2500;\ 2500/3 = 833,\ 833/3 = 278;\ 278/3 \approx 93,\ 93/3 = 31,$$

which is very close to $32 = (2)^5$.

So, to get our maximum modulation index, we need five each triplers and doublers. However, 7500×0.1 MHz = 750 MHz, and we have missed the broadcast band by about 7 times. One more thing we need is a mixer, after a certain amount of multiplication. Let's use all the doublers and one tripler to get a multiplication of $32 \times 3 = 96$, so the carrier arrives at 9.6 MHz. Suppose our final carrier frequency is 90.9 MHz, and because we have remaining to be used a multiplication of $3^4 = 81$, what comes out of the mixer must be

$$90.9/81 = 1.122 \text{ MHz}$$

To obtain an output of 1.122 MHz from the mixer, with 9.6 MHz going in, we need a local oscillator of either 10.722 or 8.478 MHz. Note that this local oscillator needs a crystal control also, or the eventual carrier frequency will wander about more than is allowed.

16.11 Quick-and-Dirty FM Slope Detection

A method of FM detection that is barely respectable, but surprisingly effective, is called "slope detection." The principle is to feed an FM signal into a tuned circuit, not right at the resonant frequency but rather somewhat off the peak. Therefore, the frequency variations due to the modulation will drive the signal up and down the resonant curve, producing simultaneous amplitude variations, which then can be detected using an envelope detector. Let us just take a case of FM and a specific tuned circuit and find the degree of AM.

Example 16.4

We have an FM signal centered at 10.7 MHz, with frequency deviation of 75 kHz. We have a purely parallel resonant circuit with a Q = 30, with resonant frequency such that 10.7 MHz is at the lower half-power frequency. Find the output voltage for Δf = +75 kHz and for –75 kHz.

Solution When we operate close to resonance, adequate accuracy is given by

$$V_o = \frac{V_i}{1 + j2Q\delta'}$$

where δ is the fractional shift of frequency from resonance. If now, 10.7 MHz is the lower half-power point, we can say that 2Qδ = 1.

$$\text{Hence, } \delta = 1/(2 \times 30) = (f_o - 10.7\ \text{MHz})/f_o;\ f_o = 10.881\ \text{MHz}.$$

Now, we evaluate the transfer function at 10.7 MHz ± 75.kHz.
We defined it as 0.7071 at 10.7 MHz. For 10.7 + 0.075 MHz, δ = (10.881 – 10.775)/ 10.881 = 9.774×10^{-3}, and the magnitude of the transfer function is $|1/(1 + j60\delta)| = 0.8626$.

Because the value was 0.7071 for the unmodulated wave, the modulation index in the positive direction would be

$$(0.8624 - 0.7071)/0.7071 = 0.2196 \text{ or } 21.96\%.$$

For (10.7 – 0.075) MHz, δ = (10.881 – 10.625)/10.881 = 0.02356, and the magnitude of the transfer function is $|1/(1 + j60)| = 0.5775$. The modulation index in the negative direction is (0.7071 – 0.5775)/0.7071 = 18.32%. So, modulation index is not the same for positive as for negative indices. The consequence of such asymmetry is that this process will be subject to harmonic distortion, which is why this process is not quite respectable.

16.12 Lower Distortion FM Detection

We will assume that the reader has been left wanting an FM detector that has much better performance than the slope detector. A number of more complex circuits have a much lower distortion level than the slope detector. One, called the Balanced FM Discriminator, is shown in Figure 16.9.

Basically, we may consider that the circuit contains two "stagger-tuned" resonant circuits, that is, they are tuned equidistant on opposite sides of the center frequency, connected back to back. The result is that the nonlinearity of the resonant circuits balance each other out, and the FM detection can be very linear.

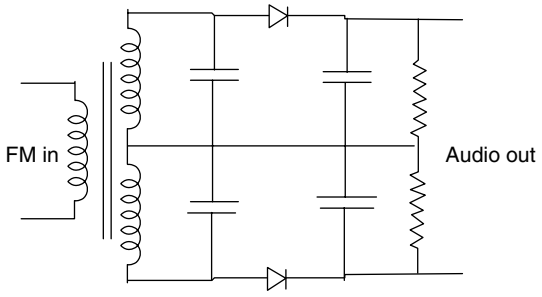


FIGURE 16.9 Balanced FM discriminator.

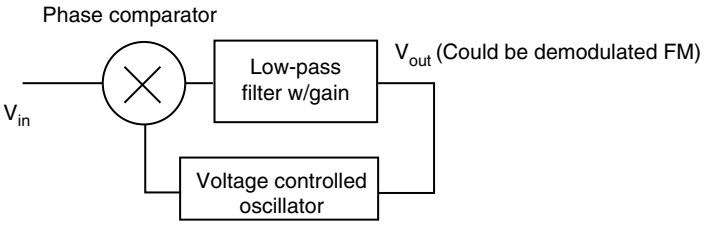


FIGURE 16.10 Basic phase-locked loop.

The engineer designing an FM receiving system has a relatively easy job to access such performance; all that he/she must do is to spend the money to obtain high-quality components.

16.12.1 Phase-Locked Loop

The phase-locked loop is an assembly of circuits or systems that perform a number of functions to accomplish several operations, any one or more of the latter, perhaps being useful and to be capitalized upon. If one looks at a simple block diagram, one will see something like Figure 16.10.

Thus, one function that will always be found is called a "voltage-controlled oscillator;" the linking of these words means that there is an oscillator which would run freely at some frequency, but that if a non-zero DC voltage is fed into a certain input, the frequency of oscillation will shift to one determined by that input voltage. Another function one will always find (although the nomenclature might vary somewhat) is "phase-comparison." The phase "comparator" will usually be followed by some kind of low-pass filter. Of course, if a comparator is to fulfill its function, it requires two inputs—the phases of which to compare. This operation might be accomplished in various ways; however, one method which might be understood from previous discussions is the analog multiplier. Suppose an analog multiplier receives the inputs $\cos \omega t$ and $\sin(\omega t + \phi)$; their product has a sine and a cosine. Now, a trigonometric identity involving these terms is

$$\sin A \cos B = 0.5\left[\sin\left(A + B\right) + \sin\left(A - B\right)\right].$$

Thus, the output of a perfect analog multiplier will be $0.5[\sin(2\omega + \phi) + \sin\phi]$. A low-pass filter following the phase comparator is easily arranged; therefore, one is left with a DC term, which, if it is fed to the VCO in such a polarity as to provide negative feedback, will "lock" the VCO to the frequency of the input signal with a fixed phase shift of 90°.

Phase-locked loops (abbreviated PLL) are used in a wide variety of applications. Many of the applications are demodulators of one sort or another, such as synchronous detectors for AM, basic FM, FM–stereo detectors, and in very precise oscillators known as "frequency synthesizers." One of the early uses seemed to be the detection of weak FM signals, where it can be shown that they extend the threshold of usable weak signals a bit.* This latter facet of their usefulness seems not to have made a large impact, but the other aspects of PLL usefulness are very commonly seen.

16.13 Digital Means of Modulation

The sections immediately preceding have been concerned with rather traditional analog methods of modulating a carrier. While the beginning engineer can expect to do little or no design in analog communication systems, they serve as an introduction to the digital methods which most certainly will dominate the design work early in the twenty-first century. Certainly, analog signals will continue to be generated, such as speech, music, and video; however, engineers are finding it so convenient to do digital signal processing that many analog signals are digitized, processed in various performance-enhancing ways, and only restored to analog format shortly before they are fed to a speaker or picture tube. Digital signals can be transmitted in such a way as to use extremely noisy channels. Not long ago, the nightly news brought us video of the Martian landscape. The analog engineer would be appalled to know the number representing traditional signal-to-noise ratio for the Martian signal. The detection problem is greatly simplified because the digital receiver does not need at each instant to try to represent which of an infinite number of possible analog levels is correct; it simply asks, was the signal sent a one or a zero? *That* is simplicity.

Several methods of digital modulation might be considered extreme examples of some kind of analog modulation. Recall amplitude modulation. The digital rendering of AM is called "amplitude shift keying," abbreviated ASK.

What this might look like on an oscilloscope screen is shown in Figure 16.11. For example, we might say that the larger amplitude signals represent the logic ones and smaller amplitudes represent logic zeroes. Thus, we have illustrated the modulation of the data stream 10101. If the intensity of modulation were carried to the 100% level, the signal would disappear completely during the intervals corresponding to zeroes. The 100% modulation case is sometimes called on–off keying and abbreviated OOK. The latter case has one advantage if this signal were nearly obscured by large amounts of noise; it is easiest for the digital receiver to distinguish between ones and zeroes if the difference between them is maximized. That is, however, only one aspect of the detection problem. It is also often necessary to know the timing of the bits, and for this one may use the signal to synchronize the oscillator in a phase-locked loop; if, for

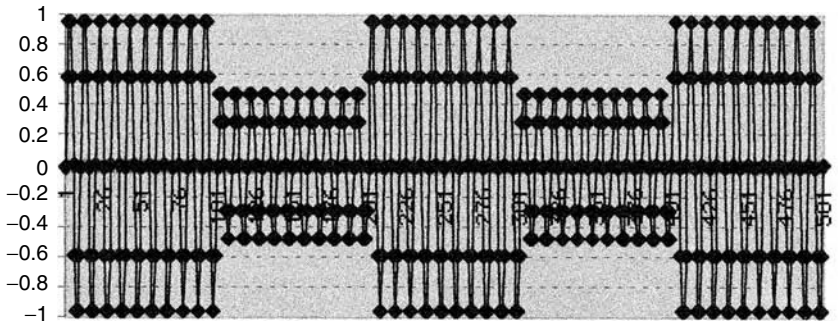


FIGURE 16.11 ASK (amplitude shift keying).

* Taub, H. and Schilling, D.L. *Principles of Communication Circuits,* 2nd edition, McGraw-Hill, New York, 1986, 426–427.

50% of the time, there is zero signal by which to be synchronized, the oscillator may drift significantly. In general, a format for digital modulation in which the signal may vanish utterly at intervals is to be adopted with caution and with full cognizance of one's sync problem. Actually, amplitude shift keying is not considered a very high performance means of digital signaling, in much the same was that AM is not greatly valued as a quality means of analog communication. What is mainly used is one or the other of the following methods.

16.13.1 Frequency Shift Keying

Frequency shift keying (abbreviated FSK) can be used in systems having very little to do with high data rate communications; for years it has been the method used in the simple modems one first used to communicate with remote computers. For binary systems, one just sent a pulse of one frequency for a logic one and a second frequency for a logic zero. If one was communicating in a noisy environment, the two signals would be orthogonal, which meant that the two frequencies used were separated by at least the data rate. Now, at first the modem signals were sent over telephone lines which were optimized for voice communications, and were rather limited for data communication. Suppose we consider that for ones we send a 1250 Hz pulse and for zeroes, we send 2250 Hz. In a noisy environment one ought not to attempt sending more than 1000 bits per second (note that 1000 Hz is the exact difference between the two frequencies being used for FSK signaling). Let us instead send at 250 bps. Twelve milliseconds of a 101 bit stream would look as in Figure 16.12.

It is not too difficult to imagine a way to obtain FSK. Assuming one does have access to a VCO, one simply feeds it two different voltage levels for ones and for zeroes. The VCO output is the required output.

16.13.2 Phase Shift Keying

Probably the most commonly used type of digital modulation is some form of phase shift keying. One might simply say there is a carrier frequency f_c and that logic zeroes will be represented by $-\sin 2\pi f_c t$, logic ones by $+\sin 2\pi f_c t$. If the bit rate is 40% of the carrier frequency, the data stream 1010101010 might look as in Figure 16.13.

In principle, producing binary phase shift keying ought to be fairly straightforward, if one has the polar NRZ (nonreturn to zero, meaning a logic one could be a constant positive voltage for the duration of the bit, zero being an equal negative voltage) bit stream. If then, the bit stream and a carrier signal are fed into an analog multiplier, the output of the multiplier could indeed be considered $\pm\cos \omega_c t$, and the modulation is achieved.

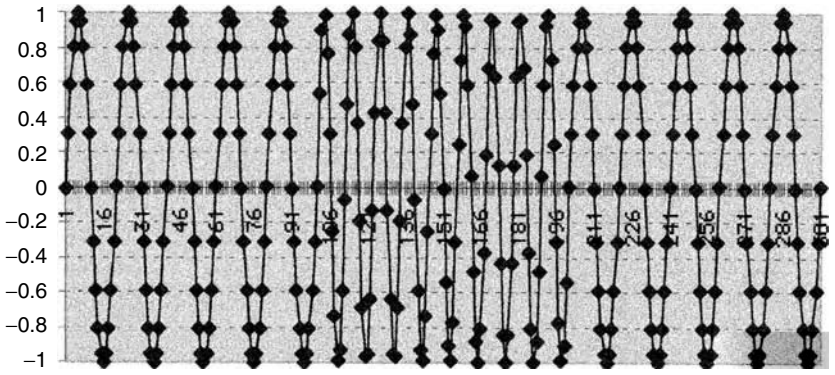


FIGURE 16.12 Frequency shift keying.

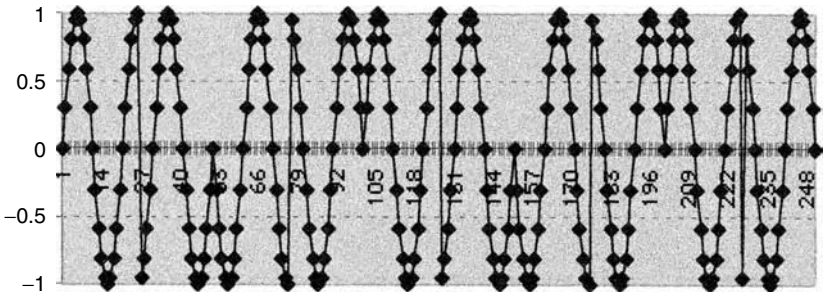


FIGURE 16.13 Phase shift keying.

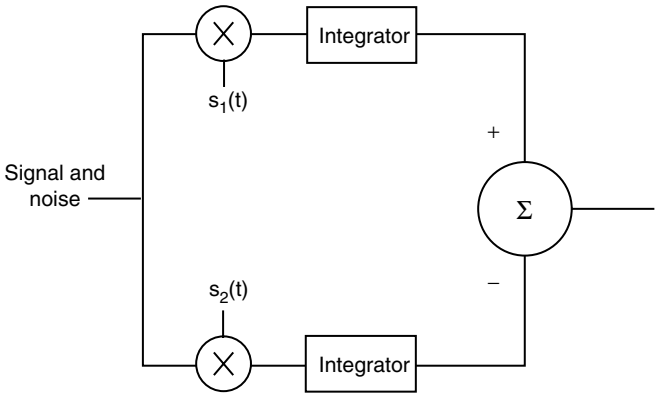


FIGURE 16.14 Correlation detector.

16.14 Correlation Detection

Many years ago, the communications theorists came up with the idea that if one could build a "matched filter," that is, a special filter designed with the bit waveform in mind, one would startlingly increase the signal-to-noise ratio of the detected signal. Before long, a practically minded communications person had the bright idea that a correlation detector would do the job, at least for rectangular bits. For some reason, as one explains this circuit, one postulates two signals, $s_1(t)$ and $s_2(t)$, which represent, respectively, the signals sent for logic ones and zeroes. The basics of the correlation detector are shown in Figure 16.14.

Now, a key consideration in the operation of the correlation detector is bit synchronization. It is crucial that the signal $s_1(t)$ be lined up perfectly with the bits being received. Then, the top multiplier "sees" sin $\omega_c t$ coming in one input, and $\pm\sin \omega_c t$ + noise coming in the other, depending upon whether a one or a zero is being received. If it happens that a one is being received, the multiplier is asked to multiply sin $\omega_c t$(sin $\omega_c t$ + noise). Of course,

$$\sin^2 \omega_c(t) = (1/2)(1+\cos 2\omega_c t).$$

In the integrator, this is integrated over one bit duration, giving a quantity said to be the energy of one bit. The integrator might also be considered to have been asked to integrate n(t) sin $\omega_c t$, where n is the noise signal. However, the nature of noise is that there is no net area under the curve of its waveform, so considering integration to be a summation, the noise output out of the integrator would simply be the last instantaneous value of the noise voltage at the end of a bit duration, whereas the signal output was bit energy, if the bit

synchronization is guaranteed. Meanwhile, the output of the bottom multiplier was the *negative* of the bit energy, so with the signs shown, the output of the summing amplifier is twice the bit energy. Similar reasoning leads to the conclusion that if the instantaneous signal being received were a zero, the summed output would be *minus* twice the bit energy. It takes a rather substantial bit of theory to show that the noise output from the summer is *noise spectral density.* The result may be summarized that the correlation detector can "pull a very noisy signal out of the mud." And, we should assert at this point that the correlation detector can perform wonders for any one of the methods of digital modulation mentioned up to this point.

16.15 Digital QAM

Once the engineer has produced carrier signals that are 90° out of phase with each other, there is no intrinsic specification that the modulation must be analog, as is done for color TV. As a start toward extending the capabilities of PSK, one might consider that one sends bursts of several periods of $\pm\cos \omega_c t$ or $\pm\sin \omega_c t$. This is sometimes called "4-ary" transmission, meaning that there are four different possibilities of what might be sent. Thus, whichever of the possibilities is sent, it may be considered to contain two bits of information. It is a method by which more information may be sent without demanding any more bandwidth, because the duration of the symbol being sent may be no longer or shorter than it was when one was doing binary signaling, sending, for example, simply $\pm\cos \omega_c t$. This idea is sometimes represented in a "constellation," which, for the case we just introduced, would look like part a of Figure 16.15. However, what is more often done is as shown in Figure 16.15b, where it could be said that one is sending $\pm\cos (\omega_c t + 45°)$ or $\pm\cos (\omega_c t + 135°)$. It seems as though this may be easier to implement than the case of part a; however, the latter scheme lends itself well to sending 4 bits in a single symbol, as in Figure 16.15c.

Strictly speaking, one might consider "a" to be the constellation for 4-ary PSK. This leads also to the implication that one could draw a circle with "stars" spaced 45° apart on it and one would have the constellation for 8-ary PSK. The perceptive or well-informed reader might have the strong suspicion that crowding more points on the circle makes it possible to have more errors in distinguishing one symbol from adjacent ones, and would be correct in this suspicion. Hence, M'ary communication probably more commonly uses "b" or "c," which should be considered forms of QAM.

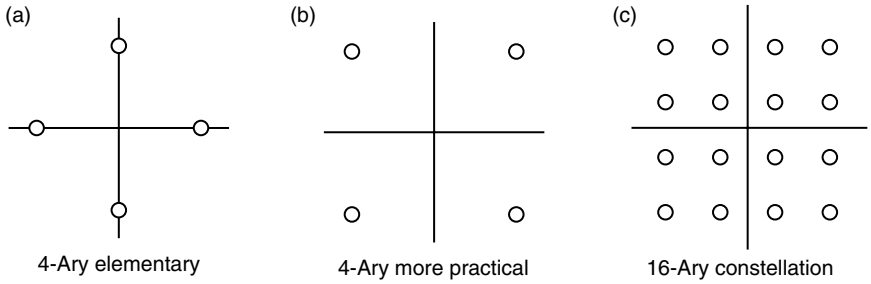


FIGURE 16.15 Constellation showing carrier amplitudes and phase for M'ary signals.

17

Power Amplifier Fundamentals

Douglas A. Teeter
Edward T. Spears
RF Micro Devices

17.1 Introduction

Power amplifiers are a critical component in any radio. Because they are the final stage of the transmit chain, they typically use the most current. Proper design of the power amplifier dictates the overall linearity and efficiency of the transmitter.

This work is divided into two chapters, Chapter 17, *Power Amplifier Fundamentals* and Chapter 18, *Handset Power Amplifier Design. Power Amplifier Fundamentals* discusses transistor operating classes, key determinants of power amplifier efficiency, and characteristics of modulated signals. *Handset Power Amplifier Design* builds on the foundation established in the *Fundamentals* chapter by walking the reader through critical considerations in designing power amplifiers for handsets.

17.2 Transistor Operating Classes

A variety of transistor operating classes may be chosen for a power amplifier. The Sokals [1] first distinguished between current and switching mode operation of power transistors. Figure 17.1 compares the basic difference between current mode and switch mode operation.

In the current mode operation, the transistor behaves as a voltage controlled current source. For the ideal case, the transconductance, g_m, is constant when the transistor is on and zero when the transistor is in cut-off mode. Common classes of current mode operation are class A, AB, B, C, and F. In switch mode operation, the transistor may be viewed as a switch. The most common class of switch mode

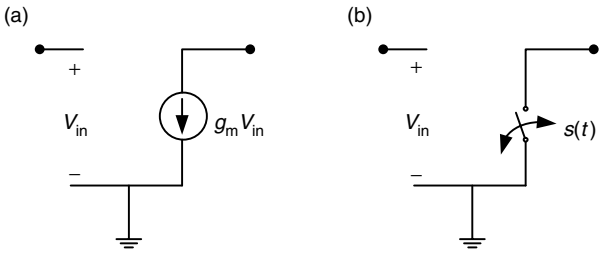


FIGURE 17.1 Ideal modes of transistor operation: (a) current mode and (b) switch mode.

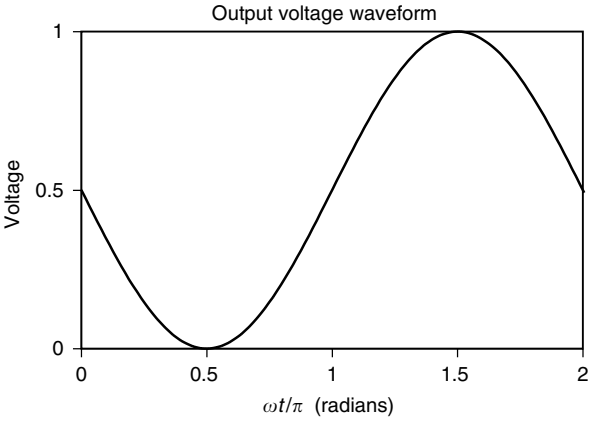


FIGURE 17.2 Output voltage waveform for idealized class A, AB, B, and C operating modes for $V_{supply} = 0.5$ V and $I_{max} = 1$ A.

operation is class E. An excellent summary and comparison of transistor operating classes may be found in References 2 and 3.

17.2.1 Current Mode Operation

Current mode power amplifiers are most commonly used for handsets due to their relative ease of design and linearity. A current mode power amplifier treats a transistor as a voltage controlled current source. This simplistic representation is the foundation for all classical power amplifier analysis regarding class A, B, C, and F modes of operation. We will touch briefly on class A and B/F since these modes are frequently used in handset power amplifiers. More detailed analysis can be found in References 2 to 13 or in many introductory amplifier design textbooks [4,14].

Figures 17.2 and 17.3 compare textbook output voltage and current waveforms for the most common unclipped current mode power amplifiers. The sinusoidal voltage waveform shown in Figure 17.2 implicitly assumes a tuned output matching network that shorts all harmonic voltages to ground [10]. For class A, the transistor conducts all the time and the corresponding conduction angle, θ, is 2π. As the transistor is biased more toward the cut-off region, the conduction angle decreases, moving from class A ($\theta = 2\pi$) to AB ($2\pi > \theta > \pi$), to B ($\theta = \pi$), and to C ($\theta < \pi$). For the waveforms shown in Figures 17.2 and 17.3, the peak efficiency may be expressed as

$$\eta = \frac{1}{4}\frac{\theta - \sin\theta}{\sin(\theta/2) - (\theta/2)\cos(\theta/2)}. \tag{17.1}$$

where θ is the conduction angle [10].

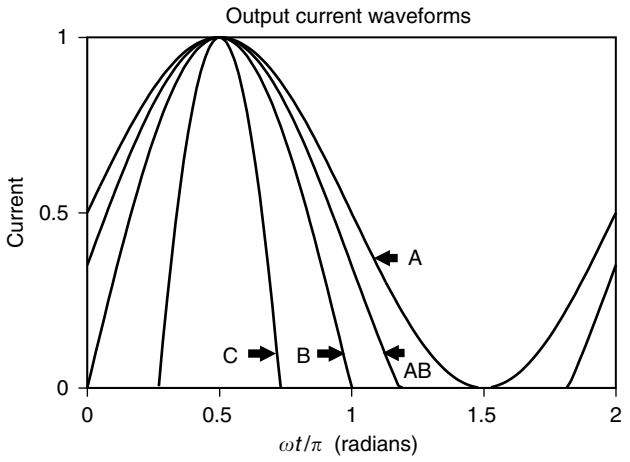


FIGURE 17.3 Output current waveforms for idealized class A, AB, B, and C operating modes for $V_{\text{supply}} = 0.5$ V and $I_{\max} = 1$ A.

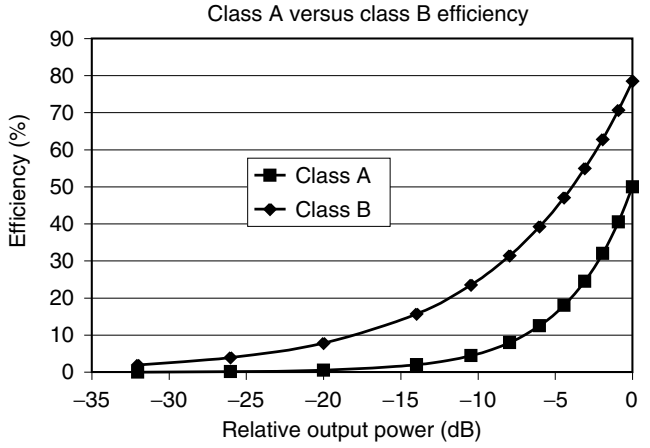


FIGURE 17.4 Output efficiency versus output power for ideal class A and B power amplifiers.

Classical analysis, typically, evaluates the case where the waveforms are not clipped. This results in a maximum class A power amplifier efficiency of 50% *assuming a constant transconductance,* g_{m} [2,4,6,10]. The assumption of a constant g_{m} is subtle, but important, because it affects the shape of the output current waveform. In Reference 15, Kushner compares class A and B efficiency of idealized power amplifiers assuming a constant g_{m} to that of a linear g_{m}. For the class A case with constant g_{m} and optimally tuned fundamental load, a peak no-clipping efficiency of 50% results (neglecting the knee voltage). Assuming instead a linear g_{m}, a peak efficiency of 67% can be obtained. The transconductance of a real transistor is neither constant nor linear, so the peak efficiency will be slightly different than the idealized cases. None the less, these textbook cases provide a qualitative understanding of basic power amplifier operation.

The peak efficiency of the class A amplifier (neglecting clipping), is 50% for constant g_{m} while the peak efficiency for class B (neglecting clipping), is 78.5% [10]. As mentioned earlier, a tuned output matching network which shorts all harmonic voltages to ground is assumed. More interesting is the comparison of efficiency versus output power for the class A and B amplifiers, as shown in Figure 17.4. Both the class A and B power amplifiers have the same RF output power as a function of drive. However, because the DC current for the class A amplifier remains constant over drive and while the current decreases as output

power goes down for the class B case, the efficiency for class A is even worse than class B at back-off power levels.

Assuming an output voltage waveform of

$$V_{\text{out}} = V_{\text{DC}} + V_{\text{m}} \sin(\omega t), \tag{17.2}$$

the resulting output power for both class A and B may be written as

$$P_{\text{out}} = \frac{V_{\text{m}}^2}{2R_{\text{L}}} \quad \text{(Class A or B)}, \tag{17.3}$$

where $R_{\text{L}} = 2V_{\text{DC}}/I_{\text{max}}$ is the fundamental frequency load resistance for maximum efficiency, V_{DC} is the DC bias voltage, and I_{max} is the maximum output current for the transistor. The corresponding DC current is

$$I_{\text{DC}} = \frac{I_{\text{max}}}{2} = \text{const} \quad \text{(Class A)} \tag{17.4}$$

and

$$I_{\text{DC}} = \frac{2I_1}{\pi} = \frac{2(V_{\text{m}}/R_{\text{L}})}{\pi} \quad \text{(Class B)} \tag{17.5}$$

where I_1 is the fundamental frequency component of the current.

This results in an associated output efficiency of

$$\eta_A = \frac{V_{\text{m}}^2}{2V_{\text{DC}}^2} \quad \text{(Class A)} \tag{17.6}$$

and

$$\eta_B = \frac{V_{\text{m}}}{V_{\text{DC}}} \cdot \frac{\pi}{4} \quad \text{(Class B)} \tag{17.7}$$

By sweeping V_{m} and utilizing Equations 17.3, 17.6, and 17.7, Figure 17.4 results [10].

17.2.2 Over Driven Current Mode Operation (Class F)

As one drives a class B amplifier harder (above 0 dB in Figure 17.4), output voltage clipping begins. With proper termination of the fundamental and harmonic impedances, optimal efficiency may be obtained [12,13]. Following analysis similar to [4], one can illustrate the evolution from class B to class F operation. The current waveform is the same as the class B case shown in Figure 17.3. A clipped voltage waveform results when the amplifier is overdriven, as shown in curves 1, 2, and 3 of Figure 17.5. Higher order harmonic voltage terms are introduced into the waveform once clipping takes place [4,12,13]. The associated efficiency is shown in Figure 17.6 where the 0 dB reference point is the ideal unclipped class B case. Up to 100% theoretical efficiency is possible if an infinite number of harmonics are included in the analysis. For constant envelope applications, such as with global system mobile (GSM) or large signal polar modulation (discussed further in the next chapter), class F amplifiers are appealing due to their high efficiency. Class F amplifiers are not appropriate for applications where a linear power amplifier is needed since significant clipping occurs in the output voltage waveform.

References 3, 11–13 give a comprehensive analysis of class A to F operation. It is interesting to note that up to 90% efficiency can be obtained by properly terminating three harmonics, but the progression to higher peak efficiency is very slow even as additional harmonics are added to the analysis [3].

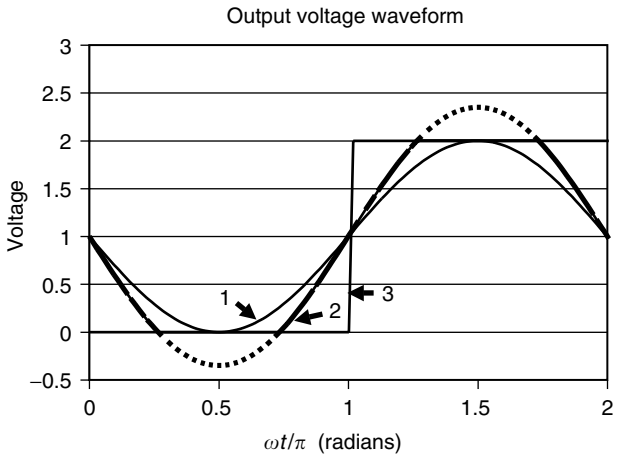


FIGURE 17.5 Output voltage waveform for overdriven class B operation (transition from class B to class F).

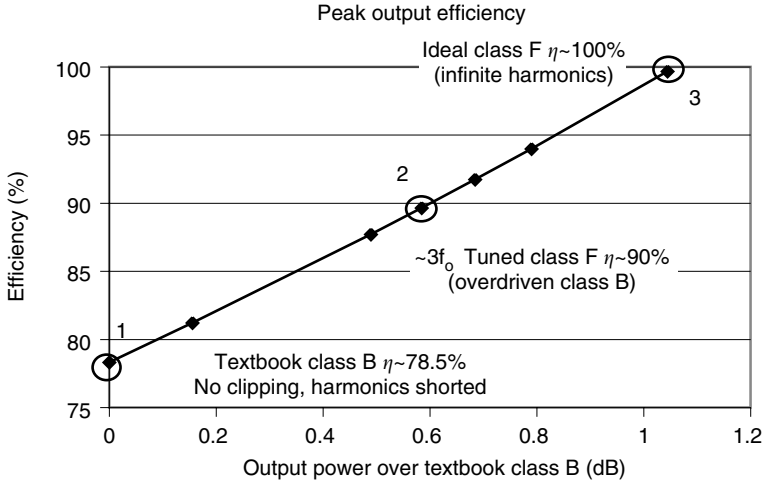


FIGURE 17.6 Output efficiency for overdriven class B operation.

17.2.3 Switch Mode Operation (Class E)

Another operating mode is class E. In this case, the transistor is operated as a switch instead of a current source with the device voltage and current waveforms determined completely by the output matching network. Figure 17.7 illustrates ideal class E operation. The capacitor, C_{tot}, includes both the device capacitance as well as added external capacitance necessary for proper class E operation. Early analysis treated this capacitor as a constant value but more recent treatments have included a nonlinear component [16–18]. Numerous class E amplifier papers have appeared in the literature over the past several years (see [1–3,16–23]). Output voltage and current waveforms at the device for the ideal optimum class E amplifier with infinite harmonics are shown in Figures 17.8 and 17.9 [24]. The voltage at the collector (or drain) of the device (V_{dev}) is 3.5 times the supply voltage during optimum ideal class E operation.

For the case of a finite number of harmonics, as is the case of most RF power amplifiers, there is little difference in efficiency between classes C, E, and F. The efficiency is controlled by the number of harmonics, not the operating class [3]. An excellent analysis showing the relationship between classes C, E, and F may be found in Reference [3].

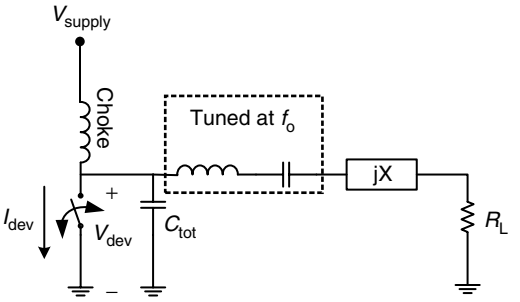


FIGURE 17.7 Idealized class E operation.

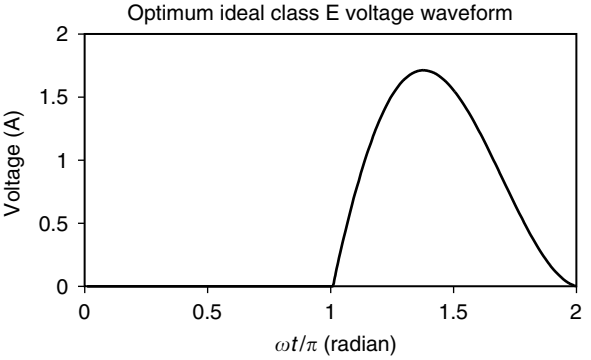


FIGURE 17.8 Ideal optimum class E amplifier device voltage waveform (V_{dev}) for $V_{supply} = 0.5$ V and $I_{max} = 1$ A [24].

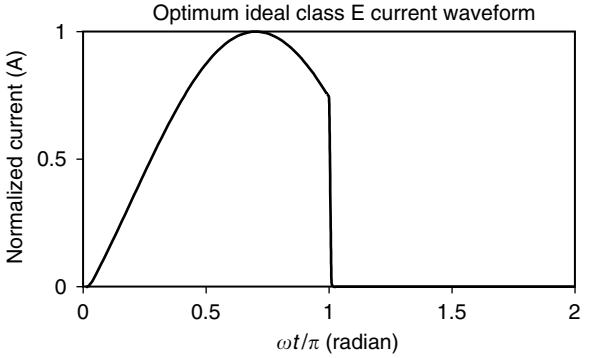


FIGURE 17.9 Ideal optimum class E device current waveform (I_{dev}) for $V_{supply} = 0.5$ V and $I_{max} = 1$ A [24].

17.2.4 Device Considerations

Real devices have nonzero on state resistance. Parasitic on state resistance will limit the maximum efficiency attainable out of a device due to the knee voltage it creates [14 (Section 12-3), 15]. The knee voltage limits the maximum voltage swing out of the transistor. This, in turn, reduces the peak efficiency attainable from the transistor, which can be expressed mathematically through the factor α, where

$$\alpha = \frac{V_{DC} - V_{knee}}{V_{DC}} \tag{17.8}$$

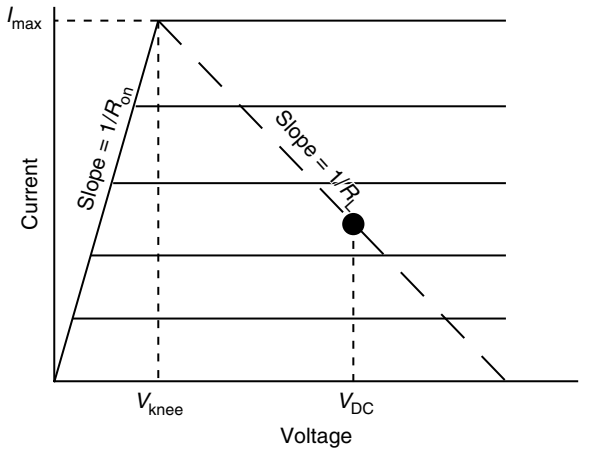


FIGURE 17.10 Idealized device I–V curve showing on-state resistance, knee voltage, and load line.

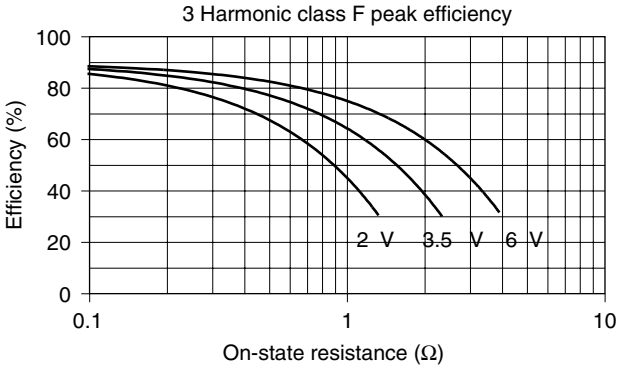


FIGURE 17.11 Output efficiency for a third harmonic tuned class F output stage as a function of device on resistance and supply voltage. (Device I_{max} normalized to 1 A.)

In Equation 17.8, V_{DC} is the DC supply voltage applied to the transistor and V_{knee} is the knee voltage of the transistor. The peak efficiency may then be written as [15]

$$\eta = \alpha \cdot \eta_{ideal} \tag{17.9}$$

where η_{ideal} is the ideal peak efficiency of the device in the absence of any knee voltage. The knee voltage is related to the on-state resistance through

$$V_{knee} = I_{max} R_{on} \tag{17.10}$$

where I_{max} is the maximum safe operating current out of the transistor and R_{on} is the on-state resistance. Figure 17.10 illustrates the on-state resistance for an idealized device along with its effect on the load line.

For example, Equations 17.8 through 17.10 may be used to calculate the impact the nonideal on state resistance would have for a third-harmonic-tuned class F output stage. The ideal peak efficiency is about 90%, as shown in Figure 17.6. Figure 17.11 illustrates how the on-state resistance of a transistor can drastically reduce this efficiency. For simplicity, I_{max} has been normalized to 1 A in Figure 17.11. The parasitic effect becomes especially noticeable at low battery voltages.

17.3 Power Amplifier Efficiency

Several factors contribute to the overall maximum efficiency of a power amplifier (PAE_{MAX}). Both device and circuit level factors are involved. At the device level, the peak output terminal efficiency* of the output stage ($\eta_{\text{term}}^{\text{MAX}}$) is controlled by the operating class, fundamental load match (tradeoff between efficiency and linearity), harmonic terminations, device knee voltage, and by on-state breakdown. At the circuit level, the peak power added efficiency is controlled by the output matching circuit loss (*output loss factor*), the current ratio between stages (*multistage ratio factor*), and the total gain of the amplifier (G). Mathematically, this relationship may be written as

$$\text{PAE}_{\text{MAX}} = \eta_{\text{term}}^{\text{MAX}} \cdot \text{output loss factor} \cdot \text{multistage ratio factor} \cdot \left(1 - \frac{1}{G}\right). \tag{17.11}$$

For a power amplifier operating at 50% collector (drain) efficiency at the collector (drain) terminal of the output stage, every 0.1 dB of output matching network loss results in 1 point reduction in efficiency. More detailed discussion of contributors to output matching network loss will be given in Section 18.3.2. Figure 17.12 plots the output match loss factor used in Equation 17.11.

A second factor in any power amplifier design is how many stages are needed. This is determined by how much gain is needed from the amplifier. Adding more stages increases gain, but at the expense of efficiency. The critical determinants for this factor, denoted "multistage ratio factor" in Equation 17.11, are the number of stages and the ratio of current in each stage. Figure 17.13 plots the multistage factor assuming the current ratio between each stage is the same. The current ratio is simply the ratio of currents between two adjacent stages, as shown on the inset to Figure 17.13. The last two stages control most of the current consumption of the power amplifier. Once the current ratio is above about 10, having more than two stages has minimal impact on overall efficiency.

17.4 Figures of Merit for Modulated RF Signals

Modulated signals place additional requirements on the power amplifier. In order to avoid significant distortion, the power amplifier must operate linearly in both amplitude and phase. This section describes

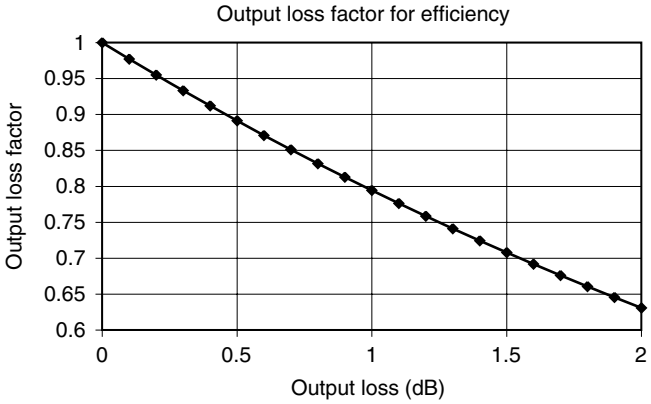


FIGURE 17.12 Efficiency reduction factor due to output loss.

*For a bipolar device, η_{term}^{MAX} would be the collector efficiency and for a FET device η_{term}^{MAX} would be the drain efficiency.

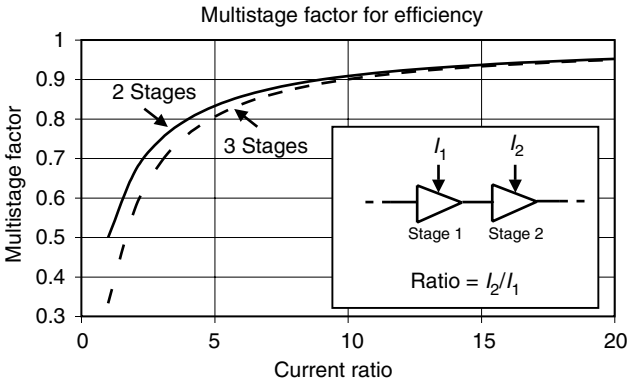


FIGURE 17.13 Efficiency reduction factor due to the current ratio for two and three stage amplifiers.

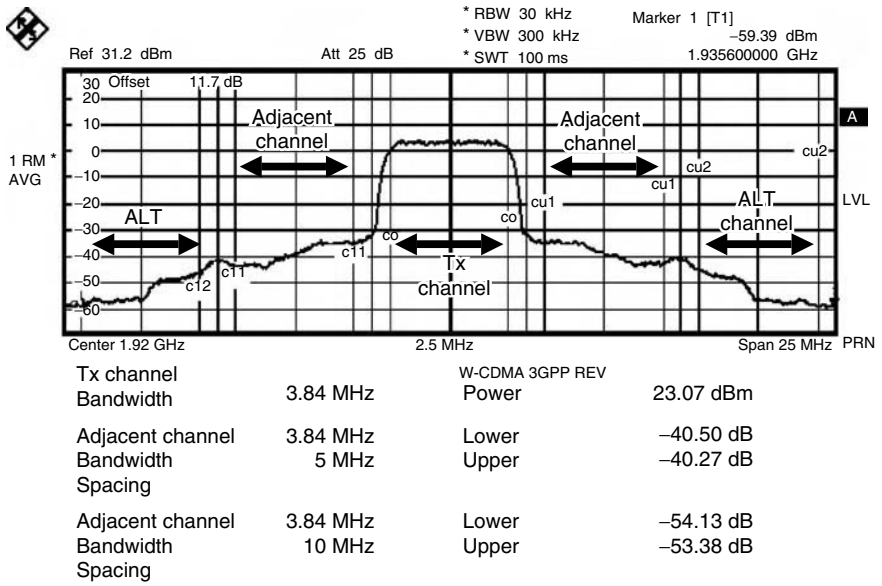


FIGURE 17.14 Typical output spectrum for a WCDMA power amplifier.

the most common figures of merit used to describe power amplifier linearity and the amount of output power back-off required for linearpower amplifier operation.

17.4.1 ACPR

One of the most critical design parameters for handset power amplifiers is linearity. A common parameter used to describe the linearity is Adjacent Channel Power Ratio (ACPR) [5,25]. Figure 17.14 shows a typical output spectrum for a WCDMA (Wideband Code Division Multiple Access) signal. The ACPR is simply the power in the adjacent channel (CL_1 or CU_1) divided by the power in the main channel (CO). Mathematically, ACPR can be expressed as

$$\text{ACPR} = \frac{\int_{\text{CU}_1} P(f)\mathrm{d}P}{\int_{\text{CO}} P(f)\mathrm{d}P} = \frac{\text{Adjacent channel power}}{\text{Transmit channel power}}. \tag{17.12}$$

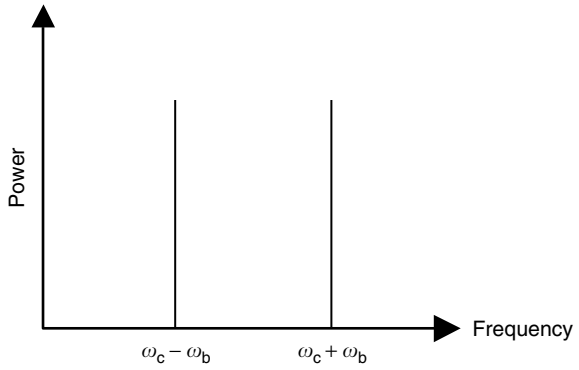


FIGURE 17.15 Frequency domain representation of a two-tone input signal.

Similarly, the alternate channel power (ALT) is simply the power in the alternate channel (CL_2 or CU_2) divided by the power in the main channel (CO). Mathematically, this may be written as

$$\text{ALT} = \frac{\int_{\text{CU}_2} P(f)\text{d}P}{\int_{\text{CO}} P(f)\text{d}P} = \frac{\text{Alternate channel power}}{\text{Transmit channel power}}. \tag{17.13}$$

The channel spacing and bandwidth for the ACPR and ALT measurement depends on the standard being measured. For WCDMA, the channel bandwidth is 3.84 MHz and the spacing is 5 MHz between channels.

17.4.2 CCDF

A second figure of merit is the complementary cumulative density function (CCDF) [26]. The CCDF tells a designer what percent of the time a signal is above a certain power level. The average power level represents 50% CCDF. A two-tone signal, shown in Figure 17.15, can be used to illustrate the CCDF concept. For a two-tone signal, the normalized voltage waveform may be written as

$$V(t) = \tfrac{1}{2}\cos[(\omega_c - \omega_b)t] + \tfrac{1}{2}\cos[(\omega_c + \omega_b)t]. \tag{17.14}$$

Rearranging terms, $V(t)$ becomes

$$V(t) = \cos(\omega_b t)\cos(\omega_c t) \tag{17.15}$$

where ω_c is the carrier frequency and ω_b is the baseband frequency. Rewriting Equation 17.14 as 17.15, it can easily be seen that the signal consists of a carrier modulated by a sine wave. The normalized signal power is simply $V^2(t)$. The two-tone envelope power* is plotted in Figure 17.16. The CCDF is the percentage of time the instantaneous envelope power is greater than "a." Due to the predictable nature of the two-tone case, this can easily be shown graphically, as denoted in Figure 17.16. A CCDF plot versus power can be generated by varying the value of "a." The result for the two-tone case is shown in Figure 17.17.

Modulation formats used for wireless communication are more complex, but the same concepts apply. Mathematically, the signal voltage for a modulated signal can be broken down into an in phase (I)

* Envelope power is defined as the power in the time varying baseband signal.

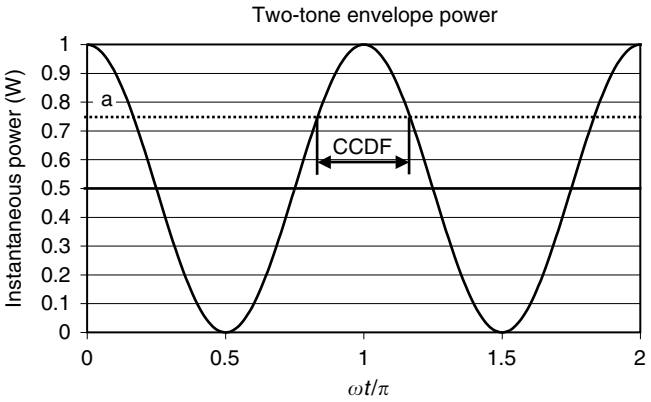


FIGURE 17.16 Normalized instantaneous envelope output power for the two-tone case.

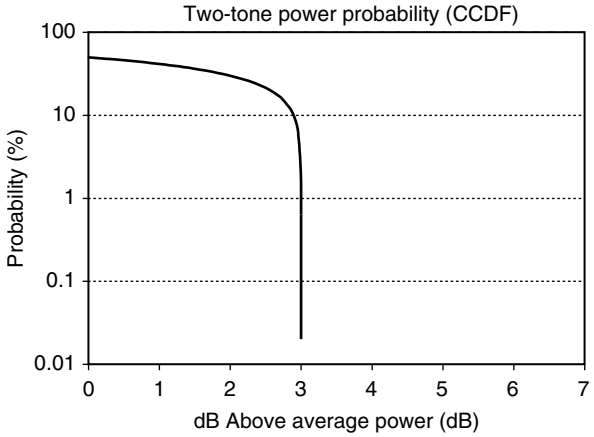


FIGURE 17.17 Two-tone power probability (CCDF).

and quadrature components (Q)

$$V(t) = A(t)\cos(\omega_c t + \phi(t)) \tag{17.16}$$

$$V(t) = I(t)\cos(\omega_c t) + Q(t)\sin(\omega_c t) = \mathrm{Re}[\{I(t) + jQ(t)\}e^{-j\omega_c t}] \tag{17.17}$$

For example, Figure 17.18 shows the I–Q constellation diagram for a standard WCDMA signal (Release 99) and Figure 17.19 shows the I–Q constellation diagram for the 12/15 high speed downlink packet access (HSDPA) modulation formats,* the worst case format for the handset PA. The corresponding CCDF is shown in Figure 17.20. Comparing Figures 17.18 through 17.20, one notices that as the number points in the constellation diagram increases (Figures 17.18 versus 17.19), one sees a corresponding increase in the amount of time the signal is above its average value (Figure 17.20). The added constellation points in the HSDPA modulation allow a higher bit rate out of the phone, but it also results in a signal with peaks that exceed the average signal level for a higher percentage of the time.†

*The worst case HSDPA format for the handset (DPCCH/DPDCH=12/15 and DPCCH/HSDPCCH=15/24) is shown.

†In the case of HSDPA, the data rate available to the handset user (reverse link) does not increase, but the amount of information that needs to be transmitted to the base station from the handset increases to maintain backward compatibility with WCDMA (Rel 99), so the total data rate out of the handset goes up.

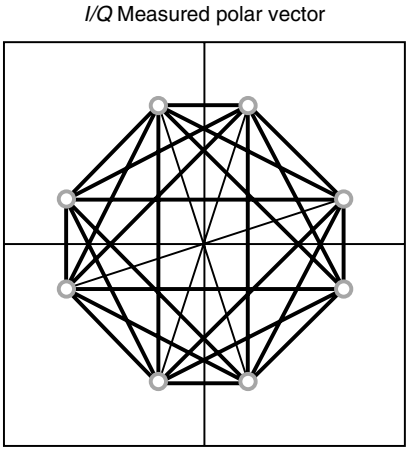


FIGURE 17.18 Unfiltered *I*–*Q* constellation diagram for WCDMA (Release 99) modulation.

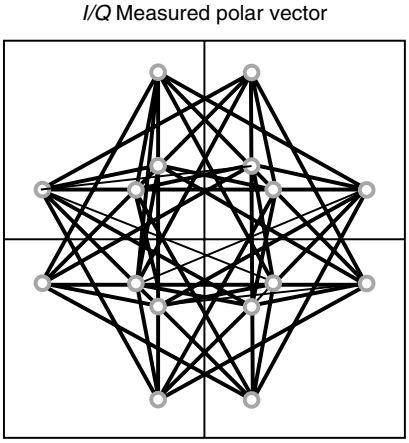


FIGURE 17.19 Unfiltered *I*–*Q* constellation diagram for DPCCH/DPDCH = 12/15. HSDPA modulation (worst case HSDPA format for handset PA).

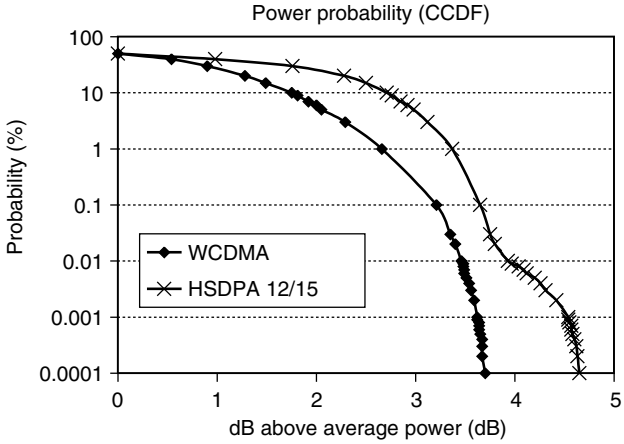


FIGURE 17.20 CCDF curve for WCDMA and DPCCH/DPDCH = 12/15 HSDPA modulation.

The CCDF curves are helpful for power amplifier design because they provide the circuit designer information about the peak-to-average ratio of a signal. As will be discussed more in the next chapter, the higher the peak-to-average ratio, the more backed off the power amplifier must operate.

17.4.3 Cubic Metric

An alternate metric has recently been proposed to allow power amplifier designers an easy way to estimate the amount of back-off required to achieve a given linearity for different modulation schemes. Motorola first proposed the cubic metric (CM) figure of merit [27]. The CM is based on the assumption that the dominant source of nonlinearity in a power amplifier is third-order. Thus, the voltage gain characteristic of the power amplifier may be written as

$$v_o(t) \approx G_1 v_i(t) + G_3[v_i(t)]^3 \tag{17.18}$$

where v_o is the output voltage, v_i is the input voltage, G_1 is the linear gain, and G_3 is the third-order gain nonlinearity. If $v_i(t)$ is a modulated signal, the cubic term in Equation 17.18 is responsible for any resulting ACPR products. The CM compares the amount of power in the cubic term of Equation 17.18 relative to a reference modulation. To compute the CM, first the input signal, $v_i(t)$, is normalized such that its rms value is 1. The normalized input signal voltage is then cubed and the rms value of the cubed term is converted to dB to calculate the third-order power. Typically, this power is compared to the third-order power of a reference modulation such as 3GPP WCDMA voice modulation (Release 99). The CM in dB may be written as*

$$\mathrm{CM} = 20\log_{10}(\{[v_{\mathrm{i\_norm}}]^3\}_{\mathrm{rms}}) - 20\log_{10}(\{[v_{\mathrm{i\_norm\_ref}}]^3\}_{\mathrm{rms}}) \tag{17.19}$$

where

$$\{[v_{\mathrm{i\_norm}}]^3\}_{\mathrm{rms}} = \sqrt{\frac{1}{N}\sum_{1}^{N}([v_{\mathrm{i\_norm}}]^3)^2} \tag{17.20}$$

is the rms value of the cubic power when v_i is normalized such that

$$\{v_{\mathrm{i\_norm}}\}_{\mathrm{rms}} = \sqrt{\frac{1}{N}\sum_{1}^{N}(v_{\mathrm{i\_norm}})^2} = 1. \tag{17.21}$$

Motorola measured several power amplifiers with a number of WCDMA (Release 6) modulation formats and found that for a constant ACPR of −33 dBc, the amount of power amplifier output power back-off compared to 3GPP WCDMA voice modulation was related to the CM of the modulation by a factor of 1/1.85. Thus, the amount of output power back-off (derating) relative to the reference modulation is approximately

$$\text{Output power derating} = \mathrm{CM}/1.85 \tag{17.22}$$

It is worth noting that some references include the factor of 1.85 directly in the CM equation while it is kept separate here. Qualcomm has suggested that the factor 1.85 should be reduced to about 1.38 to allow sufficient linearity margin for practical applications [28].

The Motorola and Qualcomm findings illustrate that, for a given linearity (ACPR), the difference in power amplifier output power between two different modulation formats track the CM much more

*Some references include the 1/1.85 factor in Equation 4.3.2 in the cubic metric. Since there is still some debate about what this multiplication factor should be [28], this factor is kept separate from the CM in Equation 4.3.2. It is included in Equation 4.3.5 instead.

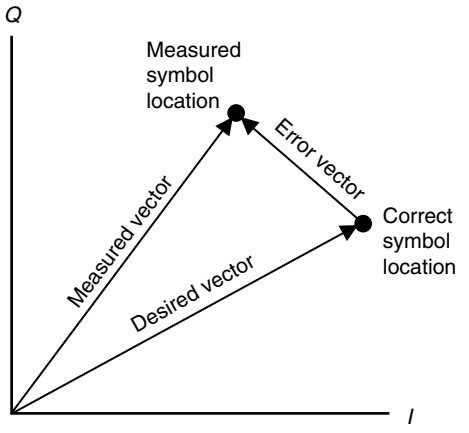


FIGURE 17.21 Error vector concept.

closely than the 99.9% CCDF peak-to-average ratio of the signal. For example, using the CM approach, the output power derating for HSDPA 12/15 modulation is 1 dB relative to WCDMA voice modulation, consistent with the 1 dB back-off observed in Figure 18.18 of the next chapter when ACPR = −33 dBc.

17.4.4 Error Vector Magnitude

Another frequently used metric for modulated signals is the error vector magnitude (EVM). Figure 17.21 illustrates the concept. In any modulated signal, the measured location of a symbol will differ from the ideal or "correct" location. This difference is caused by various system impairments including nonlinearity in the transmitter. The vector difference between the measured and desired location of the symbol represents the error vector. The magnitude of the error vector is known as the EVM. Keeping below a certain maximum EVM level is often a requirement in power amplifier designs. For example, enhanced data-rates for GSM evolution (EDGE) and wireless local area network (WLAN) systems both call for certain maximum levels of EVM.

The error vector is constantly changing over time. EVM is typically specified on a percentage basis relative to the magnitude of the desired vector. Mathematically, the EVM for the kth symbol may be calculated as

$$\text{EVM}_k = \frac{|\text{Error vector}_k|}{|\text{Desired vector}_k|} 100\%. \tag{17.23}$$

The EVM may be specified by one of several ways. One approach is to specify the rms value of the EVM for a certain number of symbols. Alternatively, EVM may be specified to being below a certain maximum level. Another possible specification is known as the 95th percentile, where only 5% of the sampled symbols are allowed to have an EVM above a certain level [29].

Vector signal analyzers are used to measure EVM by comparing the measured symbol locations in I–Q space to the nearest constellation point. Error vector data is stored for many symbols and then used to calculate one of the several EVM measurements outlined above. A direct link between signal-to-noise ratio, EVM, and bit error rate can be established [30].

17.4.5 Composite Rho and Code Domain Power

Rho (ρ) is a figure of merit used to characterize the modulation accuracy of a CDMA (code division multiple access) signal. CDMA signals are first modulated and then "spread" using one of a number of orthogonal Walsh Codes [31]. In the case of IS-95 CDMA, there are 64 possible Walsh codes. Each Walsh code defines a unique communication channel, or user. It is often useful to analyze CDMA systems in

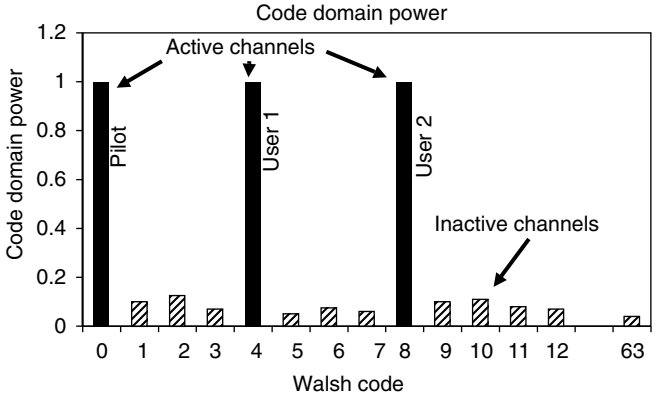


FIGURE 17.22 Power correlated (received in) each Walsh code channel for an example CDMA signal.

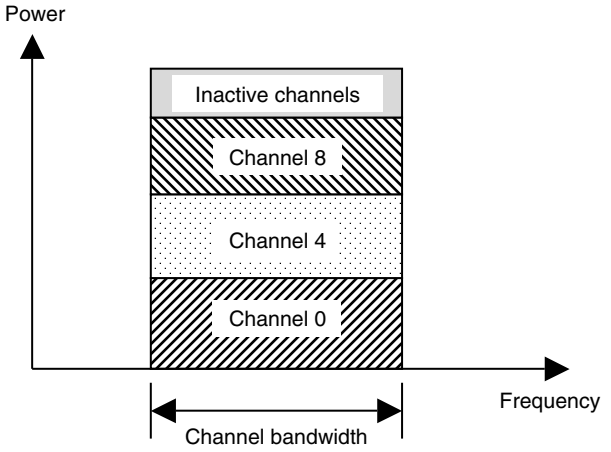


FIGURE 17.23 Simplified frequency domain representation of the CDMA modulated signal shown in Figure 17.22.

the code domain, where the power in each Walsh Code is plotted in much the same way as power is plotted over frequency in a frequency domain, multiple access system. Thus, in a CDMA system, any given frequency channel can have multiple Walsh code channels sharing the same spectrum. Figure 17.22 shows a code domain representation of a CDMA signal with three channels (a pilot signal, plus two users). The corresponding frequency domain signal is shown in Figure 17.23.

Rho refers to the amount of power that can be correlated* to the correct channels when compared to the original, or ideal, transmitted signal. A ρ value of 1 means all the transmitted power is correlated to the correct active channels, and there is no power in the inactive (undesired) channels. Mathematically, this may be written as

$$\rho = \frac{\text{Power correctly correlated to active channels}}{\text{Total power}}. \tag{17.24}$$

Any energy correlated to the inactive channels degrades the modulation quality and results in a value of rho which is less than one. For the case of CDMA, the lower system limit for ρ is 0.912. ρ can be

*In a CDMA system, power that is received in a certain Walsh Coded channel is referred to as correlated power.

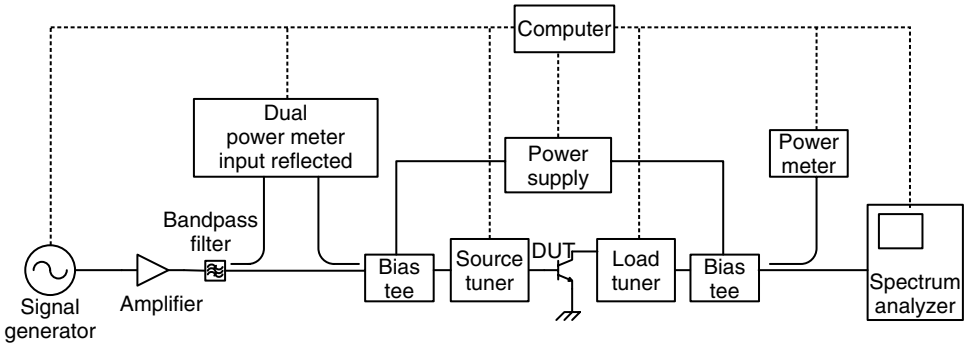


FIGURE 17.24 Block diagram of a typical loadpull measurement system.

approximately tied to EVM through the following equation [32]:

$$\rho \approx \frac{1}{1 + \text{EVM}^2}. \tag{17.25}$$

A more rigorous discussion of ρ may be found in [32].

17.5 Loadpull Characterization

A critical step in any power amplifier design is determining the proper output matching impedance. Loadpull measurements are commonly used to collect impedance matching data for power amplifier design. Figure 17.24 shows a typical block diagram for a loadpull measurement system. For handset power amplifier design, the RF source is usually able to handle the modulation formats the amplifier is designed for (such as CDMA, WCDMA, or EDGE). A linear amplifier often follows the source to boost the output power. Couplers are placed at the input to measure incident and reflected input power. Next comes a bias tee followed by the input tuner and the transistor to be tested. On the output of the transistor is another tuner, bias tee, coupler, and output power meter. A spectrum analyzer is used to measure the ACPR (described in Section 17.4.1) and harmonic content. A number of finite tuner positions are selected (typically 200–400 points) that cover the region of interest on the Smith Chart. At each tuner location, bias currents, output power, input power, ACPR, and harmonics are recorded and stored on a computer. This data is later used to plot contours of constant efficiency, gain, or ACPR.

Figure 17.25 shows some typical loadpull contours for an output stage using a heterojunction bipolar transistor. This particular example uses CDMA IS95 modulation with a transistor collector voltage of 3.2 V and bias current of 85 mA. The output impedance providing maximum PAE and linearity is shown in Figure 17.25. Usually, a compromise between efficiency and linearity is desired. So an "optimum" load (shown in Figure 17.25) is used for design purposes. One can see by the closeness of the ACPR contours that small deviations in the output matching impedance can cause large variations in amplifier linearity and efficiency. When the load impedance is in the vicinity of the "optimum" load, the tradeoff between linearity and efficiency is approximately one for one. In other words, a 1 point improvement in PAE results in a 1 dB reduction in ACPR, as can be observed by studying the contours in Figure 17.25.

Harmonic and baseband terminations can also have a significant impact on the overall performance of the output stage of a power amplifier. For example, most theoretical treatments of ideal class B and C transistor operation assume that all harmonics are shorted. Also, the most ideal classes E and F analyses assume that an infinite number of harmonics are available. In reality, only a finite number of harmonics are present due to the finite f_T of a transistor and these harmonics are frequently not all optimally terminated due to circuit or tuner constraints. Raab examines the impact of a finite number of harmonics for idealized classes E, C, and F amplifiers in [3,12,13].

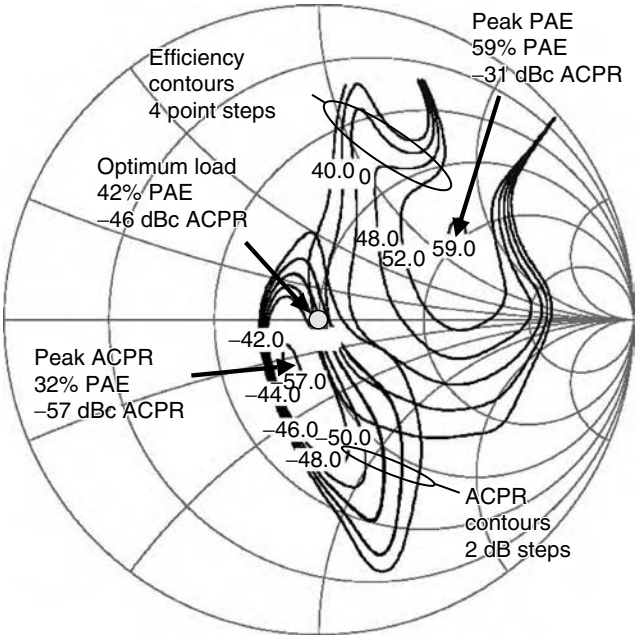


FIGURE 17.25 Loadpull contours of PAE and ACPR for a single stage HBT power cell under CDMA IS-95 modulation at 28 dBm output power with $V_{cc} = 3.2$ V and $I_{cq} = 85$ mA. Contours have been normalized to the optimum load impedance at the transistor. Efficiency contours are in 4 point steps (except where noted) and ACPR contours are in 2 dB steps (except where noted).

Several authors have examined the impact of harmonic and baseband terminations on linearity [33–36]. Asymmetry between upper and lower intermodulation distortion (IMD) and ACPR products can be attributed to low-frequency (baseband) terminations [33]. The low-frequency impedance of bias networks or bias tees can significantly impact IMD or ACPR. The second and third harmonic impedances can also affect power amplifier efficiency and linearity [34–36]. Optimization of harmonic terminations in addition to the fundamental frequency load and source impedance can improve amplifier performance at peak power levels.

17.6 Summary

This chapter presented a number of fundamental issues regarding power amplifiers. Each stage of a power amplifier can be described by one of several operating classes discussed in the first half of this chapter. Power amplifiers used in communication systems such as cellular telephones must operate under modulated RF conditions. Several figures of merit were discussed that describe the influence the power amplifier will have on a modulated signal. The concepts discussed in this chapter apply to a broad range of power amplifiers. They lay the foundation for Chapter 18.

References

1. Sokal, N. and Sokal, A., Class E—A new class of high-efficiency tuned single-ended switching power amplifiers, *IEEE Journal of Solid-State Circuits*, vol. SC-10, no. 3, 1975, p. 168.
2. Raab, F., et al., Power amplifiers and transmitters for RF and microwave, *IEEE Transactions on Microwave Theory and Techniques*, vol. 10, no. 3, 2002, p. 814.
3. Raab, F., Class-E, Class-C, and Class-F power amplifiers based upon a finite number of harmonics, *IEEE Transactions on Microwave Theory and Techniques*, vol. 49, no. 8, 2001, p. 1462.

4. Cripps, S., *RF Power Amplifiers for Wireless Communications*, Artech House, Boston, MA, 1999.
5. Raab, F., et al., RF and microwave power amplifier and transmitter technologies—part 1, *High Frequency Electronics*, vol. 2, no. 3, 2003, p. 22.
6. Raab, F., et al., RF and microwave power amplifier and transmitter technologies—part 2, *High Frequency Electronics*, vol. 2, no. 4, 2003, p. 22.
7. Raab, F., et al., RF and microwave power amplifier and transmitter technologies—part 3, *High Frequency Electronics*, vol. 2, no. 5, 2003, p. 38.
8. Raab, F., et al., RF and microwave power amplifier and transmitter technologies—part 4, *High Frequency Electronics*, vol. 2, no. 6, 2003, p. 34.
9. Raab, F., et al., RF and microwave power amplifier and transmitter technologies—part 5, *High Frequency Electronics*, vol. 3, no. 1, 2004, p. 46.
10. Raab, F, High efficiency amplification techniques, *IEEE Transactions on Circuits and Systems*, vol. 7, no. 10, 1975, p. 3.
11. Snider, D., A theoretical analysis and experimental confirmation of the optimally loaded and overdriven RF power amplifier, *IEEE Transactions on Electron Devices*, vol. ED-14, no. 12, 1967, p. 851.
12. Raab, F., Class-F power amplifiers with maximally flat waveforms, *IEEE Transactions on Microwave Theory and Techniques*, vol. 45, no. 11, 1997, p. 2007.
13. Raab, F., Maximum efficiency and output of class-F power amplifiers, *IEEE Transactions on Microwave Theory and Techniques*, vol. 49, no. 6, 2001, p. 1162.
14. Krauss, H., Bostian, C., and Raab, F., *Solid State Radio Engineering*, John Wiley & Sons, New York, 1980.
15. Kushner, L., Output performance of idealized microwave power amplifiers, *Microwave Journal*, vol. 32, no. 10, p. 103, October 1989.
16. Suetsugu, T. and Kazimierczuk, M., Analysis and design of class E amplifier with shunt capacitance composed of nonlinear and linear capacitances, *IEEE Transactions on Circuits and Systems—I: Regular Papers*, vol. 51, no. 7, 2004, p. 1261.
17. Suetsugu, T. and Kazimierczuk, M., Comparison of class-E amplifier with nonlinear and linear shunt capacitance, *IEEE Transactions on Circuits and Systems—I: Fundamental Theory and Applications*, vol. 50, no. 8, 2003, p. 1089.
18. Alinikula, P., Choi, K., and Long, S., Design of class E power amplifier with nonlinear parasitic output capacitance, *IEEE Transactions on Circuits and Systems—II: Analog and Digital Signal Processing*, vol. 46, no. 2, 1999, p. 114.
19. Raab, F. and Sokal, N., Transistor power losses in the class E tuned power amplifier, *IEEE Journal of Solid-State Circuits*, vol. SC-13, no. 6, 1978, p. 912.
20. Wong, G. and Long, S., An 800 MHz HBT class-E amplifier with 74% PAE at 3.0 volts for GMSK, *GaAs Integrated Circuit Symposium*, October 1999, p. 299.
21. Kazimierczuk, M. and Puczko, K., Exact analysis of class E tuned power amplifier at any Q and switch duty cycle, *IEEE Transactions on Circuits and Systems*, vol. CAS-34, no. 2, 1987, p. 149.
22. Kazimierczuk, M., Class E tuned power amplifier with nonsinusoidal output voltage, *IEEE Journal of Solid-State Circuits*, vol. SC-21, no. 4, 1986, p. 575.
23. Kessler, D. and Kazimierczuk, M., Power losses and efficiency of class E power amplifier at any duty ratio, *IEEE Transactions on Circuits and Systems—I: Regular Papers*, vol. 51, no. 9, 2004, p. 1675.
24. Class E waveform plots courtesy Jessel T. Xavier, graduate research assistant, University of Massachusetts, Amherst.
25. Wolf, J., Adjacent channel leakage power on 3GPP W-CDMA signals with the FSP, Application Note, Rohde & Schwarz, www.rohde-schwartz.com 1EF41_0E, Jan. 2001.
26. Agilent Technologies, Characterizing Digitally Modulated Signals with CCDF Curves, Application Note, January 2000, literature number 5968-6875E, www.agilent.com, Jan. 2000.
27. 3GPP memo R4-040367, Comparison of PAR and Cubic Metric for power derating, Motorola, http://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_31/Docs/
28. 3GPP memo R4-040721, Mapping of cubic metric to additional PA headroom, Qualcomm, http://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_33/Docs/
29. Agilent Technologies, Understanding GSM/EDGE transmitter and receiver measurements for base transceiver stations and their components, application note 1312, 5968-2320E, www.agilent.com, Aug. 24, 2002.

30. Newman, E., Optimizing receiver performance through error vector analysis, *Microwave Product Digest*, www.mpdigest.com October 2004.
31. Agilent Technologies, Understanding CDMA measurements for basestations and their components, application note 1311, 5968-0593E, www.agilent.com, June 2000.
32. 3GPP memo R4-99027, Submission of Composite Rho proposal to 3GPP RAN W4, Hewlett Packard, http://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_02/Docs/Tzips/
33. Borges de Carvalho, N. and Carlos Pedro, J., A comprehensive explanation of distortion sideband asymmetries, *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, no. 9, 2002, p. 2090.
34. Staudinger, J. and Norris, G., The effect of harmonic load terminations on RF power amplifier linearity for sinusoidal and $\pi/4$ DQPSK stimuli, *IEEE MTT-S Symposium on Technologies for Wireless Applications Digest*, Feb. 1997, p. 23.
35. Yang, Y., High pass output matching technique with enhanced third harmonic rejection for CDMA power amplifiers, *2005 IEEE MTT-S Digest*, June 2005, p. 2051.
36. Hartskeerl, D., Volokhine, I., and Spirito, M., On the optimum 2nd harmonic source and load impedances for the efficiency-linearity trade-off of RF LDMOS power amplifiers, *2005 IEEE Radio Frequency Integrated Circuits Symposium*, June 2005, p. 447.

18

Handset Power Amplifier Design

Douglas A. Teeter
Edward T. Spears
RF Micro Devices

18.1 Introduction

Mobile phones are the highest volume electronic consumer product in the world. The concept of cell based mobile radios first appeared at Bell Laboratories in the early 1970s. Mobile cellular systems were commercially available in the 1980s, but were limited due to coverage and mobile equipment.

First generation phones utilized simple frequency modulation (FM) and were limited to voice communication. However, today's mobile phone offers multiple methods of communication including voice, text messaging, web browsing, and streaming video. A typical mobile phone will cover multiple frequency bands and air standards including several mobile phone bands as well as possibly wireless local area network (WLAN) bands and Bluetooth bands, each band requiring its own radio.

Every radio, in turn, requires a power amplifier (PA) as part of the transmit chain. Figure 18.1 shows the basic block diagram for each of the radio transceivers within a handset. The advent of higher data rates combined with the addition of multiple bands within the handset has put increased demands on the PA in terms of linearity, efficiency, average current, and peak current consumption.

This chapter discusses the basic issues faced by a PA circuit designer. The emphasis will be placed on those attributes which affect the overall performance of the handset PA. Most of the concepts presented in this chapter are independent of air standard. A number of excellent papers and books have been written

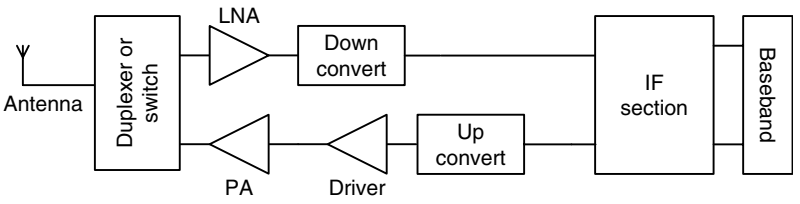


FIGURE 18.1 Basic handset radio architecture.

on this subject, so the reader will be referred to the literature for more detailed discussion of the topics addressed in this chapter.

18.2 Common Air Standards Used for Cellular Telephones

The vast majority of handset PAs in development today are for 2.5 and 3 G applications. Tables 18.1 and 18.2 summarize common frequency bands, output power levels, and information relevant to handset PA design for the global system mobile (GSM) and code division multiple access (CDMA) air standards [1–5]. Development and system improvement is on going, so readers should view the entries in the tables as an approximation that reflects the evolution of the air standards. For example, use of advanced modulation techniques, multiple carriers, dual antenna terminals, and mobility enhancements can enable further improvement of enhanced data-rates for GSM evolution (EDGE) thru-put beyond 384 kbit/s [2].

From the PA designer's perspective, output power, input drive, dynamic range, and peak to average ratio have the most influence on the circuit design. Section 18.4 illustrates how the 80 dB output power range of CDMA and WCDMA (Wideband Code Division Multiple Access) systems places significant challenges on the average current consumption of the PA. The importance of the peak to average ratio on PA design is discussed in the previous chapter and in Section 18.3.5.1. The peak to average ratio entries in Tables 18.1 and 18.2 (at 1% CCDF—complementary cumulative density function defined in Section 17.4.2) represent the peak to average for the worst case modulation format *with regard to the handset PA operation* (*reverse link*). In 3G systems, there are multiple radio configurations with differing modulation schemes and data rates. The radio configuration that requires the *highest peak output power* from the PA controls the load impedance and device area for the output stage. It is not necessarily the format with the highest peak to average ratio since many of the formats with high peak to average ratios are operated with some output power back-off. Thus, the entries in the table represent the modulation with worst case peak to average ratio that the PA is not operated at backed off power.

18.3 Power Amplifier Design

This section discusses the critical stages in designing handset PAs. A WCDMA PA is used to illustrate the issues regarding frequency domain duplex systems. A GSM/EDGE PA is used to demonstrate design factors for time domain access. Many of the concepts discussed apply universally to all types of PAs. The section begins by walking the reader through basic steps of any PA design. In the latter part of this section, specific issues faced in WCDMA and GSM/EDGE PA design are discussed.

18.3.1 Device Selection and Characterization

The initial phase of any PA design involves device selection. There are several technologies available for mobile PA design. What must be determined is which technology optimally fits the design requirements.

Depletion mode pseudomorphic high electron mobility transistor (pHEMT) devices typically have the highest power added efficiency (PAE) and maximum channel current, but require a negative gate voltage and a drain side switch. Enhancement mode pHEMT (E-pHEMT) devices exhibit similar high efficiency

TABLE 18.1 GSM/EDGE/WCDMA Evolution

System	GSM	GPRS	EDGE	WCDMA (R99)	HSDPA (R5)
Common frequency Bands	*US Cellular* UL: 824–849 MHz DL: 869–894 MHz	*US Cellular* UL: 824–849 MHz DL: 869–894 MHz	*US Cellular* UL: 824–849 MHz DL: 869–894 MHz	*US Cellular (Band V)* UL:824–849 MHz DL: 869–894 MHz	*US Cellular (Band V)* UL: 824–849 MHz DL: 869–894 MHz
	EGSM UL: 880–915 MHz DL: 925–960 MHz	*EGSM* UL: 880–915 MHz DL: 925–960 MHz	*EGSM* UL: 880–915 MHz DL: 925–960 MHz		
	DCS UL: 1710–1785 MHz DL: 1805–1880 MHz	*DCS* UL: 1710–1785 MHz DL: 1805–1880 MHz	*DCS* UL: 1710–1785 MHz DL: 1805–1880 MHz		
	PCS UL: 1850–1910 MHz DL: 1930–1990 MHz	*PCS* UL: 1850–1910 MHz DL: 1930–1990 MHz	*PCS* UL: 1850–1910 MHz DL: 1930–1990 MHz	*PCS (Band II)* UL: 1850–1910 MHz DL: 1930–1990 MHz	*PCS (Band II)* UL: 1850–1910 MHz DL: 1930–1990 MHz
				IMT2000 (Band I) UL: 1920–1980 MHz DL: 2110–2170 MHz	*IMT2000 (Band I)* UL: 1920–1980 MHz DL: 2110–2170 MHz
Multiple access	TDMA/FDMA	TDMA/FDMA	TDMA/FDMA	CDMA	CDMA
Peak-ave power ratio at 1% CCDF (rev)[a]	0	0	3.2	2.63	3.71
Output power dynamic range (dB)	30	30	30	80	80
Bandwidth (MHz)	0.2	0.2	0.2	3.84	3.84
Antenna power (dBm)	33 (Low band) 30 (High band)	33 (Low band) 30 (High band)	27 (Low band) 26 (High band)	24 (class 3)	24 (class 3)
Theoretical max. data rate (kbit/s)	14.4 (fwd & rev)	53.6 (fwd & rev)	384 (fwd & rev)	2048 (fwd) 2304 (rev)	14400 (fwd) 2304 (rev)

[a] Worst case for handset PA (reverse link) with no back-off allowance.

TABLE 18.2 CDMA Evolution

System	CDMA-one (IS-95A)	CDMA2000 (IS95C 1 × RTT)	CDMA 1 EV-DO (IS-856 Rev 0)	CDMA 1 EV-DO (IS-856 Rev A)
Common frequency Bands	US Cellular UL: 824–849 MHz DL: 869–894 MHz	US Cellular UL: 824–849 MHz DL: 869–894 MHz	US Cellular UL: 824–849 MHz DL: 869–894 MHz	US Cellular UL: 824–849 MHz DL: 869–894 MHz
	PCS UL: 1850–1910 MHz DL: 1930–1990 MHz	*PCS* UL: 1850–1910 MHz DL: 1930–1990 MHz	*PCS* UL: 1850–1910 MHz DL: 1930–1990 MHz	*PCS* UL: 1850–1910 MHz DL: 1930–1990 MHz
Multiple Access	CDMA/FDMA	CDMA/FDMA	CDMA/TDMA	CDMA/TDMA
Peak-ave power ratio at 1% CCDF (rev)[a]	3.9	4.5	4.5	4.5
Output power dynamic range (dB)	73	80	80	80
Bandwidth (MHz)	1.25	1.25	1.25	1.25
Antenna power (dBm)	24	24	24	24
Theoretical max. data rate (kbit/s)	14.4 (fwd & rev)	153 (fwd & rev)	2458 (fwd) 153 (rev)	3072 (fwd) 1800 (rev)

[a] Worst case for handset PA (reverse link) with no back-off allowance.

performance without the need for a negative gate voltage or a drain side switch [6]. The threshold voltage (V_t) of these devices ranges from 0 to +0.7 V. E-pHEMT devices also maintain higher efficiency than heterojunction bipolar transistor (HBT) devices at low supply voltages due to a lower knee voltage [6]. This allows reduction of the drain supply voltage below 2.5 V, which is the cut-off voltage for a lithium ion battery. In addition to PAs, pHEMT devices are capable of radio frequency (RF) power switches, offering the opportunity for increased integration of PA module functions using this technology.

GaAs HBT devices have excellent linearity performance and have the highest power density available, which leads to reduced die size. The biggest challenge with HBT power devices is thermal management. Proper device ballasting must be used to maintain thermal stability (see Section 18.3.3). Because the transistor is defined by the epitaxial layer doping and thickness, it has a high degree of manufacturability. This ease in manufacturing along with good RF performance has made the GaAs HBT a popular choice for handset PAs.

The technology that can provide the highest level of integration is Si CMOS. In CMOS technology, all of the power management, switching, complex biasing, and logic could be implemented in a single mainstream process.

Most of the design procedures listed in this chapter pertain to any device technology. Except where noted, one can interchange base/gate terminals and collector/drain terminals in the discussion that follows.

The air standard that the amplifier must operate in must be understood to achieve the required peak-to-average ratio. If the output stage device periphery is not large enough, the amplifier's load line will be forced low and gain compression will be introduced, negatively affecting the AM–AM, AM–PM, and linearity performance. Also, the designer must ensure that the device is going to operate within the safe current limits (I_{max}). When the device periphery is larger than required, efficiency degrades and quiescent current goes up, which increases the average current consumption.

From either measured or simulated results, the device I_{max}, V_{sat}, and g_m characteristics are determined. At peak output power, the final stage of most handset PAs is operating near class B. The case of the ideal class B amplifier with constant g_m was discussed in Section 17.2.1. A closer approximation for HBT and field effect transistor (FET) devices is to assume a linear g_m. For a bipolar transistor, the collector current is approximately

$$I_c = I_s \exp\left(\frac{qV_{be}}{nkT}\right) \tag{18.1}$$

where I_s is the saturation current and n is the base-emitter junction ideality factor. Expanding Equation 18.1 in a Taylor series about the DC bias point, V_{beo}, one obtains [7]

$$I_c = I_s \exp\left(\frac{q(V_{beo} + v_i)}{nkT}\right) = I_C^{DC} \exp\left(\frac{qv_i}{nkT}\right) \approx I_C^{DC}\left(1 + \frac{qv_i}{nkT} + \frac{1}{2}\left(\frac{qv_i}{nkT}\right)^2 \cdots\right). \tag{18.2}$$

Taking the derivative of Equation 18.2 with respect to the small signal voltage variation, v_i, yields the small signal transconductance

$$g_m = \frac{\partial I_c}{\partial v_i} \approx \frac{qI_C^{DC}}{nkT}\left(1 + \frac{qv_i}{nkT}\right) \tag{18.3}$$

A transconductance that varies linearly with input voltage, v_i, results if one only retains the first two terms.

For an idealized FET device, the I–V characteristics are often approximated as [7, p. 52]

$$I_{DS} \approx I_{DSS}\left(1 - \frac{V_{gs}}{V_P}\right)^2 \tag{18.4}$$

where I_{DSS} is the current when V_{gs} is zero and V_P is the pinch off voltage. The corresponding transconductance varies linearly with gate voltage and may be written as

$$g_m \approx g_{mo}\left(1 - \frac{V_{gs}}{V_P}\right) \tag{18.5}$$

with $g_{mo} = -2I_{DSS}/V_P$.

Reference 8 derives the following expressions for the optimum load, peak output power, and efficiency of a tuned class B amplifier with a linear transconductance for either Si or GaAs devices. For the case of a bipolar device such as an HBT, one obtains:

$$\text{Optimum load (HBT)} \qquad R_L = \frac{3\pi}{4}\frac{(V_{cc} - V_{ce}^{sat})}{I_{max}} \tag{18.6}$$

$$\text{Peak output power (HBT)} \qquad P_{out} = \frac{2(V_{CC} - V_{ce}^{sat})I_{max}}{3\pi} \tag{18.7}$$

$$\text{Peak efficiency} \qquad \eta = 85\alpha \tag{18.8}$$

$$\text{Knee voltage factor (HBT)} \qquad \alpha = \frac{(V_{CC} - V_{ce}^{sat})}{V_{CC}} \tag{18.9}$$

Similar expressions also exist for FET devices, as listed below [8]:

$$\text{Optimum load (FET)} \qquad R_L = \frac{3\pi}{4}\frac{(V_{DD} - V_{DS}^{sat})}{I_{max}} \tag{18.10}$$

$$\text{Peak output power (FET)} \qquad P_{out} = \frac{2(V_{DD} - V_{DS}^{sat})I_{max}}{3\pi} \tag{18.11}$$

$$\text{Knee voltage factor (FET)} \qquad \alpha = \frac{(V_{DD} - V_{DS}^{sat})}{V_{DD}} \tag{18.12}$$

In general, the output stage of a multi-stage design operates in class B to AB depending on whether the amplifier operates in a constant envelope or varying envelope.

With any PA design, the major concerns are achieving the desired power output at specified power added efficiency (PAE) while maintaining stability into a load VSWR. For example, given a specification for a high band GSM amplifier that can produce 32 dBm of power output with greater than 50% PAE and 32 dB of large signal gain, a designer must ensure that the output load line presented to the transistor provides the impedance which maximizes the voltage and current swing.

Device breakdown voltage must be greater than the maximum voltage swing across the device into a load VSWR. In order to achieve the large current swing present during output VSWR variation, the output device area must be optimally sized to meet the output power specifications. This can be determined through simulation or empirically through automated load pull.

As an example, a commercial harmonic balance simulator has been used to design the output stage of a GSM PA using an HBT device. A maximum safe peak current (I_{max}) of 3.6 Amps (corresponding to J_{max} of 50 kA/cm^2) is used in this example with a 3.4 V supply voltage and 0.6 V saturation voltage. A loadpull simulation of the output stage is performed to determine the optimum load and source impedance for the transistor. The bias network from Section 18.3.4 is included in the circuit, as shown in Figure 18.2. This verifies that the bias network can supply adequate base current under mismatch conditions and also includes the effects of the bias circuit impedance on the amplifier performance. The device layout and cell to cell spacing must be considered for reliable thermal operation. From Equations 18.6 to 18.9, the maximum output power should be 33.3 dBm with 70% PAE when presented with a load line of 1.83 Ω. Simulated load pull results show an output power of 33.5 dBm with 67.3% PAE and a load line of 3.1 Ω,

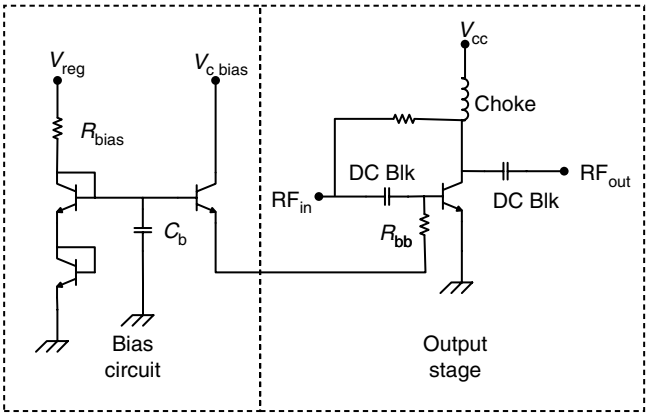


FIGURE 18.2 Output stage schematic used in the loadpull simulations.

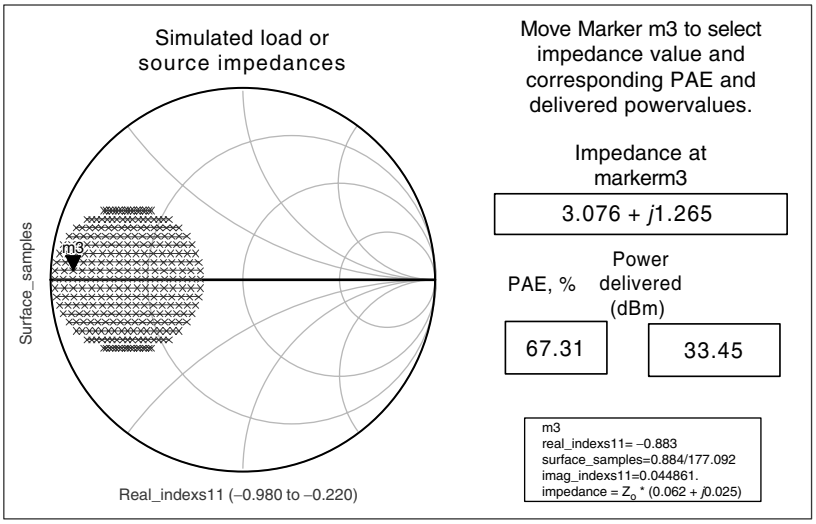


FIGURE 18.3 Simulated optimum load impedance and associated performance. Smith Chart is normalized to 50 ohms.

as shown in Figures 18.3 and 18.4. These results were obtained by performing both source and loadpull simulations to obtain the optimum source and load impedances, as shown in Table 18.3.

This procedure is continued for all proceeding stages until the large signal power gain is achieved. Once each stage's optimum source and load impedance is determined, the next step is designing the output, interstage, and input matching match.

18.3.2 Device Matching

Matching network design for high efficiency amplifiers differs from that of small signal design. For the small signal case, the optimum matching produces the maximum available gain. The output matching network of a PA is designed to present a load line to the device that will maximize the output voltage and current swing while at the same time transforming this load to the desired system impedance.

The output match is also designed to include the decoupling from the supply line and the proper harmonic ($2f_o$ and $3f_o$) and baseband terminations. Typically, a two-section low pass topology will achieve the desired bandwidth and harmonic attenuation, as shown in Figure 18.5. To help

TABLE 18.3 Power/Gain/Efficiency Budget

		Power Supply Data								
Amplifier Stage	Volts	Amps	DC (W)	P_{in} (dBm)	P_{out} (dBm)	Match Loss (dB)	Effective RF (W)	Coll Eff %	Stage PAE%	Summed Eff %
1	3.4	0.021	0.07	0.0	12.0	0.75	0.013	19.0	17.6	19.00
2	3.4	0.103	0.35	10.5	24.0	1.50	0.178	51.0	47.8	42.45
3	3.4	0.790	2.69	22.5	33.0	0.50	1.778	66.2	59.6	57.27

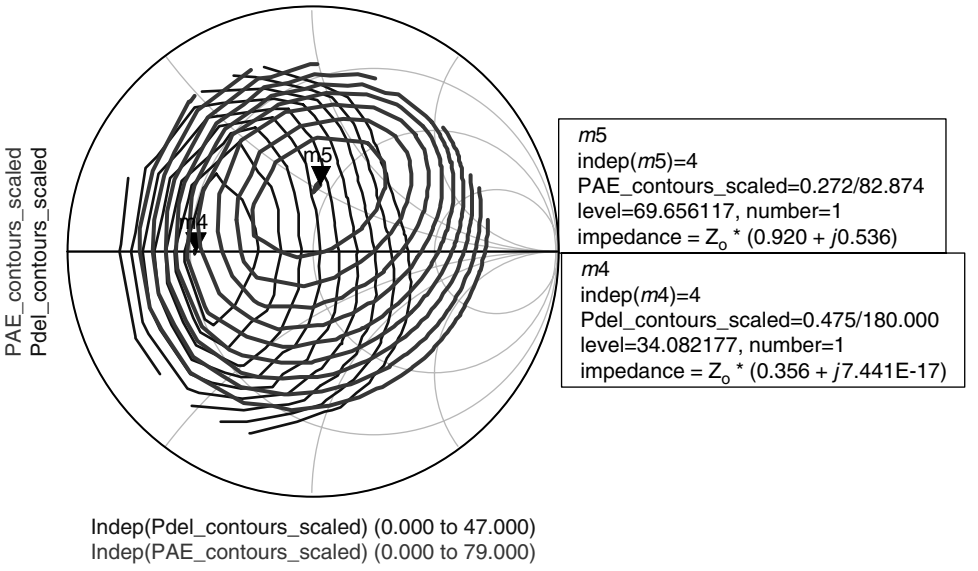


FIGURE 18.4 Simulated load pull results on an HBT output stage design. Smith chart normalized to 5 Ω.

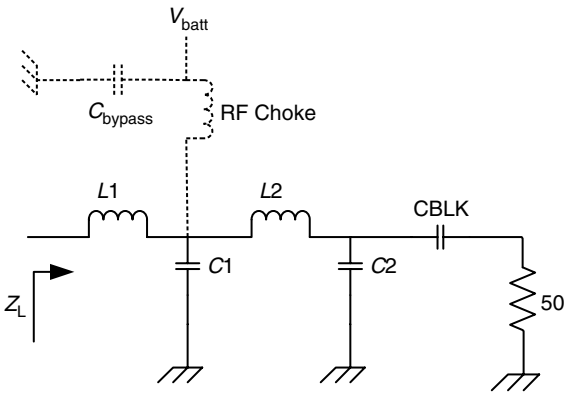


FIGURE 18.5 Simple two section output match commonly used for power amplifier design.

reduce the output stage's sensitivity to load variation, the first section is often transformed to the geometric mean,

$$\text{Geometric mean} = \sqrt{R_L Z_o} \tag{18.13}$$

where Z_o is the system impedance (usually 50 Ω) and R_L is the optimum output impedance for the final stage of the amplifier. The impedance transformation resulting from a two-section network is plotted on the Smith chart in Figure 18.6.

A critical factor in the overall amplifier efficiency is the output matching network loss (the loss between Z_L and 50 Ω in Figure 18.5). Loss in matching networks is determined by the Q of the components in the matching network, with the components carrying the most RF current being the largest contributors to overall loss. The Q of $C1$ and $L1$ (see Figure 18.5) are the most critical for the output match since this is a low impedance region of the matching network and RF currents are high. Use of high-Q components for $C1$ and $L1$ will have the greatest impact on reducing output loss.

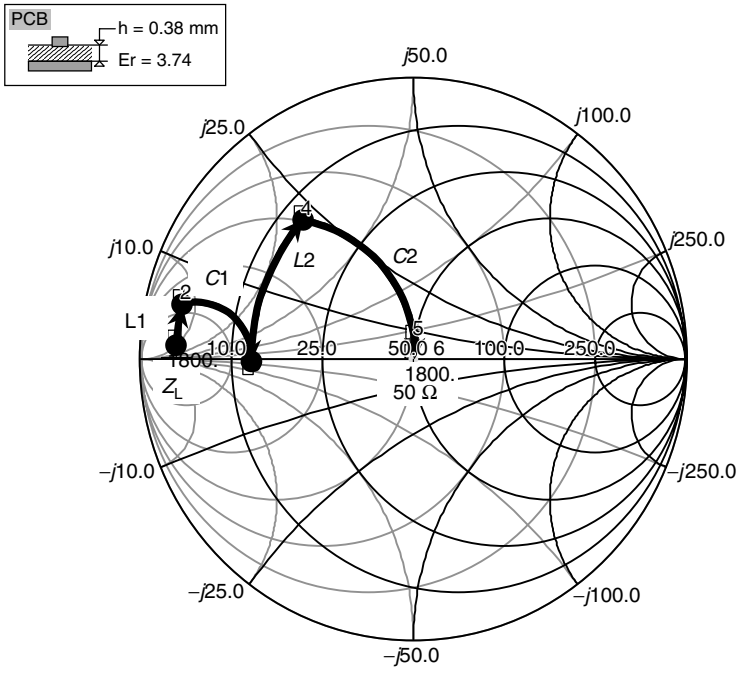


FIGURE 18.6 Output match impedance transformation for a commonly used two-section matching network.

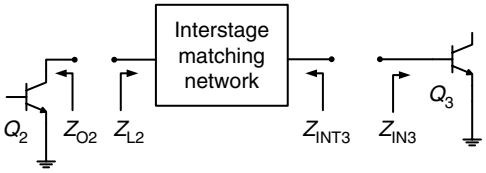


FIGURE 18.7 Interstage matching network design.

Output matching network insertion losses affect the efficiency of the amplifier. For an output stage operating at 50% collector (or drain) efficiency, efficiency drops approximately 1% for every 0.1 dB insertion loss. An output match with 1 dB insertion loss results in a 10 percentage point reduction in the overall PAE. These losses cause excessive device heating and reduce the mean time to failure (MTTF) rating of the device.

In multi-stage amplifier designs, the effect of the interstage matching on the amplifier's overall gain, efficiency, and linearity needs to be understood. For example, 2:1 VSWR source pull measurements on an output stage matched to 50 Ω have shown as much as ±3 dB variation in the adjacent channel power ratio (ACPR) for WCDMA modulation. The input impedance looking into the base of the final stage of an HBT PA is on the order of 1 Ω. FET based final stages also have similar low input impedance values. This impedance needs to be transformed to the desired impedance for the collector (or drain) of the preceding stage while minimizing the insertion loss. For example, implementing a single section high pass interstage match (shunt L, series C) using an inductor with a Q of 20 will give an insertion loss of about 1.8 dB. Figure 18.7 illustrates the key attributes of interstage matching networks.

Care must be taken in the design of matching networks between two complex source and load impedances. The interstage matching network must provide the optimum source impedance (Z_{INT3}) to the input of $Q3$, while presenting the optimum load impedance (Z_{L2}) to the output of $Q2$. It should be noted that these impedances are not necessarily the small signal conjugate match. Looking in the direction away from the output of $Q2$, the impedance transformation starts at Z_L^{opt} of $Q2$ (Z_{L2}) and goes to the complex conjugate of Z_{in}^{opt} for $Q3$ (Z_{int3}^*), as shown in Figure 18.7.

The total number of stages required for the amplifier will depend on the overall gain specification. Having too many stages will lead to poor noise performance (see Section 18.3.5.2), degrade the PAE performance, and reduce stability. Minimizing the matching network losses will help to reduce the total number of stages required.

18.3.3 Device Ballasting

Heterojunction bipolar transistors are commonly used in handset PA design. The output stage is formed with multiple HBT devices connected in parallel. Multi-cell HBT power devices require resistor ballasting to prevent thermal runaway due to the variation in V_{be} from cell to cell [9–15]. Figure 18.8 illustrates how the temperature of the center cells of a typical GaAs HBT output stage can be 40°C hotter than the edge cells. As shown in Figure 18.11, the "turn on" voltage of the hotter cells, in the center, will be lower than the cooler edge cells, resulting in a non-uniform distribution of current throughout the array. Current will eventually "crowd" to a single cell, resulting in current "collapse" in the power cell [9–15].

Emitter ballasting, shown in Figure 18.9a, reduces the effects of thermal runaway. As the device turn on voltage (V_{be}) decreases, the emitter current increases, causing a larger voltage drop across the ballasting resistor. This, in turn, starts to turn the device off, thereby "regulating" the amount of current that can flow through the device. Proper selection of the emitter ballast resistance will result in more uniform current

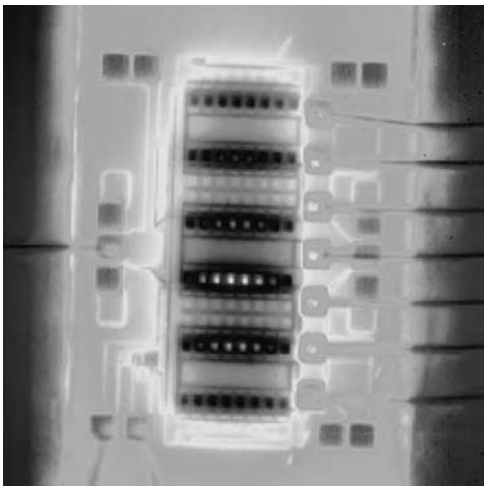


FIGURE 18.8 Thermal image from an HBT output stage with 5 Volts V_{cc} and 85°C base plate temperature. A maximum junction temperature of T_{peak} (at the center of the output stage) and a minimum of T_{peak} −40 are measured (at the edges of the output stage). Light colored areas in the thermal image photograph are hotter than darker colored areas.

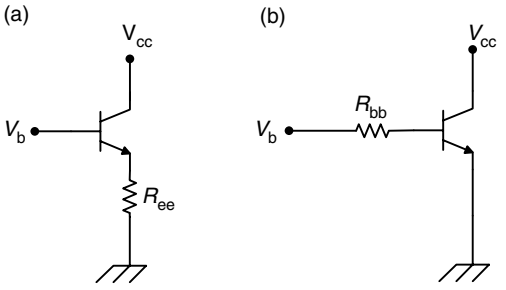


FIGURE 18.9 Implementation of (a) emitter and (b) base ballasting in HBT devices.

and temperature distribution across the array. A negative aspect of emitter ballasting is that it limits the available output voltage swing, which will lower the maximum output power, RF gain, and PAE [11].

Base ballasting, shown in Figure 18.9b, has the same thermal effect as emitter ballasting except the voltage to the base junction is reduced as the base current increases through the base ballast resistor. However, base ballasting does not degrade RF performance as much as emitter ballasting. For the case of GaAs HBT's, less equivalent base ballast resistance ($R_{bb} = R_{ee}\beta$) is required compared to emitter ballast resistance because the β decreases as temperature goes up [11,13]. The decreasing β with temperature causes base current to increase with temperature. Unfortunately, β increases with temperature for Si devices, so the same logic does not apply to Si BJTs. If the device ballasting is too large, the device will prematurely shut down as RF input drive increases. Ballasting resistor design should be verified with DC and harmonic balance simulations.

In [13, p. 440; 14, 15], Liu derives the critical collector current per emitter finger, $I_{critical}$, for the onset of current collapse in a 2-finger HBT device under constant total base current (or voltage) drive.* This critical current may be expressed as

$$I_{critical} = \frac{nkT}{q} \frac{1}{(\phi \cdot R_{th} \cdot V_{ce} - R_{ee})} \tag{18.14}$$

where n is the base-emitter junction ideality factor, T is the junction temperature, ϕ is the V_{be} change per degree celsius (Figure 18.11),† R_{th} is the effective thermal resistance,‡ and R_{ee} is the total emitter resistance (ballast and contact resistance). Using the collapse loci generated by Equation 18.14 and assuming a simple class A resistive load line swinging from BV_{ceo} to I_{max}, Lui derives the following expression for the optimal value of emitter ballast resistance [13, p. 466; 15]:

$$R_{ee}^{opt} = R_{th} \cdot \phi \cdot BV_{ceo}(1 - L) \tag{18.15}$$

where

$$L = 2\sqrt{\left(\frac{nkT}{q} \cdot \frac{1}{I_{max} \cdot \phi \cdot R_{th} \cdot BV_{ceo}}\right)} \tag{18.16}$$

is the thermal stability factor figure of merit. I_{max} in Equation 18.16 is the maximum safe operating current per emitter finger and BV_{ceo} is the collector–emitter breakdown voltage when the base is open. The optimal value of emitter ballast resistance occurs when the collapse loci from Equation 18.14 is tangential to the loadline [13, p. 466; 15]. When $L < 1$, emitter ballast resistance is required to prevent current collapse in the HBT device. When $L > 1$, no external emitter ballast resistance is needed for thermal stability. Reference 12 illustrates the impact varying the amount of emitter ballast has on the thermal stability of a multi-finger HBT device. The equivalent value of base ballast is approximately [15]

$$R_{bb} = \beta R_{ee} \tag{18.17}$$

where β is the current gain of the transistor. Liu's analysis may be extended to multi-device HBT power transistors by letting R_{th} and R_{ee} in Equations 18.14 and 18.15 represent the emitter resistance per transistor instead of the resistance per emitter finger.

---

* Lui examines both the case of constant base voltage and base current drive in [13, p. 440] and [14].

† We use $-\partial V_{be}/\partial T = \phi$, the same as used by Lui.

‡ R_{th} is the effective thermal resistance. For cases where thermal coupling between emitter fingers is significant, $R_{th} = R_{th11} - R_{th12}$ for constant base current drive and $R_{th} = R_{th11} + R_{th12}$ for constant base voltage drive. The constant base current case is more frequently encountered in PA design. R_{th11} is the self heating thermal resistance of each finger (the thermal resistance of each finger if perfectly isolated from other fingers) and R_{th12} is the coupling thermal resistance (thermal coupling between adjacent fingers) [13–15].

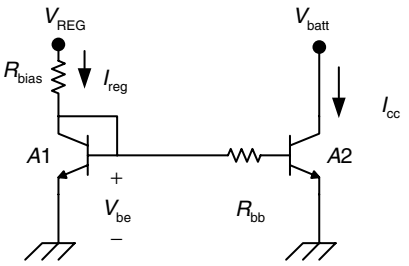


FIGURE 18.10 Basic current mirror bias circuit building block.

18.3.4 Bias Networks

One of the most critical yet often overlooked aspects in any PA design is the bias network. The bias network determines the amplifier performance over temperature as well as RF drive. Class B output stages biased with current sources behave differently as a function of output power compared to stages biased with voltage sources. Thus, the equivalent impedance of the bias network influences the overall PA performance. Many handset PAs utilize current mirrors on the PA die as a basic building block [7]. In some cases, additional bias control circuitry may be included in the module, either on die or as a separate control chip [16–18]. An in-depth discussion of current mirrors can be found in any introductory analog circuit design text (see, e.g., [7] Chapter 4). It is assumed that the reader is already familiar with the basic principles behind current mirrors. This section will focus in the essential elements of the current mirror that affect overall PA performance.

Thermal performance of any bias network is determined by the circuit topology, device technology, and material system. A simple HBT based current mirror will be used as an example since this is a common building block for many handset PAs. Figure 18.10 shows a schematic of the DC circuit for a simple current mirror. Neglecting the voltage drop across the base ballast resistor, R_{bb}

$$I_{cc} \approx \frac{A2}{A1} I_{reg} \tag{18.18}$$

where $A1$ is the area of the reference device and $A2$ is the area of the power transistor to be biased.

The variation in I_{cc} over temperature then depends on the variation in I_{reg} over temperature, which in turn is controlled by the voltage drop across resistor R_{bias}.

$$I_{reg} = \frac{V_{reg} - V_{be}}{R_{bias}} \tag{18.19}$$

The temperature dependence of V_{be} dominates the thermal behavior of I_{reg}.

For an npn HBT device, the collector current density (J_C) may be approximated as [19]

$$J_C \approx J_s e^{\frac{V_{be}}{nkT}} = q\left(\frac{D_n n_{i,p}^2}{w_B N_A}\right) e^{\frac{V_{be}}{nkT}} \tag{18.20}$$

where [20]

$$n_{i,p}^2 \propto T^3 e^{-E_g/kT} \tag{18.21}$$

E_g in Equation 18.21 is the bandgap of the p doped base material, w_B is the base width, D_n is the electron diffusion constant, and k is Boltzmann's constant in eV. Plugging Equation 18.21 into Equation 18.20 and

making use of the Einstein relationship ($D_n = kT\mu_n$), Equation 18.20 may be approximated as

$$J_C \propto T^4 e^{\frac{V_{be}-E_g}{kT}} \approx e^{\frac{V_{be}-E_g}{kT}} \tag{18.22}$$

if an ideality factor of 1 ($n = 1$) is assumed. For the case of a constant collector current density (J_C) over temperature, $\left(V_{be} - E_g/kT\right)$ must also be constant over temperature. Thus,

$$\left(\frac{V_{be} - E_g}{T}\right) = -\phi \approx \text{const} \tag{18.23}$$

Solving for V_{be} gives

$$V_{be} = -\phi \cdot T + E_g \tag{18.24}$$

The value of ϕ in Equation 18.23 is the junction voltage coefficient over temperature. Figure 18.11 plots V_{be} for an example AlGaAs/GaAs HBT with current density of 0.7 kA/cm^2. The value ϕ is about 1.415 mV/C for a current density of 0.7 kA/cm^2. Thermal variation of V_{be} for AlGaAs/GaAs HBT devices depends current density as well as somewhat on device layout, but typically ranges from -1.4 to -1.7 mV/C [13, p. 484; 14,15]. For a $2 \times 50\,\mu m^2$ AlGaAs/GaAs HBT device, Lui experimentally found [14]

$$\phi = 0.00079054 - 8.5937 \times 10^{-5} \ln(I_c) \tag{18.25}$$

The fact that V_{be} increases as temperature decreases for GaAs pn junctions is a critical factor in bias circuit design for handset PAs that use HBT devices. Referring back to Equation 18.19, if V_{be} increases with decreasing temperature, I_{reg} goes down and I_{cc} goes down. The temperature variation becomes more pronounced as V_{reg} approaches V_{be}. Typical handset components are specified for operation between -30°C and 100°C, a wide enough temperature range that the characteristics of the bias circuit over temperature affects the amplifier performance over temperature.

While the simple current mirror shown in Figure 18.10 can be found on some handset PAs, the base current drive into the power stage (transistor *A*2) is limited by resistor R_{bias}. This can cause early saturation of the PA. To circumvent this problem, an emitter follower can be added to the bias circuit, an example

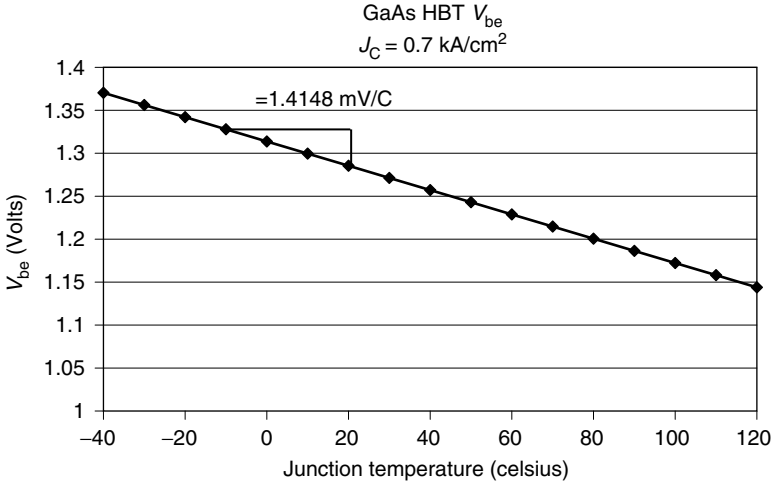


FIGURE 18.11 AlGaAs/GaAs HBT base-emitter junction voltage versus temperature for $J_c = 0.7$ kA/cm^2.

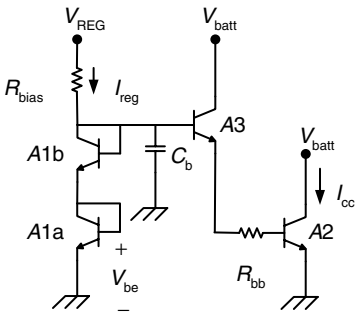


FIGURE 18.12 Example bias network used in WCDMA power amplifiers similar to Reference 21.

of which is shown in Figure 18.12 [21]. The addition of the emitter follower requires an extra V_{be} drop, making the design more sensitive to variation over temperature. For the bias circuit shown in Figure 18.12,

$$I_{reg} = \frac{V_{reg} - 2V_{be}}{R_{bias}} \tag{18.26}$$

and

$$I_{cc} \approx \frac{A2}{A1} I_{reg} \tag{18.27}$$

with $A1$ being the area of the diodes ($A1 = A1a = A1b$).

For example, in many WCDMA applications, the PA must operate with V_{reg} down to 2.75 V at −30°C. Referring to Figures 18.11 and 18.12, one can see that 2 V_{be} drops at −30°C is about 2.7 V, not much less than 2.75 V. This results in a small voltage drop across R_{bias}, resulting in low I_{reg} and correspondingly low I_{cq}, a condition which often degrades the linearity of the PA due to clipping.

In addition to thermal performance, bias networks also control the power saturation characteristics of the PA. As an example, the authors in [21] illustrate how varying the capacitance, C_b, in Figure 18.12 can alter the gain versus output power characteristics of a two stage PA. This, in turn, affects the power saturation and linearity of the amplifier.

This section presented a brief overview of some of the main issues designers face with bipolar bias networks. Similar trends are also observed with bias circuits using enhancement mode pHEMT devices. The circuits shown in Figures 18.10 and 18.12 could be implemented using an E-mode pHEMT. In this case, the base ballast resistor, R_{bb}, would not be required. The gate voltage (V_{gs}) under constant drain current (I_{DSS}) conditions decreases with increasing temperature at a rate of about −1 mV/C [22], which is about 35% lower than the −1.4 to −1.7 mV/C observed for HBT devices. The associated V_{gs} is typically in the vicinity of +0.4 V. More detailed discussion of bias circuits can be found in References 23 to 28.

18.3.5 CDMA/WCDMA PA Example

A generic class AB two stage amplifier designed for CDMA and WCDMA applications will be used to illustrate some of the basic issues in designing a handset PA. Class AB power amplifies are frequently used for CDMA/WCDMA applications due to the combination of relative ease of implementation and associated linearity.

18.3.5.1 Data Rate Implications on Amplifier Efficiency

As discussed in the previous chapter, when the bit rate increases for a given bandwidth, the peak to average ratio of the signal also increases. This requires more backed off operation of the PA to avoid distortion of the signal. Figure 18.13 shows an idealized power transfer characteristic of a PA designed with an ideal

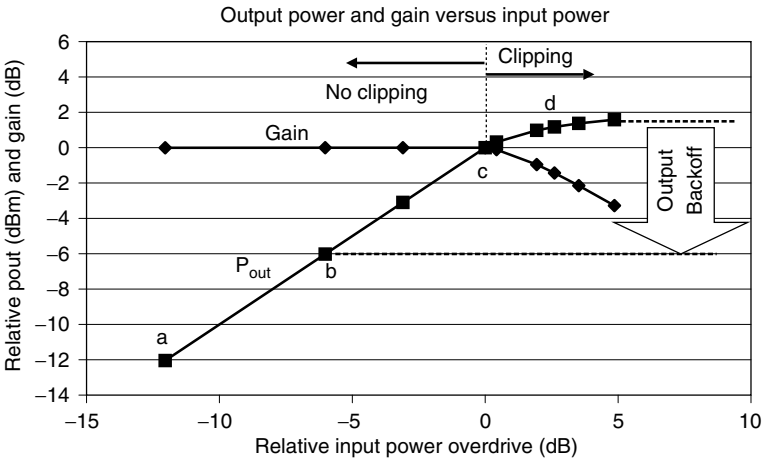


FIGURE 18.13 Relative output power and gain sweep for an idealized transistor (constant g_m) operating in Class B mode with constant load resistance. Points a, b, c, and d refer to the output current and voltage waveforms in Figures 18.14 and 18.15.

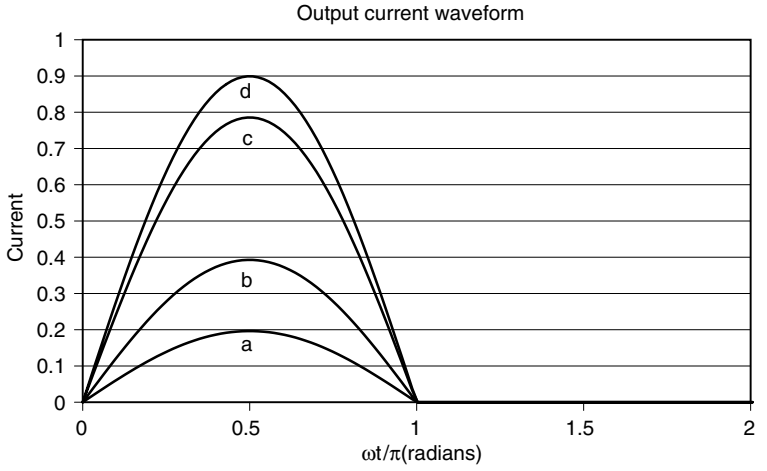


FIGURE 18.14 Corresponding normalized output current waveform for Figure 18.13. An ideal transistor is assumed with a maximum output current, I_{max}, of 1 Amp.

transistor. Figures 18.14 and 18.15 show the associated current and voltage waveforms. Once clipping of the output waveform begins, distortion occurs and gain begins to compress. The PA must be operated at significant output backoff to avoid distorting the signal.

The same concepts can be applied to a commercial WCDMA PA. Migration from WCDMA (Release 99) to HSDPA (Release 5) and HSDPA/HSUPA (Release 6)* results in an increased bit rate out of the handset for the same bandwidth. This section illustrates the impact transition to higher bit rates has on handset PA design. More detailed comparison of WCDMA and HSDPA can be found in Chapter 17, *Power Amplifier Fundamentals.*

Figure 18.16 shows the gain as a function of output power for a typical 2 stage GaAs HBT WCDMA PA for several modulation formats. The corresponding complementary cumulative density function (CCDF) is shown in Figure 18.17 and the associated linearity is depicted in Figure 18.18. As the peak to average ratio increases, the onset of gain compression occurs earlier. At an ACPR of −35 dBc,

*HSDPA = high speed downlink packet access (basestation to handset), HSUPA = high speed uplink packet access (handset to basestation).

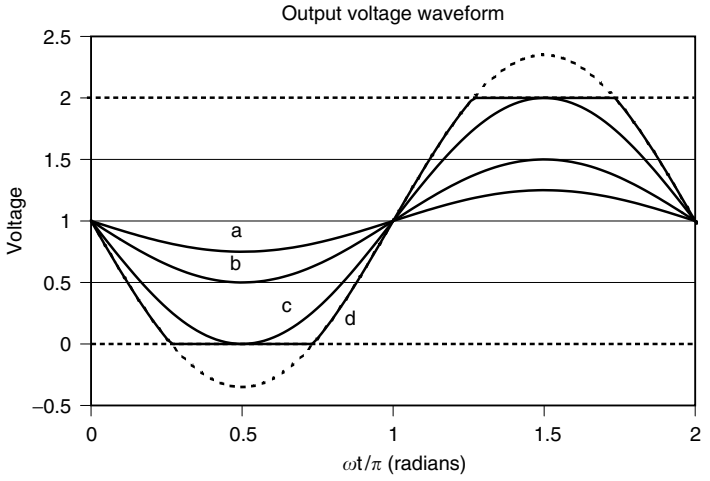


FIGURE 18.15 Output voltage waveform corresponding to Figure 18.13 showing clipping with overdrive. An idealized transistor is assumed with a minimum V_{ce} (or V_{ds}) of 0 V and a maximum voltage of 2 V. The output voltage swing is limited to be between 0 and 2 V.

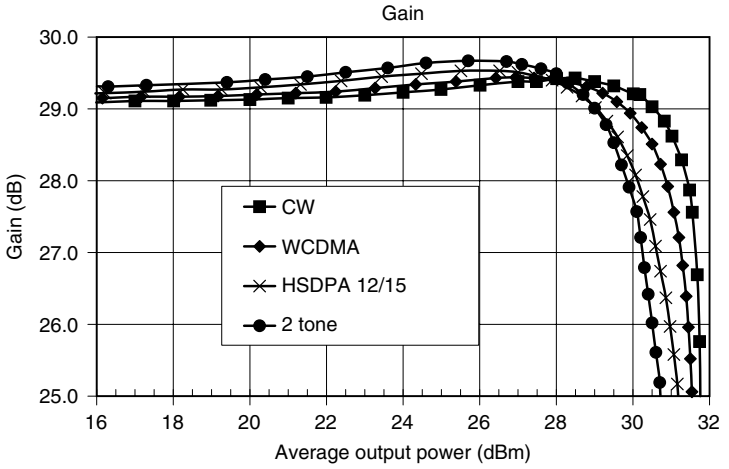


FIGURE 18.16 Gain versus output power for a typical 2 stage HBT linear power amplifier under different signal modulation.

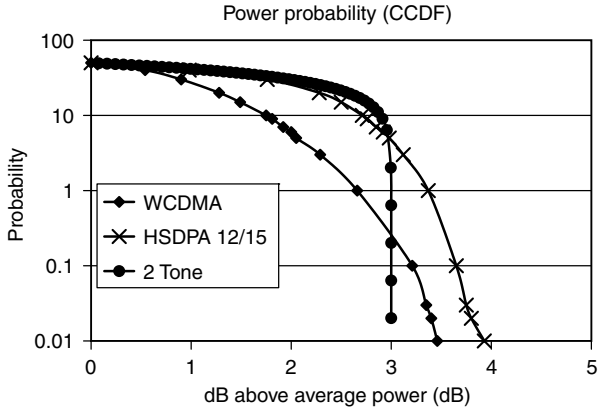


FIGURE 18.17 Corresponding CCDF characteristics for Figures 18.16, 18.18, and 18.19.

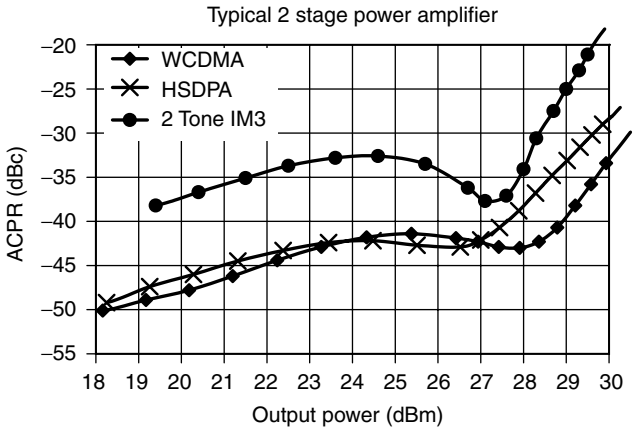


FIGURE 18.18 Measured linearity for a commercial 2 stage GaAs HBT WCDMA power amplifier.

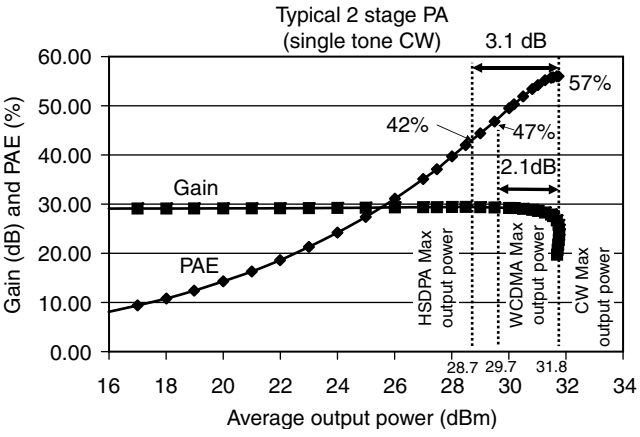


FIGURE 18.19 Output power back-off requirements for WCDMA and HSDPA modulation with ACPR $= -35$ dBc. The corresponding CCDF is shown in Figure 18.17.

the maximum output power under HSDPA conditions is about 1 dB lower than under WCDMA, which correlates with the 1 dB difference in the CCDF curves at the 10% level. This has severe implications on amplifier efficiency because for a given linearity (ACPR), an amplifier must operate at a higher output power back-off for HSDPA modulation (3.1 dB) compared to WCDMA modulation (2.1 dB), as illustrated in Figure 18.19. Thus, for a given output power and associated linearity, an amplifier tuned for HSDPA modulation will consume more peak current than a WCDMA PA because it must be tuned for higher peak saturated output power.

One interesting observation in Figure 18.18 is the presence of a dip in the ACPR at about 27 dBm. This drop is caused by cancellation, or self linearization, between the first and second stages of the amplifier. By optimally tailoring the gain versus power of each stage, one can self-linearize the PA. References 21 and 23 further discuss methods for analyzing and designing self linearization into handset PAs by keeping AM–AM (insertion gain) and AM–PM (insertion phase) distortion flat over as wide a range of output power as possible.

18.3.5.2 Noise Power Considerations

Because CDMA and WCDMA systems use frequency division duplex (FDD), transmit and receive paths are active at the same time. The amount of noise power at the output of the PA in the receive band

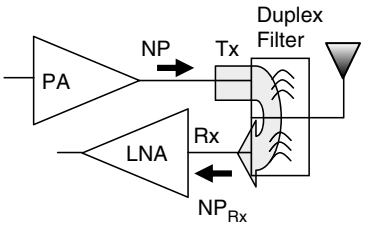


FIGURE 18.20 Noise power consideration for CDMA/WCDMA power amplifier design.

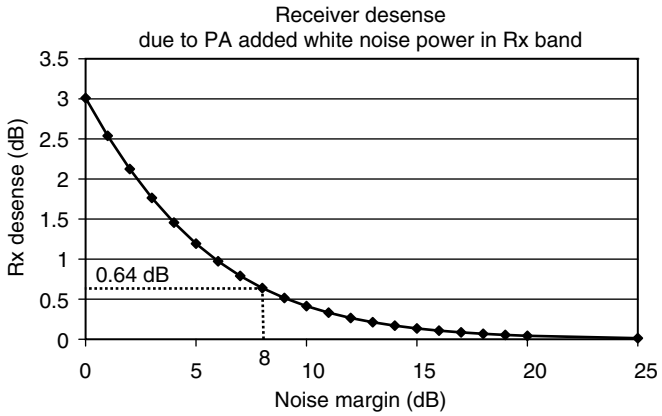


FIGURE 18.21 Receiver desense due to noise power contribution from the PA.

frequency, typically 45–90 MHz separated from the transmit band, must be kept low in order to avoid desensitizing the receiver.

A CDMA PA can be used as an example. Due to imperfect duplexer Tx–Rx isolation, some noise power at the transmit port (Tx) leaks through to the receive port (Rx), as illustrated in Figure 18.20. This power leakage looks like white noise to the low noise amplifier (LNA), raising the noise floor and desensitizing the receiver. Thus, the effective noise at the input to the receiver is the sum of the thermal noise at the antenna due to the receive path* ($kT \cdot \mathrm{NF_{rcvr}}$) and PA noise power in the receive band ($\mathrm{NP_{Rx}}$). The level of receiver desense can be expressed mathematically as

$$\mathrm{Desense_{dB}} = 10 \cdot \log\left(\frac{NP_{\mathrm{Rx}} + kT \cdot NF_{\mathrm{rcvr}}}{kT \cdot NF_{\mathrm{rcvr}}}\right) = 10 \cdot \log\left(1 + \frac{1}{\mathrm{Noise\ margin_{lin}}}\right) \tag{18.28}$$

where NP_{Rx} is the amount of PA noise power in a 1 Hz bandwidth which leaks through the duplexer into the LNA and NF_{rcvr} is the cascaded noise figure of the receiver at the antenna. Equation 18.28 is shown graphically in Figure 18.21. Noise margin represents how much the PA noise power at the receive port is below the thermal noise of the receiver.

$$\text{Noise margin (dB)} = 10\log(kT \cdot NF_{\mathrm{rcvr}}) - 10\log(NP_{\mathrm{Rx}}) \tag{18.29}$$

A common rule of thumb in receiver design is that the noise margin should be 8 dB or better so that the PA noise power does not degrade the receiver sensitivity by more than about 0.6 dB. One can clearly see in Figure 18.21 that as the noise margin falls below 8 dB, the noise power from the PA

*For the calculation of thermal noise in this section, use k = Boltzmann's constant = $1.381 \cdot 10^{-20}$ mW/Hz/K and T = 290 K.

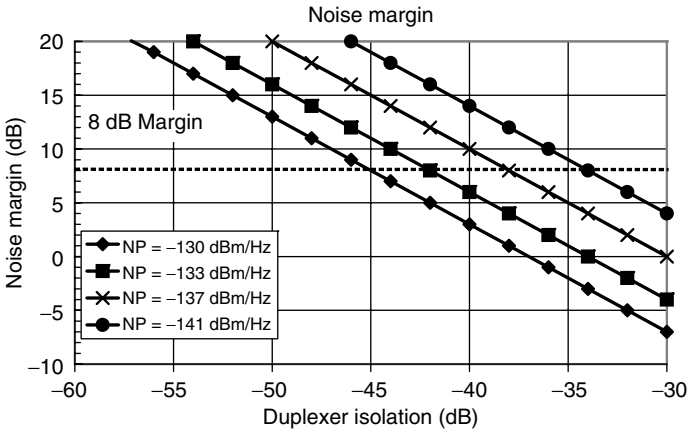


FIGURE 18.22 Noise margin as a function of duplexer isolation for different power amplifier noise power levels.

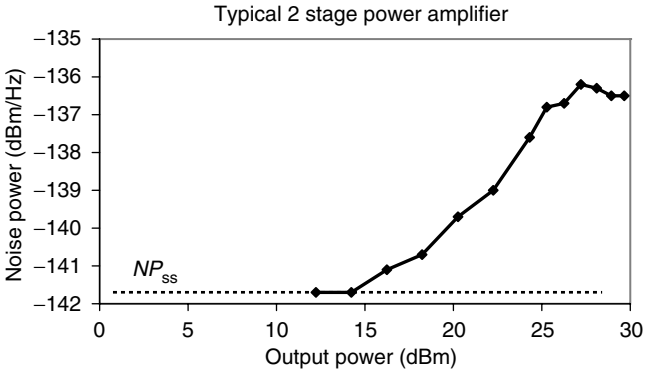


FIGURE 18.23 PA output noise power as a function of output power for a typical 2 stage 28 dBm CDMA power amplifier.

begins to significantly degrade the sensitivity of the receiver. Using the 8 dB rule of thumb, one can calculate the minimum allowable noise power out of the PA for various levels of duplexer isolation. Assuming the receiver has 7 dB cascaded noise figure, the thermal noise at the input to the LNA is $10\log(kT \cdot NF_{\mathrm{Rx}}) = -174 + 7 = 167$ dBm/Hz. If the isolation of the duplexer is 40 and 8 dB minimum noise margin is desired, the minimum acceptable noise power from the PA is $-167 - 8 + 40 = -135$ dBm/Hz. Figure 18.22 shows noise margin as a function of duplexer loss for PAs with different noise power levels.

As form factors of duplexers shrink to allow integration of the duplexer into the PA module, the isolation and return loss of the duplexer generally degrades. Thus, radio designers are becoming increasingly sensitive to PA noise power performance. PA noise power, to first order, is controlled by the gain and noise figure in the Rx band of the PA during large signal operation. At low power levels, one may easily compute the small signal PA noise power (NP_{ss}) based on the small signal noise figure ($NF_{\mathrm{ss}}^{\mathrm{Rx}}$) and small signal gain ($G_{\mathrm{ss}}^{\mathrm{Rx}}$) of the PA in the receive band

$$NP_{\mathrm{ss}} = 10 \cdot \log(kT \cdot G_{\mathrm{ss}}^{\mathrm{Rx}} \cdot NF_{\mathrm{ss}}^{\mathrm{Rx}}) \quad (\mathrm{dBm/Hz}). \tag{18.30}$$

As the output power of the amplifier increases, the noise power increases from the small signal value in Equation 18.30. Noise power of a typical 2 stage HBT CDMA PA is shown in Figure 18.23. A major contributor to the PA noise power in bipolar PAs is the upconversion of low frequency base shot noise, which is proportional to the base current. As the PA is driven harder, the base current increases though

rectification (self bias). This in turn increases the base shot noise and hence the PA noise power. Additional discussion of basic noise properties of HBT devices may be found in Reference 29.

GSM/EDGE PAs also have limits on the amount of noise power allowed. A detailed discussion of the relationship between noise power and receiver sensitivity for WCDMA as well as GSM radios may be found in Reference 30. The authors start from the thermal noise floor and illustrate how receiver noise figure, S/N margin, and spreading gain determine the sensitivity of receiver for a given bit error rate.

18.3.6 GSM/EDGE PA Design

Many of the concepts used for WCDMA PA design are also applicable for GSM/EDGE PAs. However, because the GSM/EDGE system uses time domain multiple access (TDMA), the PA must be switched on and off. This switching causes additional design challenges.

In TDMA systems, the PA must go from standby state to full output power in a very short period of time. GSM handsets require output power to ramp up from −5 to 35 dBm (full rated PA output power) in less than 10 μs and also ramp down in the same time period. This causes the PA to produce transmitted power outside the operation bandwidth of the system, commonly referred to as output RF spectrum (ORFS). By examining the Fourier transform of a trapezoidal pulse, one can easily understand the source of this spectral content. The ramp up and ramp down control voltage must have a smooth ramp function to reduce the AM–AM and AM–PM of the PA, thereby minimizing ORFS. In addition, the impedance seen looking into the PA will cause VCO frequency pulling and wideband noise, which also contribute to ORFS. Table 18.4 shows an example switching ORFS requirement for a GSM system.

Adding modulation to a pulsed system, such as $3\pi/8$ shifted 8-PSK used in EDGE, contributes further spectral content. Nonlinearity in the PA will cause additional spectral re-growth on this modulated signal just as it does for WCDMA signals. Excessive spectral re-growth will cause interference to other users operating in different frequencies bands. It is also indicative of waveform distortion which can increase the bit error rate. Wideband noise will also contribute to the spectral re-growth along with the phase noise of the VCO. Modulation ORFS can also be thought of as adjacent channel power leakage. Table 18.5 shows an example modulation ORFS specification for EDGE systems. The 400 kHz requirement is usually the most difficult spec to meet. In addition, error vector magnitude (discussed in the previous chapter) specifications are often provided for EDGE PAs.

Power amplifiers designed for GSM/EDGE applications must pass two ORFS specifications; switching ORFS shown in Table 18.4 and modulation ORFS shown in Table 18.5. A 5–6 dB back off in output power is permitted in the EDGE mode to allow the same PA to operate in both saturated (GSM) and linear (EDGE) modes. More detailed discussion of the design specifications for GSM/EDGE applications may be found in Reference 32.

Figure 18.24 shows a fully integrated quad band GSM/EDGE module with Si control chip. The module consists of a quad-band GaAs PA die and CMOS bias control die. An organic laminate substrate is used for routing and inclusion of surface mount components. The output matching is realized using high-Q surface mount capacitors. Lengths of transmission line printed on the substrate form the inductors. Bypass

TABLE 18.4 Example Switching ORFS Specification for a GSM System with PA Output Power 37 dBm or Less

Offset (kHz)	Limit (dBm)
400	−23
600	−26
1200	−32
1800	−36

Measurements use a 30 kHz resolution band width, a 100 kHz video bandwidth, and a zero frequency span. Detailed information may be found in References 31 and 32.

TABLE 18.5 Example Modulation ORFS (EDGE) Specification for the GSM850 and GSM900 Frequency Bands

Offset (kHz)	Limit (dBc)	Resolution Bandwidth (kHz)
0 (reference)	-	30
100	0.5	30
200	−30	30
250	−33	30
400	−54	30
$600 \leq f < 1800$	−60	30
$1800 \leq f < 3000$	−63	100
$3000 \leq f < 6000$	−65	100
$6000 \leq f$	−71	100

Power is measured at the specified offset with the specified resolution bandwidth relative to the power in a 30 kHz channel centered about the carrier [31,32].

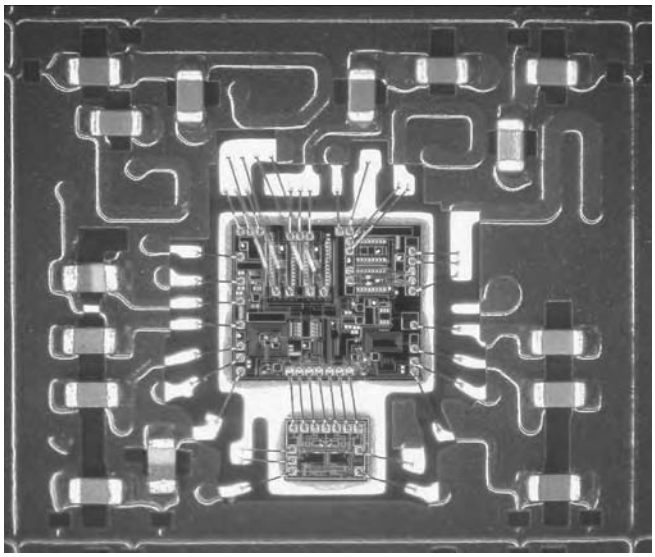


FIGURE 18.24 Example multi-chip GSM PA module.

capacitors are also included inside the module using surface mount components. Today's GSM PA designs have evolved into multi-chip modules that have silicon power controllers, antenna switches, and harmonic filters on the module, as shown in Figure 18.25.

18.3.7 Amplifier Stability

Stability is a common yet rarely discussed design issue for PAs. Small signal stability is relatively straight forward to address using K-factor analysis combined with normalized determinant function (NDF) techniques [33–35]. One point worth mentioning is that simple K-factor analysis alone, frequently discussed in many introductory texts, is necessary but not sufficient to guarantee stability [35]. Concerned readers are referred to References 33 to 35 for discussion of rigorous methods for determining small signal stability.

An additional concern in PAs is that fact that under drive and VSWR, especially at low temperatures (−30°C), there is an increased probability the amplifier will become unstable. At low power drive, the amplifier may appear stable, but as the amplifier is driven harder, there may be certain bias points and VSWR conditions that cause the amplifier to break out into an oscillation. Some detailed examples

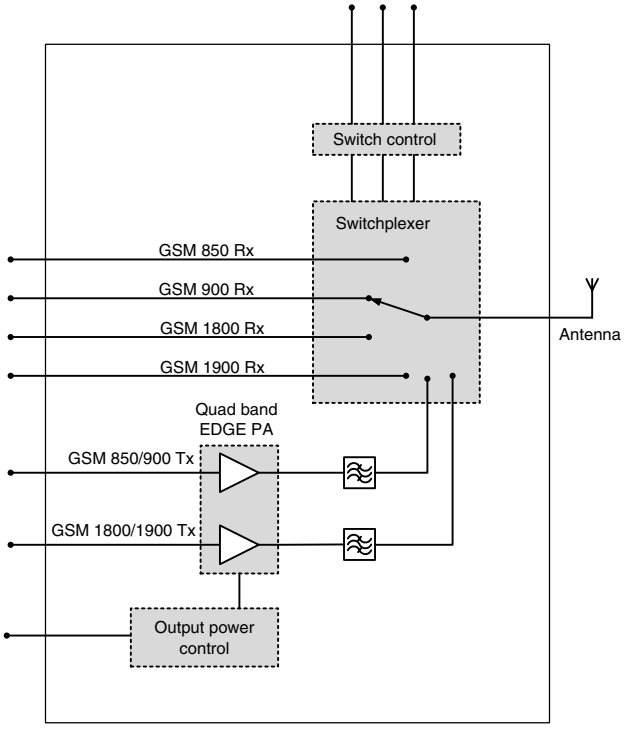


FIGURE 18.25 Example quad-band EDGE transmit module block diagram.

may be found in References 36 and 37. Drive dependant oscillations are referred to as parametric oscillations.

Parametric oscillations can be categorized into two basic types, subharmonic and spurious. Subharmonic oscillations are fractionally related to the carrier frequency by an integer fraction (n/m where n and m are integers). The most common subharmonic oscillations are at $f/2$ and $3f/2$ frequencies. Spurious oscillations are not fractionally related to the carrier. They can often be low frequency oscillations that are mixed by the carrier, such as 100 MHz tones on each side of the carrier. It is even possible to have both spurious and subharmonic oscillations at the same time. Figures 18.26 and 18.27 give examples of subharmonic and spurious oscillations.

Several authors have studied large signal oscillations over the last century [38–48]. Van Der Pol investigated instabilities arising from a nonlinear resistance in 1927 [46]. The subject of subharmonic oscillations due to the nonlinear I–V characteristic found in tubes was studied by Mandelstam and Papalaxi [47, p. 469].

Many authors have demonstrated how nonlinear capacitance, such as found in a pumped varactor diode, can result in subharmonic ($f/2, 3f/2, \ldots$) oscillations [38,41–45,47,48]. The simple series RLC circuit shown in Figure 18.28 may be used to explain the existence of subharmonic oscillations by assuming a sinusoidal time varying capacitance

$$C(t) = C_o - \Delta C \sin(\omega_p t) = C_o \left(1 - \frac{\Delta C}{C_o} \sin(\omega_p t)\right) \tag{18.31}$$

with a pump frequency ω_p.

Applying Kirchoff's voltage law to the simple circuit in Figure 18.28

$$V_R + V_L + V_C = 0 \tag{18.32}$$

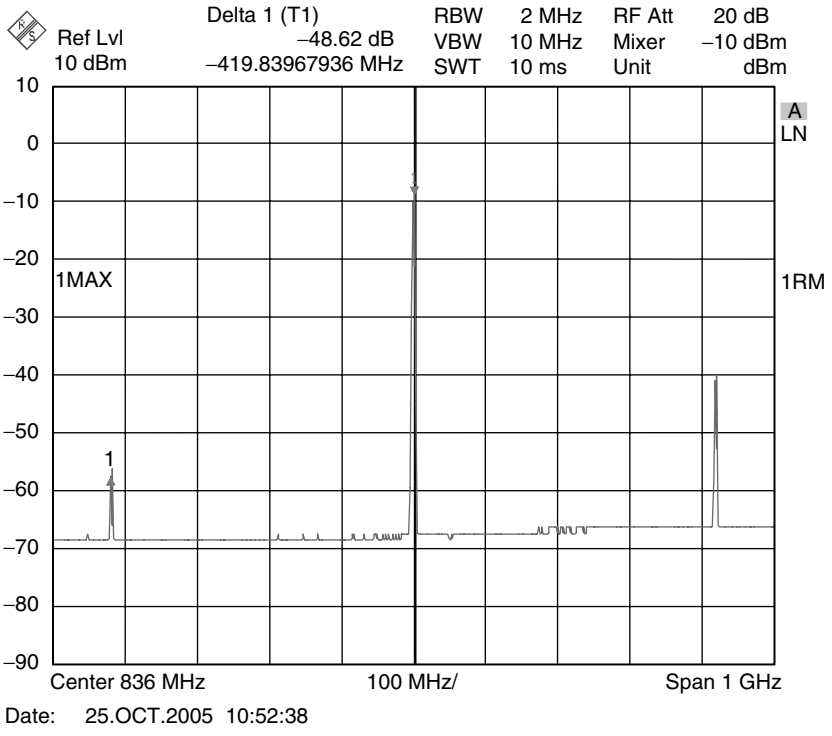


FIGURE 18.26 Example 418 MHz f/2 and 1254 MHz 3f/2 oscillation at −30°C.

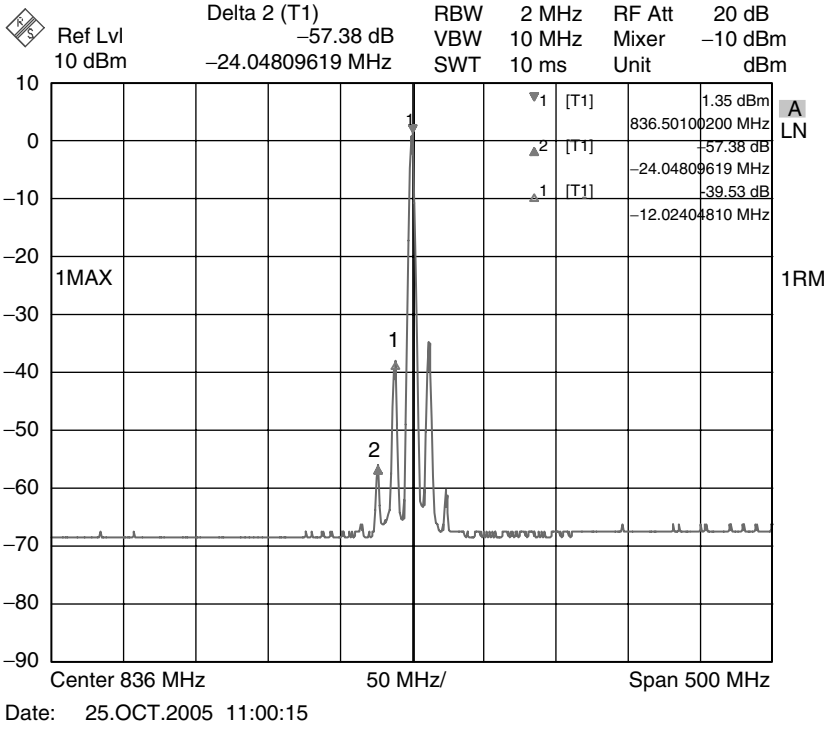


FIGURE 18.27 Example 12 MHz spurious oscillation at −30°C.

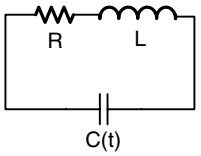


FIGURE 18.28 Simple RLC circuit with time varying capacitance.

results in

$$iR + L\frac{\mathrm{d}i}{\mathrm{d}t} + \frac{q}{C} = 0 \tag{18.33}$$

where q is charge, V_R is the voltage across the resistor, V_L is the voltage across the inductor, and V_C is the voltage across the capacitor.

Multiplying by C and taking the derivative with respect to time yields

$$CL\frac{\mathrm{d}^2 i}{\mathrm{d}t^2} + L\frac{\mathrm{d}i}{\mathrm{d}t}\frac{\mathrm{d}C}{\mathrm{d}t} + \frac{\mathrm{d}i}{\mathrm{d}t}RC + iR\frac{\mathrm{d}C}{\mathrm{d}t} + i = 0 \tag{18.34}$$

where $i = \dfrac{\mathrm{d}q}{\mathrm{d}t}$. Assuming $\frac{\Delta C}{C_\mathrm{o}} \ll 1$, the $\frac{\mathrm{d}C}{\mathrm{d}t}$ terms can be neglected. Equation 18.34 then reduces to

$$\frac{\mathrm{d}^2 i}{\mathrm{d}t^2} + \frac{\mathrm{d}i}{\mathrm{d}t}\left(\frac{R}{L}\right) + i\left(\frac{1}{LC}\right) = 0. \tag{18.35}$$

Equation 18.35 describes the natural response of a series RLC circuit when C is constant. Solutions at $n(f/2)$ are possible by assuming the time varying capacitance in Equation 18.31 [40,41]. Plugging Equation 18.31 into 18.35 and making use of the fact that

$$\frac{1}{1-\varepsilon} \approx 1 + \varepsilon \tag{18.36}$$

when $\varepsilon \ll 1$ results in

$$\frac{\mathrm{d}^2 i}{\mathrm{d}t^2} + \frac{\mathrm{d}i}{\mathrm{d}t}\left(\frac{R}{L}\right) + i\left(\frac{1}{LC_\mathrm{o}}\right)\left(1 + \frac{\Delta C}{C_\mathrm{o}}\sin(\omega_\mathrm{p} t)\right) = 0 \tag{18.37}$$

Phillips and the references therein analyze Equation 18.37 in detail [41–44]. Equation 18.37 is equivalent to the damped Mathieu equation [40]

$$\frac{\mathrm{d}^2 i}{\mathrm{d}t^2} + \gamma\frac{\mathrm{d}i}{\mathrm{d}t} + i(\omega_\mathrm{o}^2 + a\sin(\omega_\mathrm{p} t)) = 0 \tag{18.38}$$

if

$$\gamma = \frac{R}{L} \tag{18.39}$$

$$\omega_\mathrm{o}^2 = \frac{1}{LC_\mathrm{o}} \tag{18.40}$$

and

$$a = \frac{\Delta C}{C_o}\omega_o^2 \tag{18.41}$$

A closed form solution to Equation 18.38 is not possible in general, but assuming $\Delta C/C_o \ll 1$ and $\gamma \ll 1$ (small capacitive pumping in a nearly lossless circuit), an approximate solution results [40]. Oscillations will occur at half the pump frequency, ω_p, provided [40]

$$\omega_o = \frac{\omega_p}{2} \tag{18.42}$$

and

$$\frac{\Delta C}{C_o}\frac{\omega_o}{2} > \frac{R}{L} \tag{18.43}$$

This oversimplified theory does a reasonable job of qualitatively describing the presence of subharmonic spurs under drive [38]. For low drive levels (ΔC small) or if the resistance, R, is high enough, subharmonic oscillations will not appear.

For HBT devices, the impedance at the base is typically responsible for subharmonic oscillations. The resistance, R, in Equation 18.43 includes the base resistance as well the source impedance presented to the base of the transistor. C_o represents the effective small signal input capacitance and ΔC represents the variation in that capacitance under drive.

Sources of nonlinear capacitance in HBT devices include the base emitter and base collector capacitance. Adding resistance to the base of the HBT will eventually eliminate a subharmonic oscillation. Parametric oscillations can also be observed in pHEMT devices. In this case, nonlinearities in the gate-drain and gate-source capacitance are the primary contributors to subharmonic spurs.

Muller and Figel [38] explain the onset of parametric oscillations by representing the PA as two equivalent parallel amplifiers: a nonlinear amplifier and a linear amplifier. The quiescent current bias (I_{cq}) of the linear amplifier at each drive level is set by the input drive. Subharmonic oscillations are the result of device nonlinearities in the nonlinear amplifier. Spurious oscillations are caused by the effective operating conditions of the linear amplifier. Reference 38 provides numerous examples of drive dependant instabilities in PAs.

Rigorous analysis of large signal PA stability is extremely difficult, so most designers resort to guaranteeing small signal stability over a wide range of bias conditions and using empirical methods to solve the nonlinear problem. Common "cures" for stability problems are trying different bypass capacitor combinations as well as adding resistance to the base (or gate) of the transistor or adding loss to a matching network. In some cases, resorting to using a lower Q bypass capacitor can add just enough loss to dampen an oscillation.

18.4 Power Management Issues*

Energy conservation in handsets is becoming more critical as phone manufactures add features at a faster rate than battery technology can keep up with. Handset designers are placing increased emphasis on reduced current consumption for all components in the phone to preserve talk time while at the same time increasing features such as color displays, streaming video, Bluetooth radios, and MP3 players. A number of "enhancements" have been suggested over the past several years to reduce the average current consumption of the PA [49].

*This section includes excerpts from Teeter, D. et. al. Average current reduction in (W)CDMA PAs, *2006 IEEE RFIC Symposium Digest*, p. 429, 2006, copyright 2006 IEEE, included with permission.

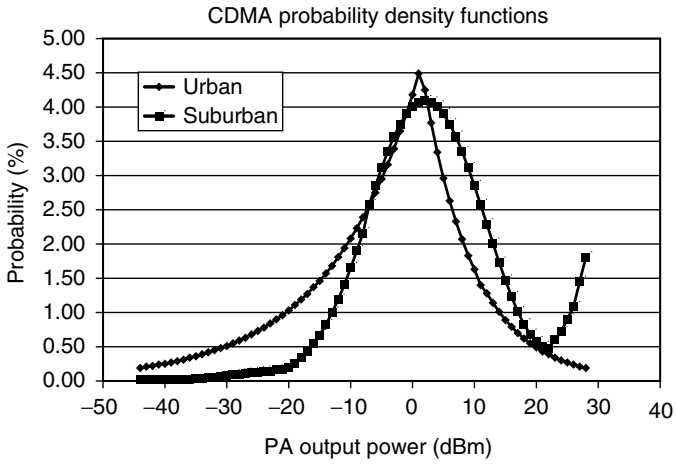


FIGURE 18.29 Power amplifier output power probability density function for urban and suburban cases for a CDMA handset. (From Teeter, D., et al., Average current reduction in (W)CDMA power amplifiers, *2006 IEEE RFIC Symposium Digest*, pp. 429–432, 2006, © 2006 IEEE, reprinted with permission.)

In the case of CDMA and WCDMA, the PA spends the majority of its time at backed off powers where the PA is very inefficient. Figure 18.29 shows the probability density function for CDMA PAs [50,51]. The average PA output power is about 30 dB backed off from the peak power of 28 dBm.

The net effect the PA current has on overall talk time can be understood by first computing the statistical average current, I_{ave}. Average current may be calculated by multiplying the current consumed by the PA by the usage statistics shown in Figure 18.29 and integrating over power. Mathematically,

$$I_{ave} = \int_{-\infty}^{P_{max}} I_{cdg}(P_{out}) \cdot dP_{out} \tag{18.44}$$

where

$$I_{cdg}(P_{out}) = I_{batt}(P_{out}) \cdot PDF(P_{out}) \tag{18.45}$$

In Equations 18.44 and 18.45, I_{batt} is the total battery current consumed by the PA at a given output power, P_{out}, and PDF, shown graphically in Figure 18.29, is the probability the PA will operate at that output power. I_{cdg} in Equation 18.45 represents the statistically weighted current for a given output power. Example plots of I_{cdg} versus output power are provided later in this section (see Figure 18.40). Reducing the average current consumption boils down to reducing I_{cdg} over output power.

Various techniques for reducing average current in PAs have been discussed in the literature [51–61]. These approaches can be grouped into two general categories: switch load techniques and dynamic bias control.

The switch load approach involves changing the output load between low and high power modes, as shown in Figures 18.30 and 18.31. Common switch load circuit techniques include the Doherty amplifier [55–57], load switching [60], parallel amplifier [53,54], and switch stage approaches [61]. All of these techniques improve back off efficiency by increasing the output load line at low power drive. These approaches will be referred to collectively as the low power enhanced efficiency (LPEE) PA.

A second approach to reduce average current consumption involves changing the bias point on the PA, either by adjusting the quiescent current or the supply voltage. Figure 18.32 illustrates the concept. Figure 18.33 depicts a typical implementation utilizing closed loop bias control based on output power. The simplest technique is to lower the quiescent current bias point on the PA at low output power

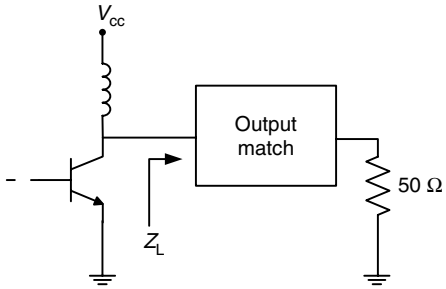


FIGURE 18.30 Switch load approach to low power efficiency enhancement.

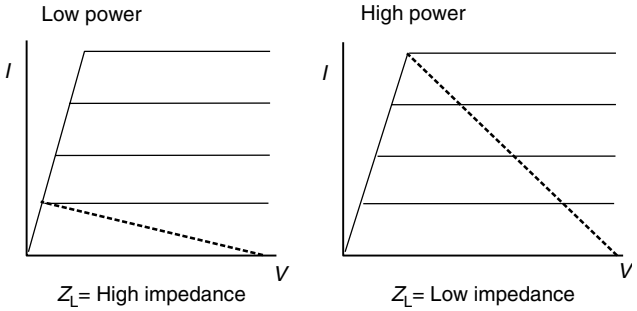


FIGURE 18.31 Illustration of switch load concept.

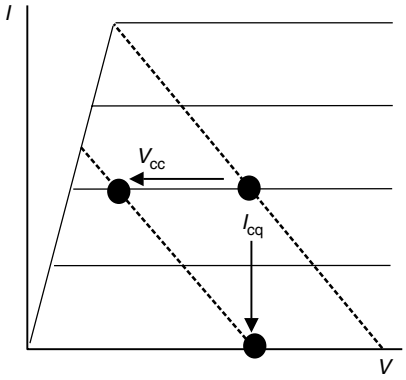


FIGURE 18.32 Illustration of bias control.

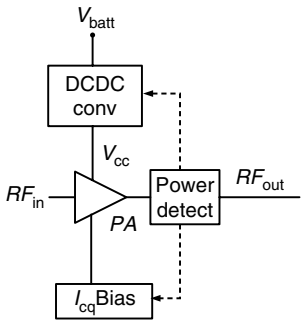


FIGURE 18.33 Closed loop bias control.

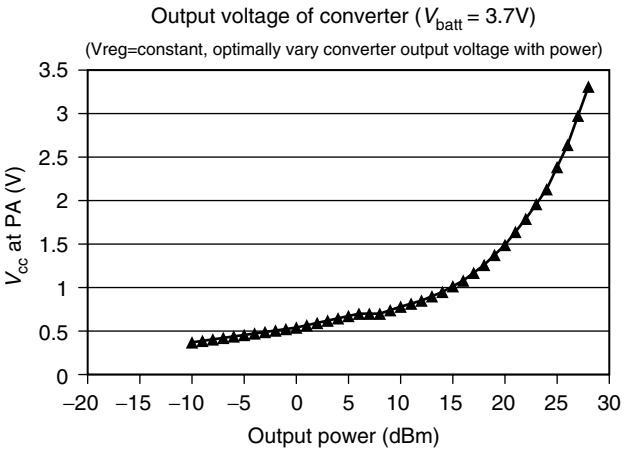


FIGURE 18.34 Optimal V_{cc} versus P_{out} for CDMA ACP = −48 dBc. This collector voltage profile was used in the optimal DCDC results shown in Figures 18.39 to 18.41. (Modified from Teeter, D., et al., Average current reduction in (W)CDMA power amplifier, *2006 IEEE RFIC Symposium Digest.*, pp. 429–432, 2006, © 2006 IEEE, reprinted with permission.)

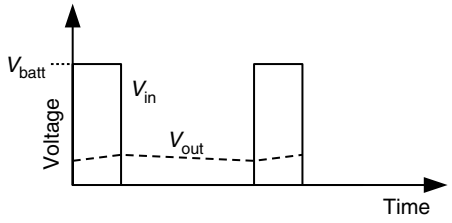


FIGURE 18.35 Input and output voltage waveforms for an ideal Buck DCDC converter.

levels [16,18,51]. This approach is referred to as analog bias control (ABC). Alternatively, one can vary the V_{cc} supply voltage on the PA, applying a high voltage at full power and reducing the voltage to the PA as the output power decreases [18]. A typical V_{CC} versus output power profile for an HBT PA is illustrated in Figure 18.34. The same approach can be used for a pHEMT amplifier by varying the drain voltage.

Supply voltage variation is accomplished through the use of a DCDC converter. Often, a Buck DCDC converter is used. Figure 18.35 illustrates the buck converter input and output voltages (refer also to Figure 18.36). A Buck converter may be viewed as a pulse width modulator passed through a low pass filter. By varying the duty cycle, the average voltage, and hence the PA supply voltage, changes. Figure 18.36 shows a simplified representation of a Buck DCDC converter.

DCDC converters can greatly reduce battery current consumption. The relationship between battery current (I_{batt}) and PA current (I_{load}) may be expressed as

$$I_{batt} = I_{load} \cdot \left(\frac{V_{cc}}{V_{batt}}\right) \cdot \left(\frac{1}{\eta}\right) \tag{18.46}$$

where I_{batt} is the current consumption from the battery, I_{load} is the the current delivered to the PA, V_{cc} is the collector voltage supplied to the power amplifier, V_{batt} is the battery voltage (generally 3.2 to 4.2 V), and η is the efficiency of the DCDC converter. For the case of a pHEMT PA, V_{dd} would be substituted for V_{CC} in Equation 18.46 and Figure 18.36.

All converters contain some losses. The dominant sources of power loss in Buck DCDC converters are power dissipation in the series resistance of the inductor, power dissipation in the on state resistance of

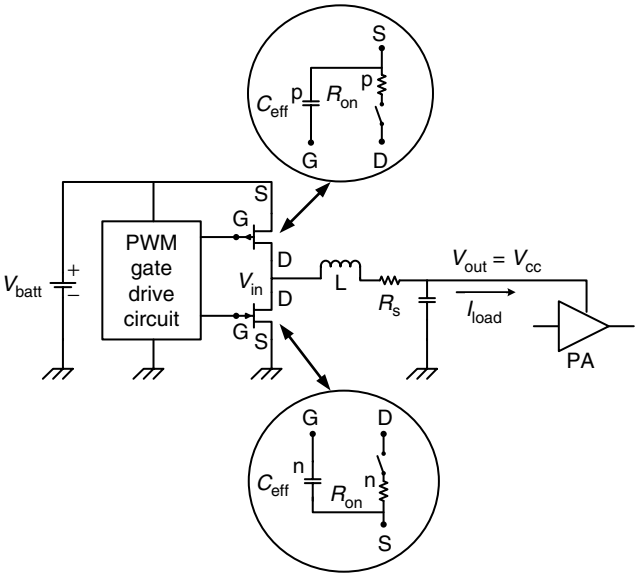


FIGURE 18.36 Simplified block diagram of a Buck DCDC converter.

the MOSFETS, and charging/discharging the effective gate capacitance of the switching MOSFETs [62]. Referring to Figure 18.36 and assuming the gate voltage on the MOSFETs swings from 0 to V_{batt}, one may approximate the losses in a converter as [63,64]

$$P_{\text{loss}}^{\text{Tot}} \approx (C_{\text{eff}}^{n} + C_{\text{eff}}^{p}) V_{\text{batt}}^{2} f_{\text{sw}} + I_{\text{load}}^{2}(DR_{\text{on}}^{p} + [1 - D]R_{\text{on}}^{n} + R_{\text{s}}^{\text{ind}}) \tag{18.47}$$

where D is the duty cycle, C_{eff} is the effective gate capacitance of the n or p channel MOSFET, R_{on}, is the on state resistance of the corresponding MOSFET, f_{sw} is the switching frequency of the converter, R_{s} is the series resistance of the inductor, and I_{load} is the current delivered to the PA (the load). The critical point of Equation 18.47 is that the dissipated power of the converter has two basic components; one due to the charging and discharging of the MOSFET gate capacitance which depends on the switching frequency and a second component which varies with load current caused by parasitic resistances [62–64]. At low load currents, the MOSFET capacitance charging/discharging related losses dominate. At high load currents, the resistive losses dominate. Also, as the switching frequency goes up, so do the charging/discharging losses [62]. Figure 18.37 shows the efficiency of a typical buck converter as a function of output voltage for a typical PA loading of 10 Ω. Figure 18.38 illustrates the impact finite converter losses have on average current consumption of a 2 stage HBT CDMA PA by comparing the average current one would obtain from a lossless converter to that obtained from a real converter. For the urban case illustrated (the worst case), converter losses typically account for about 23% of the total average current.

Average current consumption of different PA architectures can be compared by measuring the current versus output power for the amplifier and then applying Equation 18.44 [49]. Example calculations for a standard class AB PA (std PA) and a standard class AB PA with an optimized DCDC converter and analog bias control (std PA + DCDC + ABC) are shown in Figures 18.39 and 18.40. Use of low power efficiency enhancement techniques as well as DCDC converters significantly reduces average current consumption. Figure 18.41 compares average current for several different power management techniques. The standard PA (std PA) case refers to a conventional class AB PA tied directly to the battery. The LPEE PA makes use of the switch load concepts discussed at the beginning of this section. The final two cases in Figure 18.41 use a standard class AB PA along with a Buck DCDC converter. Use of a DCDC converter with an optimal V_{CC}

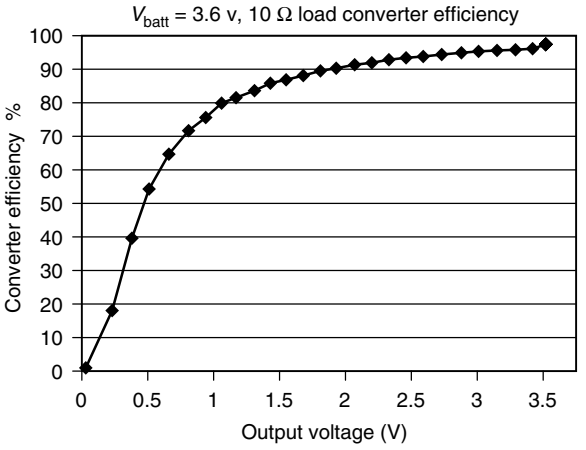


FIGURE 18.37 Typical buck DCDC converter efficiency. (Modified from Teeter, D., et al., Average current reduction in (W)CDMA power amplifiers, *2006 IEEE RFIC Symposium Digest*, pp. 429–432, 2006, © 2006 IEEE, reprinted with permission.)

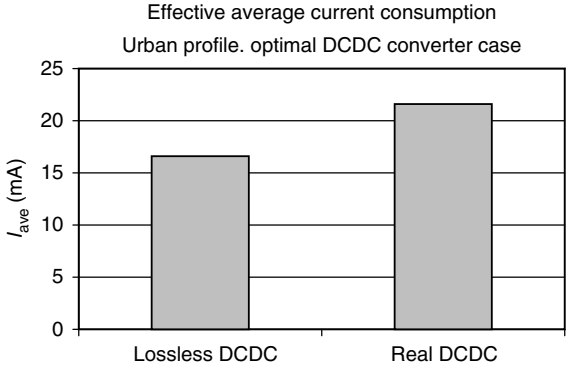


FIGURE 18.38 Comparison of urban average current consumption of a 2 stage HBT CDMA PA using a lossless and a real DCDC converter. Losses account for about 23% of the average current. (Modified from Teeter, D., et al., Average current reduction in (W)CDMA power amplifiers, *2006 IEEE RFIC Symposium Digest*, pp. 429–432, 2006, © 2006 IEEE, reprinted with permission.)

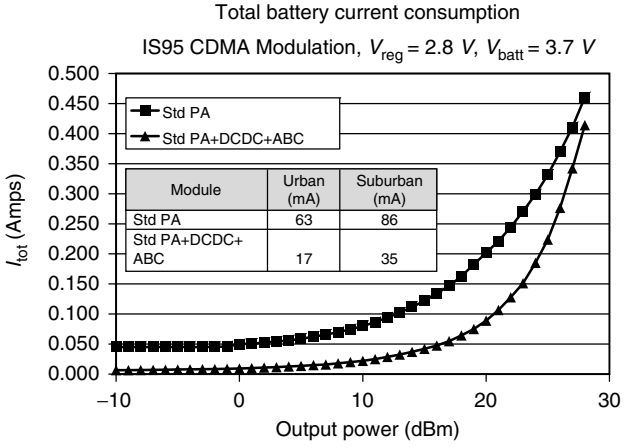


FIGURE 18.39 Total battery current comparison for different power amplifier architectures. (Modified from Teeter, D., et al., Average current reduction in (W)CDMA power amplifiers, *2006 IEEE RFIC Symposium Digest*, pp. 429–432, 2006, © 2006 IEEE, reprinted with permission.)

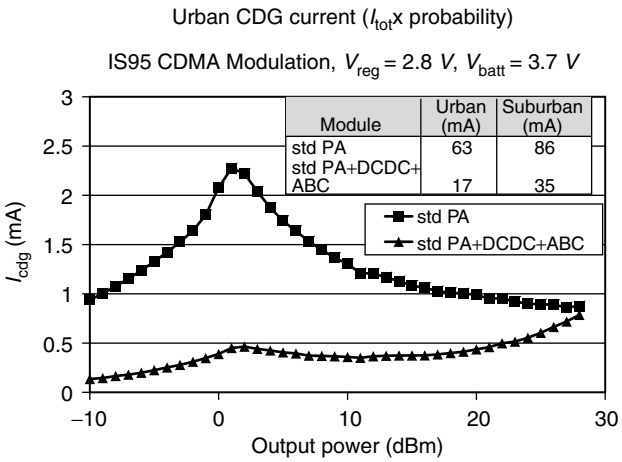


FIGURE 18.40 Current in figure 18.39 multiplied by CDMA urban usage statistics in Figure 18.29. (Modified from Teeter, D., et al., Average current reduction in (W)CDMA power amplifiers, *2006 IEEE RFIC Symposium Digest*, pp. 429–432, 2006, © 2006 IEEE, reprinted with permission.)

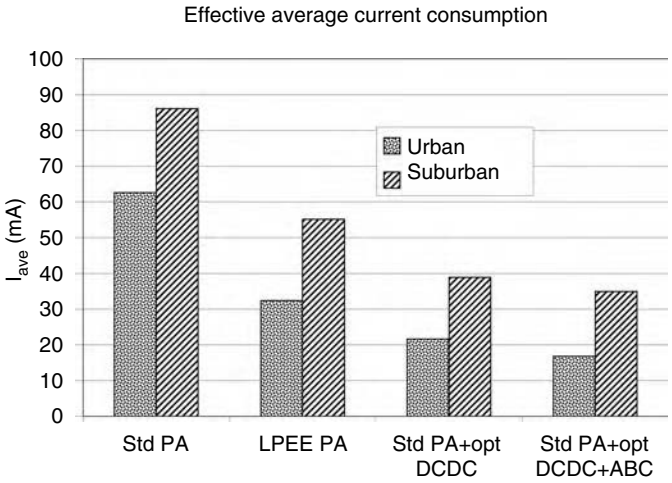


FIGURE 18.41 Average current consumption for different power management techniques. (From Spears, E., Teeter, D., Ngo, D., and Martin, S., Modulation trends and key specifications of mobile transmitters, *2006 IEEE MTT-S Workshop Notes – Workshop WSC – Advanced Power Amplifier IC's for High Efficiency Mobile Transmitters*, pp. 1–15, 2006, © 2006 IEEE, reprinted with permission.)

profile (Figure 18.34) in conjunction with analog bias control can result in a more than 70% reduction in average urban current usage with a standard class AB PA.

18.5 Large Signal Polar Modulation

As demands for lower cost of implementation and reduction in current consumption have increased, the use of large signal polar modulation has generated great interest for EDGE and multi-mode handsets [65–79]. Large signal polar modulation offers several advantages compared to standard transmitter architectures. Because the PA is always in saturation, polar modulation allows one to reduce the number of SAW filters required in the transmitter by reducing output noise and spurious signals [70]. In addition, a

saturated PA operates at peak efficiency, thereby reducing current consumption. By adding predistortion (AM–AM and AM–PM compensation) to the amplitude and phase modulation, improved PA linearity can also be achieved [65,68,75]. Finally, large signal polar modulation techniques allow one to efficiently operate transmitters with multiple air standards such as GMSK, EDGE, and WCDMA [70]. Detailed discussion of various polar modulation architectures may be found in References 1, and 65 to 79.

The Polaris 2 Total Radio^{TM}, developed by RF Micro Devices, is an example of an open loop large signal polar transmitter. The open loop architecture was chosen over a closed loop system to avoid the possibility that close range spurious jamming signals could break the lock of the phased lock loop [65]. A simplified diagram of the open loop polar architecture is shown in Figure 18.42. Basic polar modulation begins by converting the output of the vector modulator from rectangular $(I(t),\ Q(t))$ to polar components $(A(t), \theta(t))$ [68]. The RF output voltage may be written as

$$v_{\mathrm{RF}}(t) = I(t)\cos(\omega t) + Q(t)\sin(\omega t) \tag{18.48}$$

$$v_{\mathrm{RF}}(t) = A(t)\cos(\omega t + \theta(t)) \tag{18.49}$$

where $I(t)$ and $Q(t)$ are the in phase and quadrature components, respectively, and $A(t)$ and $\theta(t)$ are the corresponding amplitude and phase components [68]. Equation 18.49 can be viewed as amplitude modulating, $A(t)$, a constant envelope signal, $\cos(\omega t+\theta(t))$. Thus, the PA can be operated in saturation for a large signal polar modulated system since the RF input signal to the PA, $\cos(\omega t + \theta(t))$, has a constant envelope. Since the output voltage swing of the PA varies linearly with supply voltage, the amplitude information contained in Equation 18.49 is obtained by varying the collector (or drain) voltage on the PA by $A(t)$.

Use of large signal polar modulation offers significant current savings [1,72,73]. For example, the large signal polar modulation found in Polaris 2 reduces PA current consumption by about 25% compared to conventional *I*–*Q* modulation methods [72,73].

18.6 Output Power Control

Today's PAs have power control from very simple input drive level control to very sophisticated closed loop control systems, such as shown in Figure 18.33. W(CDMA) PAs are typically controlled by input drive level from a variable gain amplifier. The system implements both an open loop and a closed loop power control. Initial output power is adjusted using an open loop based on the received signal power at the base station. Tighter closed loop power control is then utilized that is capable of stepping the power in 1 dB increments every 666.7 μs*. The amplifier may incorporate some stage bypassing or DC–DC converter to improve the low power efficiency and lower the amplifier gain. The W(CDMA) specification requires a minimum power output of −50 dBm at the antenna port and a maximum of 24 dBm.

Power control methods for GSM vary from collector control, base control, or a combination of both. The objective of the control loop is to compensate for any variation in output power performance caused by temperature, supply voltage, or load VSWR. Power control is also used to implement the amplitude modulation used in large signal polar modulation illustrated in Figure 18.42.

When selecting a control loop, the overall system efficiency needs to be considered along with stability, dynamic range, and ease of implementation. Either voltage or current feedback can be employed, as shown in Figure 18.43 [68,70,74–80] or RF power detection, shown in Figure 18.44 [66,68]. Output power is varied in Figure 18.43a by varying the PA supply voltage through a MOSFET. Adjusting the gate voltage on the MOSFET causes it to act as a voltage controlled resistor. The PA supply voltage (V_{CC} for an HBT PA or V_{dd} for a FET PA) is sensed and compared to V_{ramp} in a closed loop by using an operational amplifier. The gate voltage of the MOSFET is adjusted accordingly [77,78,80]. Current feedback, shown in Figure 18.43b, works in much the same way as voltage feedback except supply current is sensed instead

*IS-95 CDMA systems use a similar power control scheme with the handset output power updated every 1.25 ms.

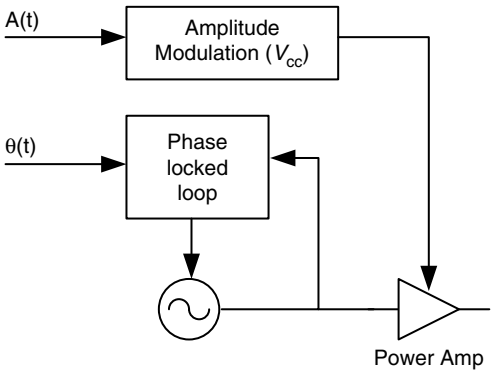


FIGURE 18.42 Block diagram of open loop polar modulated power amplifier.

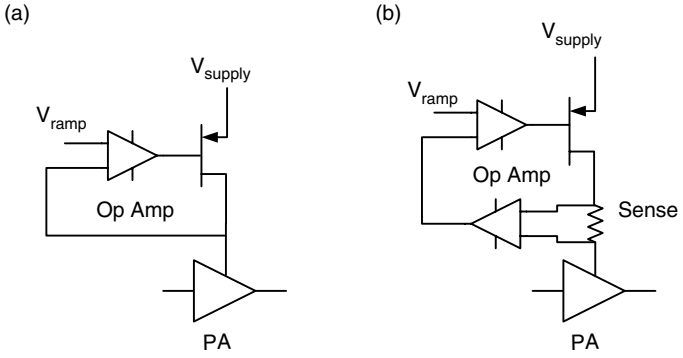


FIGURE 18.43 Power control using (a) voltage sense and (b) current sense.

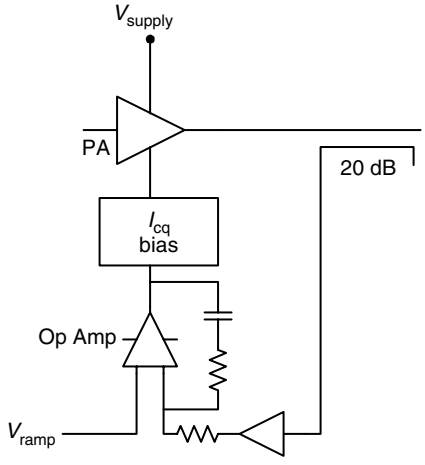


FIGURE 18.44 RF power detection closed loop control.

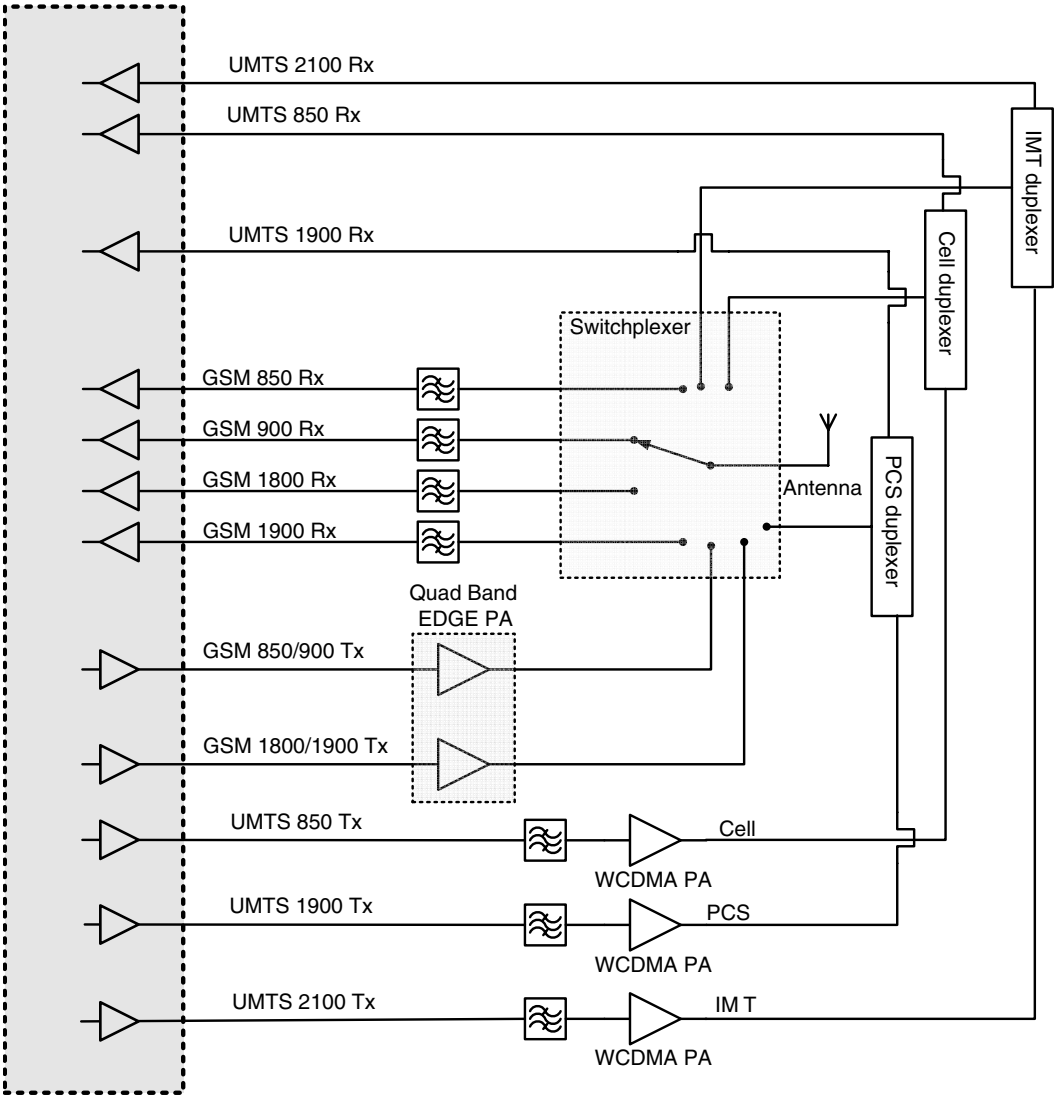


FIGURE 18.45 Simplified RF front end of multi-mode handset.

of supply voltage [78]. Alternatively, Figure 18.44 illustrates how output power may be sensed through a coupler and used to adjust the quiescent current bias (I_{cq}) of the amplifier [66].

The output power control range for GSM is from −5 dBm to rated output power and the dB/Volt control slope should be linear to assure that the output spectrum complies with the ETSI specification [80]. Collector (or drain) control keeps all stages of the PA in saturation while the base (or gate) voltage is held constant. To maintain forward power into a load VSWR, either voltage or current feedback can be implemented. These power control methods are indirect as there is no direct measurement of the PAs output power.

18.7 Multi-Mode Considerations

The migration to higher data rates from the GSM/EDGE air standard to HSDPA has resulted in the need for multi-mode phones to assure backward compatibility with the existing GSM/GPRS/EDGE network. Present implementations of WCDMA/EDGE phones (referred to as WEDGE) require multiple

transmit and receive chains. A simplified illustration of a common multi-mode front end architecture is shown in Figure 18.45. Commercial examples include Qualcomm's MSM6280 and 7200 chipsets [81,82]. Five PAs are used for this implementation. The GSM/EDGE bands utilize a standard quad band PA. The WCDMA/HSDPA bands utilize three separate PAs. The added amplifier chains result in higher PA to antenna losses compared to a conventional GSM/EDGE or CDMA phone because switch loss typically goes up as the number of poles increases.

The higher loss increases the output power requirements of the PA to achieve the same power out of the antenna. PA output power requirements on the order of 1–2 dB higher are not uncommon. Thus, more current draw from the battery is required, reducing talk time and increasing heat dissipation. A secondary challenge is that the PA die are larger and more phone board area is required for the RF function due to the large number of PA modules. Detailed discussion of the many design challenges for multi-mode radios may be found in References 1, 30, 83, and 84. The roll-out of multi-mode handsets offers enormous challenges and opportunities for RF engineers.

18.8 Conclusion

Design of PAs for mobile applications places numerous demands on the designer. On the one hand, he or she is being asked to design a high performance circuit with acceptable linearity and minimal current consumption. At the same time, there is an increased push for higher levels of integration to reduce size and simplify manufacturing. This frequently comes at the "expense" of PA performance through increased loss or poorer output matching. In addition, the desire for handsets with multiple radio standards and frequency bands is placing pressure on designers to develop PAs which are multi-mode. The result is an enormous technical challenge and opportunity for PA designers.

All of the concepts presented in this chapter can be applied equally well to PA designs for any modulation format. While the specific air standards will continue evolving to allow faster data rates, the basic principles used for PA design remain the same.

Acknowledgments

The authors would like to thank Jim Bao, Eric Yeon, Hong Jiang, Hannes Rahn, and Rob Simrell from RF Micro Devices for technical guidance during the writing of this manuscript. We would also like to thank Elizabeth Glass from RF Micro Devices for helpful editing suggestions.

References

1. McCune, E., High-efficiency, multi-mode, multi-band terminal power amplifiers, *IEEE Microwave Magazine*, vol. 6, no. 1, p. 44, March 2005.
2. Axelsson, et al., GSM/EDGE continued evolution, *Ericsson Review*, no. 1, p. 20, http://www.ericsson.com/ericsson/corpinfo/publications/review, 2006.
3. EV-DO Snapshot, Airvana, http://www.airvananet.com/solutions/solutions_evdo_snapshot.htm
4. Delivering voice and data: comparing CDMA2000 and GSM/GPRS/EDGE/UMTS, Dec.ember 2005, CDMA Development Group, www.cdg.org/resources/white_papers.asp
5. Thelander, M., The 3G Evolution—taking CDMA2000 into the next decade, Oct. 2005, prepared by Signals Research Group, LLC & CDMA Development Group, www.cdg.org/resources/white_papers.asp
6. Tkachenko, Y. et al., Performance comparison of HBTs and quasi E-mode PHEMTs for single supply high efficiency power amplifiers, *2001 Gallium Arsenide Applications Symposium (GAAS 2001)*, Sept. 2001, London.
7. Gray, R. and Meyer, R., *Analysis and Design of Analog Integrated Circuits*, 2nd Edn., John Wiley & Sons, New York, 1984.
8. Kushner, L., Output performance of idealized microwave power amplifiers, *Microwave Journal*, vol. 32, no. 10, p. 103, Oct. 1989.

9. Liu, W. et al., Current gain collapse in microwave multifinger heterojunction bipolar transistors operated at high power densities, *IEEE Transactions on Electron Devices*, vol. 40, no. 11, p. 1917, Nov. 1993.
10. Liu, W. and Khatibzadeh, A., The collapse of current gain in multi-finger heterojunction bipolar transistors: its substrate temperature dependence, instability criteria, and modeling, *IEEE Transactions on Electron Devices*, vol. 41, no. 10, p. 1698, Oct. 1994.
11. Liu, W. et al., The use of base ballasting to prevent the collapse of current gain in AlGaAs/GaAs heterojunction bipolar transistors, *IEEE Transactions on Electron Devices*, vol. 43, no. 2, p. 245, Feb. 1996.
12. Lu, K. and Snowden, C., Analysis of thermal instability in multi-finger power AlGaAs/GaAs HBT's, *IEEE Transactions on Electron Devices*, vol. 43, no. 11, p. 1799, Nov. 1996.
13. Liu, W., *Handbook of III–V Heterojunction Bipolar*, John Wiley & Sons, Inc., New York, 1998.
14. Liu, W., Thermal coupling in 2-finger heterojunction bipolar transistors, *IEEE Transactions on Electron Devices*, vol. 42, no. 6, p. 1033, June 1995.
15. Liu, W., Thermal management to avoid the collapse of current gain in power heterojunction bipolar transistors, *1995 IEEE GaAs IC Symposium Digest*, p. 147, 1995.
16. Lentini, P. and DeFalco, J., Adaptive power management increases cell phone talk time, *Applied Microwave & Wireless*, vol. 14, no. 10, p. 34, Oct. 2002.
17. Denning, D., Setting bias points for linear RF amplifiers—part 1, *Microwaves & RF*, vol. 41, no. 6, p. 71, June 2002.
18. Denning, D., Setting bias points for linear RF amplifiers—part 2, *Microwaves & RF*, vol. 41, no. 7, July 2002.
19. Wang, K. et al., Heating effects on the accuracy of HBT voltage comparators, *IEEE Transactions on Electron Devices*, vol. 34, no. 8, p. 1729, Aug. 1987.
20. Sze, S., *Physics of Semiconductor Devices*, John Wiley & Sons, New York, 2nd Edn., 1981.
21. Noh, Y. and Park, C., Linearized high efficient HBT power amplifier module for L-band application, *2001 IEEE GaAs IC Symposium Digest*, p. 197, 2001.
22. Wohlmuth, W. et al., E-/D-pHEMT technology for wireless components, *2004 IEEE Compound Semiconductor Integrated Circuit Symposium Digest*, p. 115, 2004.
23. Kim, J., Noh, Y. and Park, C., High linear HBT MMIC power amplifier with partial RF coupling to bias circuit for W-CDMA portable application, *2002 IEEE International Conference on Microwave and Millimeter Wave Technology Proceedings*, p. 809, 2002.
24. Kim, J. et al., MMIC power amplifier adaptively linearized with RF coupled active bias circuit for WCDMA mobile terminals applications, *2003 IEEE MTT-S Digest*, p. 2209, June 2003.
25. Noh, Y. and Park, C., An intelligent power amplifier MMIC using a new adaptive bias control circuit for WCDMA applications, *IEEE Journal of Solid-State Circuits*, vol. 39, no. 6, p. 967, June 2004.
26. Kim, J. et al., A power efficient WCDMA smart power amplifier with emitter area adjusted for output power levels, *2004 IEEE MTT-S Digest*, p. 1165, June 2004.
27. Jarvinen, E., Kalajo, S. and Matilainen, M., Bias circuits for GaAs HBT power amplifiers, *2001 IEEE MTT-S Digest*, p. 507, June 2001.
28. Jarvinen, E., GaAs HBT power amplifier with smooth gain control characteristics, *1998 RAWCON Proceedings*, p. 321, Aug. 1998.
29. Niu, G. et al., Transistor noise in SiGe HBT RF technology, *2000 IEEE Bipolar Circuit Technology Meeting*, p. 207, Oct. 2000.
30. Minnis, B. and Moore, P., A highly digitized multimode receiver architecture for 3G mobiles, *IEEE Transactions on Vehicular Technology*, vol. 52, no. 3, p. 637, May 2003.
31. 3GPP TS 45.005 v4.1.1 (2000-11), Third Generation Partnership Project, Technical Specification Group GERAN, Digital cellular telecommunications system (Phase 2+), Radio transmission and reception (Release 4), www.3gpp.org
32. Understanding GSM/EDGE Transmitter and Receiver Measurements for Base Transceiver Stations and their Components, Agilent Technologies, Application Note 1312, August 2002, www.agilent.com
33. Struble, W. and Platzker, A., A rigorous yet simple method for determining stability of linear n-port networks, *1993 GaAs IC Symposium Digest*, p. 251.

34. Platzker, A. and Struble, W., Rigorous determination of the stability of linear n-node circuits from network determinants and the appropriate role of the stability factor K of their reduced two-ports, *Third International Workshop on Integrated Nonlinear Microwave and Millimeterwave Circuits, 1994.* p. 93, Oct. 5–7, 1994.
35. Platzker, A., Struble, W. and Hetzler, K., Instabilities diagnosis and the role of K in microwave circuits, *1993 IEEE MTT-S Digest*, p. 1185, 1993.
36. Imbornone, J. et al., New insight into subharmonic oscillation mode of GaAs power amplifiers under severe output mismatch condition, *IEEE Journal of Solid-State Circuits*, vol. 32, no. 9, p. 1319, Sept. 1997.
37. Teeter, D., Platzker, A. and Bourque, R., A compact network for eliminating parametric oscillations in high power MMIC amplifiers, *1999 IEEE MTT-S Digest*, vol. 3, p. 967, 1999.
38. Muller, O. and Figel, W., Stability problems in transistor power amplifiers, *Proceedings of the IEEE*, vol. 55, no. 8, p. 1458, Aug. 1967.
39. Mumford, W., Some notes on the history of parametric oscillations, *Proceedings of the IRE*, May 1960, vol. 48, no. 5, May 1960, p. 848.
40. Elmer, Franz-Josef, Parametric resonance, *The Pendulum Lab*, University of Basel, Switzerland, http://monet.physik.unibas.ch/∼elmer/pendulum/parres.htm
41. Phillips, R., Parametric oscillation in a damped resonant system, *IEEE Transactions on Circuit Theory,* vol. 10, no. 4, p. 512, Dec. 1963.
42. McLachlan, N., *Theory and Application of Mathieu Functions*, Dover, New York, 1964.
43. Bura, P., The degenerate and quasi-degenerate mode of parametric amplification, *IRE Transactions on Circuit Theory*, vol. 7, no. 3, p. 200, Sept. 1960.
44. Bura, P., Correspondence on: the degenerate and quasi-degenerate mode of parametric amplification, *IRE Transactions on Circuit Theory*, vol. 8, no. 4, p. 487, Dec. 1961.
45. Manley, J. and Rowe, H., Some general properties of nonlinear elements—part 1. General energy relations, *Proceedings of the IRE*, vol. 44, no. 7, p. 904, July 1956.
46. Van Der Pol, B., Forced oscillations in a circuit with non-linear resistance (reception with reactive triode), *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, series 7, vol. III, no. 13, p. 65, Jan. 1927.
47. Minorsky, N., *Nonlinear Oscillations*, D. Van Nostrand Company, Princeton, NJ, 1962.
48. Penfield Jr., P. and Rafuse, R., *Varactor Applications*, The MIT Press, Cambridge, MA, 1962.
49. Teeter, D. et al., Average current reduction in (W)CDMA power amplifiers, *2006 IEEE RFIC Symposium Digest*, p. 429, 2006.
50. Sevic, J., Statistical characterization of RF power amplifier efficiency for CDMA wireless communication systems, *1997 IEEE Wireless Communications Conference*, p. 110, 1997.
51. Fowler, T. et al., Efficiency improvement techniques at low power levels for linear CDMA and WCDMA power amplifiers, *2002 IEEE RFIC Symposium Digest*, p. 41, June 2002.
52. Deng, J., et al., A high average-efficiency SiGe HBT power amplifier for WCDMA handset applications, *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, no. 2, p. 529, Feb. 2005.
53. Kim, J.H. et al., An InGaP-GaAs HBT MMIC smart power amplifier for W-CDMA mobile handsets, *IEEE Journal of Solid State Circuits*, vol. 39, no. 6, p. 905, June 2003.
54. Hau, G. et al., A 20 mA quiescent current 40% PAE WCDMA HBT power amplifier module with reduced current consumption under backoff power operation, *2005 IEEE RFIC Symposium Digest*, p. 243, June 2005.
55. Nam, J., Shin, J. and Kim, B., A handset power amplifier with high efficiency at low level using load modulation technique, *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, no. 8, p. 2369, Aug. 2005.
56. Iwamoto, M. et al., An extended Doherty amplifier with high efficiency over a wide power range, *IEEE Transactions on Microwave Theory and Techniques*, vol. 49, no. 12, p. 2472, Dec. 2001.
57. Kim, J. et al., A highly integrated Doherty amplifier for CDMA handset applications using an active phase splitter, *IEEE Transactions on Microwave and Wireless Components*, vol. 15, no. 5, p. 333, May 2005.
58. Staudinger, J. et al., High efficiency CDMA RF power amplifier using dynamic envelope tracking technique, *2000 IEEE MTT-S Digest*, vol. 2, , p. 873, June 2000.

59. Staudinger, J., An overview of efficiency enhancements with application to linear handset power amplifiers, *2002 IEEE RFIC Symposium Digest*, p. 45, June 2002.
60. Kim, S. et al., CDMA handset power amplifier with a switched output matching circuit for low/high power mode operations, *2004 IEEE MTT-S Digest*, vol. 3, p. 1523, June 2004.
61. Staudinger, J., Applying switch gain concepts to improve efficiency and linearity for mobile CDMA power amplification, www.microwavejournal.com
62. Yao, K. et al., A novel winding-coupled buck converter for high frequency, high step down DC-DC conversion, *IEEE Transactions on Power Electronics*, vol. 20, no. 5, p. 1017, Sept. 2005.
63. Mulligan, M., Broach, B. and Lee, T., A constant-frequency method for improving light-load efficiency in synchronous buck converters, *IEEE Power Electronics Letters*, vol. 3, no. 1, p. 24, Mar. 2005.
64. Musunuri, S. and Chapman, P., Improvement of light-load efficiency using width-switching scheme for CMOS transistors, *IEEE Power Electronics Letters*, vol. 3, no. 3, p. 105, Sept. 2005.
65. Hietala, A., A quad-band 8PSK/GMSK polar transceiver, *IEEE Journal of Solid State Circuits*, vol. 41, no. 5, p. 1133, May 2006.
66. Sowlati, T. et al., Quad-band GSM/GPRS/EDGE polar loop transmitter, *IEEE Journal of Solid State Circuits*, vol. 39, no. 12, p. 2179, Dec. 2004.
67. Elliott, M. et al., A polar modulator transmitter for GSM/EDGE, *IEEE Journal of Solid State Circuits*, vol. 39, no. 12, p. 2190, Dec. 2004.
68. Reynaert, P. and Steyaert, M., A 1.75-GHz polar modulated CMOS RF power amplifier for GSM-EDGE, *IEEE Journal of Solid State Circuits*, vol. 40, no. 12, p. 2598, Dec. 2005.
69. Hadjichristos, A. et al., A highly integrated quad band low EVM polar modulation transmitter for GSM/EDGE applications, *IEEE 2004 Custom Integrated Circuits Conference*, p. 565, Oct. 2004.
70. Hadjichristos, A., Transmit architectures and power control schemes for low cost highly integrated transceivers for GSM/EDGE applications, *Proceedings of the 2003 International Symposium of Circuits and Systems*, vol. 3, p. III-610, May 2003.
71. Sander, W., Schell, S., and Sander, B., Polar modulator for multi-mode cell phones, *IEEE 2003 Custom Integrated Circuits Conference*, p. 439, Sept. 2003.
72. Wilkins, B. and Della Corte, F., Polar modulator saves power in EDGE sets, *Microwaves & RF*, www.microwavesrf.com, Nov. 2004.
73. Wilkins, B. and Della Corte, F., Large signal polar modulation reduces heat dissipation and increases battery life in EDGE handsets, *Microwave Product Digest*, www.mpdigest.com, Feb. 2005.
74. Wang, R., Introduction of polar modulation on EDGE (Chinese), *Electronic Products*, p. 46, June 2004.
75. Seegerer, G. and Ulbricht, G., EDGE transmitter with commercial GSM power amplifier using polar modulation with memory predistortion, *2005 IEEE MTT-S Digest*, p. 1553, June 2005.
76. Pehlke, D., Hadjichristos, A. and Justice, S., High performance open-loop AM modulator designed for power control of E-GPRS polar modulated power amplifier, *2004 IEEE Custom Integrated Circuits Conference*, p. 1553, June 2005.
77. Epperson, D. and Quist, L., Power amplifier control, US Patent US 6,701,138 B2, RF Micro Devices, March 2, 2004.
78. Epperson, D., Increased dynamic range for power amplifiers used with polar modulation, US Patent US 6,701,134 B1, RF Micro Devices, March 2, 2004.
79. Jau, J. K. et al., Polar modulation-based RF power amplifiers with enhanced envelope processing technique, *2004 European Microwave Conference*, vol. 3, p. 1317, 2004.
80. Madsen, U., 3G GMSK/EDGE power control with enhanced switching transient performance, *RF Design*, vol. 29, no. 1, p. 22, Jan. 2006.
81. Qualcomm MSM 6280 Multi-Mode chipset, http://www.cdmatech.com/products/msm6280_chipset_solution.jsp
82. Qualcomm MSM 7200 Multi-Mode chipset, http://www.cdmatech.com/products/msm7200_chipset_solution.jsp
83. Bakkaloglu, B. and Fontaine, P., Multi-mode, multi-band RF transceiver circuits for mobile terminals in deep-submicron CMOS processes, *2005 IEEE Radio Frequency Integrated Circuits Symposium*, p. 483, June 2005.
84. Wang-Chang Albert Gu, RF Front-end modules in cellular handsets, *2004 IEEE Compound Semiconductor Integrated Circuits Digest*, p. 225, Oct. 2004.

19

Class A Amplifiers

Warren L. Seely
Ubidyne, Inc.

19.1 Introduction

This derivation looks at class A amplifier electrical performance versus power based on a minimal parameter set. This allows an engineer to make accurate performance predictions (the results of this derivation are not intended to replace simulation) without having to actually execute a design. Additionally, the results can be used to great advantage in estimating system performance. Much of the derivation is based on compression characteristics, making it a unique approach to analyzing power amplifier performance. A fast and easy to use engineering tool can be created by programming the results into an Excel spreadsheet.

The single stage class A amplifier derivations herein are carried out with respect to power. All matching is presumed to be infinite bandwidth and constant impedance unless otherwise stated. The bias circuits are assumed to have infinite isolation to the amplifier at all but the DC frequency (ideal chokes). The active device in Figure 19.1 is a FET, but it could just as easily be a BJT, or any other controlled current source.

In a class A amplifier (Figure 19.2), the active device acts as a modulated current source biased midway between saturation and conduction cutoff. As the amplitude of an applied sinusoidal signal is increased, the output will start to clip at both ends simultaneously. This is equivalent to a conduction angle of 360° as long as the output signal is not clipping, which is avoided. The output waveform is a replication of the input with fixed gain and phase shift.

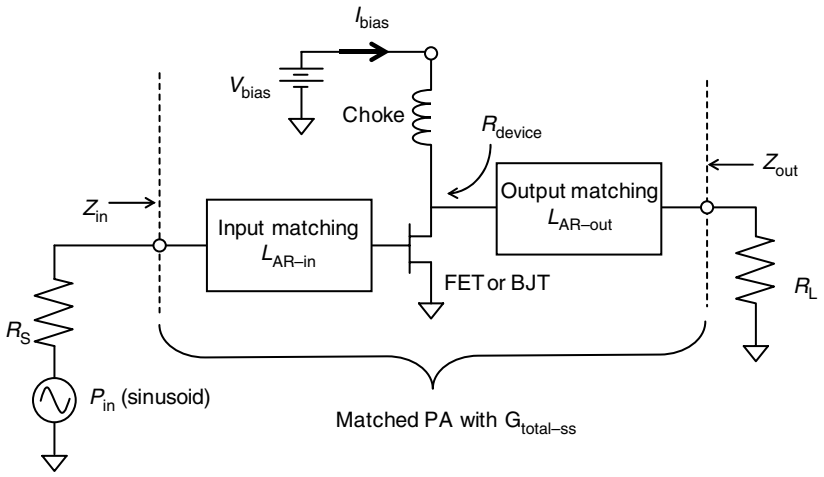


FIGURE 19.1 Class A amplifier nomenclature.

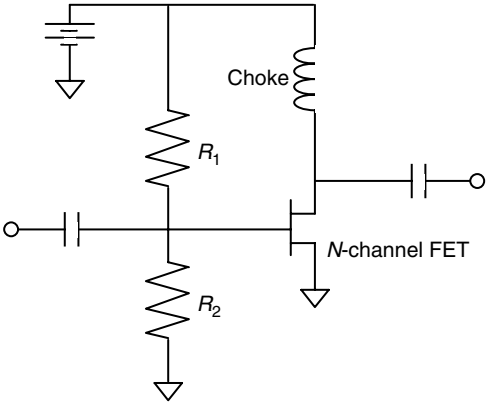


FIGURE 19.2 Example class A amplifier.

In the equations used and derived herein, it is important to note the basic nomenclature system. Parameters that are capitalized are in Volts, Amps, decibels referenced to a Watt (dBW), decibels referenced to a milli-Watt (dBm), power ratio in decibels (dB), and so on, as the situation dictates. Parameters that are lower case are in Watts (W), milli-Watts (mW), power ratio (dimensionless), and so on, again as the situation dictates.

There are two major performance groupings used throughout this derivation, which also are indicated in the subscript nomenclature: small signal (ss) and large signal (LS). Small signal refers to any linear operation or situation. The only signals present are those that are supplied, and no new tones are created. These signals can only be amplified, attenuated, and/or phase shifted. Although there are no truly linear devices, the small signal approximation can be made when the nonlinear effects are negligible, such as when the output power is well below (at least 15–20 dB) the achievable saturated output power. Large signal refers to operation or situations that generate harmonics and/or mixing products. These signals can be modified in any way desired. For small signals, the large signal and small signal performance converge to the same values, but as operation approaches saturated conditions, large signal operation and conditions must be included. Thus, for this derivation, small signal conditions indicate linear operation, whereas large signal includes the entire range of operation, both linear (or small signal) and nonlinear.

All DC and radio frequency (RF) power calculations herein are based on root mean square (RMS) currents and voltages, which when multiplied result in average power. The average RF power (p_{AVE-RF}) using an RMS current or voltage into a real load is given below, where T is the period of integration, θ is the angle, and f is the signal form (sine or cosine wave). Substituting simple sine or cosine dependant peak–peak and peak currents ($I_{pk-pk} = 2I_{pk}$) or voltages ($V_{pk-pk} = 2V_{pk}$) and integrating over a complete cycle ($T = 2\pi$) results in the following:

$$I_{RMS} = \sqrt{\frac{1}{T}\int_{\theta 1}^{\theta 2} f^2(\theta)\,\mathrm{d}\theta} = \frac{I_{pk}}{\sqrt{2}} = \frac{I_{pk-pk}}{2\sqrt{2}}$$

where $f = I_{pk}\cos(\phi + \theta)$ for any ϕ

$$V_{RMS} = \sqrt{\frac{1}{T}\int_{\theta 1}^{\theta 2} f^2(\theta)\,\mathrm{d}\theta} = \frac{V_{pk}}{\sqrt{2}} = \frac{V_{pk-pk}}{2\sqrt{2}}$$

where $f = V_{pk}\cos(\phi + \theta)$ for any ϕ

$$\therefore\ p_{AVE-RF} = I_{RMS}V_{RMS} = \frac{I_{pk}V_{pk}}{2} = \frac{I_{pk-pk}V_{pk-pk}}{8}.$$

Substituting fixed DC currents and/or voltages into the average expression based on an RMS form above results in the simple average DC power (p_{AVE-DC}) over the period as given below:

$$I_{RMS} = I_{AVE-DC} = \sqrt{\frac{1}{T}\int_{\theta 1}^{\theta 2} f^2(\theta)\,\mathrm{d}\theta} = I_{DC} \qquad V_{RMS} = V_{AVE-DC} = \sqrt{\frac{1}{T}\int_{\theta 1}^{\theta 2} f^2(\theta)\,\mathrm{d}\theta} = V_{DC}$$

where $f = I_{DC}$ or V_{DC} respectively.

$$\therefore\ p_{AVE-DC} = I_{RMS}V_{RMS} = I_{DC}V_{DC}.$$

19.2 The Assumptions

The following assumptions are made in the derivations and calculations below. An understanding of these assumptions is fundamental to understanding the final results obtained herein. For the most part, these assumptions are made to simplify and clarify the analysis with little or no effect on the final calculated results versus measured results. The exceptions include the last three items on the list:

(a) Pure sinusoidal inputs or tones.
(b) Independent noninterfering RF sources for multitone derivations.
(c) Ideal matching at all power levels (this can **not** be physically realized) except when specifically stated otherwise.
(d) Ideal matching at the fundamental, all harmonics, and all intermodulation (IM) (mixing) products.
(e) Bias circuits are fully isolated from the active devices (ideal choke).
(f) Input bias is not considered in any efficiency derivations.
(g) Higher-order terms do not significantly contribute to power levels.

Bias isolation is the isolation from the amplifier output to each bias injection port. Good bias isolation (e) is a design requirement that is often overlooked, and can significantly affect the results

in many ways. Signals reflecting back from or coupling between poorly designed or poorly implemented bias circuits can make for all kinds of problems including gain and stability issues, impedance variations, harmonic variations, and linearity problems. The key here is to pay attention to detail and verify good bias isolations with both the simulator and in the lab. It is a good idea to achieve a minimum bias isolation equal to the forward amplifier gain plus 10–15 dB. Thus an amplifier with 20 dB of gain should have 30–35 dB minimum isolation from the output port to each bias port. This simple step will save all kinds of development time and result in amplifiers with far fewer problems, both in development and in the field.

Although input bias (f) is not included in the efficiency calculations herein, it should be relatively easy to compensate or add these into the calculations if warranted. There are many different configurations for controlling or injecting bias into both the input and output side of an amplifier. The specific requirements for any given application will vary, some significantly while some others can be ignored. For example, the bias circuitry may include voltage regulators or pass transistors that must be accounted for. With the information given, these situations should be easily accounted for in a systematic way.

Higher-order terms (g) are a not so easy to deal with. Harmonic reflections at the input can have significant impacts on linearity. Output harmonic reflections generally affect the output power levels, and thus many other parameters including efficiency. These reflections can enhance or detract from performance, so it is important to control them. In fact, various other classes of amplifiers such as a class *F* (harmonically tuned) amplifier specifically target the use of harmonic power to significantly enhance amplifier performance. For the purposes of this derivation, these effects are controlled by ideal terminations that inhibit any reflections, resulting in statistically neutral or "average" performance calculations. However, some harmonic information is generated, which provides a path for further exploration on the effects of harmonic power.

19.3 Basic Math Relationships

Some basic trigonometric relationships and mathematical identities are used throughout this derivation. They are given below in order to provide a review and separate the math from the concepts under derivation:

$$\cos^2\alpha = \tfrac{1}{2}(1+\cos 2\alpha)$$

$$\cos^3\alpha = \tfrac{1}{4}(3\cos\alpha+\cos 3\alpha)$$

$$\cos^4\alpha = \tfrac{1}{8}(3+4\cos 2\alpha+\cos 4\alpha)$$

$$\cos^5\alpha = \tfrac{1}{16}(10\cos\alpha+5\cos 3\alpha+\cos 5\alpha)$$

$$\cos^6\alpha = \frac{1}{32}(10+15\cos 2\alpha+6\cos 4\alpha+\cos 6\alpha)$$

$$\cos\alpha\cos\beta = \tfrac{1}{2}\cos(\alpha-\beta)+\tfrac{1}{2}\cos(\alpha+\beta)$$

$$(\cos\alpha+\cos\beta)^2 = 1+\tfrac{1}{2}\cos 2\alpha+\tfrac{1}{2}\cos 2\beta+\cos(\alpha-\beta)+\cos(\alpha+\beta)$$

$$(\cos\alpha+\cos\beta)^3 = \tfrac{9}{4}\cos\alpha+\tfrac{9}{4}\cos\beta+\tfrac{1}{4}\cos 3\alpha+\tfrac{1}{4}\cos 3\beta+\tfrac{3}{4}\cos(\alpha-2\beta)$$
$$+\tfrac{3}{4}\cos(2\alpha-\beta)+\tfrac{3}{4}\cos(2\alpha+\beta)+\tfrac{3}{4}\cos(\alpha+2\beta)$$

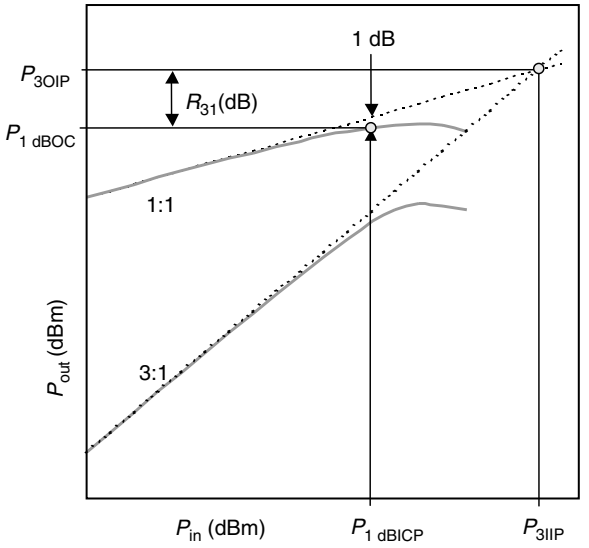


FIGURE 19.3 Two tone 3OIP to single tone 1dBOCP ratio.

19.4 The Parameter Set

The following parameters will be used as the basis from which all calculations are made (the example includes specified values). From this minimal parameter set, much of the class A performance versus power can be determined, as will be shown. These parameters will be defined as the derivations proceed.

Amplifier parameters		
V_{bias} =	28.00	Volts
$I_{bias-ss}$ ($\sim I_{on}/2$) =	2.000	Amperes
R_{on} =	1.000	Ohms
$G_{total-ss}$ =	18.00	dB
Comp @ P_{sat} =	3.00	dB
Input L_{AR} =	1.20	dB
Output L_{AR} =	0.50	dB
Back-off from P_{sat} ($\geq$1dB) =	15.00	dB

19.5 Ratio of Two-Tone 3OIP to Single-Tone 1 dBOCP

The objective of this section is to derive R_{31}, the ratio in dB between the third-order output intercept point (3OIP, a measure of the amplitude distortion) and the 1 dB output compression point (1 dBOCP) as shown in Figure 19.3. This rather tedious derivation will result in a single number representing a power ratio, given in decibels.

The first step is to determine a generalized expression for the single-tone 1 dBOCP. Thus, a single-tone excitation is applied to a matched amplifier (Figure 19.4).

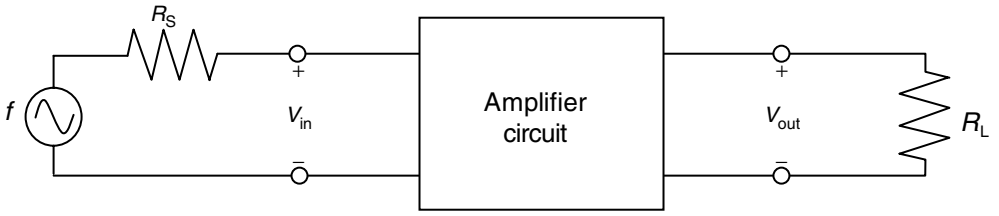


FIGURE 19.4 Single tone excitation.

Let $V_{in} = A\cos(\omega t) = A\cos\theta$, thus $A = V_{pk}$ and $2A = V_{pk-pk}$.

$$\text{Linear circuit} \Rightarrow V_o = K_1 V_{in} = K_1 A\cos\theta$$

$$\text{Nonlinear circuit} \Rightarrow V_o = K_o + K_1 V_{in} + K_2 V_{in}^2 + \cdots = \sum_{n=0}^{\infty} K_n V_{in}^n$$

In the above linear circuit equations, the input magnitude is $V_{in} = A$, and the output magnitude is $V_o = K_1 A$, where K_1 is voltage gain. Continuing with the derivation using the nonlinear circuit gives

$$V_o = K_o + K_1 A\cos\theta + K_2 A^2\cos^2\theta + \cdots = \sum_{n=0}^{\infty} K_n A^n \cos^n\theta$$

$$V_o = K_o + K_1 A\cos\theta + K_2 A^2 \tfrac{1}{2}(1 + \cos 2\theta) + K_3 A^3 \tfrac{1}{4}(3\cos\theta + \cos 3\theta)$$

$$+ K_4 A^4 \tfrac{1}{8}(3 + 4\cos 2\theta + \cos 4\theta) + \cdots$$

Gathering the terms gives

$$V_o = A_{DC} + A_{fund}\cos\theta + A_{2nd}\cos 2\theta + A_{3rd}\cos 3\theta + A_{4th}\cos 4\theta + \cdots = \sum_{n=0}^{\infty} A_{nth}\cos(n\theta)$$

where DC, the fundamental, and the second-, third-, and fourth-harmonic amplitudes are given by the following. The only concerns are with the fundamental and third harmonics:

$$A_{DC} = K_o + \tfrac{1}{2}K_2 A^2 + \tfrac{3}{8}K_4 A^4 + \tfrac{10}{32}K_6 A^6 + \cdots$$

$$A_{fund} = K_1 A + \tfrac{3}{4}K_3 A^3 + \tfrac{5}{8}K_5 A^5 + \cdots$$

$$A_{2nd} = \tfrac{1}{2}K_2 A^2 + \tfrac{1}{2}K_4 A^4 + \tfrac{15}{32}K_6 A^6 + \cdots$$

$$A_{3rd} = \tfrac{1}{4}K_3 A^3 + \tfrac{5}{16}K_5 A^5 + \tfrac{7}{32}K_7 A^7 + \cdots$$

$$A_{4th} = \tfrac{1}{8}K_4 A^4 + \tfrac{3}{16}K_6 A^6 + \tfrac{5}{32}K_8 A^8 + \cdots$$

At the fundamental, the RMS powers are given by

$$\text{Fundamental input power} \Rightarrow p_{\text{in}} = \frac{V_{\text{pk-pk}}^2}{8R} = \frac{(2A)^2}{8R} = \frac{A^2}{2R}$$

$$\text{Fundamental linear power} = p_{\text{out-ss}} = \frac{V_{\text{pk-pk}}^2}{8R} = \frac{(2K_1A)^2}{8R} = \frac{K_1^2A^2}{2R}$$

$$\text{Fundamental nonlinear power} = p_{\text{out-LS}} = \frac{V_{\text{pk-pk}}^2}{8R} = \frac{(2A_{\text{fund}})^2}{8R}$$

$$= \frac{(K_1A + 3/4K_3A^3 + 5/8K_5A^5 + \cdots)^2}{2R}$$

The power gains at the fundamental are then given by

$$g = \frac{p_{\text{out}}}{p_{\text{in}}} \Rightarrow G = 10 \cdot \text{LOG}\left(\frac{p_{\text{out}}}{p_{\text{in}}}\right) = P_{\text{out}} - P_{\text{in}}$$

$$\text{Linear power gain} \Rightarrow g_{\text{ss}} = \frac{p_{\text{out-ss}}}{p_{\text{in}}} = \frac{K_1^2A^2}{2R} \cdot \frac{2R}{A^2} = K_1^2$$

$$\text{Nonlinear power gain} \Rightarrow g_{\text{LS}} = \frac{p_{\text{out-LS}}}{p_{\text{in}}} = \frac{(K_1A + 3/4K_3A^3 + 5/8K_5A^5 + \cdots)^2}{2R} \cdot \frac{2R}{A^2}$$

$$= \left(K_1 + \frac{3}{4}K_3A^2 + \frac{5}{8}K_5A^4 + \cdots\right)^2$$

All of the higher-order terms in the equation for g_{LS} above represent compression (compare to g_{ss} above). The input power for the single-tone 1 dB compression point is then given by

$$G_{1\text{ dBOCP}} = G_{\text{ss}} - 1\text{ dB} \Leftrightarrow g_{1\text{ dBOCP}} = g_{\text{ss}} \cdot 10^{-1/10} = K_1^2 \cdot 10^{-1/10} = (K_1 \cdot 10^{-1/20})^2$$

Setting the large signal gain equal to the gain at the 1 dBOCP results in

$$g_{\text{LS}} = g_{1\text{ dBOCP}} \Rightarrow \left(K_1 + \tfrac{3}{4}K_3A^2 + \tfrac{5}{8}K_5A^4 + \cdots\right)^2 = (K_1 \cdot 10^{-1/20})^2$$

Assuming higher-order terms have little effect on the end result, this can be simplified to the following:

$$K_1 + \tfrac{3}{4}K_3A^2 = K_1 \cdot 10^{-1/20}$$

Solving for the input amplitude and power at the 1 dB compression point results in

$$A_{1\text{ dBOCP}} = \sqrt{\frac{4K_1(10^{-1/20} - 1)}{3K_3}}$$

$$\therefore\ P_{1\text{ dBICP}} = 10 \cdot \text{Log}\left(\frac{V_{\text{pk-pk}}^2}{8R}\right) = 10 \cdot \text{Log}\left(\frac{A_{1\text{ dBOCP}}^2}{2R}\right) = 10 \cdot \text{Log}\left(\frac{2K_1(10^{-1/20} - 1)}{3K_3R}\right)$$

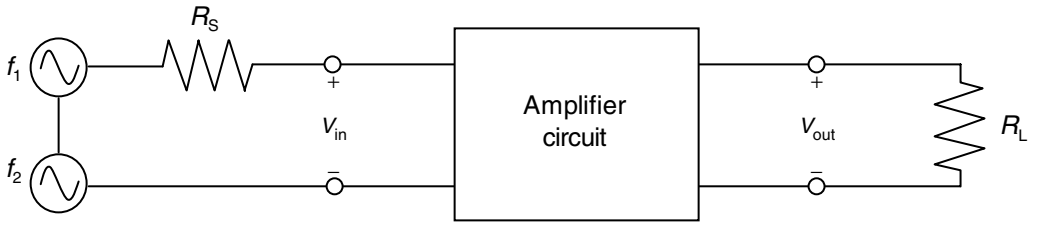


FIGURE 19.5 Two tone excitation.

The output power at the 1 dB compression point for a single-tone input is now easily found:

$$P_{1\,\mathrm{dBOCP}} = P_{1\,\mathrm{dBOCP}} + G_{ss} - 1\ \mathrm{dB}$$

$$\therefore\ p_{1\,\mathrm{dBOCP}} = p_{1\,\mathrm{dBICP}} \cdot g_{ss} \cdot 10^{-1/10} = \frac{2K_1(10^{-1/20} - 1)}{3K_3R} \cdot K_1^2 \cdot 10^{-1/10} = \frac{2K_1^3(10^{-1/20} - 1) \cdot 10^{-1/10}}{3K_3R}$$

$$\therefore\ P_{1\,\mathrm{dBOCP}} = 10 \cdot \mathrm{LOG}\left(\frac{2K_1^3(10^{-1/20} - 1) \cdot 10^{-1/10}}{3K_3R}\right)$$

This intermediate result will be combined with the following two-tone results. For the two-tone analysis, two different tones, each with the same amplitude as used for the single-tone input amplitude above, are applied to the amplifier (Figure 19.5).

Let $V_{in} = A\cos(\omega_1 t) + A\cos(\omega_2 t) = A\cos\theta_1 + A\cos\theta_2$, thus $A = V_{pk}$ and $2A = V_{pk-pk}$

Linear circuit $\Rightarrow V_o = K_1 V_{in} = K_1 A\cos\theta_1 + K_1 A\cos\theta_2$

Nonlinear circuit $\Rightarrow V_o = K_o + K_1 V_{in} + K_2 V_{in}^2 + \cdots = \sum_{n=0}^{\infty} K_n V_{in}^n$

Continuing with the derivation using the nonlinear circuit gives

$$V_o = K_o + K_1 A(\cos\theta_1 + \cos\theta_2) + K_2 A^2(\cos\theta_1 + \cos\theta_2)^2 + \cdots = \sum_{n=0}^{\infty} K_n A^n(\cos\theta_1 + \cos\theta_2)^n$$

The following expansion of the equation for V_o results in harmonics (multiples of θ_n) and mixing or IM products (sums and differences of the two input tones and their harmonics). The order of the IMs is determined by simply adding together the harmonic number of the sum or difference product (i.e., $\cos(3\theta_1 - 4\theta_2) \Rightarrow$ seventh-order IM). Applying trigonometric identities, then doing the math and regrouping results in

$$\begin{aligned} V_o = {} & K_o + K_2 A^2 \quad \mathrm{DC} \\ & + \cos\theta_1 \left(K_1 A + \frac{9}{4} K_3 A^3\right) \quad \mathrm{Fundamentals} \\ & + \cos\theta_2 \left(K_1 A + \frac{9}{4} K_3 A^3\right) \end{aligned}$$

$$
\begin{aligned}
&+ \cos 2\theta_1 \left(\frac{1}{2}K_2A^2\right) \quad \text{Second Harmonics}\\
&+ \cos 2\theta_2 \left(\frac{1}{2}K_2A^2\right)\\
&+ \cos 3\theta_1 \left(\frac{1}{4}K_3A^3\right) \quad \text{Third Harmonics}\\
&+ \cos 3\theta_2 \left(\frac{1}{4}K_3A^3\right)\\
&+ \cos(\theta_1 - \theta_2)(K_2A^2) \quad \text{Second-order IMs}\\
&+ \cos(\theta_1 + \theta_2)(K_2A^2)\\
&+ \cos(\theta_1 - 2\theta_2) \left(\frac{3}{4}K_3A^3\right) \quad \text{Third-order IMs}\\
&+ \cos(\theta_1 + 2\theta_2) \left(\frac{3}{4}K_3A^3\right)\\
&+ \cos(2\theta_1 - \theta_2) \left(\frac{3}{4}K_3A^3\right)\\
&+ \cos(2\theta_1 + \theta_2) \left(\frac{3}{4}K_3A^3\right)\\
&+ \cdots \quad \text{Higher order terms}
\end{aligned}
$$

Note that there are four different third-order IMs in the equation above, all at different frequencies. The power that is being derived is the power in any one of the four, not the sum. Therefore, the output power for one of the third-order IM products is given by

$$
\begin{aligned}
p_{\text{out}-3\text{rdIM}} &= \frac{V_{\text{pk}-\text{pk}}^2}{8R} = \frac{(3/4K_3A^3)^2}{2R}\\
\therefore P_{\text{out}-3\text{rdIM}} &= 10 \cdot \text{Log}\left(\frac{(3/4K_3A^3)^2}{2R}\right)
\end{aligned}
$$

The third-order intercept point requires by definition that the extrapolated small signal output power of the single tone must equal the output power of the third-order IM products:

$$
p_{\text{out}-\text{ss}} = p_{\text{out}-3\text{rdIM}} \Rightarrow \frac{K_1^2A_{3\text{IIP}}^2}{2R} = \frac{\left(3/4K_3A_{3\text{IIP}}^3\right)^2}{2R} \Rightarrow K_1A_{3\text{IIP}} = \frac{3}{4}K_3A_{3\text{IIP}}^3 \Rightarrow A_{3\text{IIP}} = \sqrt{\frac{4K_1}{3K_3}}
$$

which is the amplitude at the input to the two-tone circuit when at the 3IIP. It is now an easy matter to find the output powers:

$$
\therefore\ P_{3\text{IIP}} = 10 \cdot \text{Log}\left(\frac{V_{\text{pk}-\text{pk}}^2}{8R}\right) = 10 \cdot \text{Log}\left(\frac{4K_1}{3K_3} \cdot \frac{1}{2R}\right) = 10 \cdot \text{Log}\left(\frac{2K_1}{3K_3R}\right)
$$

The output power at the third-order intercept point for the two-tone input is now easily found:

$$P_{3\text{OIP}} = P_{3\text{IIP}} + G_{\text{ss}} \Rightarrow p_{3\text{OIP}} = p_{3\text{IIP}} \cdot g_{\text{ss}} \qquad \text{where } g_{\text{ss}} = K_1^2$$

$$\therefore p_{3\text{OIP}} = \left(\frac{2K_1}{3K_3R}\right) K_1^2 = \frac{2K_1^3}{3K_3R}$$

$$\therefore p_{3\text{OIP}} = 10 \cdot \text{LOG}\left(\frac{2K_1^3}{3K_3R}\right)$$

The power ratio R_{31} in dB between $P_{3\text{OIP}}$ and $P_{1\text{ dBOCP}}$ can now be found:

$$R_{31} = P_{3\text{OIP}} - P_{1\text{ dBOCP}} \Rightarrow r_{31} = \frac{p_{3\text{OIP}}}{p_{1\text{ dBOCP}}}$$

$$\therefore r_{31} = \left(\frac{2K_1^3}{3K_3R}\right)\left(\frac{3K_3R}{2K_1^3(10^{-1/20}-1)\cdot 10^{-1/10}}\right) = \frac{10^{-1/10}}{10^{-1/20}-1} = -11.576$$

$$\therefore R_{31} = 10\,\text{Log}(r_{31}) = 10.64\text{ dB}$$

The result for the ratio (R_{31}) above is independent of the rate at which compression occurs. It is important to remember that this ratio applies to the difference in output powers between the two-tone 3OIP and the single tone 1 dBOCP for input signals that are all set to the same amplitude.

19.6 The Load Line Construct

The idealized load line concept is based on a graphical method of solving for the currents and voltages in a series connection of two devices, regardless of what type of devices they are. For example, they could be two resistors, a diode and a resistor, a diode and a transistor, or two transistors. There are two different load lines commonly associated with power amplifiers: one for DC and one for RF. The DC load line construct is discussed below and presented as a springboard to the RF load line construct. The derivation herein focuses on the RF load line, primarily as a visual aid for setting up the analysis.

For the single stage amplifiers being considered herein, the static or DC load line requires examination of only the simplified DC circuit (Figure 19.6). The DC load resistance (total resistance in the bias network) is connected in series with the transistor. The DC load resistor and the transistor have known I–V characteristics as shown below, which are plotted using the same vertical and horizontal scales. It is important to remember that the transistor I–V characteristic is a two-dimensional representation of a three-dimensional I–V characteristic in which the individual curves are simply the I_{ds} values when the three-dimensional representation is sliced at specific V_{gs} values, somewhat like a forest service topo-map.

Referring to Figure 19.6, the current from the supply through the resistor and through the transistor are all identical in magnitude (Kirchoff's Current Law). The voltage across the load resistor and across the output of the transistor must equal the supply voltage (Kirchoff's Voltage Law). The low side device (transistor connected to ground or low side) characteristic is reproduced without change, while the high side device (resistor connected to the supply or high side) characteristic is flipped horizontally. When the transistor is off, the node voltage between the transistor and load resistor is the supply voltage. Thus, superimposing the two curves over each other requires that the resistor characteristic at zero current be the supply voltage, resulting in a static or DC load line characteristic as shown in Figure 19.7. The transistor saturates and thus limits the current to I_{max}, which can not be exceeded, explaining why the load line ends at I_{max}.

The static load line graph can now be used to analyze the DC currents and voltages in the circuit. Applying a fixed gate-to-source voltage (V_{gs}) sets the transistor drain-to-source current (I_{ds}) as indicated

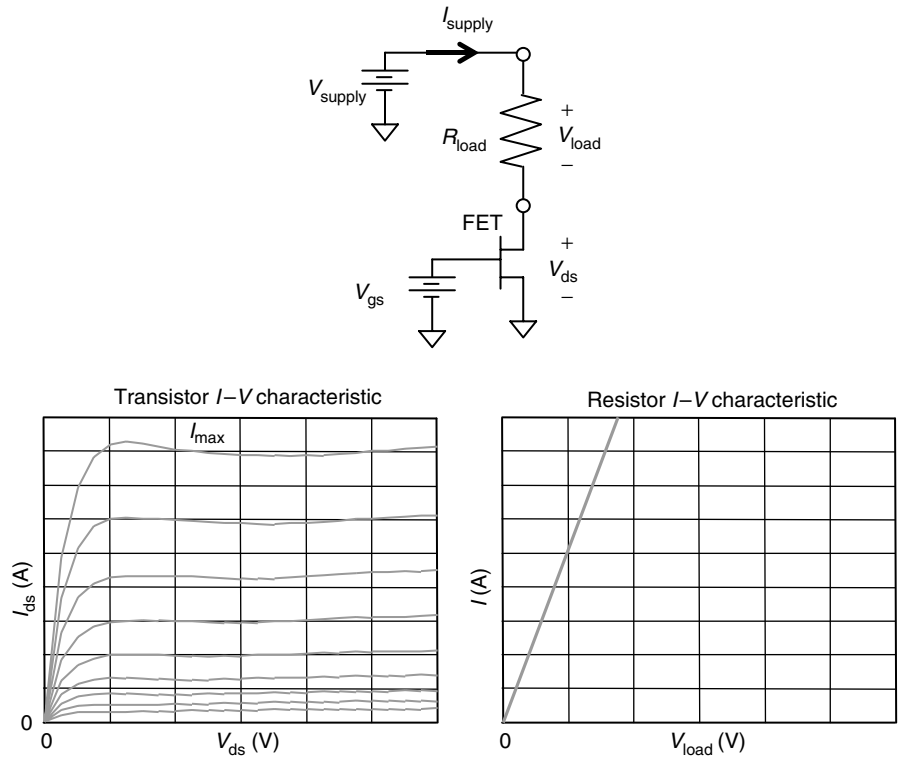


FIGURE 19.6 Device I–V characteristics.

by the transistor highlighted I–V curve on the static load line (Figure 19.7). The intersection of this curve with the resistor curve (black dot) is the sought after solution. The current through all elements can be read directly from the vertical axis, and are all equal to each other. The drain-to-source voltage (V_{ds}) across the transistor is read directly from the horizontal axis, while the voltage across the load resistor (V_{load}) is the difference between the supply voltage (V_{supply}) and the transistor voltage (V_{ds}), as indicated on the plot below. Any change in the gate supply voltage will change the currents and voltages.

The class A dynamic or RF load line construction follows the same procedure, except that the load resistance now represents the resistance of the RF circuit looking from the transistor into the RF output circuitry (Figure 19.8). In the presence of matching circuitry, this resistance has been transformed to a higher or lower level than the device output resistance. Note that the RF load resistance is not equal to the DC load resistance. The intersection of the resistance characteristic with zero current must also be modified, which is now set to V_{max}, the sum or the maximum RF peak-to-peak voltage ($V_{pk–pk\text{-}max}$) and Von of the device. The supply voltage for the class A amplifier is mid-way between V_{max} and V_{on}, and the associated output current is $^1/_2$ of I_{on}. These two known points thus determine the desired load line geometry. Once again, I_{max} limits the circuit current and thus truncates the RF load line at I_{max}.

The class A RF signal moves up and down the load line with each RF cycle, but is always centered on the bias point. At saturation, the RF signal cycles the full length of the RF load line. The instantaneous position on the load line gives the sought after solution at that instant in time. Finally, to reiterate, the load line does not represent the transistor output resistance, but rather the in-band resistance presented to the transistor. Unlike the input impedance to the transistor, the load line does not change with changes in operating point or signal level since the output matching circuits are fixed. In order to increase the output power, either the voltage swing must be increased (requiring a higher breakdown voltage) and/or the current swing must be increased (requiring a larger device).

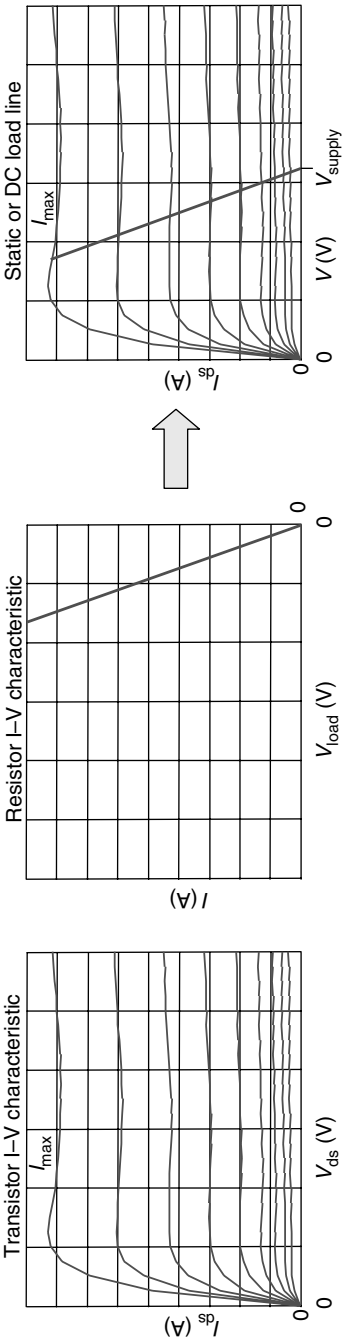


FIGURE 19.7 Static or DC load line.

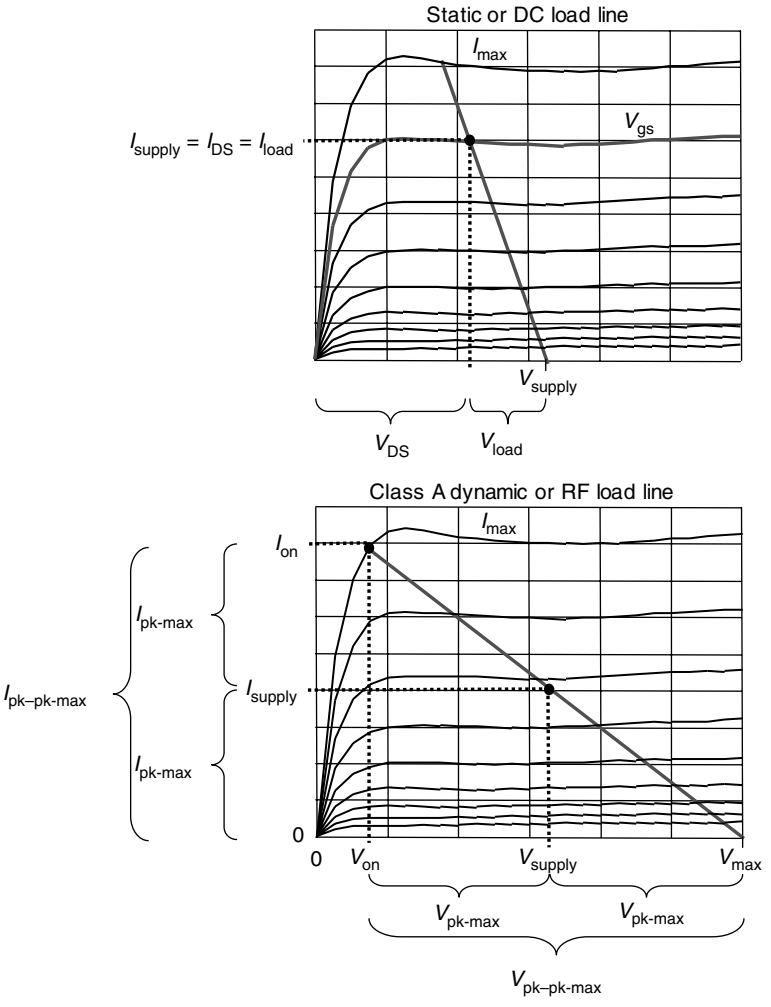


FIGURE 19.8 DC and RF load lines.

19.7 Saturated Output Power (P_{sat})

The maximum or saturated output power that can be generated occurs when the product of the peak-to-peak voltage and peak-to-peak current is maximized for a given device or I–V characteristic and load line. In order to accomplish this, the load line should push up to the knee of the I_{max} curve, where the output power will be maximized and reasonable flat, varying only a few tenths of a dB over a wide range of load line values (Figure 19.9).

Typical device I–V characteristics and load lines are shown in Figure 19.10. R_{on} is the lowest resistance that the output of the device can be driven to at the top end of the load line. It is determined from the linear region of the transistor I–V characteristic, and is given by V_{on}/I_{on}. Depending on the device, R_{on} might need to be replaced with the effective R_{on} (R_{on-eff}). For example, depletion mode GaAs MESFETs can operate with some enhancement, while enhancement mode Si MOSFETs cannot. This does not affect any of the derivations, but it does have a significant effect on the numerical results. Once the class A load line and bias point (V_{bias} and I_{bias}) have been selected, the maximum or saturated power that the device is capable of generating is fixed.

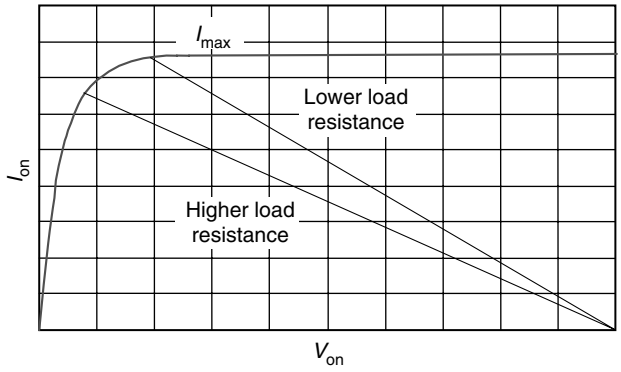


FIGURE 19.9 Selecting the load line.

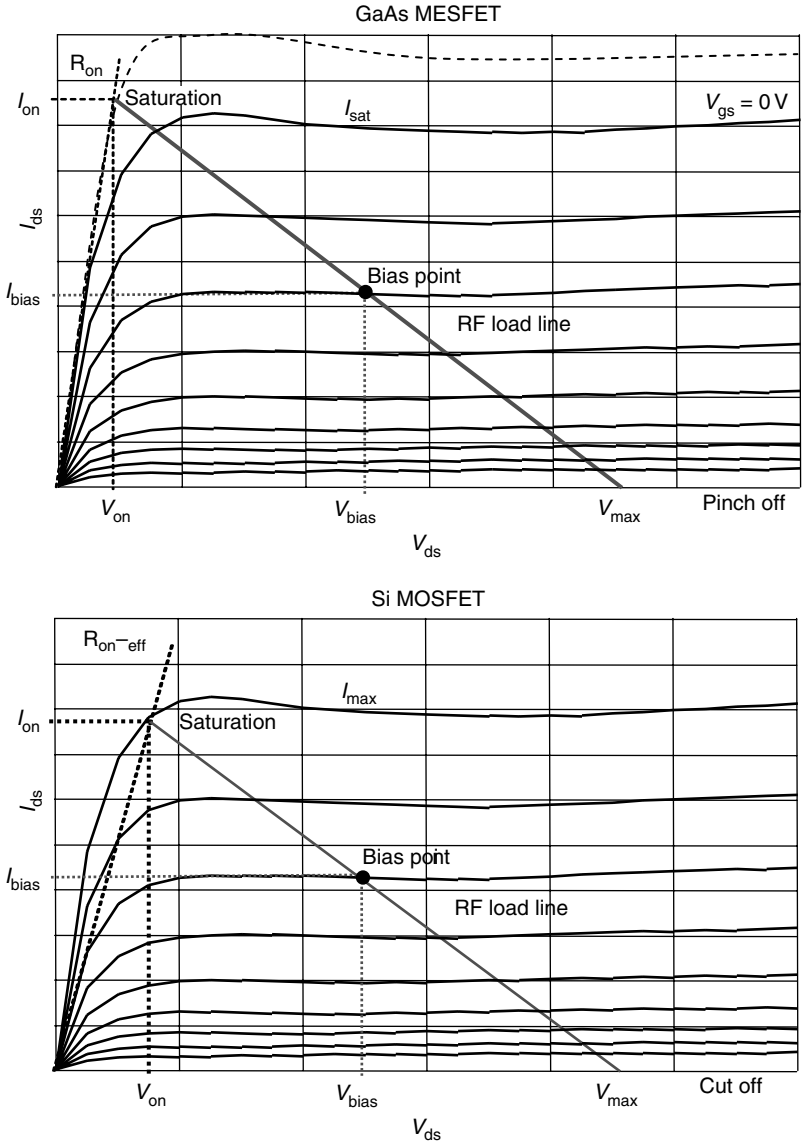


FIGURE 19.10 Typical device load lines.

For this derivation, the device is biased at $I_{on}/2$ for class A operation. At saturation or maximum output power, the output voltage swing is from V_{on} to V_{max} and the output current swing is from 0 to I_{on}.

$$I_{bias} \cong \frac{I_{on}}{2}$$

Referring to the load line depiction above, the maximum RF RMS voltage and current are then given by

$$V_{RMS} = \frac{V_{max\ pk-pk}}{2\sqrt{2}} = \frac{2\,(V_{bias} - V_{on})}{2\sqrt{2}} = \frac{V_{bias} - I_{on}R_{on}}{\sqrt{2}} \cong \frac{V_{bias} - 2I_{bias}R_{on}}{\sqrt{2}}$$
$$I_{RMS} = \frac{I_{max\ pk-pk}}{2\sqrt{2}} = \frac{I_{on}}{2\sqrt{2}} \cong \frac{2I_{bias}}{2\sqrt{2}} = \frac{I_{bias}}{\sqrt{2}}$$

The saturated RF power (p_{sat}, see the figure below) delivered to the load (i.e., across the load line) is then given by

$$p_{sat} = I_{RMS} \cdot V_{RMS} = \frac{I_{bias}}{\sqrt{2}} \cdot \frac{V_{bias} - 2I_{bias}R_{on}}{\sqrt{2}} = \frac{I_{bias} \cdot V_{bias} - 2I_{bias}^2 \cdot R_{on}}{2}$$

Converting to dBm and accounting for the loss in the output match (L_{AR}) results in the maximum deliverable saturated power (P_{sat}):

$$P_{sat} = 10 \cdot \mathrm{LOG}(p_{sat}) - L_{AR-out} + 30\,\mathrm{dB}$$
$$\therefore P_{sat} = 10 \cdot \mathrm{LOG}\left(\frac{I_{bias} \cdot V_{bias} - 2I_{bias}^2 \cdot R_{on}}{2}\right) - L_{AR-out} + 30\,\mathrm{dB}$$

This expression for the output power at saturation takes into account the major losses associated with a class A amplifier, including R_{on} and the output circuit losses. By setting R_{on} to zero, idealized device performance can be examined. By setting L_{AR} to zero, actual device level performance can be evaluated.

19.8 Breakdown Voltage

The device output reverse breakdown voltage (V_{BR}) is a very important parameter. The required breakdown voltage is a function of the output bias voltage, the RF voltage swing, and the operational requirement for ruggedness, which is operation into a substantial mismatch (usually a 10:1 VSWR) that is presented looking back into the power amplifier output at all possible phases (0–360°). Also, a significant factor is the class operation utilized in the amplifier. In order to market a packaged RF power device, it is necessary to determine the worst case and design accordingly. It is essential to realize that the active device voltage breakdown should be lower than the packaged device dielectric breakdown voltage. If the package dielectric breaks down first, permanent damage is incurred and the part is usually damaged beyond use. Thus, there is a voltage window that device breakdown should be confined to.

The highest voltages will be generated at the drain under severe mismatch conditions using "class A" bias schemes. The voltage reflection coefficient and return loss (RL) are given by

$$\rho_v = \frac{V_{SWR} - 1}{V_{SWR} + 1} \qquad RL_{dB} = -20\,\mathrm{Log}(\rho_v)$$

The loss in attenuation ratio for the signal is the sum of the output matching circuit loss and the RL as the signal travels out to the mismatched load and then back again to the device

$$L_{AR} = L_{AR-match\_incd} + L_{reft\_mismatch} + L_{AR-match-reft} = 2 \cdot L_{AR-match} + RL_{dB} \quad \text{in dB}$$

Converting to the attenuation ratio in voltage form gives

$$l_{AR} = 10^{-L_{AR}/20}$$

The minimum required breakdown voltage for "class A" operation is given as

$$V_{BR-min} = V_{quiescent} + \frac{V_{swing}}{2} + V_{reft}$$

Substitutions give the voltage at maximum swing on the I–V characteristic, and the added voltage due to a mismatch.

$$\therefore V_{BR-min} = V_{Bias} + \left(V_{Bias} + \frac{V_{on}}{2}\right) + l_{AR}\left(V_{Bias} + \frac{V_{on}}{2}\right) = V_{Bias}(2 + l_{AR}) + \frac{V_{on}}{2}(1 + l_{AR})$$

If there are no losses, then $V_{on} = 0$ V and $L_{AR} = 0$, or $l_{AR} = 1$. Substituting these into the equation above results in the following:

$$V_{BR-min} = 3V_{Bias}$$

Since the dielectric breakdown voltage must be understood, some discussion is warranted. Dielectric breakdown voltage is a measure of an insulator's ability to withstand a high electric field without arcing. Gauss's Law governs the formation of the E-fields. Charge likes to concentrate linearly on the least curved surfaces of a conductor, but the associated field strength is strongest, increasing quadratically, on the most sharply curved conductor surfaces. Therefore, sharp points on a charged electrical conductor have extremely strong fields. When an insulator's critical electric field is exceeded, microscopic conduction paths grow in microseconds, exhibiting intense ionization. Therefore, conductors with sharp points and bends should be avoided where high E-fields are expected.

The dielectric breakdown voltage is not a constant characteristic of a material, rather it is a statistical characteristic. Therefore a number of material samples should be measured to determine dielectric breakdown voltage, which is highly dependant the materials, geometries, applied waveforms, ambient conditions, and also the highly chaotic "initial conditions." Even the best insulators, such as diamond and quartz, arc when subjected to a high enough E-field. Insulators are only insulating under relatively weak E-fields, and tend to cross over from that weak field regime to a conducting state catastrophically.

For many situations, the dielectric breakdown of air is an important factor. The dielectric strength of air is approximately 3 Volts/micron. Its exact value varies with the shape and size of the electrodes, especially curved or pointed conductors, and ambient air pressure.

19.9 Compression

One of the fundamental characteristics of a class A amplifier is power compression. As the input power is increased, the fundamental output signal will begin to clip, eventually limiting at P_{sat} (Figure 19.11). This assumes a perfect match at both the input and output for all power levels under consideration (this is a physical impossibility, but easily conceptualized). Under this condition, P_{sat} is maintained for all higher input powers (excluding breakdown).

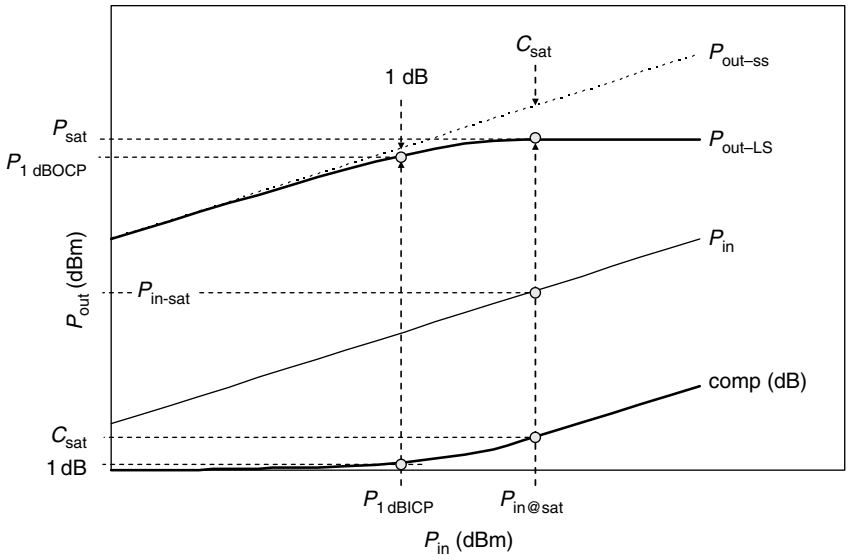


FIGURE 19.11 RF power compression.

The compression curve (Comp(dB)) is defined as the difference between the projected small signal output power ($P_{\text{out}-\text{ss}}$) and the actual large signal fundamental output power ($P_{\text{out}-\text{LS}}$):

$$\text{Comp(dB)} = P_{\text{out}-\text{ss}} - P_{\text{out}-\text{LS}} \quad \text{or} \quad \text{comp} = \frac{p_{\text{out}-\text{ss}}}{p_{\text{out}-\text{LS}}}$$

The definition for P_{sat} and C_{sat} are then given as occurring at the input power for which the compression slope (as given above) is equal to 1, which occurs at $P_{\text{in@sat}}$.

$$\frac{\text{dComp(dB)}}{\text{d}P_{\text{in}}} = 1$$

The large signal fundamental output power ($P_{\text{out}-\text{LS}}$) is easily found from simple geometry as follows:

$$\therefore P_{\text{out}-\text{LS}} = P_{\text{in}} + G_{\text{ss}} - \text{Comp(dB)}$$

Or equivalently,

$$p_{\text{out}-\text{LS}} = \frac{p_{\text{in}} \cdot g_{\text{ss}}}{\text{comp}}$$

The saturated output power at the onset of saturation (P_{sat}) is easily found from simple geometry as follows:

$$\therefore P_{\text{sat}} = P_{\text{in@sat}} + G_{\text{ss}} - C_{\text{sat}}$$

Or equivalently,

$$p_{\text{sat}} = \frac{p_{\text{in@sat}} \cdot g_{\text{ss}}}{c_{\text{sat}}}$$

At or above the saturated input power $P_{\text{in@sat}}$, the compression is found by a line with a 1:1 slope that intersects with the compression curve at the saturation point $P_{\text{in@sat}}$. From simple geometry, this is given by

$$\therefore \text{Comp(dB)} = C_{\text{sat}} + P_{\text{in}} - P_{\text{in@sat}} \quad \text{for } P_{\text{in}} \geq P_{\text{in@sat}}$$

From the expression for the large signal output power ($p_{\text{out-LS}}$) above,

$$\text{comp} = \frac{p_{\text{in}} \cdot g_{\text{ss}}}{p_{\text{out-LS}}}$$

Substituting the expressions from Section 19.5 above gives

$$\text{comp} = \frac{K_1^2 \cdot A^2}{\left(K_1 A + 3/4 K_3 A^3 + 5/8 K_5 A^5 + \cdots\right)^2}$$

Assuming higher-order terms are insignificant, results in

$$\text{comp} \cong \frac{K_1^2 \cdot A^2}{\left(K_1 A + 3/4 K_3 A^3\right)^2} = \frac{K_1^2}{\left(K_1 + 3/4 K_3 A^2\right)^2}$$

Solving for the amplitude gives

$$\left(K_1 + 3/4 K_3 A^2\right)^2 = \frac{K_1}{\text{comp}}$$

$$\frac{3}{4} K_3 A^2 = \frac{K_1}{\sqrt{\text{comp}}} - K_1 = K_1 \cdot \left(\frac{1}{\sqrt{\text{comp}}} - 1\right)$$

$$\therefore A^2 = \frac{4K_1}{3K_3} \cdot \left(\frac{1}{\sqrt{\text{comp}}} - 1\right)$$

Substituting this into the equation for $p_{\text{out-LS}}$ above,

$$p_{\text{out-LS}} = \frac{A^2 K_1^2}{2R \cdot \text{comp}} = \frac{4K_1}{3K_3} \cdot \left(\frac{1}{\sqrt{\text{comp}}} - 1\right) \cdot \frac{K_1^2}{2R \cdot \text{comp}}$$

$$\therefore p_{\text{out-LS}} = \frac{2K_1^3}{3K_3 R \cdot \text{comp}} \cdot \left(\frac{1}{\sqrt{\text{comp}}} - 1\right)$$

At saturation the boundary condition is known (comp $= c_{\text{sat}}$). Applying this boundary condition to the result just achieved results in

$$\therefore p_{\text{sat}} = \frac{2K_1^3}{3K_3 R \cdot c_{\text{sat}}} \cdot \left(\frac{1}{\sqrt{c_{\text{sat}}}} - 1\right)$$

$$\therefore \frac{2K_1^3}{3K_3 R} = p_{\text{sat}} c_{\text{sat}} \left(\frac{1}{\sqrt{c_{\text{sat}}}} - 1\right)^{-1}$$

Substituting this into the intermediate expression above gives

$$\therefore p_{\text{out-LS}} = \frac{p_{\text{sat}} c_{\text{sat}}}{\text{comp}} \left(\frac{1}{\sqrt{\text{comp}}} - 1\right) \cdot \left(\frac{1}{\sqrt{c_{\text{sat}}}} - 1\right)^{-1}$$

However, the output powers in terms of the input powers are given above. Substituting these into the expression results in

$$\therefore p_{\text{out-LS}} = \frac{p_{\text{in}} \cdot g_{\text{ss}}}{\text{comp}} = \frac{p_{\text{in@sat}} \cdot g_{\text{ss}}}{c_{\text{sat}}} \cdot \frac{c_{\text{sat}}}{\text{comp}} \left(\frac{1}{\sqrt{\text{comp}}} - 1 \right) \cdot \left(\frac{1}{\sqrt{c_{\text{sat}}}} - 1 \right)^{-1}$$

Simplifying gives

$$\therefore p_{\text{in}} = p_{\text{in@sat}} \cdot \left(\frac{1}{\sqrt{\text{comp}}} - 1 \right) \cdot \left(\frac{1}{\sqrt{c_{\text{sat}}}} - 1 \right)^{-1}$$

Solving for compression results in

$$\text{comp} = \left(\frac{p_{\text{in@sat}}}{\left(p_{\text{in}} / \sqrt{c_{\text{sat}}} \right) - p_{\text{in}} + p_{\text{in@sat}}} \right)^2$$

The final result is then given by the following expression, where the quantities inside the brackets must be converted from dB:

$$\therefore \text{Comp(dB)} = 20 \cdot \text{LOG} \left(\frac{p_{\text{in@sat}}}{\left(p_{\text{in}} / \sqrt{c_{\text{sat}}} \right) - p_{\text{in}} + p_{\text{in@sat}}} \right)$$

This equation describes the class A amplifier compression characteristic in terms of two saturation parameters that are easily determined or estimated ($p_{\text{in@sat}}$ and c_{sat}), and also the input power (p_{in}). Thus, the entire saturation characteristic versus input power is known! All large signal gain and linearity amplifier characteristics are tied to this one important equation, and as well the large signal DC performance. Dropping back to the equation for p_{in} above, the powers at the 1 dB compression can easily be found:

$$\therefore p_{1\,\text{dBICP}} = p_{\text{in@sat}} \cdot \left(\frac{1}{10^{1/20}} - 1 \right) \cdot \left(\frac{1}{\sqrt{c_{\text{sat}}}} - 1 \right)^{-1}$$

$$P_{1\,\text{dBOCP}} = P_{1\,\text{dBICP}} + G_{\text{ss}} - 1\,\text{dB}$$

Again, these simple equations, dependant only on known or easily determined parameters ($p_{\text{in@sat}}$, c_{sat}, and G_{ss}) allow calculation of the 1 dB compression point! The linearity can be estimated by combining this with the results for the ratio to the third-order intercept point (R_{31}) in Section 19.5. The 1 dB compression point is a commonly used industry performance reference power used to compare amplifier performance and for specifying operational parameters such as the back-off (BO) power required to meet linearity requirements for various modulation schemes.

19.10 Output Bias Current

From the introduction above, small signal conditions indicate linear operation, whereas large signal includes the entire range of operation, both linear (or small signal) and nonlinear. As the RF signal amplitude increases, the output will progress from small signal uncompressed conditions to large signal compressed conditions, and the DC component generated by signal compression will increase the bias current from the class A setting of $I_{\text{on}}/2$, eventually limiting out at the saturated current of the device.

In the derivation for compression above, it was assumed that higher-order terms did not significantly contribute to the solution, allowing the compression to be determined with sufficient accuracy for estimating class A performance. Continuing with this line of reasoning, the power difference between the

extrapolated small signal conditions ($p_{\text{out}-\text{ss}}$) and the large signal conditions ($p_{\text{out}-\text{LS}}$) must be due to a change in the DC bias conditions. Thus, compression is a direct ratio of the extrapolated small signal bias power ($p_{\text{bias}-\text{ss}}$) and large signal bias power ($p_{\text{bias}-\text{LS}}$):

$$\therefore \text{Comp} \cong \frac{p_{\text{bias}-\text{ss}}}{p_{\text{bias}-\text{LS}}}$$

Since the class A applied bias voltage (V_{bias}) does not change under any conditions, the DC current must be changing as a function of compression. The small signal bias current (I_{ss}) is the DC current in the output side of the amplifier when operated without any RF input up to RF signals well below P_{sat} (i.e., 15–20 dB or more below P_{sat} and thus $I_{\text{ss}} = I_{\text{on}}/2$). The small signal DC current I_{ss} is a single fixed value for any given amplifier. Large signal current in the output side of the amplifier (I_{LS}) is the current over the full range of power, from no RF signal to fully compressed conditions. Under small signal conditions, $I_{\text{ss}} = I_{\text{LS}}$, but as the signal compresses, I_{LS} will grow as a function of the compression:

$$p_{\text{bias}-\text{ss}} = I_{\text{ss}} V_{\text{bias}} \qquad p_{\text{bias}-\text{LS}} = I_{\text{LS}} V_{\text{bias}}$$

$$\therefore \text{Comp(dB)} = P_{\text{bias}-\text{ss}} - P_{\text{bias}-\text{LS}} = 10\,\text{LOG}\left(\frac{p_{\text{bias}-\text{ss}}}{p_{\text{bias}-\text{LS}}}\right) = 10\,\text{LOG}\left(\frac{I_{\text{ss}} V_{\text{bias}}}{I_{\text{LS}} V_{\text{bias}}}\right) = 10\,\text{LOG}\left(\frac{I_{\text{ss}}}{I_{\text{LS}}}\right)$$

Solving for the large signal RMS current gives the following result:

$$I_{\text{LS}} = I_{\text{ss}} \cdot 10^{\text{Comp(dB)}/10}$$

Thus the large signal current is a simple function of the compression characteristic and the small signal bias current.

19.11 Additional Performance Parameters

The following equations are readily discerned for the given conditions, therefore no further derivations are presented.

$$G_{\text{ss}} = G_{\text{device}-\text{ss}} - L_{\text{AR}-\text{in}} - L_{\text{AR}-\text{in}} \quad \text{in dB}$$

$$G_{\text{LS}} = G_{\text{ss}} - \text{Comp(dB)} \quad \text{in dB}$$

$$P_{\text{in}} = P_{\text{out}} - G_{\text{LS}} \quad \text{in dBm}$$

$$P_{\text{diss}} = V_{\text{bias}} \cdot I_{\text{LS}} + p_{\text{in}} - p_{\text{out}} \quad \text{in W}$$

$$P_{3\text{OIP}} = P_{1\,\text{dBOCP}} + 10.64\,\text{dB} \quad \text{in dBm}$$

$$R_{\text{out}-\text{LS}} = \frac{V_{\text{bias}}}{I_{\text{LS}}} \quad \text{in } \Omega$$

19.12 Efficiency

The following efficiency calculations are all expressed in percent (%). They do not include any input DC bias in the calculations, but can be easily modified to account for any bias considerations specific to an application. As the RF output power approaches zero, so do each of the three efficiencies. As the gain gets very large, all three efficiencies converge to the same values. From the introduction above, remember that large signal current (I_{LS}) includes the entire range of operation, both linear (or small signal) and nonlinear.

The power added efficiency (PAE) is given by the ratio of the difference in RF output and RF input power to the DC power required to generate the output power. PAE is the most commonly used, as it is an excellent gauge of overall efficiency performance.

$$\text{Power added efficiency} \Rightarrow \eta_a = 100\left(\frac{p_{\text{out}} - p_{\text{in}}}{V_{\text{bias}} \cdot I_{\text{LS}}}\right)$$

The output efficiency is simply the fundamental RF output power divided by the DC power required to generate the output power. Output efficiency (collector or drain efficiency) only addresses the output switching efficiency of the amplifier, which can be very good even if there is no gain. If only the switching efficiency is presented in a set of data, it should be an automatic flag to check PAE and power gain.

$$\text{Output efficiency} \Rightarrow \eta_{\text{out}} = 100\left(\frac{p_{\text{out}}}{V_{\text{bias}} \cdot I_{\text{LS}}}\right)$$

True efficiency is almost never used. It is a ratio of the RF output power to the total input power (RF and DC) into the amplifier.

$$\text{True efficiency} \Rightarrow \eta_{\text{true}} = 100\left(\frac{p_{\text{out}}}{V_{\text{bias}} \cdot I_{\text{LS}} + p_{\text{in}}}\right)$$

All three of the major efficiencies deal with ratios of output power to input power from different perspectives. It should be noted that the three basic efficiencies are interrelated. Any two are sufficient to determine the third.

19.13 Output Reflection Coefficient

The following derivation gives the reflection coefficient versus power based on a known perfect match at a specified back-off (BO) power from P_{sat}. The term "large signal" covers all operating power levels. Thus, by specifying the back-off power at which a perfect match occurs, and sweeping the large signal through a range of power, the reflection coefficient, RL, mismatch loss (ML), and voltage standing wave ratio (VSWR) can be determined for all power levels. The specified back-off power is usually determined by user requirements.

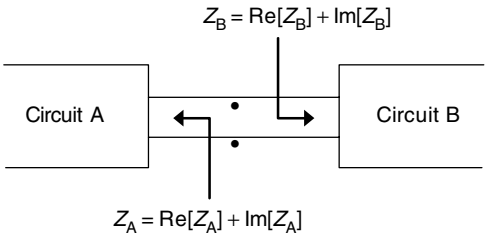


FIGURE 19.12 Circuit impedances.

The magnitude (ρ) of the reflection coefficient (Γ) is generically defined as the absolute (ABS) value of the difference ratioed to the sum of two impedances, where Z_A and Z_B are the two complex RF impedances looking from a central node in a system (Figure 19.12). Each complex impedance has orthogonal real (Re) and imaginary (Im) components that vary with frequency and power.

$$\rho = |\Gamma| = \text{ABS}\left[\frac{Z_A - Z_B}{Z_A + Z_B}\right] = \sqrt{\frac{\text{Re}|Z_A - Z_B|^2 + \text{Im}|Z_A - Z_B|^2}{\text{Re}|Z_A + Z_B|^2 + \text{Im}|Z_A + Z_B|^2}}$$

For the derivation herein, circuit B and hence Z_B is fixed by the matching circuitry. Circuit A and the associated Z_A looking into the transistor is a function of power. Since capacitance is not being considered in this derivation there is no imaginary component, thus simplifying the equation above:

$$\rho = |\Gamma| = \text{ABS}\left[\frac{Z_A - Z_B}{Z_A + Z_B}\right] = \sqrt{\frac{\text{Re}|Z_A - Z_B|^2}{\text{Re}|Z_A + Z_B|^2}}$$

Let Z_{BO} be the optimum matched device impedance (circuit A) at the specified back-off condition, formed by a shunt parallel RC circuit,

$$Z_{BO} = \text{Re}|Z_{BO}| = R_{BO} = \frac{R_{ss} \cdot I_{ss}}{I_{BO}}$$

Thus, circuit B is the conjugate of Z_{BO} (i.e., matched) that is designated by superscript asterisks ($*$). Since only real components of the impedances are being considered, there is no change.

$$Z_B = Z_{BO}^* = \text{Re}\left|Z_{BO}^*\right| = R_{BO} = \frac{R_{ss} \cdot I_{ss}}{I_{BO}}$$

Similarly, let Z_{LS} be the impedance (circuit A) modified by the operating conditions as the power is varied (i.e., impedance under large signal or LS conditions). Z_{LS} will exactly equal Z_{BO} at the specified back-off power, but will move away from that impedance for any other power level:

$$Z_A = Z_{LS} = Re|Z_{LS}| = R_{LS} = \frac{R_{ss} \cdot I_{ss}}{I_{ss}}$$
$$Z_A - Z_B = (R_{ss} \cdot I_{ss}) \cdot \left(\frac{1}{I_{LS}} - \frac{1}{I_{BO}}\right)$$
$$Z_A + Z_B = (R_{ss} \cdot I_{ss}) \cdot \left(\frac{1}{I_{LS}} + \frac{1}{I_{BO}}\right)$$

Substituting these into the reflection coefficient equation above gives the following:

$$\rho = |\Gamma| = \sqrt{\frac{(R_{ss} \cdot I_{ss})^2 \cdot \left(\frac{1}{I_{LS}} - \frac{1}{I_{BO}}\right)^2}{(R_{ss} \cdot I_{ss})^2 \cdot \left(\frac{1}{I_{LS}} + \frac{1}{I_{BO}}\right)^2}}$$
$$\rho = |\Gamma| = \sqrt{\frac{(I_{BO} - I_{LS})^2}{(I_{BO} + I_{LS})^2}} = \frac{I_{BO} - I_{LS}}{I_{BO} + I_{LS}}$$

The final result is obtained by generalizing for any values, which simply requires taking the absolute value:

$$\rho = |\Gamma| = \text{ABS}\left|\frac{I_{BO} - I_{LS}}{I_{BO} + I_{LS}}\right|$$

The RL, ML, and VSWR are then easily found from the magnitude of the reflection coefficient (ρ):

$$\text{RL} = -10 \cdot \text{LOG}(\rho^2) \quad \text{in dB}$$

$$\text{ML} = -10 \cdot \text{LOG}(1 - \rho^2) \quad \text{in dB}$$

$$\text{VSWR} = \left(\frac{1+\rho}{1-\rho}\right)$$

By specifying the back-off, and sweeping the large signal through a range of power, the reflection coefficient, RL, ML, and VSWR can be determined across that power range as long as a perfect match is assumed at the specified back-off power. Applying the mismatch results to the idealized "matched at all power level" power curves above, results in a good estimate of the actual output power that can be achieved for any back-off, even for a power sweep.

19.14 Gain Expansion

Gain expansion is the term applied to amplifiers in which the large signal gain is greater than or expanded above the expected large signal gain (as calculated above) over some region of input power between small signal conditions and P_{sat}, usually just above the 1 dB compression point (Figure 19.13). Gain expansion is given in dB, and is usually less than a few dB.

$$G_{\text{expansion}} = G_{\text{expanded LS}} - G_{\text{LS}}$$

There are several factors that can contribute to gain expansion, but the dominant cause for any appreciable gain expansion can be discerned directly from the device I–V characteristic curves. Active devices that exhibit higher gain at higher current will result in some form of gain expansion. As the applied signal increases, the device starts to compress and, as shown above, the DC current also increases, continuously rebiasing the device at higher levels. If the I–V characteristic for the device has higher gain at higher current, then the gain will also increase for signal levels below saturation but pushing into compression. The equations developed above do not include this undesirable effect, since well-designed amplifier devices will have constant gain over the vast majority of the I–V characteristic. However, it is easy to envision how they could be included, simply by making the gain a function of the bias current instead of a constant as used herein. Secondary gain expansion effects can be attributed to the power match, harmonic terminations, reflections, and reconversion, and even to the input match. As an example, the

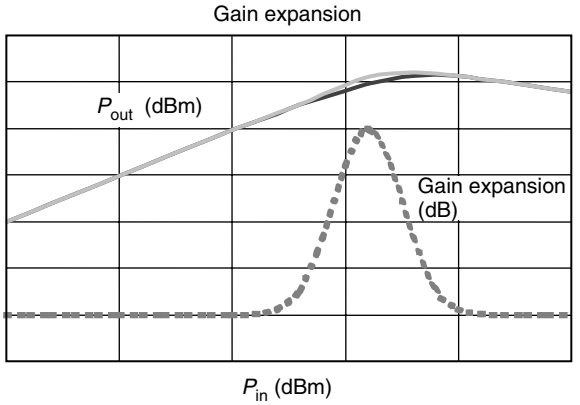


FIGURE 19.13 Gain expansion.

expected large signal output power and the measured large signal output power in the plot below are nearly identical over most of the input power range, with the measured power curve slightly higher in the compressed region. The gain expansion (dashed curve) is also shown using a much larger scale.

19.15 Excel Spreadsheet

An Excel spreadsheet was created incorporating all of the concepts above to create the following example. The parameter set in the first block is used to generate the summary block and all of the graphs shown below. All calculations are directly taken from the closed form equations developed herein with one exception; the input power under the specified "lossy back-off" condition requires an iterative solution to converge on the correct value. P_{out} is plotted in each of the four graphs (Figure 19.14) to provide a consistent reference.

Parameter Set

Amplifier parameters		
V_{bias} =	28.00	Volts
$I_{bias-ss}$ ($\sim I_{on}/2$) =	2.000	Amperes
R_{on} =	1.000	Ohms
$G_{total-ss}$ =	18.00	dB
Comp @ P_{sat} =	3.00	dB
Input L_{AR} =	1.20	dB
Output L_{AR} =	0.50	dB
Back-off from P_{sat} ($\geq$1dB) =	15.00	dB

Summary Block

Performance Summary Estimates	Lossless P_{sat}	Lossy P_{sat}	Lossy 1 dBOCP	Lossy Back-off	Units
P_{in} =	29.27	29.49	25.20	11.53	dBm
P_{in} =	0.846	0.890	0.331	0.014	Watts
P_{out} =	45.97	44.49	42.20	29.49	dBm
P_{out} =	39.551	28.137	16.605	0.890	Watts
$G_{total-Is}$ =	16.70	15.00	17.00	17.96	dB
PAE =	48.93	34.45	25.90	1.56	%
Drain efficiency =	50.00	35.57	26.43	1.58	%
True efficiency =	49.47	35.17	26.29	1.58	%
I_{bias-I_s} =	2.825	2.825	2.244	2.009	Amperes
$P_{dissipated}$ =	40.397	51.855	46.559	55.387	Watts
R_{out} @ DC =	9.91	9.91	12.48	13.93	Ohms
Compression =	3.00	3.00	1.00	0.041	dB
1-dB output comp pt =	42.48	42.20	42.20	42.20	dBm
3rd order OIP (sq law) =	53.12	52.84	52.84	52.84	dBm

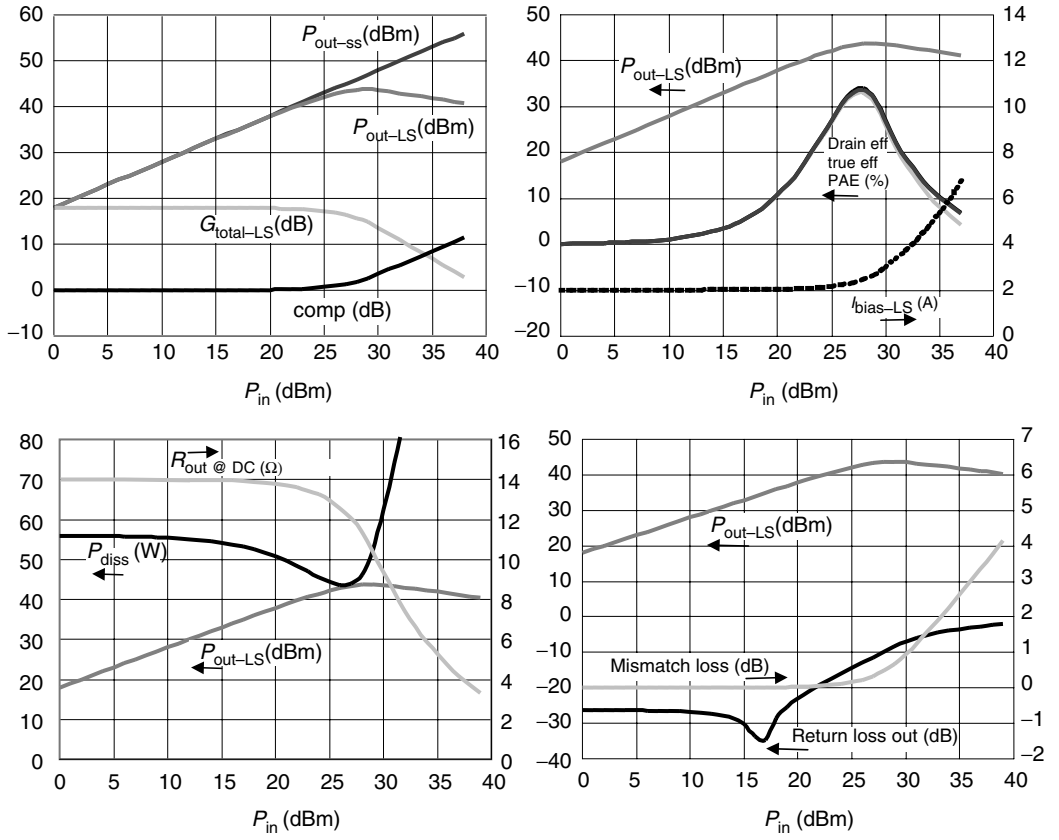


FIGURE 19.14 Performance versus power estimates.

20

High Power Amplifiers

Brent Irvine
Todd Heckleman
MKS Instruments, Inc.

Aaron Radomski
Harris RF Communications

20.1 Introduction

High power amplifiers (HPAs) are mostly found in the context of a larger system. Any time a large amount of radio frequency (RF) power is required in a system, there will be some sort of HPA found at the end of a signal generation and amplification chain. HPAs are found in a diverse number of applications, and include a large number of markets and purposes. A short list of some of the more popular applications are listed below:

- Cellular networks
- Two-way radio
- TV and radio broadcasting
- RADAR
- Medical systems (e.g., magnetic resonance imaging or MRI)
- Laser Pumps
- Semiconductor RF power supplies (i.e., plasma generators)
- Industrial heating

Because of the diversity of applications, it is difficult to categorize them by their applications absolutely. Thankfully, there is a great deal of similarity amongst all HPA designs as well as groupings of specific application types, making a discussion of design approaches and considerations fruitful.

The author arbitrarily defines the HPA into three broad categories: Linear, CW, and pulsed applications, detailed in Table 20.1.

In each category of application, the merits of the amplifier are measured slightly differently. The techniques used to maximize the measured performance are slightly different and sometimes mutually exclusive. For example, the approach used to supply collector or drain voltage and bias feed between a CW amplifier and a linear amplifier can be different enough that one cannot be used on the other without significant performance degradation or redesign.

In this chapter, the author will attempt to delineate the tradeoffs, commonalities, and differences when designing such an amplifier type for a number of applications as defined above. While there has been a

TABLE 20.1 Application Classifications for HPA's

Mode	Description
Linear	The amplifier is meant to amplify and pass a modulated signal of some sort, with minimal distortion
CW	The amplifier is meant to amplify a CW signal at a fixed power level or levels.
Pulsed	The amplifier is meant to amplify a momentary pulse, with a specific range or type of repetition rate (duty cycle) and pulse duration.

great deal of effort on analyzing generalized HPAs in the literature (see sources), this chapter will focus upon introducing the reader to a number of concepts and tradeoffs that are required for a more effective specific design. Unfortunately, this guide will only be able to take a step in that direction. It is a worthwhile step, and coupled with additional resources it is hoped that a PA designer will be able to launch into an effective design without all of the traditional pain of a rather steep learning curve.

20.2 Specifications

As with all engineering pursuits, the specification of the targeted application is the first consideration. As mentioned in the introduction, since there are a number of diverse end uses for an HPA, this step is especially important. Due to the similarities and differences between these types of applications, the author has arbitrarily divided them into universal specifications—the sorts of specifications that any HPA will have to have—and application specifications—the sorts of specifications that are determined by the end use of the HPA.

20.3 Universal Specifications

All HPA specifications will have a number of items that are universal.

Output Power: Amount of output power required in the application, as well as a detailed description of how it is to be measured. Obviously, this is one of the most important specifications, though for each application type, the way it is measured needs to be detailed. One popular definition used in many cases is the amount of CW power output at a specific gain compression point, such as Power at 1 dB of gain compression from the gain peak. It is not the only definition; other example methods of power specification are shown in Table 20.2.

Efficiency: Essentially, efficiency is a measure of the amount of power put out by a circuit divided by the amount of power supplied. In practice, it is measured in two ways: PAE and DC efficiency as shown in Figure 20.1. The main difference between the two methods is that the PAE includes both the power used to drive the circuit as well as all DC inputs to the circuit and the collector efficiency is where the efficiency of just the output circuit is considered (called collector after the bipolar transistor. Technically, a FET will have drain efficiency, this chapter will refer to both as η_c). It is interesting to note that as the amount

TABLE 20.2 Power Specification Examples

Amplifier Mode	Examples of Specifications
Linear	*P*1 dB, peak power, average power (under modulation type such as two tone or other signal)
Pulsed	Peak power at a particular compression under a specific duty cycle
CW	Saturated output power, *P*1 dB, or compression greater than 2 dB

Power added efficiency	DC efficiency
$\eta_{PAE} = \frac{P_{out}}{P_{DC}+P_{RFin}}$	$\eta_c = \frac{P_{out}}{P_{DC}}$
Note: If $P_{RFin} \ll P_{DC}$ then $\eta_{PAE} \approx \eta_c$	

FIGURE 20.1 Calculation of DC efficiency.

of gain increases the PAE approaches the collector efficiency asymptotically—in systems with high gains (e.g., 25 dB or more), or in efficiency is often specified without distinguishing between power-added or collector efficiency.

Efficiency is also, sometimes indirectly, specified as a gross power consumption figure, or the amount of current drawn off of a supplied input voltage, even if the HPA under consideration has to go through a conversion step in order to use the input. It is obvious (but sometimes overlooked) that losses due to any converter must be taken into account when creating the efficiency targets for individual circuits or stages in a HPA system, regardless of the specifications format.

Gain: Amount of gain at a specified power level or compression level, it is also sometimes indirectly specified as the minimum input signal required to reach a certain output power. The amount of gain specified will determine if more than one stage of gain is required. Due to stability as well as device constraints, it usually difficult to have more than 15–20 dB of gain with a single stage, and it may be advantageous to have more than one gain stage (using an input attenuator or a lower power input signal of some sort to trim the gain appropriately).

Frequency Range: The frequency of operation of the circuit (e.g., Frequency of operation = 1.93 − 1.99 GHz). This specification will determine the components used, the device used, and other factors—as in most other RF designs.

Inputs and Outputs: As simple as this may seem, this is fairly important in order to define how the HPA interfaces with the rest of the system. Will the amplifier be required to provide samples of RF input or output? Also, how will the voltage required to power the amplifier be supplied to the transistor stages? Will there be some sort of metrology required on the amplifier to report back to a control system (typically temperature, current and rail voltages are popular, though some very sophisticated and high-speed sensing is required in some applications). The RF output and input connectors have to be specified as well, and compared to the power requirements. This is sometimes overlooked, and sometimes specified by the system, though it is terribly important for a reliable and well-designed system.

Nature of the Load: Most HPAs will be specified into a nominal load impedance (usually 50 ohm). Sometimes the load into which a HPA must operate deviates from this. To this end, depending upon the application, frequency range, or other factors, a circulator or isolator is employed to prevent excessive load reflections from affecting the reliability or power delivery of a HPA. In applications where such a device is impractical, either technically (as in the case of a broad band amplifier) or economically (at $<$200 MHz), a power fold-back is sometimes employed as a control technique. One example of power de-rating under VSWR is shown in Figure 20.2, where the power is backed off uniformly with rising VSWR. Power back off can be achieved by reducing the primary supply voltage to the stage, decreasing the input drive circuitry, or a combination thereof. Depending upon the application and class of operation, the designer must choose the exact method balancing dissipation and high voltage and currents on the HPA circuitry. It is important to note that the de-rating is significantly greater than would be by simple reflection. When exposed to VSWRs greater than 1:1, even when backed off, the HPA will be subject to great stress—note in the figure that dissipation, peak voltage, and currents are all greatest into non 50 ohm impedances, even when backed off. Also, the gain and output power will be affected. If there is a specified amount of output power into a series of loads, usually the amplifier will have to be made capable of larger output power than the rating into 50 Ω to accommodate reasonable power delivery into a high VSWR.

Ambient Conditions: Ambient conditions are mainly defined as the range of temperatures the HPA is expected to operate at, either at full specification or at a specific de-rating condition. Typically, this will be an ambient air temperature range, and may also include mounting temperature if the HPA is to be mounted to an external surface. Also, in the case of liquid cooling, this will also have to include coolant temperature range. Humidity and altitude need to also be specified or considered—especially

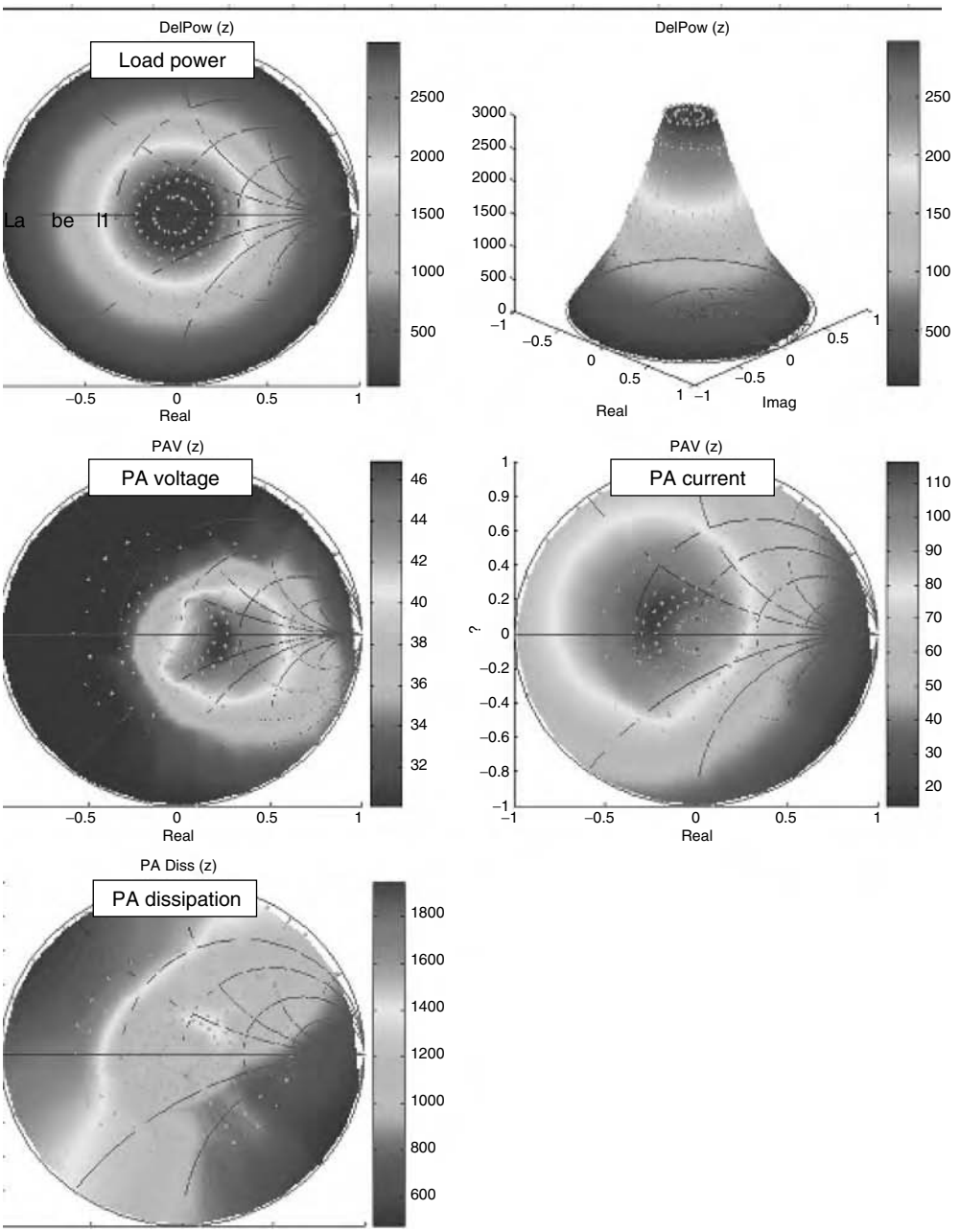


FIGURE 20.2 Example of VSWR foldback and stress. (Courtesy MKS Instruments. Used with permission.)

since excessive humidity may include a risk of condensation depending upon the cooling method, and high voltages at higher impedance points can carry the risk of an arc if not carefully approached.

Isolation: A parameter related to input impedance is signal isolation from either system power supply and ground or RF output signal. Signal isolation specifications are typically used for safety of the user, input circuit protection, or to prevent unwanted noise and oscillations of the integrated system. For differential input signals, it is common to have a resistance or voltage breakdown specification to system ground. The coupling of a signal between the RF output port and the RF input port is referred to as reverse isolation.

Reverse isolation may be of importance in such applications as signal boosters, RF broadcast amplifiers, radar systems, etc. where the output signal may distort RF measurements, signal integrity, or cause oscillations. For example, the RF output signal load impedance can affect or change the input signal source impedance, thereby changing the amplifier efficiency. Output to input isolation is typically specified in decibels.

Signal and Distortion Parameters: Details of the signal to be amplified must be defined, since the designer may need to develop specific tests or special designs in some cases. The bandwidth of the signal envelope to be amplified, as well as the amount of allowed distortion or clipping on the output of the amplifier is important.

Amplitude Distortion: Amplitude distortion occurs when the gain of the amplifier changes with frequency or input signal level. Amplitude distortion is typically specified as gain flatness or the gain change for a given signal bandwidth. Input signal level distortion is normally specified as a compression or maximum input signal level as mentioned previously.

Phase Distortion: Phase distortion occurs when phase does not change proportionately to frequency within the pass band of the amplifier.

Group Delay: Phase slope through a "perfect amplifier" should be a straight line—which means the delay a signal has through the amplifier is constant with frequency. In active devices this is rarely the case and HPAs are no exception. Many types of systems, especially with linearization techniques, this is a critical parameter. Group delay is typically specified as the amount of time a signal will take to get through the amplifier over a frequency band—the deviation from a constant delay, is also specified (measured as the deviation from absolute flatness of the slope across a frequency band).

Harmonic Distortion: Harmonic distortion is of concern in many sorts of amplifiers, especially those that are greater than one octave wide in bandwidth. Harmonic distortion occurs due to amplifier nonlinearity which generates signals that are multiples of the fundamental input signal. Harmonic distortion is typically measured by comparing the logarithmic difference between the fundamental power signal level to the harmonic signal level in dBc as shown in Figure 20.3.

Intermodulation Distortion: Intermodulation distortion is another method of measuring and specifying an amplifier [1,2]. Figure 20.4 depicts a typical spectrum analyzer measurement of this sort of distortion.

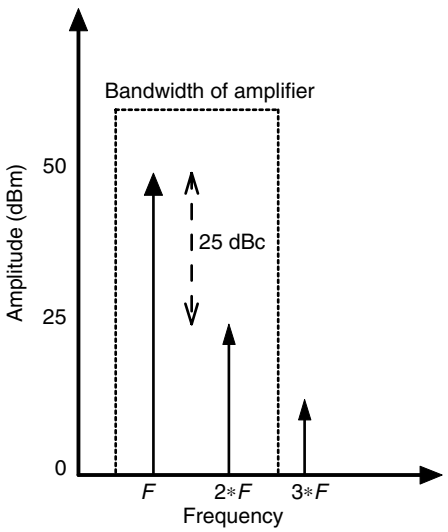


FIGURE 20.3 Example of harmonic distortion.

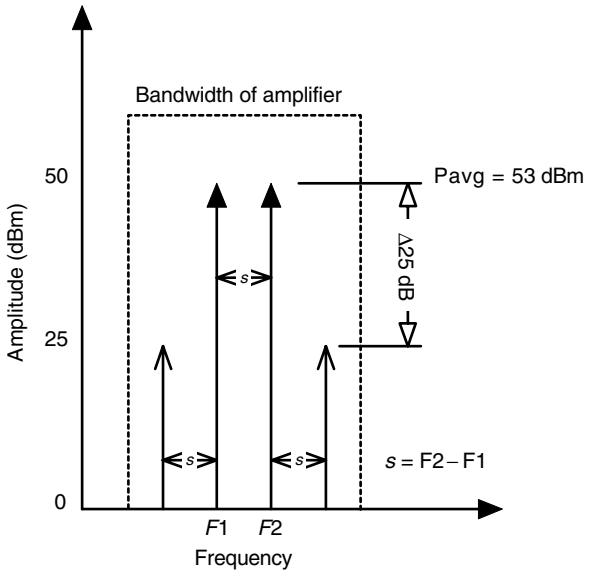


FIGURE 20.4 Example of intermodulation distortion.

When two signals of equal level are put into the amplifier mixing products above and below the original signal frequencies is a result in addition to the amplified input signal. For this measurement, the third-order intermodulations (depicted as the signals to the right and left of F1 and F2 in the figure) are present, and usually are the most significant. Sometimes higher order, such as 5^{th}, 7^{th}, 9^{th}, and so on are measured or specified. While a simple predictive calculation is used in class A circuitry, in large signal applications empirical data are measured. With modern signals and digital systems, two-tone tests are used less, and multi-tone testing or application-specific linearity tests are used to specify and characterize some amplifiers.

Reliability: Due to the large amounts of heat and high voltages and currents in a typical amplifier, all amplifiers of this type will have to have careful scrutiny paid to components and component de-rating in order to have acceptable reliability, as well as how the individual modules in a HPA contribute to overall HPA failure rates [3]. Usually, this will be a system level failure rate or mean time between failure (MTBF) that will have to be translated into HPA MTBF. Sometimes for extremely long life amplifiers (such as 250,000 h or larger) special consideration will have to be given to component de-rating and component selection.

Cost: Owing to the expense of RF power devices, the high quality required of many RF components, as well as the significant thermal transfer requirements, the cost of an HPA can easily be the single largest expense in an overall system, and in most cases will be the single largest expense. This fact alone requires some important considerations in the design. Since each specific customer or application has certain requirements and budgets, there is little specific advice that has to be given except that HPAs are, generally, fairly expensive compared to many other RF subassemblies, and balancing the specifications against cost is very difficult and a large challenge in many HPA designs.

20.4 Application Specific Specification Considerations

As defined above, the author has chosen to divide the numerous application types into three general, broad categories, and specific considerations for specifications salient to each is detailed below. It is important

to note that there are very few "pure" amplifiers that fit into only one category, though in many cases one of the three categories will be dominant.

Linear: A linear HPA is usually used for communications tasks, and many other applications (such as MRI) will require a linear response. Typical linearity distortion allowances can be expressed as specific versions of AM–AM as well as AM–PM distortion [1]. AM–AM is amplitude distortion due to amplitude at the output—classic examples will be the amount of compression of output power due to input power (Pin vs. Pout curve). AM–PM is the amount of phase shift that occurs when an amplifier is driven into compression—this can be a significant issue in some tightly controlled digital modulations or in applications that are very sensitive to phase shift. In the overall system in which an HPA will be used, some sort of correction scheme is sometimes employed to minimize the effect of the native HPA response.

Most linear amplifiers are designed to amplify a specific signal or family of signals, and might have more specialized means of measurement—indicating acceptable levels of spectral regrowth, intermodulation distortion, or other parameters. In a specification, this sort of measurement as well as the type of signal must be indicated. Also, if the amplifier design requires some sort of linearization technique, the specification must reflect the technique that is expected [1,4].

CW: The basic goal of a CW amplifier is to generate a lot of power. It is important to specify if the output power is fixed or variable, the expected dynamic range of the amplifier, and the P_{in}/P_{out} deviation from a straight line allowed (sometimes called "linearity", though it is not the same as that for a linear application). Additionally, the HPA may have to amplifier phase or freequency deviation accurately such as a FM or PM amplifier. Also important to note is that in some applications, stability into a varying complex loads may be required as well.

Pulsed: Some applications of pulsed amplifiers include radar systems, RF industrial heating, laser exciters, medical imaging, and special forms of digital communications. A pulsed amplifier produces high RF power for short durations. Characteristic of pulsed amplifiers is the shortness of the pulse relative to the off time—a duty cycle of less than 10% has typically been labeled pulse amplifiers. Beyond 10% duty cycle, the advantages of a specific pulsed technique diminish. In cases where the duty cycle is very long, the application may be better served by a CW style of amplifier. Since the duty cycle is so low on a pulsed amplifier, the pulsed power can be three to four times the CW power rating of the device. Some semiconductor devices designed for pulsed amplifier operation can operate at as high as 8–10 times the CW operating power rating. The designer's goal is not to exceed the maximum junction temperature of the device chosen while getting as much power out of the amplifier as possible.

Pulsed amplifiers are essentially amplifiers that can handle transients well without overshoot, ringing, phase distortion, or time delay. Figure 20.5 shows the common parameters by which pulsed amplifiers are characterized.

Rise and fall time are not indicated in the figure but are also important. Rise and fall are controlled by amplifier bandwidth and large reactive elements. Pulse droop can be caused by junction heating, insufficient bulk capacitors near the RF amplifier, excessive source, or emitter lead inductance.

BJTs demonstrate higher peak powers than MOSFET devices in general, though the gap is closing, and MOSFET devices are improving significantly [5].

20.5 Design Concepts

20.5.1 Gain, Power Lineup

Once the specification is defined, the most fruitful approach to the design of the power amplifier is to next map out the power lineup. In this context, the device requirements and system requirements are fleshed out, and the specific application requirements are met. The gain lineup, class of operation, and the combining scheme skeletons are all laid out in order to assure compliance to the requirements. A sample

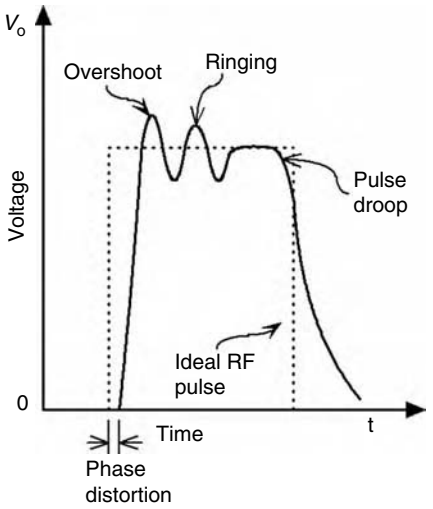


FIGURE 20.5 Pulsed signal with distortions.

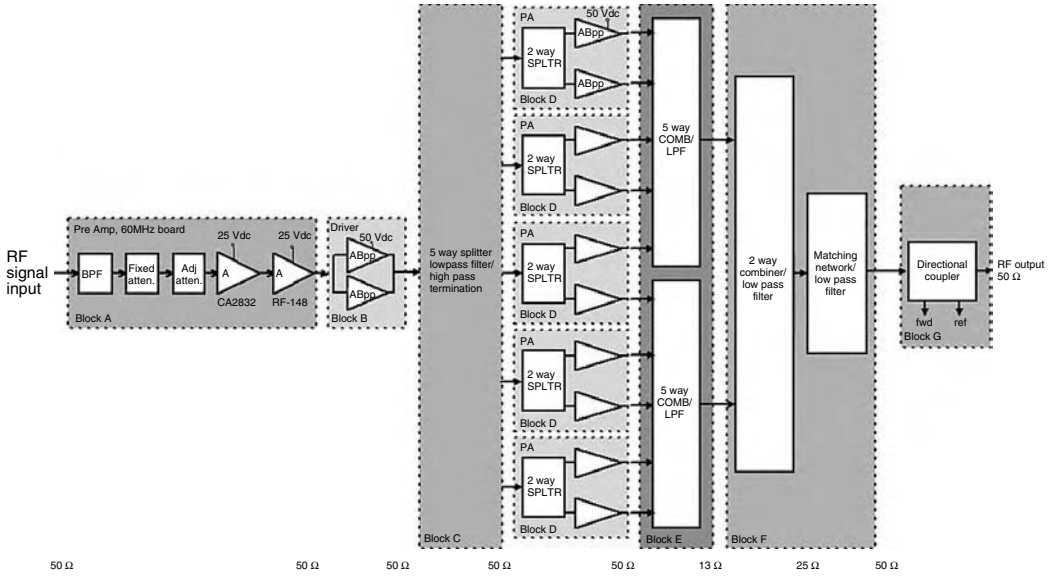


FIGURE 20.6 Block diagram of HPA. (Courtesy MKS Instruments. Used with permission.)

block diagram is shown in Figure 20.6. A discussion of gain and power budgeting along with the theory of corporate combining to higher power levels was pioneered in the 1970s and is detailed in Ostroff et al. [3] in Chapter 4.

A simple gain and power budget of the block diagram in Figure 20.6 is shown in Figure 20.7. The basic gain, power, and currents are all shown, and the amount of drive available is shown to be enough to bring the output parts to the required output power. Note that power margin is noted specifically on the spreadsheet. In the figure it is important to note that several output stages have been combined in parallel in order to get the right amount of output power. Also, important to note is that the margin in the driver is larger, preventing a doubling up of compression.

This type of gain and power budget is called a power architecture or power lineup, and its layout and conception are an important elements in successful HPA design. In any line up, the designer will define

CNTRL BD input power for 3KW (dBm): 4.20

	Block A, Preamplifier					Block B	Block C	Block D		Block E	Block F		Block G	
Average case (Based on measured data statistics)	Bandpass filter	Fixed atten.	Adjacent atten.	Amplifier CA2832	Amplifier MRF 148	Driver	5 way splitter/LPF-HP termination	2 Way splitter	Amplifier P–P	10 way input combiner/ low pass filter	2 way combiner/ LPF-HP termination	Output matching network/LPF	Directional coupler	RF output
Gain min (dB)	–0.5	–9.63	–5	34	13	15	–5	–0.25	15	–0.2	–0.2	–0.1	–0.05	60.57
Rated power (W)				1.26	7	520			520					
Efficiency (%)				6%	35%	50%			50%					
Required attenuation for given input power:			0.30											
P_{in}/path (W)	*0.003*	*0.002*	*0.000*	*0.000*	*0.203*	*4.0*	*128.0*	*22.8*	*10.8*	*340.5*	*1625.9*	*3105.4*	*3034.7*	*3000.0*
P_{in}/path (dBm)	*4.20*	*3.70*	*–5.93*	*–10.93*	*23.07*	*36.07*	*51.07*	*43.58*	*40.32*	*55.32*	*62.11*	*64.92*		
P_{in}/total (W)	*0.003*	*0.002*	*0.000*	*0.000*	*0.203*	*4*	*128*	*114*	*108*	*3405*	*3252*	*3105*	*3035*	*3000*
P_{in}/total (dBm)	*4.20*	*3.70*	*–5.93*	*–10.93*	*23.07*	*36.07*	*51.07*	*50.57*	*50.32*	*65.32*	*65.12*	*64.92*	*64.82*	*64.77*
P_{out} (W)	*0.0*	*0.0*	*0.0*	*0.20*	*4.0*	*128.0*	*22.8*	*10.8*	*340.5*	*1625.9*	*3105.4*	*3034.7*	*3000.0*	*0.0*
Pwr diss/device (W)	*0.0*	*0.0*	*0.0*	*3.2*	*7.5*	*128.0*	*13.9*	*6.4*	*340.5*	*153.3*	*146.4*	*70.7*	*34.7*	
Power margin (%)				*84%*	*42%*	*75%*			*35%*					

Total preamp gain (dB): 31.9

FIGURE 20.7 Sample gain and power budget. (Courtesy MKS Instruments. Used with permission.)

all the subportions of the entire HPA, both in terms of gain (or loss in the case of passive elements) and power as well as compression, linearity, and so on. Each stage will show the minimum amount of gain and power output required in order to supply the next stage with enough power to reach the ultimate goal. Also, the lineup when considered in light of available devices will show the scalability of a particular design. For instance, in the gain and power lineup sheet in Figure 20.7 for the amplifier in Figure 20.6, the lineup shows several modules in parallel with a variety of combiners and dividers as well as other ancillary items.

Additional benefits are modularity (allowing graceful degradation in the case of failure [6]) that will allow elements of the design to be used in systems producing different power levels. Size, reliability, cost, and flexibility in part selection are all factors to be weighed in using modular design.

Power Amplifier Blocks: The main portion of the block diagram consists of power modules. In many cases, by themselves, they would qualify as a HPA, though the most common usage of these modules is in conjunction with other power modules to achieve high power levels. Some design consideration is given later in the chapter to these modules, and it is the core of successful HPA design.

Combiners and Dividers: With current technology, it is typical to have devices of several hundred watts per device. Powers up to 1 kW may be possible, but are not typically common at present. In order to reach really large output powers, a method of adding powers of multiple devices together must be used. Through the use of combiners and corporate combining methods, large amounts of power can be generated with much smaller, lower power devices. Another important portion of these power and gain lineups is combiners and dividers. They allow multiple modules to be used in parallel enabling an arbitrary power level to be reached, provided the losses do not get too high.

It is also important to note that in the above illustrations there is no load accommodation method shown in the lineup. In many cases, especially at moderate to higher RF frequencies (100 MHz and higher) circulators and isolators are technically feasible for at least narrow band applications, though the cost of these devices upto the UHF band may be prohibitive (again, application dependent). Also couplers on the input and output of the power lineup may be required if some sort of automatic level control (ALC) or automatic gain control (AGC) is required by the larger system—their losses to the system need to be taken into account in a lineup design.

20.6 Special Considerations in Some Typical Power Architectures

Linear: In a linear HPA, the overall linearity can sometimes be estimated in the budget analysis. There are many methods of estimating the overall linearity of a HPA lineup, and the exact method depends upon what sort of signal is used for the analysis. Depending upon the nature of the signal to be amplified, a simple two-tone analysis may be used, multiple tone analysis or even a more elaborate simulation based on the actual modulation if the response of the circuit to the modulation is known. Sometimes a part vendor will have data based upon bias level and power output of the part in an example application circuit as part of the data sheet or application note that can be used. In a linear design it may not be always possible to have a completely accurate estimate of the linearity of the target under actual modulation—it is usually fruitful to have some sort of linearity estimate as well as an accounting for the expected peak headroom required by the application in order to make the proper tradeoffs before expending the time and money in order to build and test circuits. Figure 20.8 shows a hypothetical device inter-modulation curve versus power output. In the case of a gain, power and linearity budget, the line up designer would need to reference average power versus the curve to estimate two-tone linearity. This would be for class AB, and it is interesting to note that due to a number of physical factors as well as selecting bias levels correctly, once the power is reduced to minimize compression effects (the right side of the curve), there is a broad area of reasonably linear operation which can be advantageous to exploit in linear applications [5].

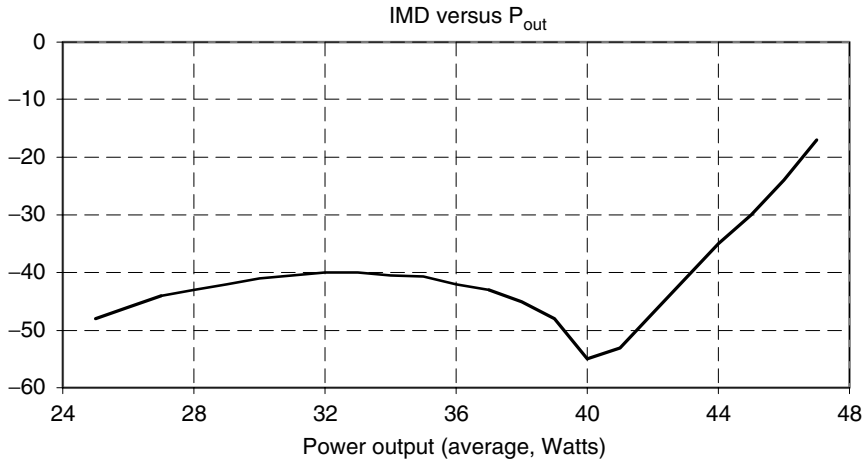


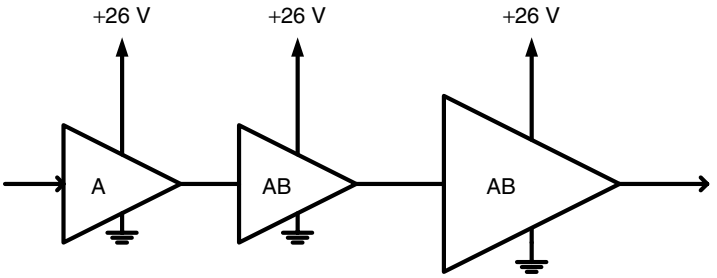
FIGURE 20.8 Hypothetical device IMD response in class AB.

For class A devices while such curves are possible, most devices have an *extrapolated/calculated* "IP3" figure—which is the power level where the third-order intermodulation product would equal the main power out—if it were possible to increase the power to the level and not destroy it. Approximately for every dB of power reduction from this figure, the IMD improves by 3 dB as in Equation 20.1.

$$IMD_3 = -3 \times (IP_3 - P_{avg}) \tag{20.1}$$

Headroom: For instance, if a power stage under development must be capable of producing 100 W of peak power output (50 dBm), and it is defined that the signal to be amplified has a 7 dB peak to average, the largest average power of the stage can support without clipping would be 43 dBm. If the amount of average power that is expected to be put out by the device is 40 dBm, there is 3 dB of additional headroom. This fact alone should give the designer the idea that the device will be able to produce the required amount of power. If, all that is available is two-tone data (as is the case most of the time) a two-tone analysis and lineup may give the designer confidence to accept or reject the device under question. Conversely, if the average power is 44 dBm, the signal is expected to be clipped by 1 dB. If the signal must not be clipped, the designer would know right away that at least 1 dB more power is needed. An excellent discussion of crest factor and its implications to RF amplifiers is in Kennington's book [1] in Chapter 2.

If some performance data is known, such as distortion under actual modulation conditions, or data taken with many tones that would give the ability to measure headroom as well as linearity approximating actual conditions, the lineup will be easier still. (Each linear signal that has high and momentary peaks, a probability density function as well as exact signal specifications is helpful to judge how close a substitute signal is to estimate the performance under actual modulation conditions.) For example, a 16 tone signal specification would need to specify tone spacing, relative tone power levels, and the phase relationship between each tone with the other tones. When performing this analysis, the ability for a digital modulation to survive clipping is best understood so that economical tradeoffs may be made (early forms of digital modulations such as CDMA would allow significant clipping without significant degradation, especially if peak restoration or predistortion techniques were used). Since power devices tend to be rather expensive, understanding the particulars of the amplified signal to present the minimum acceptable design is key to being able to produce an affordable amplifier. A sample linear application lineup is presented in Figure 20.9, assuming a two-tone analysis is used for linearity analysis.



	Stage 1 Predriver Class A	Stage 2 Driver Class AB	Stage 3 Final Class AB	Isolator	
Input power (dBm)	14.50	29.50	40.75	50.25	
Gain/loss (dB)	15.00	11.25	9.50	–0.25	
Output power (dBm)	29.5	40.8	50.3	50.0	
Output power (Watts)	0.9	11.9	105.9	100.0	
IP3	55.0				
IMD3 (dBc)	–51.0	–40.0	–40.0		
IM rolled up (voltage)	–51.0	–37.8	–32.8		
IM rolled up (RMS)	–51.0	–39.7	–36.8		
P1 dB (dBm)	33	43	50.8		
P1 dB (Watts)	2.0	20.0	120.2		
Power margin (dB)	3.5	2.3	0.5		
B+ voltage (Volts)	26	26	26		
Current (Amps)	1	1.8	9.5		
PAE (%)					
DC efficiency	7.7%	42.6%	48.7%		
HPA gain				35.50	dB
Total current				12.3	A
Total efficiency				31.3%	
Total IMD, worst case				–32.8	dBc

FIGURE 20.9 Gain and power budget with two-tone linearity information.

In Figure 20.9 a hypothetical lineup is used to estimate the suitability in a linear application. Two-tone linearity is used to calculate the linearity because in this example, this is the data that is known. Be careful when using two two-tone analysis to predict the effect of many tones or digital modulation. All important parameters are calculated for this lineup, except for required headroom. Assuming that the power numbers indicate peak power for a 3 dB peak to average, there is 0.5 dB of headroom in the example. When estimating the overall linearity of the lineup, it is important to be aware that the phase of the intermodulation products of each stage will determine the overall intermodulation product of the lineup. It is easy to imagine that one stage could produce intermodulations at a phase unequal to intermodulation products of other stages, making simple calculation difficult. Indeed, there are some rare amplifiers that are built around one stage canceling out the IMD of the next (though this practice is fraught with problems, if one or more of the devices experiences a process shift that would prevent maintaining the phase relationship, or if modulation is changed and the overall phase relationship between stages change. Also the device types must be selected very carefully as well). Unfortunately, before actual part measurements are made, such characteristics are usually unknown, though calculation of the worst case IMD stack-up is possible if the individual IMD's are known for amplitudes (see Figure 20.8). Some experience has shown that it is best to calculate the estimated distortion in two ways, one with a "voltage

stack-up"—assuming the worst possible phase alignment of intermodulation distortion, followed by one that assumes a random phase alignment that will stack up by RMS values:

Calculation of IMD stack-up based upon in-phase voltage addition is in Equation 20.2.

$$IM_{3,\text{total}} = -20\log_{10}(10^{-\frac{IM_31}{20}} + 10^{-\frac{IM_32}{20}} + \cdots + 10^{-\frac{IM_3n}{20}}) \tag{20.2}$$

Calculation of IMD stack-up based upon RMS voltage addition is shown in Equation 20.3.

$$IM_{3,\text{total}} = -10\log_{10}(10^{-\frac{IM_31}{10}} + 10^{-\frac{IM_32}{10}} + \cdots + 10^{-\frac{IM_3n}{10}}) \tag{20.3}$$

Typically, the actual IMD stack-up will fall between these numbers; *it needs to be understood that the worst case stack-up can happen and the designer needs to be aware of this possibility* (see Figure 20.7, where both methods are used). When a substitute linearity estimation is used (such as two tone for CDMA), when actual measured data are used and the lineup is recalculated, the distortion products will behave slightly differently than with the substitute analysis (due to signal clipping, the AM–AM and AM–PM characteristics under that particular load, and the phase relationship between the distortion products between the stages).

CW or Saturated Mode of Operation: Since the prime consideration of this type of application is output power with maximized efficiency, overall RF signal amplitude linearity is not much of a concern. However, the relative compression of each stage is of importance. As in the first figure, each stage is producing power at 1 dB of compression, making the output compressed 3 dB. This may be undesirable in some applications (though perfectly acceptable in some others), though there will not be much margin to accommodate variation in output power and gain due to temperature changes as well as device lot variations. The overall $P_{\text{in}}/P_{\text{out}}$ curve of the amplifier is important if the power is supposed to vary in operation (e.g., a FM transmitter that has several output powers or semiconductor RF generator that specifies 10–30 dB dynamic range). In some applications, due to the need for accurate power delivery over dynamic range, special techniques to extend the dynamic range of the HPA lineup are employed (see below). In systems that have limited dynamic range by themselves, varying the drive may not present enough range for the amplifier (such as class C, or one of the switch mode classes). In such systems, it is common to vary the B+ (or rail) voltage in order to extend the dynamic range. A more advanced version of this technique used especially in switch mode classes of operation is the "class S" modulator that strips the signal envelope from the incoming signal and modulates the rail voltage to restore the signal while preserving excellent linearity.

20.7 Designing the Building Blocks

Once the entire HPA system is laid out and specified with the gain and power lineup and the general performance requirements are known, the next step is to begin the design of the building blocks of this detailed block diagram. The two main portions of the system, the basic power modules (BPM) and the combiners and dividers need to be designed, and the design considerations for each are what make up the meat of HPA Engineering.

20.8 Basic Power Module Design

In a gain and power budget, the BPM are the pieces that make all the gain and power delivery for the system. It is the most complicated portion of the system, and since it is an active circuit, it has the largest effect on reliability, efficiency and gain of the system as compared to the other sections. It is a challenging task, and an entire career can be based upon the skillful design of these modules. While it may be straightforward,

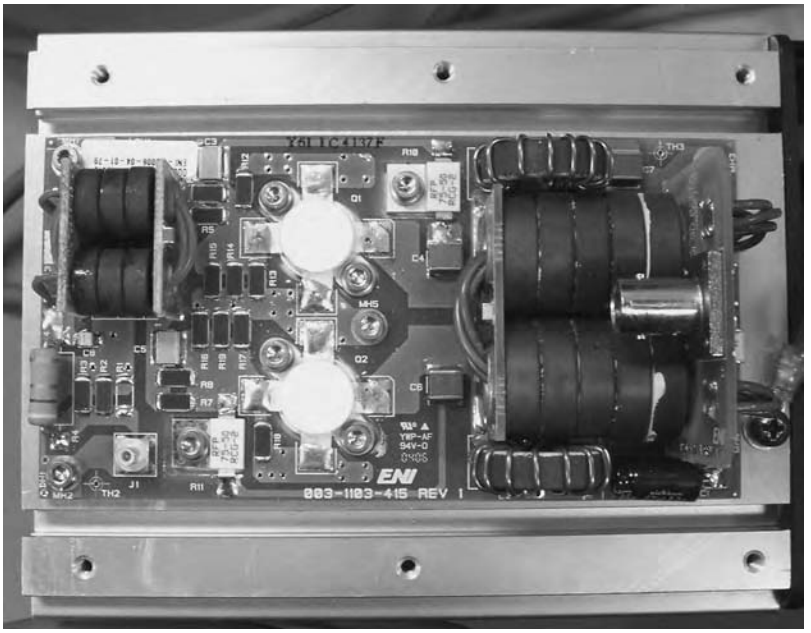


FIGURE 20.10 Basic power module. HF frequency, push-pull configuration. (Courtesy MKS Instruments. Used with permission.)

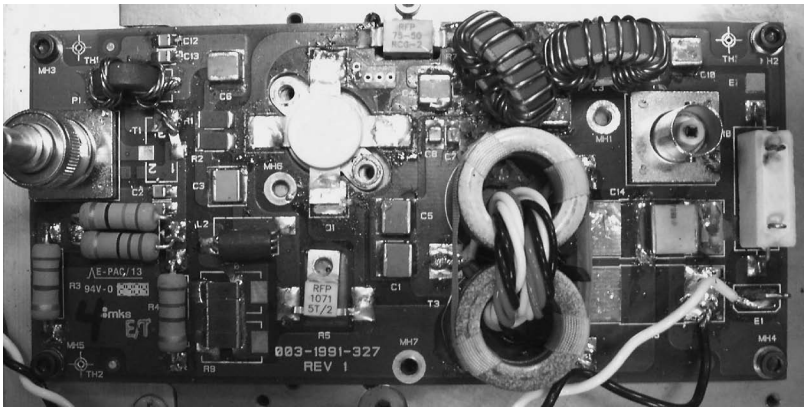


FIGURE 20.11 Basic power module. HF frequency, single-ended configuration. (Courtesy MKS Instruments. Used with permission.)

to get basic performance, design of these modules, the finesse in the execution is the difference between a mediocre and truly excellent design. Figures 20.10 and 20.11 show examples of HF BPMs.

20.9 Device Selection

When a satisfactory target lineup is reached, the designer should have already been evaluating candidate power RF devices to realize the simulated lineup. It is important to realize that RF devices will determine the "granularity" of the power lineup—meaning the basic building blocks of power if the HPA is to be scaled up or down. Each application and frequency will have a variety of parts, and the tradeoff between size and fine grain scalability will be a tradeoff in both part size and HPA complexity [5].

In the last 30 years, there have been many advances in materials and device structures. As a result there are a bewildering array of devices and an equally bewildering array of materials (Si, GaAs, SiC, GaN, etc.) for a HPA designer to choose them. Modules and RFIC's (monolithic IC's or modules that put an entire HPA or large portions of the function on a chip) present attractive options to a designer, though they can have some of their own individual issues. Given that a significant portion of the cost of an HPA is invested in the devices, correct device selection is critical if the HPA is to be a commercial success.

Rather than review each individual material technology and every variation of device structure, it may be easier to break device technologies into two general families: bipolar and field-effect transistors (FETs). For much of the RF band, silicon is the most economical material for generating large power and hence a silicon-based device will be the most likely choice faced by the designer at the time of this publication. Also, bipolar transistors and FETs based upon different materials and specific structures are similar enough that a general discussion is justified.

In the front end of a HPA, there may be an attractive module or RFIC that will save time and cost. Looking at many designs, an observer will see they are employed in the RF world effectively and in economical designs since it allows the concentration of effort on the final stages, where small errors or non-optimum solutions can cost the design efficiency, power output, or some other equally important parameter. Suffice it to say, that the use of such a module when done for the right reasons is effective and, provided it does what is expected, a good solution. There have been a number of companies that have offered larger powered modules and may be worth investigating. It can be seen from Table 20.3 that there are clear advantages and disadvantages to each technology; and the author would like to note that in most cases, when equivalent devices are available, the most popular choice tends to be a FET.

20.10 Class of Operation

The class of operation used has large implications for the rest of the design since it determines the maximum efficiency, gain, and output power. Once a specification is defined and parts selected, the class of operation has a large influence on the overall performance and RF "signature" of the final product (note that many devices will have recommended classes of operation that optimize the device performance).

TABLE 20.3 Comparison of Bipolar and FET

Factor	Bipolar	FET
Input impedance	Modeled as a diode feeding an emitter resistor (mostly real, but low impedance)	Gate capacitor (highly reactive)
Output load impedance (nearly identical techniques between technologies)	Depends upon class of operation, parasitics dominated by Cob and bondwires	Depends upon class of operation, parasitics dominated by Coss and bondwires
Linearity (generally in silicon, FETs will be more linear in class AB)	Typical BJT devices in class AB will have −30 to −35 dBc 2 tone measured with power back off	Typical DMOS or LDMOS FET devices will have −40 to −45 dBc 2 tone measurements with power back off
Common failure modes	Overdissipation, hFE rises with temperature causing avalanche. Secondary breakdown	Overdissipation under high voltage conditions, avalanche unlikely as Rds(on) (channel resistance) rises with temperature, input gates can have voltage exceeded
Common instabilities	Instabilities at half the output frequency (aka "half f-ing"), low-frequency instabilities, natural internal feedback is low	Natural internal feedback is high (Crss, Miller effect)
Advantage over competing technologies	Cost, easy to push higher in frequency	Better stability, ruggedness, linearity, easy to broadband

TABLE 20.4 Class of Operation Summary

Class of operation	Definition
Class A	The device is "on" for the entire cycle—meaning current flows for the full RF wave
	Device is always on and at the same operating point even with no RF present
	Theoretical efficiency maximum 50%
Class AB	Operating point moves with drive
	Device conducts more than 50% of the RF cycle
	Device is on with DC bias with RF removed (barely)
	Theoretical maximum efficiency 50–78.5% depending upon amount of RF cycle device remains conductive
Class B	Operating point changes with drive level
	Device conducts exactly 50% of the time
	Device is (barely) off with no RF drive
	Theoretical maximum efficiency 78.5%
Class C	Operating point changes with drive level, and a certain amount of drive is required to activate the device (hence gain is lower than in other classes)
	Device is on for less than 50% of the RF cycle
	RF drive turns on the device, is "off" with no RF drive
	Theoretical maximum efficiency is 100%, though typically much less
Class D/E/F	Switch mode device is either fully on or fully off—minimum time is spent in the resistive region
	Class E and F are like class D, though with resonant circuits to prevent current and voltage from overlapping during switching. Theoretical efficiency is 100%

Typically, a linear application will lean toward a more linear class of operation, such as class A or AB, though with a sufficiently loose specification and aggressive correction and control system; other less linear classes may be used, though at usually increased expense and complexity. When linearity is less of a concern, and efficiency is more important, the less linear modes of operation (C/D/E/F) become more attractive since successful implementation will allow a cheaper cooling scheme to be implemented and might allow a more compact form factor. Specifications such as dynamic range, simplicity, or cost may dictate a different choice. Table 20.4 briefly summarizes the differences between the classes of operation, and a good discussion of classes of operation is found both in Cripps [2] and Dye [5].

Class A: A class A device has two defining characteristics: it is "always on" meaning the device never turns off during the entire cycle of the input waveform (see Figure 20.12)—that is, the device conducts current for the entire RF cycle, hence power is produced for the entire RF waveform, and thereby gain is the highest of all classes of operation. Linearity is also the best of all the modes, but power efficiency is poor. It also has an operating point that does not move with RF excitation (fixed operating point). It requires a bias circuit (usually with some form of feedback) in place that assures the device is fixed in operating point. When it is used in a HPA design, it will be either where linearity is the overriding concern or where the power is low or is used as a high gain driver for more efficient final stages. This is also popular with low signal level parts used to drive later parts where gain is more important than efficiency. Efficiencies of 10–25% maximum are not unheard of, though extremely difficult to achieve. Mathematically, 50% is the maximum theoretical efficiency, with real world devices this number is unachievable.

Class AB: A class AB device allows the operating point to move with RF excitation, essentially the RF energy becomes part of the bias. It amplifies more than half the waveform, but less than 100% (see Figure 20.13). Note the negative excursion of the current—with a more neutral bias feed, there would be none of this overshoot, and it is a result of the bias circuit design and stored energy. Also noteworthy is that the Vcc is 5 V, and the peak of voltage is higher than this voltage—this is a characteristic of class AB and important when deriving its load line. A simple bias network is in place in order to have a small amount

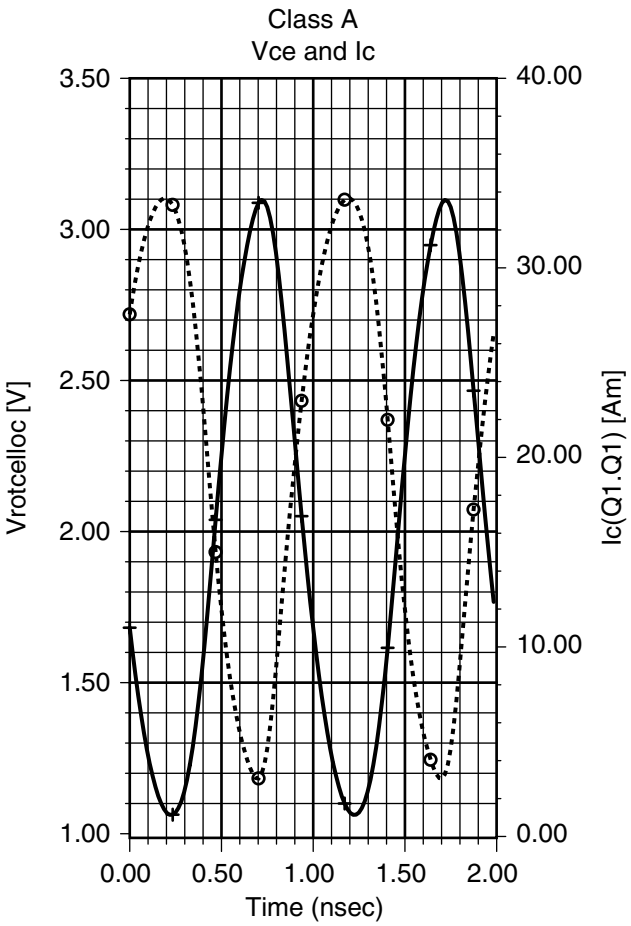


FIGURE 20.12 Class A voltage and current conduction angle (circles are current, crosses are voltage).

of current through the device when no RF is applied. While not as linear as class A, it is usually acceptable in most linear applications, though some sort of external correction (e.g., predistortion) may be required. Efficiency is much higher, 40–60% is typical, theoretical maximum is 78.5% .

Class B: Class B has a bias circuit to raise the input bias voltage to the point where any RF signal will turn the device on. It will pass precisely 50% of the input RF waveform, and will generally have better efficiency than class AB, though a lightly biased class AB will have similar efficiency. Gain is generally lower than class AB, and theoretical maximum efficiency is the same as class AB.

Class C: Class C will pass less than half of the input waveform (see Figure 20.14), and will either have a circuit in place to provide an appropriate bias, or in the case of positive bias devices (such as a npn BJT or n-channel FET) it may simply have a resistor to ground and use the bias voltage as the "negative bias." Class C is highly nonlinear, and it is typically used where efficiency is the primary concern. A class C stage has low dynamic range and low gain (since the input waveform is biasing the device)—so a system incorporating this type of stage must take that into account.

Class D/E/F: The classes of operation of D/E/F are all similar in that they use the transistor in switch-mode operation. The device is driven such that it spends as much time as possible fully on or fully off. As little time as possible is spent in the intermediate state where all the other classes of operation previously mentioned make most of their power. Different methods are used in classes E/F in order to increase

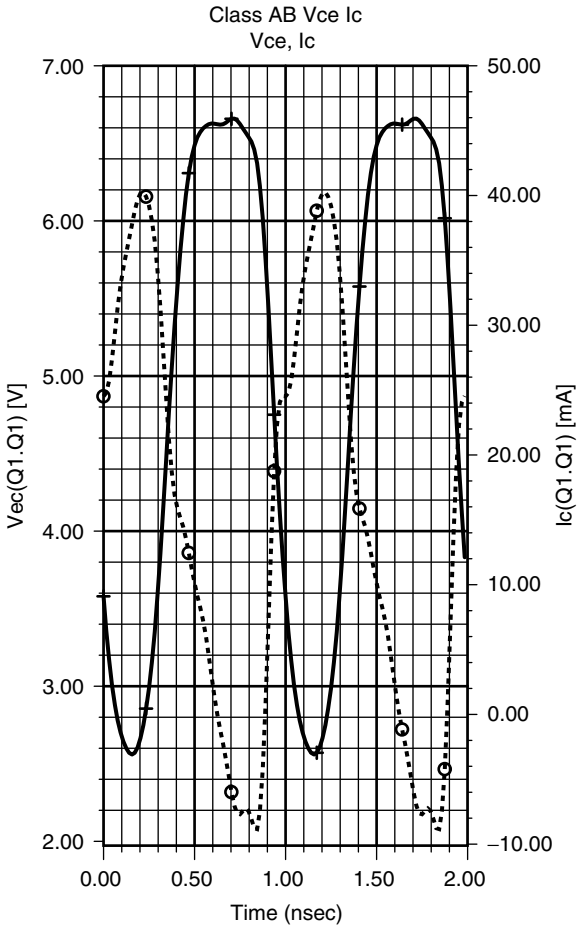


FIGURE 20.13 Class AB current and voltage (current is circles, crosses is voltage).

the efficiency as the frequency rises. The theoretical efficiency maximum of these modes is 100%, though the realities of the device physics and the frequency of operation will cause a design to fall short of this mark. The stage itself has very little dynamic range, and as the frequency rises, E and F have bandwidth limitations. Methods of rail modulation and phase and frequency shifts on the input of the stage can produce linear output. A good discussion of this exciting mode of operation is found in Cripps [2,4] and Kenington [1].

20.11 Circuit Topology

Once the specification, general class of operation, and device are chosen, the general circuit topology must be chosen. This involves selecting the load and source impedances that best accomplishes the task for the power amplifier—and designing matching networks to present that to the devices. In most designs the output load will optimized for maximum power, and the input will change targets depending upon the specification (gain, bandwidth, flatness, linearity, or other parameter), and in some cases less conventional topologies will need to be chosen (such as outphasing, Doherty peaking methods, or other methods described in Cripps [4]), but in most cases, due to system complexity and cost constraints, these methods must be carefully selected.

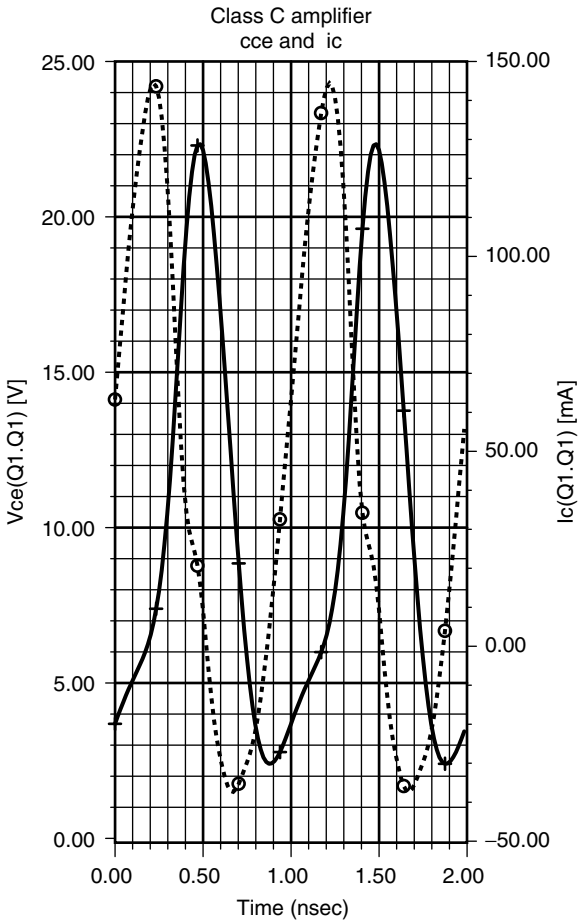


FIGURE 20.14 Class C voltage and currents (current is circles, voltage is crosses).

20.12 Determining Source and Load Impedances

In small signal linear devices, *S*-parameter characterization is usually sufficient to design an optimal circuit. The user will be able to use this characterization in order to create an input and output matching circuit that will give the maximum gain and optimal output power. When designing a HPA, however, due to their nonlinearities and varying operating point, *S*-parameters are usually insufficient in order to determine the matching circuit (the exception is class A if the application does not greatly compress the device). The designer has to determine the load and source impedances for the device that will give the optimal performance under the signal conditions of the application.

For low frequencies when package parasitics are negligible, using a class AB class of operation, the optimal output load for maximum power transfer is calculated below.

Examining Figure 20.13, one will readily observe that compared to a 5 V line, the voltage excursion is much higher. In Class AB, the peak voltage is limited to two times the rail voltage and limited from reaching zero by the saturation voltage, making the peak-to-peak voltage.

Showing the Maximum voltage in Equation 20.4

$$V_{p-p} = 2V_{cc} - V_{sat} \tag{20.4}$$

An approximation of the RMS voltage, useful for calculation of the average power is shown in Equation 20.5

$$V_{\text{rms}} = \frac{V_{p-p}}{2\sqrt{2}} \tag{20.5}$$

From Ohm's Law, the power is Equation 20.6

$$P = \frac{V^2}{R} \tag{20.6}$$

Making a simple substitution of V_{rms} for the Voltage and rearranging the above equation into Equation 20.7

$$R = \frac{V_{\text{rms}}^2}{P} \tag{20.7}$$

Therefore Equation 20.8 shows that the load line for class AB.

$$R_{\text{load}} = \frac{(V_{\text{cc}} - \frac{V_{\text{sat}}}{2})^2}{2P} \tag{20.8}$$

If the device is a 26 V device, with a 2 V V_{sat} and designed for an output power of 150 W—the load would be approximately 2.1 Ω.

If the designer looks at class A as a comparison, where peak voltage is limited to the rail voltage, the same load equation becomes Equation 20.9, which shows that all else being equal, the maximum power of class A is one quarter that of class AB (n.b. that thermal considerations may limit the device output even further).

$$R_{\text{load,class A}} = \frac{(V_{\text{cc}} - \frac{V_{\text{sat}}}{2})^2}{8P} \tag{20.9}$$

While this derivation and calculation is illustrative and under the right circumstances (low frequencies) directly useful, in actual practice this calculated number cannot be directly used to design a circuit. There are many factors that conspire to transform the basic "load line" resistance into a different, complex impedance at the device package output. Especially at higher frequencies, the device and package parasitics can make the load impedance that the designer must present to the device significantly different from the above calculation. Simple single-ended parasitics are shown in Figure 20.15 along with possible effects upon a load line of 3 ohm when the output capacitance is 100 pF with 2 nH of lead inductance—as shown parasitic reactance can bring the impedance so low that a part manufacturer is sometimes forced to put some sort of pre-matching inside the device to increase the impedance to something more easily worked within a circuit. Even at lower RF frequencies and without pre-matching, the parasitics will tend to rotate the ideal load impedance enough where special characterization is required in order to determine the optimal load for a circuit and obtaining these impedances is a nontrivial task [7].

Obtaining load impedances from scratch, as mentioned above, is nontrivial. Fortunately, there is a variety of ways from which they can be determined:

Application Data: A helpful device vendor may have already selected impedances to extract the intended performance for the application. While, not universally the case, many vendors will have this information for most mature parts and some newer ones. The benefit of using vendor supplied data is that the data is readily available for the asking. The limitation of using this approach is that if the designer's application isn't

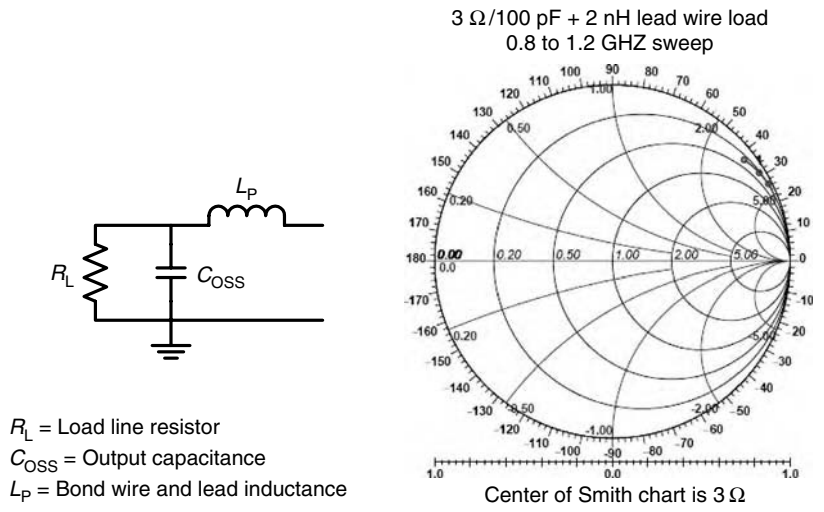


FIGURE 20.15 Simple model of RF parasitic reactance and effects on Z (At 950 MHz the impedance is $0.24 + j3.55$).

the same as the intended one on the data sheet, several circuit iterations or an additional load pull may be required to assure optimum performance (such as the impedance point giving optimal performance or a digitally modulated signal may vary slightly from a CW application) [5].

Simulation: Starting from the basic principles of the load line (such as the one calculated above) and a simulation of the packaging parasitics, one may calculate the likely optimal load spot. The benefit of this method is that you can determine a likely set of impedances quickly, but it relies upon proper simulation to get good results. Steven Cripps developed this method and it has been used effectively in many designs. Informally this technique is called the "Cripps Load Pull" method and is detailed in Cripps [2], Chapter 2.

Load Pull: To perform a load pull of the device—essentially varying the input and output match to get ideal source and load impedance—points for the device for desired operation. This method can either be manual or automated. The output can be simply a list of impedances that are ideal for an application. It may also be a variety of source and load points, and when exhaustive searches are performed, performance contours mapped to a Smith Chart. While, this can involve some up front investment in time and equipment or the use of a specialized testing facility, when properly executed will give a comprehensive map of device performance and may reduce the number of iterations required in order to produce an optimized, high performance circuit. It also will allow the selection of some points that may have nonoptimum power transfer, but have efficiency or other characteristics that might be desirable in an application. Load pulls also allow some application-specific characteristics to be optimized when determining the impedances. Occasionally a part manufacturer has this information, becoming more common in higher volume applications. Figures 20.16 and 20.17 show example of load and source pull contours. A discussion of load and source contours and their implications for amplifier design can be found in Wood [7], Cripps [2], and Ostroff [3].

It is important to note that with the load impedances the points of maximum power transfer and efficiency are not the same point—this is always the case and the designer must choose between these spots. Most of the time, maximum power transfer is the most desirable load impedance. Also, when the source is pulled with a fixed load, there are several source impedance points of varying impedance. Depending upon the results, there may be a variety of points to optimize for the application. Sometimes it is also important to perform source and load pulls under the actual modulation used in the application

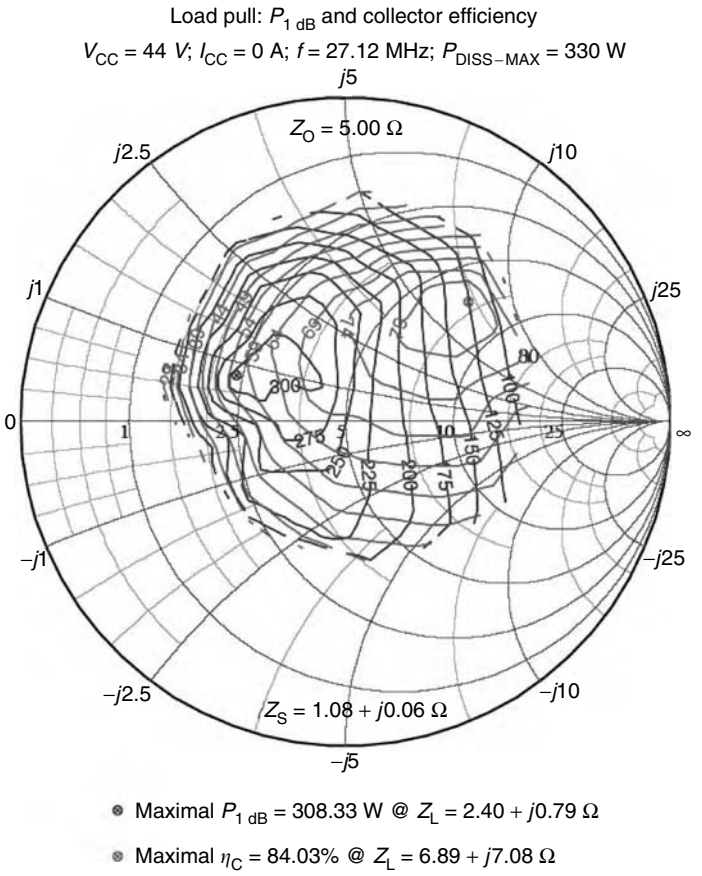


FIGURE 20.16 Example load pull contours, output load. (Courtesy MKS Instruments. Used with permission.)

such as pulsed applications which will have significantly more power under pulsed power than under CW conditions and will require slightly different load impedances to achieve this.

20.13 Matching

When target impedances have been determined, creating a circuit network to bring the impedance of the device to the characteristic impedance of the system is the next step. This activity, the creation of RF matching networks, is one of the central activities in any kind of RF engineering. Hence, there are numerous sources on the subject, as the theory and application of matching in-general is a topic in and of itself. Matthaei et al. [8] offer a number of refinements on the subject, and their work is very influential in modern RF matching networks up through millimeter wave. This chapter will not attempt to rehash the basic theory, but will point out specific things a HPA designer must watch out for when designing the building blocks that go into a HPA.

Simulation: It must be mentioned that simulation software is very useful in trying to realize the networks once the basic theory is understood and the engineer is ready to design a matching network. Many years ago, use of these tools were sporadic due to their difficulty and mixed results. In the intervening years, the accuracy and usefulness of the tools have improved to the point where they are indispensable in the design. Infact RF engineers are expected to be skilled in their use in most establishments. A number of programs are available, at the time writing of this chapter, with variety of

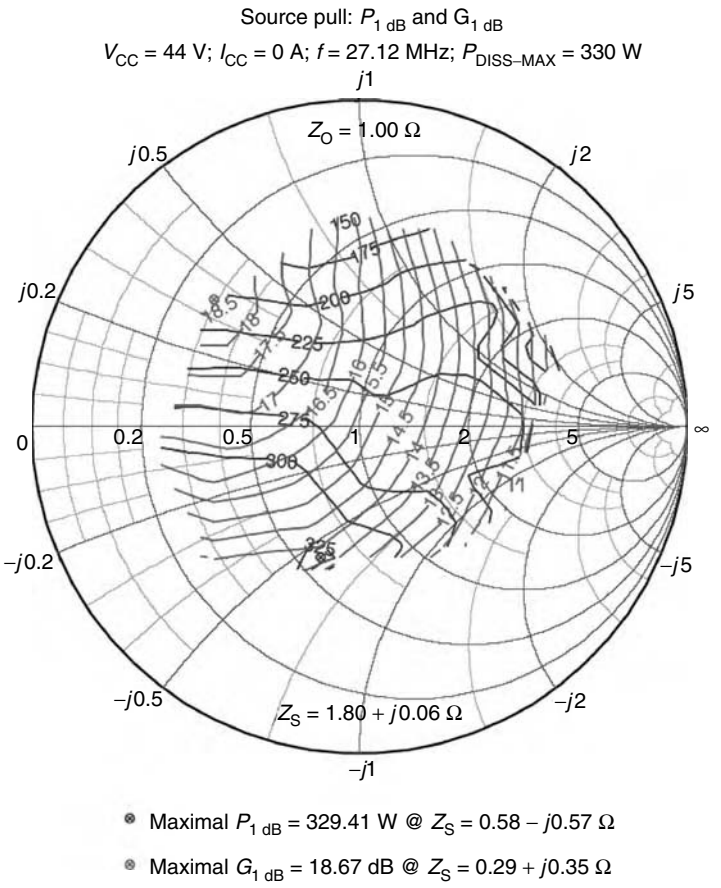


FIGURE 20.17 Example source pull on bipolar transistor. (Courtesy MKS Instruments. Used with permission.)

capabilities and prices. A program used properly and with skill will save many design iterations, time, and money in creating the circuitry to be used in a HPA. Unfortunately there can be no recommendation as to the "best" tool since both the types of circuits and the software budget will be such large factor in the choice of the tool to be used. Also the various programs change rapidly over years, so good advice today may no longer be applicable in the following year. It should be noted that software has proven to be so useful in the design, that some of the most sophisticated programs with the highest cost to buy will be inexpensive when compared to those with a long time to market or a failed design attempt.

Matching Networks: As in any specialized sub-field, there are a number of rules of thumb to be observed when designing a matching network—some of them due to the large circulating currents and voltages on the matching network, while others have to do with the typically large amount of transformation required in transistors stages.

Reactive, Not Resistive: HPAs will require most, if not all, of the matching to be done reactively. While resistors as the matching network as well as inter-stage pads are popular topologies at low powers, they quickly become impractical at higher powers. The reason for this is two-fold. At higher powers, the resistors will grow in size significantly due to the amount of power that will be dissipated in them in the process of matching. With more dissipation, a lot of power will be discarded as waste heat, decreasing efficiency and losing power that will require additional power devices to make up for this loss. Given the

expense of high power devices, this will rapidly become economically infeasible. The topology section of this chapter will, therefore, deal strictly with reactive matching techniques.

20.14 General Matching Techniques

As mentioned previously, matching for a high power device, as in other RF matching tasks, resembles filters as much as it does straight transformation. Components are determined by the frequency of operation and actual power level (voltage and current stresses), but the principles are similar no matter what frequency or device one has chosen [9].

Nearly any match for a HPA will consist of three sections (Figure 20.18) each doing a slightly different job to match the part to its final impedance (typically 50 Ω). The first section (called "Reactive Tune" in the figure) brings the device impedance to the "real" axis on the Smith Chart, the second section transforms the impedance to an intermediate impedance or to 50 Ω depending on whether the circuit is to be single-ended or part of a dual pair (combined with a 90° hybrid, balun or other transformative combiner). The third section either involves the DC block only or the transformative combiner. Given that, this section will take the match from an intermediate impedance to the impedance at the outside world, and provides any required DC isolation it needs to be designed to work with the match in question.

Section 1: The "impedance of the device" (i.e., the conjugate load or source impedance) is usually presented in the form $R + jX$. If one were to take the reactive component X and divide it by the real portion, R, one can obtain the "Q" of the match in Equation 20.10.

$$Q = \frac{X}{R}$$

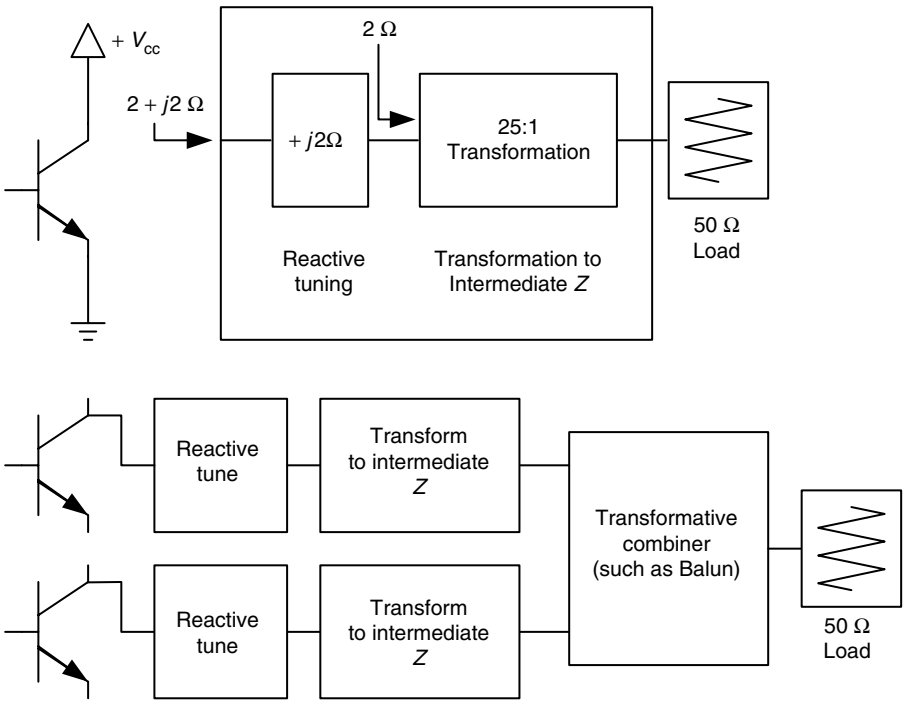


FIGURE 20.18 Block diagrams of matching circuits.

This will generally inform the user the maximum Q of the matching network—the goal of the match is to never exceed this Q (or exceed by as little as possible) throughout the entire transformation without "over-rotation." This also makes the job of the first stage crystal-clear: the first component must tune out—or eliminate—as much of the reactive (jX) term as possible. This transformation is usually done right at the device output, to make this transition as simply and as rapidly as possible, with the only limitations being the amount of current, voltage, and what is physically realizable. A couple of things happen when executed properly—the device has its first real return current physically close to the device, which enhances repeatability, provides a low Q (non-reactive) matching impedance, and allows for a more compact matching network. Lower Q matching networks allow for wide amplifier frequency bandwidths, "no tune" production tolerant amplifier designs using standard part variations, and a robust HPA that is cost effective compared to higher Q designs. Since there will be a range of frequencies and impedances, in which performance is desired, the designer must choose which frequency will be precisely zero, and how much will be slightly capacitive and slightly inductive. (n.b. Experience sometimes teaches that when the match bandwidth does not have to be large, having the interface impedance at this stage either inductive or capacitive with the upper or lower frequency reactance nulled out ($jX = 0$) will generally easily allow maximally flat responses). No matter how this is precisely done, the resulting Q is minimized which makes it possible to maintain an overall lower matching Q.

Section 2: The matching network at this point needs to transform the impedance from the "de-Q-ing" portion to the interface impedance. This is to be accomplished with as low a Q as possible—meaning that the deviations from the real axis are to be as small as possible with the maximum reactance divided by the R at that point must be minimized. Loss through this portion of the matching network must have as little loss as possible and the final impedance of the network over its operating frequency needs to be as close to the target impedances as possible.

The desire to maintain a low deviation from the real axis—or Q—will push the designer to have as many small transformation stages as possible. The desire for low loss will push the designer to have as few as possible. The desire to hit the target as closely as possible across the entire range of the frequency of operation will influence the type of match one wants to use (high pass or low pass topologies) as well as to have as many matching sections as possible. This is a difficult process, and the entire match is a tradeoff of these three criteria. There are limits to the quality of the match as well as trading off ripple and bandwidth for quality of match, and this tradeoff is treated exhaustively by Matthaei et al. [8].

When transforming impedances with several stages of matching, the most robust networks have continuously increasing impedances until the target interface impedance is reached. Sometimes, an inexperienced designer might find out that one may find a way to apparently minimize Q and have a very wideband match, but will require impedance nodes above the target impedance effectively "circling" the interface impedance. This is a valid way of matching (sometimes unavoidable), especially at lower powers, however, it will usually create a very touchy match that will be hard to reproduce in a production environment. Such matching networks have areas of large voltages and high circulating currents making component stress high. It is usually best to avoid this sort of matching if at all possible.

Pure Transformation [10,11]: At lower frequencies the only practical way to transform from the de-Q'ed stage to the interface impedance, will be with a transformer—this is usually constructed by winding wire or a transmission line with a magnetic or air core hooked up to give a straight sort of transformation. One is usually limited to fixed transformation ratios, which may limit how optimal a match one may achieve. A picture of one such transformer is in Figure 20.19.

At higher frequencies, lumped element LC matching sections may be used, and at higher frequencies still, circuit elements will be replaced by printed transmission lines (first replacing lumped inductors, then replacing lumped capacitors). Tapers, quarter wave transformers, and multistage quarter wave transformers can all be employed at higher frequencies—allowing much greater flexibility at the expense of increased losses and increased effects of parasitic inductances and capacitances. At higher frequencies, much of the matching circuit may be implemented monolithically on a single chip due to the increasing

FIGURE 20.19 HF wideband transformer. (Courtesy MKS Instruments. Used with permission.)

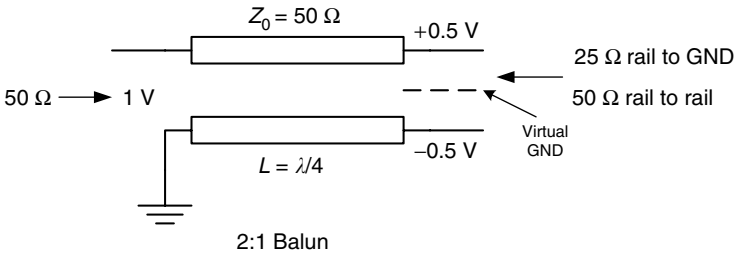


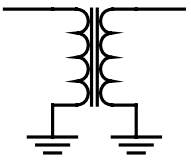
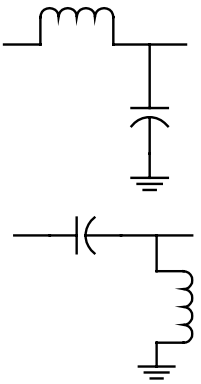
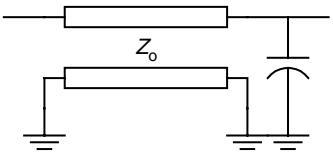
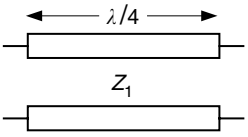
FIGURE 20.20 Diagram of 2:1 Balun based upon quarter wave transmisison line.

role of parasitics, and the smaller dimensions of the elements of a matching network. A summary of types of transformers and typical frequencies of use are shown in Table 20.5.

Section 3: After the interface impedance is achieved by the main portion of the transformer, the designer can either place a DC blocking capacitor (usually is well beyond self-resonance) if the impedance is the system characteristic impedance, or this section can also transformatively combine two transistor stages. The most popular transformative combiner is the Balun (see Figure 20.20) [12–14], which offers a number of additional advantages that explain its popularity. A balun structure is an equal power splitter where the two signals are 180 degrees out of phase with each other. In effect, if used to combine two devices, one device will be "on" when the other is "off" or approximately the case. Between the devices is a "virtual earth"—meaning if one device is putting out +20 V the other side is putting out −20 V. Between the two is the zero crossing. This will mean that side-to-side capacitors can be employed (reducing parts count), and the "effective" impedance is doubled in the circuit. This technique has become so useful that some devices are packaged specifically to be used in this configuration.

In general, all matching networks involving a power device will have the above general characteristics. Specifically, the input and output match sections have specific, different jobs to perform,

TABLE 20.5 Transformation Table

Type	Frequency Range	Notes
Pure transformer 	DC–VHF	Magnetic core Inherently broadband Fixed transformation amount (e.g., 1:4, 1:9, etc.)
LC lumped element 	HF–VHF	Inductor lumped element
Transformer + lumped capacitance 	VHF–UHF	Use higher impedance line to create approximation of inductor with lumped capacitor
Transformer based, etc. 	UHF and above	Quarter wave matching Sub-quarter wave matching in multiple sections Taper LC analog with transmission lines

and require separate treatment in order to describe the optimizations and tradeoffs inherent to the designer.

20.15 Output Match

The job of the output match is to realize the performance determined in any of the ways described above. Typically, as mentioned previously, the designer will be interested in the maximum power transfer for the output of the part. The main points of optimization and tradeoff for the output match are primarily:

1. Maximum output power of the circuit.
2. Flatness of compression across the band or bands of operation: In other words, the same maximum power, compression, and efficiency across the band.

3. Efficiency.
4. Secondary effects.
   - Gain (though it may be trimmed and optimized on the output, it is best, first addressed on the input in most cases, or can be traded off for output power). By changing the load line and delivered power, gain can be affected greatly.
   - Linearity: Changing the load line alters how the device goes into compression: the voltage saturates or the current saturates first, and it will affect how linear the device is, and so on.

20.16 Input Match

While much attention is rightly paid to the power one can extract from a device, proper optimization of the input match will help with many other items, such as linearity, bandwidth, gain, and input return loss. Depending upon the input match for the device, with nearly identical output powers, a circuit can be made to have high gain, good linearity, broad-band operation, or other characteristics. The amount of flexibility and the types of parameters that can be optimized is determined by the device technology selected. Since there are two basic device technologies that are employed (FETs and BJTs) there are two approaches to this problem, but first the tasks of an input circuit must be detailed.

The main tasks of an input circuit are the following:

Provide an impedance for the previous stage to drive. This is very important, as a poor input return loss may require some sort of isolation, or a 3 dB hybrid or other method to control the load line of circuits driving the stage. Failure to account for this may result in serious issues. For instance, if a driver stage that is optimized to work into a 1.2:1 VSWR (appx. return loss of 20 dB), and the stage it is driving presents −6 dB return loss or −6 dBr, there may be problems with gain flatness and the output power of the driver stage without some sort of mitigating circumstance (larger driver, optimized circuit, tuned circuit, interstage isolation).

Provide the gain and gain flatness across the band (primarily, assuming the output load line is making the circuit flat). In narrowband circuits, a high, yet flat gain may be achievable, and a broadband circuit with some sort of progressive mismatch may need to be employed to flatten the gain, as much as one can, across the specified bandwidth.

In the case of a FET, a great deal of system linearity may be gained if some gain and input return loss is sacrificed. For example, one particular LDMOS FET may be able to achieve a –45 dBc Intermodulation distortion when the match was optimized and traded off against input return loss (−15 to −6 dBr), and gain (15–11 dB).

20.17 Input Match: Bipolar Transistors

A bipolar transistor is probably the least flexible in the use of the input match. In fact, most circuits will simply create an input match that will have as good as an input return loss as a realizable circuit can achieve. Since the bipolar transistor is a current driven current source at its heart, the input match impedance tends to be low (see Figure 20.21), made lower and more reactive by the die and package parasitics. At UHF and higher frequencies, it is common to have some pre-matching in the package of the device, which makes a good input return loss realizable. Since the input behaves as a diode in series with a resistor, the input impedance is easier to drive than a FET (described in Section 20.18).

20.18 Input Match: FETs

A FET is extremely flexible within limits, largely due to its (lack of) isolation and the input capacitance (see Figure 20.22). Driving the input of a FET is a lot like driving a capacitor, and is the source of its

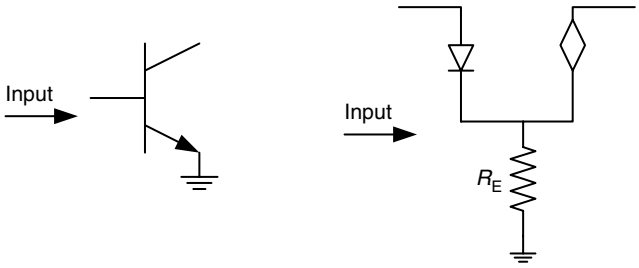


FIGURE 20.21 Input model (without parasitics) for bipolar transistor.

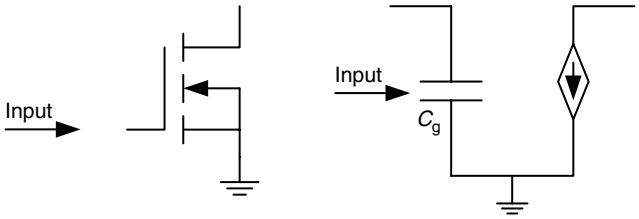


FIGURE 20.22 Input model (without parasitics except gate capacitance) for FET transistor.

flexibility, as well as frustration. One has to be aware that during the RF cycle, the input capacitance will change, and due to its high reactance, can form a resonance with the circuits one places for the match—out of band and at lower frequencies. Numerous techniques have evolved for dealing with this reactance—all involve placing fixed elements in parallel with the gate or in series with it, and reducing the potential for unwanted resonances with the input capacitance due to the matching topologies or bias feed. For sufficiently high impedances, a shunt capacitor is sometimes used to "swamp out" the variability of the input capacitance—reducing the input impedance as a tradeoff. Another method is to place a resistor from gate to ground—which has the benefit of reducing the gain at lower frequencies, though this can affect the gain in band. Other method includes placing a resistor in series as well, which increases stability and makes a matching network a resistive divider (this can result in very large and expensive resistors and significantly reduced gain).

Regardless of the difficulties of FETs, they offer tremendous flexibility in the input match. With the right source impedance, one can maximize gain and input return loss. With some intelligent tradeoff, one can sacrifice some gain and introduce some additional mismatch and maximize linearity instead. As long as the design avoids introducing an unwanted resonance with the input capacitor, FETs offer the designer a large menu of choices to optimize for the application.

Harmonic Terminations: A very important practical consideration for matching networks is the way the match terminates its harmonics. For narrow band circuits, where all the harmonics are outside the band of operation, and the device has enough inherent bandwidth to be able to support harmonics, the designer has additional flexibility to present specific impedances at the harmonics. Properly executed, harmonic termination affects circuit more than the level of harmonics measured at the output of the HPA. It can have large impact upon efficiency and linearity of the amplifier. There are a number of methods of terminating the harmonics usually involving some sort of resonant reactive circuit that is terminated in a resistor or is entirely reactive. The harmonic termination circuit must be able to reliably handle the high circulating currents and peak voltages that is the stored energy in such a resonant circuit (components can get quite warm if the designer is not careful and the harmonic level is high). When using devices with low gain or no gain near the harmonic frequencies, this technique may not be useable or have reduced effectiveness. A good discussion of various devices and their ranges is in Dye [5].

20.19 Bias Network and DC Feed Design

The bias network has two jobs. It provides both the main voltage used to power a device as well as the amount of DC required at the input to achieve a particular quiescent operating point in a BPM. Bias networks vary in circuit topologies, complexity, and features depending upon the device, operating environment, stability, cost, and class of operation. Biasing a device may seem a simple task, at first glance, and in many ways it *is* though it is not without its pitfalls for the unwary.

It is usually good to break a discussion of the feed structures into two areas. First, is the feed structure itself—the way to bring the voltages to operate the power transistor to the transistor with maximum positive effect. Second, how the power is to be prepared needs to be contemplated (specifically the input power, and the circuitry required for each class of operation) [1] Chapter 3.

20.20 DC Feeds

A DC feed is the circuit that takes the bias voltage or rail voltage and connects it to the device. This structure will source energy, and must present enough charge storage at the bandwidth rate of the signal to be passed, or the circuit will have distortion due to transient voltage sag. The structure of a feed is similar to a filter, and should have the impedance as high as possible in the RF pass band, and as low as possible at the video (i.e., signal) termination.

For proper bias feeds, consideration of the application as well as the intended bandwidth of the signal that is to be passed. Different applications will require different feed structures since the bias feed has great impact upon circuit stability and function. The ability to pass the entire signal bandwidth without distortion (video bandwidth). The ability to sustain a pulse without creating large voltage or current transients on the device. Theser are several of the considerations to be made while designing a biasing network.

20.21 Output Feed

Regardless of class of operation, at minimum, a circuit must supply at least one voltage to the power transistor output section creating the potential difference between the collector and emitter of a bipolar transistor, or drain and source of a FET allowing the device to operate and amplify. The circuitry that does this task must be able to source enough current while maintaining constant voltage to amplify the input signal with minimal distortion. Three simple output feed circuit topologies are depicted in Figure 20.23. Common among all of these (as well as all feeds of this type) is to provide a low impedance from DC to the maximum signal envelope bandwidth while maximizing the impedance at the operating frequency of the BPM. While NPN transistors are shown in Figure 20.23, the device can involve any sort of active amplifying device including FETs, RFICs or even vacuum tubes—but the principles remain the same. A low impedance source from DC to the maximum signal frequency allows the

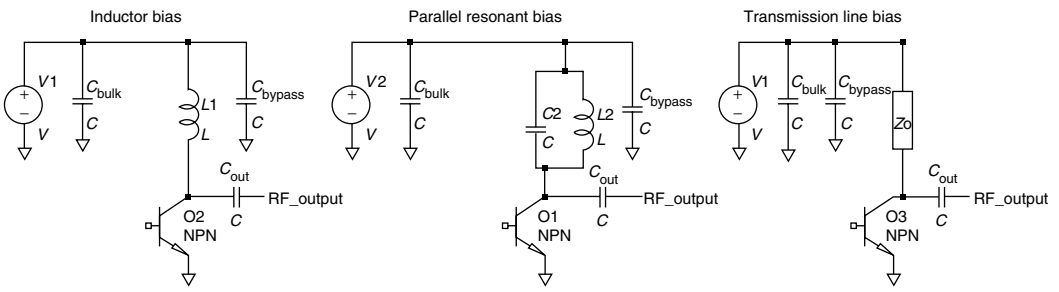


FIGURE 20.23 Three examples of output bias networks.

device to draw current with minimum voltage drop as required for the desired output power. At the RF operational frequency, as much isolation as possible is required to prevent unwanted RF feedback (to prior stages or adjacent modules) or EMI in other parts of the system as well as changing the loading characteristics of the device. A discussion of this is in Cripps [2] Chapter 10, Ostroff [3] Chapter 3, and Kenington [1]. Simply stated, examining Figure 20.23, the coupling circuit from the rail to the device should behave as an open circuit with the bypass capacitor, C_{bypass}. Due to parasitic properties of capacitors, the bypass capacitor(s) should ideally be at or near self resonance (where the impedance is close to zero) at the operating frequency of the amplifier. The bypass capacitance should be connected as *physically close* as possible to the main feed component (L_1, L_2 or the transmission line in the examples in Figure 20.23). This minimizes parasitics that could de-tune the feed circuit. The bulk capacitors source localized charge at various frequencies to maintain the rail voltage under envelope frequencies (sourcing a large current at some envelope frequencies from a remote power supply would involve very large inductive transients causing very large transient voltage drops). The bulk capacitors should be as close to the feed component as possible (right after the bypass capacitor) to minimize even this small inductance.

In Figure 20.23, the first example has a single inductor as the feed component. This feed structure is possibly the simplest of all the output bias circuits presented here. L_1 is a short at DC and should be designed to have a reactive impedance of approximately 6–10 times the intended load impedance at the operating frequency. The self resonance of an inductor should be kept in mind when designing it for this sub-circuit. At self resonance the impedance is limited only by the "Q" of the component and the circuit and is usually quite high (hundreds of Ohms) and can be susceptible to noise. This is why in many implementations, there will be a resistor in parallel with the inductor (not shown) to control its behavior during resonance. Relying on self-resonance is tricky as this characteristic can change in production lots as well as ambient conditions for some component types, so many designers will make the self-resonance much higher than the band of interest, use a parallel resistor with the inductor or use a different topology altogether. It should be noted that for wide bandwidth signals, this sort of feed can have considerable inductive reactance, and in most cases limits this technology to more narrowband applications.

The second topology "fixes" some concerns with the first by adding a capacitor in parallel with an inductor C_2. The value of C_2 should be selected to provide a resonance with L_2 at or above the RF frequency of operation (this topology usually has a resistor in parallel with this tank, omitted for clarity in Figure 20.23). A slightly wider signal envelope can be supported than the first example since the inductor can be made much smaller than if it were the sole component.

The third example uses a transmission line one quarter wave long. With the bypass capacitor at RF frequencies where the transmission line is a full quarter wave length, the short of the bypass is transformed into a near open circuit, and at lower frequencies it will resemble a low value inductance. This technique can work well at higher frequencies (especially UHF and Microwaves), and can be very wideband, though for some applications whose signal envelope involve really high transient currents, such as in a pulsed application, other topologies must be pursued. Ostroff [3] has a discussion of feeds and pulsed applications.

20.22 Input Bias

The input bias feed of a BPM uses some of the same concepts as the output feed, though due to the special requirements of particular devices, the actual implementation will be slightly different. As above, a bias feed needs to provide a high impedance at RF frequencies, and a low one at the envelope frequency of the signal that is amplified. The main difference is that the feed must be designed with the type of device in mind. For instance, the designer will have a different sort of feed for a bias circuit designed to create a current source (as in a bipolar device) or a circuit used to create a voltage source (as in a FET type of transistor).

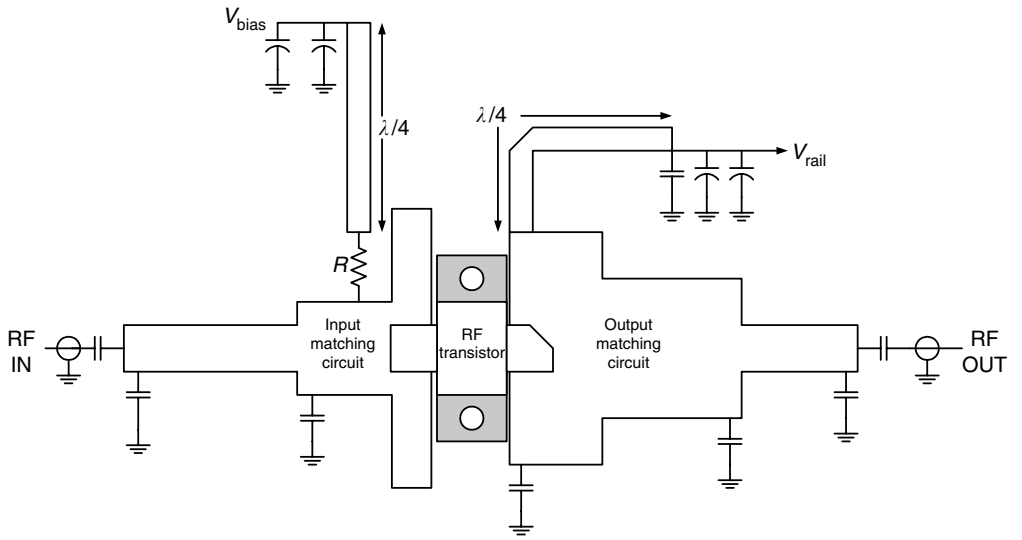


FIGURE 20.24 Example bias feed structures for FET.

20.23 FET Feeds

A FET structure does not require a particularly low impedance to its gate, so a resistor is commonly used (e.g., shown in Figure 20.24). In some applications, such as a pulsed or CW applications, special attention needs to be paid to both its effects on transients (pulsed might create a RC time constant with the bypass capacitors), and the effects of inductance on the gate capacitor and possible stability issues (see stability section). Also, excessive gate resistance may cause other unwanted effects since at RF and high signal envelope frequencies, there needs to be *some* current flow into the input capacitor.

20.24 Bipolar Feeds

A bipolar transistor requires a very low impedance feed (also known as "stiff" voltage source that is capable of a very low impedance current sourcing given the exponential relationship between I_b and V_{be}) so the feed to the base will not resemble that of the FET, and the main characteristic will be that it must be a low impedance current source at DC and up to the envelope bandwidth (one example is Figure 20.25).

20.25 Special Application Considerations

20.25.1 Modulated Signals

For a modulated signal, such as a communications amplifier, a good rule of thumb is to make sure the circuit can pass two to three times the amplitude and bandwidth envelope of the intended signal. In a corrected system the requirements on this may be wider or narrower depending upon the overall system design [1,2].

20.25.2 Pulsed Signal

A pulsed signal is a very sensitive signal. Of special importance is to the leading the training edge, so the amount of transient has to be very low—a large coil or length of transmission line will generate large spikes in voltage during the ramp up and ramp down which causes unacceptable stresses on the device as well as distortion. Pulse applications usually also require enough charge storage to be able to sustain the

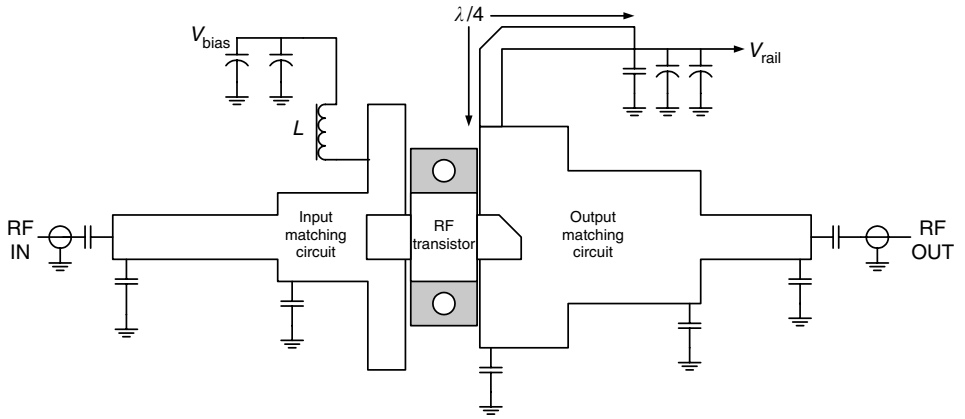


FIGURE 20.25 One example of bipolar bias feed.

pulse width without significant droop due to discharge of the capacitor bank supplying the high frequency energy of the pulse [3].

20.25.3 CW/Narrow Band Signals

These signals require very little video bandwidth since the bandwidth of a CW signal is, by definition, narrow. Depending upon the application, there may be some start-up charge storage required, and consideration of stability in nonideal loads needs to be considered. Otherwise the feed should be as narrow as possible.

20.26 Bias Circuitry

As discussed above, the designer must select the class of operation and then design a circuit to provide the correct operating point for this. Each class of operation has its own considerations, and the application will determine the type of temperature compensation. An excellent discussion of bias circuitry is in Dye [5].

Class A will generally require a constant current bias source to fix the operating point regardless of the RF drive and output. The circuit will have to have some sort of feedback to keep the output current at a fixed level, or a circuit must be created whose current is large compared to the amount of output power required (i.e., it is quasi class A in that the operating point movement is minimal).

Class AB or B operations require some form of positive biasing—though the operating point will move with RF drive. This will require an "open loop" circuit with some sort of compensation over ambient conditions. An example of a class AB bias is shown in Figure 20.26. For a FET device a simple resistive divider is sufficient in most cases, nowhere near the complexity of the bipolar bias circuitry, except that provisions need to be made for stable input voltage to the bias circuitry in some manner (e.g., local regulation) [5].

Class C will generally require a negative bias of some kind—or in most cases, the input of the transistor is tied to ground with an inductor or resistor, which is sufficient to keep the conduction angle correct. This is the simplest bias.

20.27 Thermal Compensation

Thermal compensation is a concern for most bias circuitries because the bias point for a particular quiescent current will change with temperature. All devices experience this phenomenon, and in some

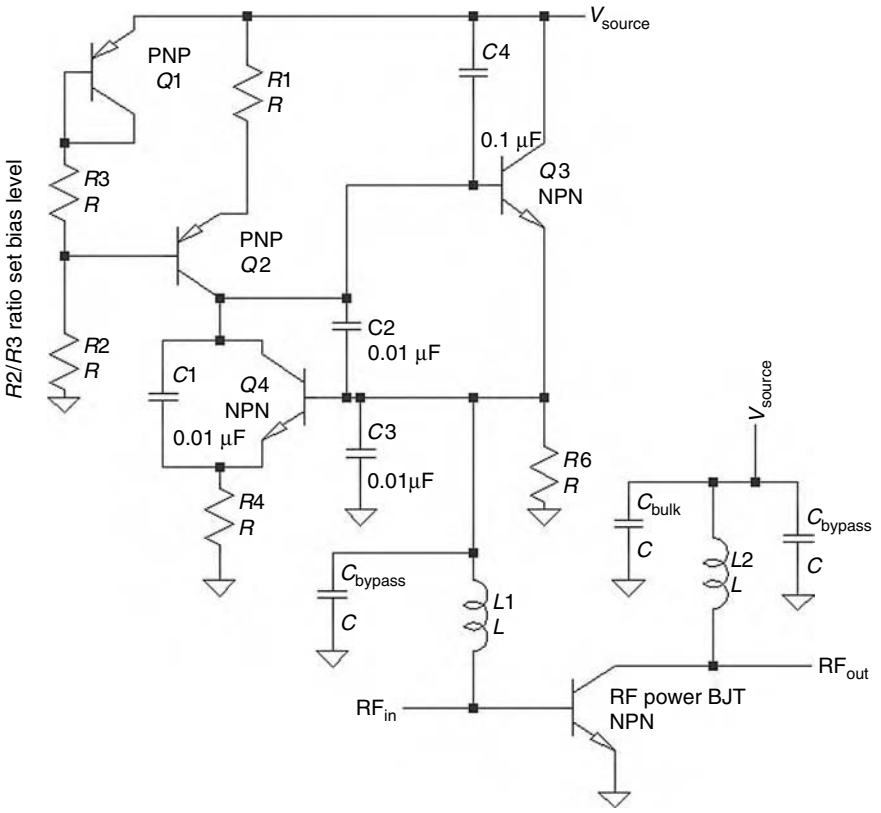


FIGURE 20.26 Example of bias circuitry for biploar transistor, class AB.

cases (Bipolar transistors) the current gain will change with a positive coefficient eventually causing device destruction if the thermal effects are not accommodated in a bias circuit. While a BJT will change at fixed rate of −2.4 mV/C, a FET has an inconsistent bias point rate of change with temperature and therefore will have to be empirically measured over the expected operating temperature range for the device chosen.

For a FET, a discussion of methods of temperature compensation can be found in Dye [5], Chapter 4.

20.28 Dividers and Combiners

Power combiners and dividers are all similar devices, with the exception of the direction of power flow; they are also referred to as couplers. Power combiner parameters that are important to the designer are: insertion loss, isolation, phasing and amplitude balance between ports, frequency bandwidth, and port return loss. Isolation is an important factor in combining and dividing. It is thought that isolation should always be maximized, but in some applications low complexity will be more important than the benefits of isolation (graceful degradation, stability, effects of one amplifier on its neighbor). Linear amplifiers will typically benefit from higher isolation, but in cost and size sensitive applications, such as industrial heating, this may not be as important. Figure 20.27 shows an example of a three way combiner implemented at HF frequencies.

High power combiners tend to be based upon quarter wave transformation and therefore can be realized by transformers, lumped elements, planar, coaxial or any number of transmission line derivatives. Table 20.6 lists the general attributes for combiners used in amplifier system design. In general, there are

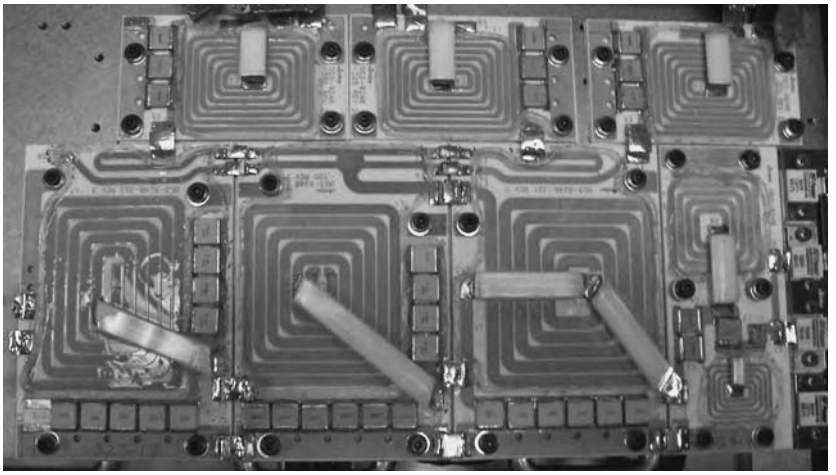


FIGURE 20.27 Three way HF coupler. (Courtesy MKS Instruments. Used with permission.)

TABLE 20.6 Table of Combiner/Divider Types

Combiner Type	Attributes
0°, In phase combiner	• Wilkinson style is the most popular • Can provide equal or arbitrary power division • 3 dB insertion loss for equal power division • Provides high port-to-port isolation with isolation resistor (more loss) • Can be implemented in planar, lumped element, transmission line and transformer configurations. • More than 2 ports are straightforward (though isolation resistor interconnects can be problematic) • Bandwidths of 20% are typical • Broader bandwidths can be achieved by using multiple sections
180° Combiner	• Magic T., rat race hybrid or Baluns are common forms • Provides high port-to-port isolation • Equal power split (3 dB) • A variety of topologies allowing many bandwidths • Can be implemented using planar, lumped element, transmission line or transformer configurations • Can suppress second harmonic on output
90° Combiner	• 90° hybrid, branch line hybrid, Lange coupler, quadrature coupler or directional coupler • Bandwidths of 20% are typical, use multiple sections for wider bandwidths • Equal power split (3 dB) • Provides good input return loss when output ports are terminated in the same impedance • Can suppress 3rd harmonic on output

three main types of combiner/dividers: in-phase (0°), quadrature (90°) and 180° types all summarized in Table 20.6 [15,16].

A typical Wilkinson combiner and design equations for transmission line structure are given in Figure 20.28.

A lumped element version of the Wilkinson combiner or any structure involving quarter wave transmission lines for that matter, can be synthesized in lumped element form using quarter wave length transmission line equivalent circuits indicated in Figure 20.29.

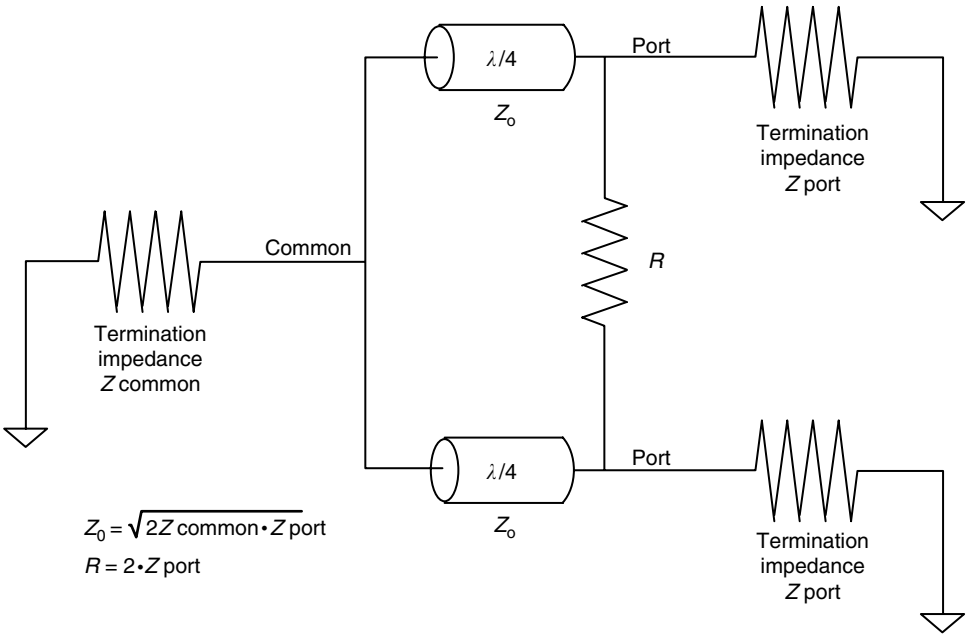


FIGURE 20.28 Wilkinson power coupler structure.

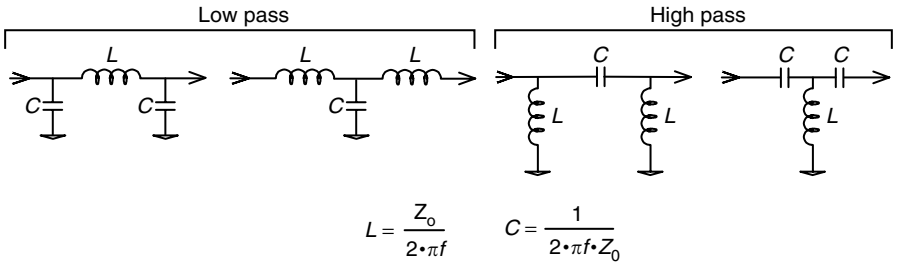


FIGURE 20.29 Transmission line lumped equivalent.

A *N*-way Wilkinson combiner with design equations is depicted Figure 20.30; it works under the same principles as a two-way Wilkinson.

A 180° combiner can be implemented in many ways aside from the traditional balun. One wideband method is using a magic-t configuration shown in Figure 20.31 or a rat-race configuration, an example depicted in Figure 20.32. A rat-race combiner can be set up either as a 0° or 180° coupler. The transmission line sections can also be implemented using lumped element equivalents.

For HF up to UHF frequency applications, magnetic core transformers have been used with great success. These types of combiners provide excellent power to size density compared to other combiner technologies (an example is Figure 20.33). Lot-to-lot consistency over a large production run of the magnetic material can be a problem and can be solved with a good vendor. There are other configurations for input and output impedance, number of ports, and phase relationships between ports.

The 90° coupler is a very popular coupling structure and there are a large variety of their implementation types. It offers a stable and controlled impedance at the combined port provided the isolation is good and the load on the split ports are identical. Transmission line, transformer, and lumped element implementations of the 90° hybrid are shown in Figure 20.34.

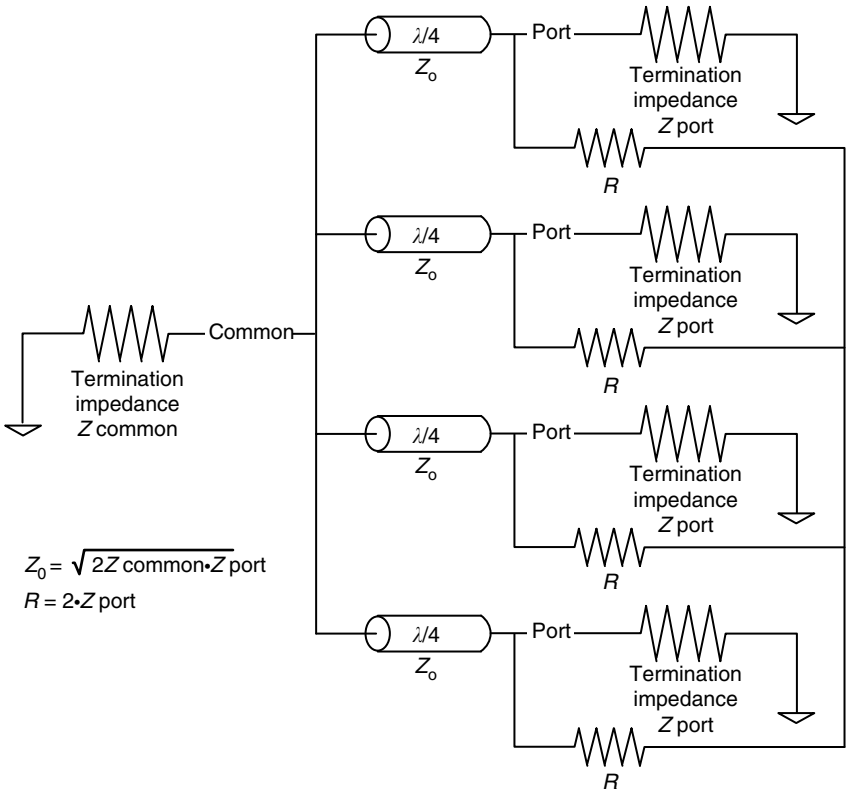


FIGURE 20.30 *N*-way Wilkinson.

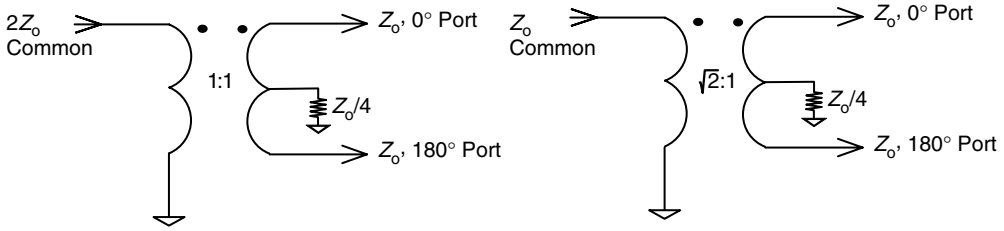


FIGURE 20.31 Magic-t 180-degree coupler.

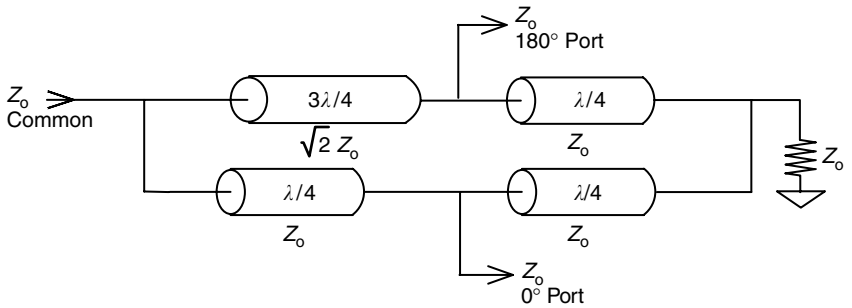


FIGURE 20.32 Rat-race coupler.

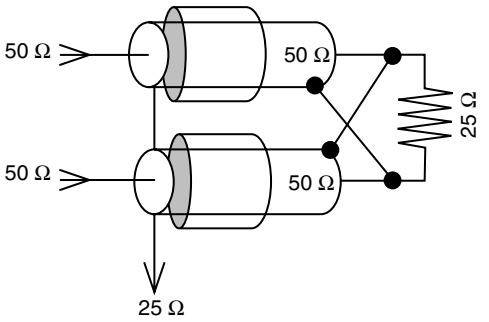


FIGURE 20.33 Magnetic coupler example.

20.29 Design for Reliability

The reliability of RF power amplifiers is a complex subject, primarily due to the large influence of the application on the failure modes. The RF transistors are often thought to be the problem, but they are commonly "innocent victims" of some other issue. A taxonomy of PA failure modes is shown in Figure 20.35.

Reliability must be considered early in the PA design cycle. Many designers use standard component derating guidelines [17] developed for the industry in which the PAs are used. These guidelines are usually based on experience from field failures or reliability testing at the system level. A communications amplifier will experience a very different environment from that of an RF plasma generator or MRI amplifier.

20.30 Transistor Qualification

Transistor selection and qualification is an extremely important ingredient to a reliable PA design. This is mainly due to the relative complexity of the transistor compared to passive components such as capacitors, resistors, and magnetics. A transistor datasheet is not capable of describing the survival of the device under stressful conditions. After a transistor is selected for performance, it should be qualified through a series of additional reliability tests [17], including:

- Die attach x-ray or ultrasonic scan
- DC safe operating area
- RF limit test
- RF hot spot junction temperature/thermal resistance
- High temperature reverse bias (HTRB) test
- Power cycling (especially if required by application)

These tests can be performed in-house, or by an outside lab. It is possible to use manufacturer's data, but it is better to have an objective set of test results (especially when comparing multiple vendors). The failed devices should be analyzed to root cause if possible. The results of these tests will prove invaluable during subsequent issues found during system level tests or root cause analysis of field failures.

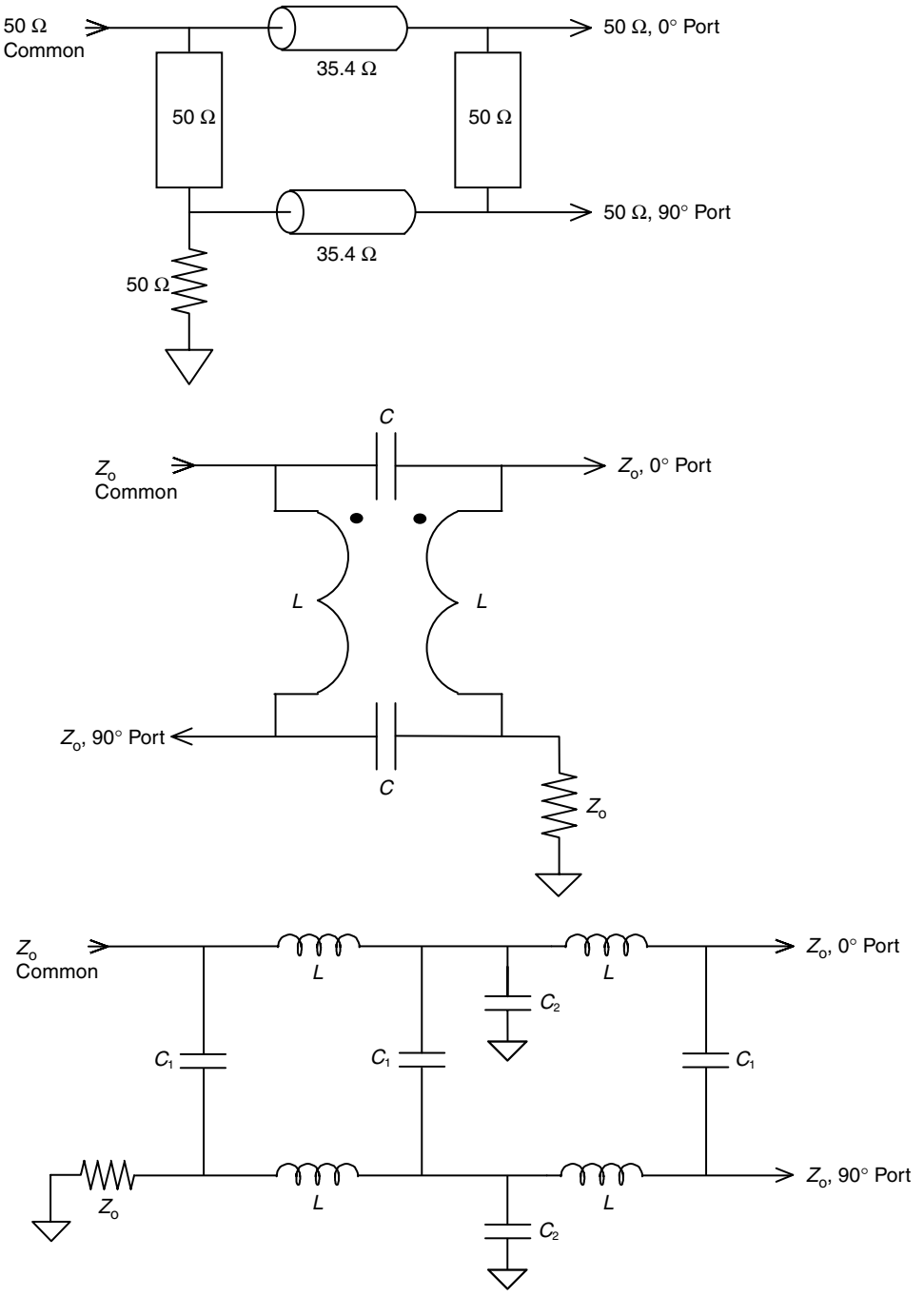


FIGURE 20.34 90° Coupling structures.

20.31 Transistor Failure Modes

Most catastrophic failure modes of the transistor die (the active silicon chip) start with a short circuit caused by local melting of the silicon. This can lead to an open circuit in some packages when the bond wires fuse open. This local melting can be brought about in several ways.

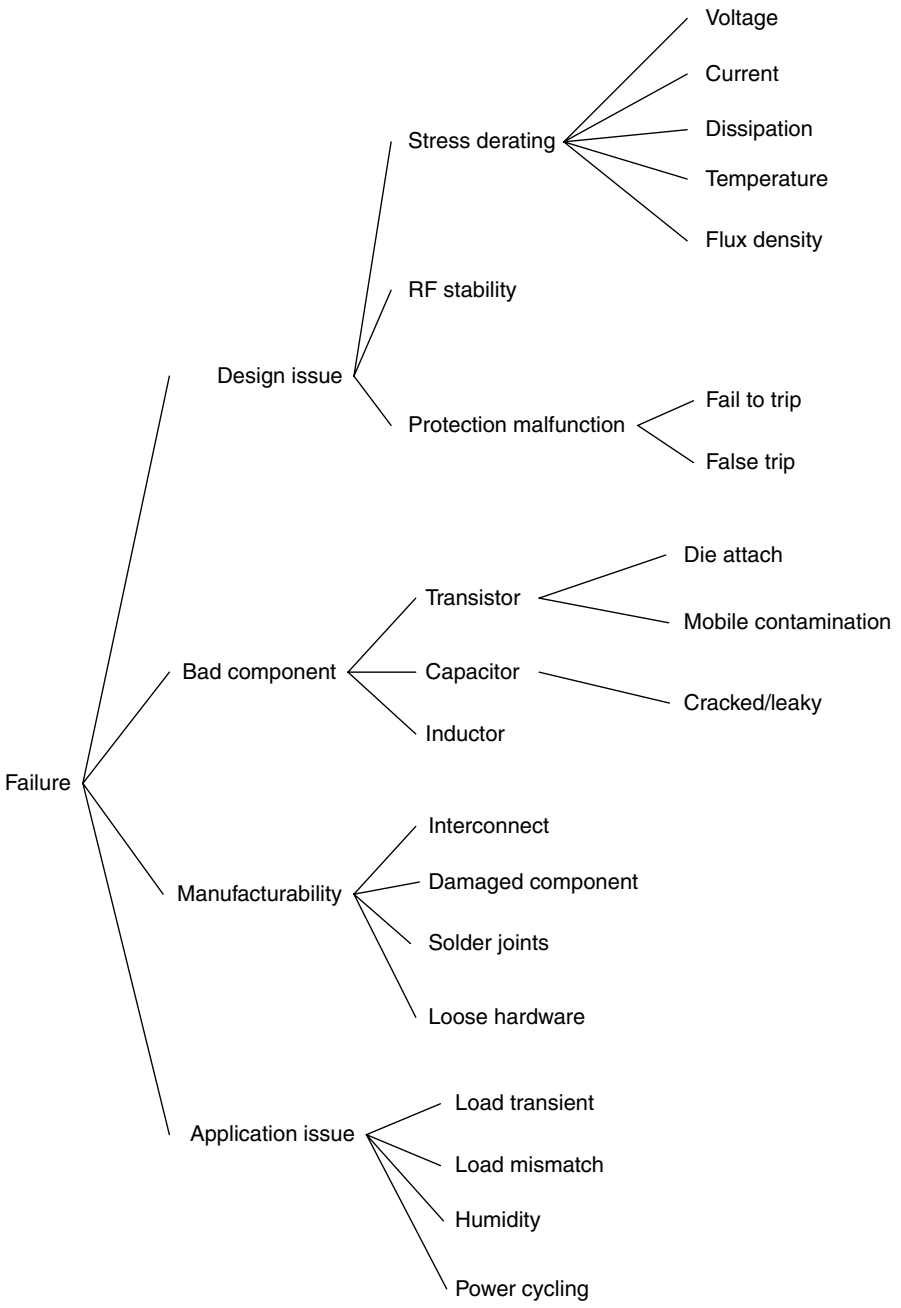


FIGURE 20.35 RF PA failure taxonomy.

20.31.1 Over Voltage Breakdown

Under high stress where voltage is maximized, a transistor may break down on the input or output. A data sheet will usually specify the maximum DC voltage that can be sustained without a breakdown—this number should not be exceeded by the peak voltage or the device may accumulate damage (or fail instantaneously). For MOSFETs, exceeding the gate-source voltage limit will cause oxide damage [18]. For bipolars, exceeding the base-emitter breakdown voltage will cause hFE degradation [19].

On the drain-source junction (for a MOSFET) or collector–emitter junction (for a bipolar), avalanche breakdown on these terminals will lead to local hot-spotting of the device. If the breakdown continues long enough, the local hot-spot temperature will exceed the temperature at which the silicon becomes intrinsic (i.e., non-semiconducting). In this state, the terminals will be connected by a very low impedance region which funnels most of the device current into a narrow "pipe" of soon-to-be molten silicon. This failure mode may look indistinguishable from overdissipation except for the region of the die where the silicon is melted (often near regions of high electric field such as edges of the active area).

20.31.2 Over Dissipation

For some output VSWRs, the device dissipation will be very high, raising the junction temperature above safe limits which will cause thermal runaway in the case of a bipolar transistor, or overdissipation in a FET. The amount of dissipation and the amount of time in the condition are both important factors in determining the proper response to the event. Transistor devices have a number of thermal time constants, and if the time is short (such as under 50 μs), the transistor may be able to handle the situation without much difficulty. If the dissipation is sustained, the power needs to be either removed (switching off the power) or reduced to safe levels (such as with power fold back).

20.31.3 Over Current

Each device is rated for an amount of current, and if exceeded, may cause damage to the device. In general, the first two mechanisms tend to be more likely, but in pulsed applications, a large current transient can cause damage in a part without creating an average junction temperature that exceeds the rating. In most cases, this is indistinguishable from overdissipation. Sustained high current operation can also cause weakening of the metal interconnect via a mechanism known as electro-migration [20].

Transistors can also be subject to environmental stresses such as moisture or corrosive agents. Mobile ions introduced during or after transistor manufacture can lead to gradually increasing leakage currents under voltage bias [21]. High temperature reverse bias (HTRB) testing can usually detect this type of defect.

20.32 Other Considerations

It is not enough to consider only design-related issues (e.g., derating, stability, protection circuits). A good designer must also understand the methods used to build the amplifier and potentially stressful aspects of the application.

20.33 Manufacturability

High power RF amplifiers tend to use a combination of surface mount and through-hole components. A typical manufacturing process solders the surface mount components first (often with automated equipment), followed by the through-hole devices (usually soldered by hand). It is very important to make sure that the hand-solder operations do not damage the delicate SMT components (e.g., ceramic chip capacitors) via asymmetric application of high temperatures. Many companies have PCB design rules restricting the location of SMT capacitors with respect to mounting holes, board edges, and large hand-soldered components. In general, it is a good idea to minimize the number of hand-assembly operations, as they tend to be more error prone.

Many designers like to use the backside of the PCB as a stiff multi-point ground which is directly bolted to a chassis or heatsink. This precludes the use of through-hole components, causing fairly heavy magnetic components to be "lap-soldered" to PCB lands. If this is necessary, it is important to ensure that there is a mechanical mounting scheme (do not rely on the solder as a mechanical fastener) and that the leads are not soldered in tension.

Other areas for reliability improvement include the reduction of cable interconnects and mounting hardware. These items tend to be overlooked until the PA is analyzed at the system level. At this point, it is usually too late to make significant changes.

20.34 Application

Finally, it is extremely important for an RF designer to thoroughly understand the normal and abnormal conditions placed upon the PA during operation and standby modes. The PA specification tends to focus on the normal operating conditions, so it may require some additional research to determine what happens when failures occur at the system level. If a system-level failure mode and effects analysis (FMEA) exists, it can be an important resource. If the FMEA does not exist, developing one is a worthwhile exercise. This extra investigation when used to correct found issues will pay big dividends in reliability. Pay extra attention to the following:

- Ambient temperature and humidity (including standby mode)
- Power-up and power-down sequence and transients
- Failure modes of controls and power supply subsystems
- Load impedances presented to the PA during combiner, filter, antenna coupler, and so on failure

20.35 Tactics to Increase Reliability

20.35.1 Isolators/Circulators

At higher frequencies and narrow band applications, (UHF+) the size and cost of isolators and circulators are low enough that they are a good consideration for HPAs. They form extremely cheap insurances, and will prevent load reflections from making it back to the device. Consequently, the amount of headroom required for ruggedness in a system protected by one of these devices is much reduced. The main consideration in a system of this sort will be the overall rating of the circulator or isolator—depending upon how long a circulator will be operating in a non ideal load, and to what degree, will determine the size of the water load. It is also rather simple to used the third port in order to measure the reverse power to help monitor the VSWR directly.

20.35.2 Instabilities

One of the largest concerns most people have regarding HPA design is instability. While engineering of parts, simulation tools and other methods have improved, diagnosing instability in a weak or strong nonlinear circuit remains very difficult to solve. Using traditional small signal tools, such as *s*-parameters, *k*-factors and other related methods, generally do a good job of determining stability in a small signal or very linear circuit, but do not do a good job of predicting or solving many instabilities in typical HPAs. There are many forms of instabilities and strategies for addressing them, in many ways, this is one area of amplifier design that remains largely empirical. In general, instabilities can be exacerbated by the following non-comprehensive list:

- Wide load ranges in PA operation where circulators or isolators are not used
- High efficiency amplifiers (due to the large amount of transient voltages and currents) as well as low loss
- Low loss combining structures or combining structures with low isolation
- Variable DC supply powering the PA
- Uncontrolled out of band impedances

A good paper dealing with very practical and theoretical treatments is Reference 22.

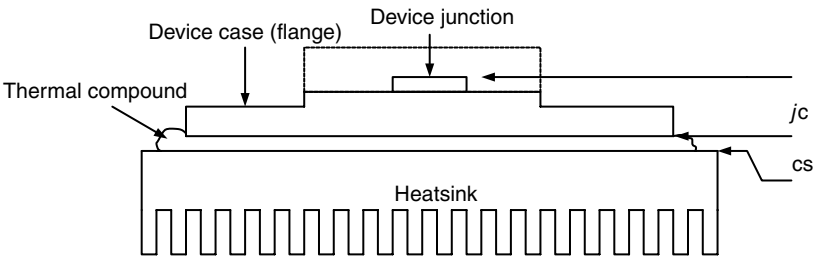


FIGURE 20.36 Diagram of transistor mounting and thermal resistances.

20.35.3 LF instabilities

A low- frequency instability is an instability that appears below the band of operation, usually much lower. As the natural gain of transistors rises about 6 dB per octave of frequency, there is usually a large amount of gain well below the band of interest. A number of techniques exist for reducing this potential source of oscillation, the most popular being the application of feedback.

20.35.4 HF Instabilities and Bias Feed Interactions

A high-frequency instability is an instability in band or above band, usually due to an undesirable resonance caused by circuit elements, feedback or unwanted RF coupling.

20.35.5 Thermal Considerations (Minimizing Junction Temperature)

One of the most difficult issues with HPA design is getting rid of the heat lost due to inefficiencies. In large HPA architectures, this can be quite a lot of heat—kilowatts are heard of in large designs. Good thermal design is key to keeping temperatures down, and therefore reliability. There are many approaches to keeping the amplifier cool enough to deliver its function without undue reliability issues [3].

When an overall design is considered, the main issue facing most electrical designers is the junction temperature versus the heatsink temperature. Transistors must be mounted to a heatsink, and usually require some sort of interface materials, such as thermal compound (a mixture of silicone and oxide material designed to improve upon having a dry mounting contact), or some types of thermal pads (indium foil, or other form of conductive material). The heatsink is designed to be able to remove a certain amount of heat—and can be simulated for purposes of analysis as a "worst case" thermal mounting.

The primary source of heat is the transistor, and its ability to get rid of waste heat is specified by term Theta jc, which is specified in degrees centigrade per watt of dissipated power. Once the heat is in the case, it must be transmitted to the heatsink, and then removed from the heatsink through the transfer medium—usually air or liquid.

In a sample design (see Figure 20.36), a part may dissipate 100 Watts worst case, have a theta jc of 0.5°C/W, making the junction temperature 50°C above the case temperature. The next junction is the thermal compound, estimated at 0.1°C/W, adding 10° to the total. The rise from the sink to the junction is thereby 60°. If the mounting surface of the heatsink is at a maximum of 80°C, the junction temperature would be 140°C. Depending upon the maximum junction temperature allowed by the device this might or might not be acceptable.

References

1. Kenington, P.B., *High Linearity RF Amplifier Design.* Artech House, Norwood, MA, 2000.
2. Cripps, S.C., *RF Power Amplifiers for Wireless Communications.* Artech House, Norwood, MA, 1999.
3. Ostroff, E.D., Borkowski, M., Thomas, H., Curtis, J., *Solid State Radar Transmitters.* Artech House, Dedham, MA, 1985.

4. Cripps, S.C., *Advanced Techniques in RF Power Amplifier Design*. Artech House, Norwood, MA, 2002.
5. Dye, N., Granberg, H. *Radio Frequency Transistors: Principles and Applications*. Butterworth-Heinemann, Stoneham, MA, 1993.
6. Saleh, A.A.M., "Improving the graceful-degradation performance of combined power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, Vol. MTT-28, No. 10, Oct. 1980, pp. 1068–1069.
7. Wood, A., Davidson, B., "RF power device impedances: practical considerations," *Proceedings RF Expo West*, Feb. 1991 (Rev 2.1), pp. 385–405.
8. Matthai, G., Young, L., Jones, E.M.T., *Microwave Filters, Impedance Matching Networks, and Coupling Structures*. Artech House, Norwood, MA, 1980.
9. Purviance, J., Brakensiek, W., Monteith, D., Ferguson, T., "Matching Structures for High Yield Amplifier Design" *Proceedings of the 1988 IEEE Trans. Microwave Theory and Techniques*, Vol. 36, No. 12, Dec. 1988.
10. Sevick, J., *Transmission Line Transformers*. 4th edn., Noble Publishing Corporation, Raleigh, NC, 2001.
11. Octavius, P., Course, T.P., "Broadband transformer design for RF transistor power amplifiers," *1968 Electronic Components Proceedings* (May 1968).
12. Lee, B.M., "Balanced transistor: a new option for RF design," *MicroWaves*, Vol. 16, No. 6, pp. 42–46, June 1977.
13. Nathan, M., "Transmission line conversion," *Electronics*, pp. 142–145, Dec. 1945.
14. Bassett, R., "Three Balun designs for push pull amplifiers," *Microwaves*, pp. 47–52, July 1980.
15. Pozar, D.M., *Microwave Engineering*, 2nd edn. John Wiley & Sons, New York, 1998.
16. Russell, K.J., "Microwave power combining techniques," *IEEE Transactions on Microwave Theory and Techniques*, Vol. MTT-27, No. 5, pp. 472–478, May 1979.
17. LaCombe, D.J., *Reliability Control for Electronic Systems*, New York, Marcel-Dekker, Inc., p. 126, 1999.
18. Dimitrijev, S., *Understanding Semiconductor Devices*, New York, Oxford University Press, p. 480, 2000.
19. McDonald, B.A., "Avalanche degradation of hFE," *IEEE Transactions on Electron Devices*, Vol. ED-17, No. 10, pp. 871–878.
20. Black, J.R., "Electromigration failure modes in aluminum metallization for semiconductor devices," *Proceedings of IEEE*, Vol. 57, p. 1587, Sept. 1969.
21. Osenbach, J.W., Comizzoli, R.B., and Voris, S.S., "Controlled surface contamination to determine surface sensitivity of HVICs," *24th Ann. Proc. Reliability Physics*, pp. 239–246, Sept. 1969.
22. Sokol, N., *Parasitic Oscillation in Solid State RF Amplfiers—Causes and Cures—Demystefied*, August, pp. 1–8.

Other Sources

1. Vizmuller, P., *RF Design Guide System, Circuits, and Equations*. Artech House, Norwood, MA, 1995.
2. Sayre, C.W., *Complete Wireless Design*. McGraw-Hill, New York, 2001.
3. Sabin, W.E. and Schoenike, E.O., *Single Sideband Systems and Circuits*, 2nd edn. McGraw-Hill, New York, NY, 1995.
4. Wadell, B.C., *Transmission Line Design Handbook*. Artech House, Norwood, MA, 1991.

21

Oscillator Circuits

Alfy Riddle
Finesse, LLC

Figure 21.1 shows a variety of styles and packaging options for RF and microwave oscillators. Oscillators serve two purposes: (1) to deliver power within a narrow bandwidth, and (2) to deliver power over a frequency range (i.e., they are tunable). Each purpose has many subcategories and a large range of specifications to define the oscillator. Table 21.1 gives a summary of oscillator specifications.

Fixed oscillators can be used for everything from narrowband power sources to precision clocks. Tunable oscillators are used as swept sources for testing, FM sources in communication systems, and the

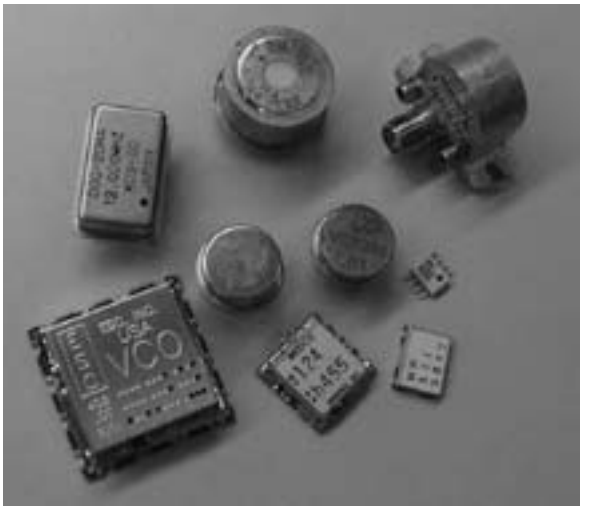


FIGURE 21.1 A picture of various RF and microwave oscillators. Top to bottom and across from the left there is a crystal oscillator and two YIG oscillators, two chip and wire oscillators in TO-8 cans, a microwave IC oscillator, and three discrete PCB VCOs in packages of decreasing size.

TABLE 21.1 Oscillator Specifications

Specification	Characteristic
Power	Minimum output power (over temperature)
	(Flatness over tuning band if tunable)
Frequency	Accuracy (in Hz or ppm)
	Drift over temperature in MHz/°C
	Aging in ppm/time
	Phase noise in dbc/Hz (or jitter in picoseconds)
	Pulling in Hz (due to load variation)
	Pushing in Hz/V (due to power supply variation)
	Vibration sensitivity in Hz/g acceleration
Tunable	Bandwidth
	Modulation sensitivity in MHz/V
	Modulation sensitivity ratio (max/min sensitivity)
	Tuning range voltage
	Tuning speed in MHz/microsecond
Power Consumption	V, I DC
Package Style	

controlled oscillator in a PLL. Fixed tuned oscillators will have a power supply input and the oscillator output, while tunable sources will have one or more additional inputs to change the oscillator frequency. Some tunable oscillators, particularly those using YIG resonators, will have a second tuning port for small deviations. The theory section will provide the background for understanding all the oscillator specifications.

21.1 Specifications

21.1.1 Power Output

Power output and frequency of oscillation are the most basic oscillator specifications [1]. Oscillators with maximal output power are used in industrial applications and usually have more noise due to their extracting as much power as possible from the resonator and thereby lowering the loaded resonator Q. Power output will vary over temperature, so some designs use a more saturated transistor drive or pass the oscillator signal through a limiter to achieve greater amplitude stability. Both of these actions also increase the oscillator noise and cost. Oscillators optimized for low noise, or jitter, usually have low output power to minimize resonator loading and so these designs rely on post-amplification stages to bring the oscillator power up to useful levels for transmitters and radars. As discussed in the theory section, oscillators create more near-carrier noise than amplifiers, so post-amplifiers usually have a minor impact on total oscillator noise. When an oscillator is tunable, the power flatness over the tuning range must also be specified.

21.1.2 Frequency Accuracy and Precision

The frequency accuracy of an oscillator encompasses a large number of sub-specifications because so many things affect an oscillator's frequency. Temperature, internal circuit noise, external vibration, load variations, power supply variations, as well as absolute component tolerance all affect frequency accuracy. We can consider only component tolerances for frequency accuracy and lump all the variations into oscillator precision.

The accuracy of the fundamental frequency of an oscillator is usually specified in ppm or parts per million. So a 2.488 GHz oscillator which is accurate to ± 10 ppm will have an output frequency within ± 24.88 kHz of 2.488 GHz at the stated temperature, supply voltage, and load impedance. Ambient temperature changes also change the oscillator frequency. The perturbation in a oscillator frequency from

temperature is often given in MHz/degree C or ppm/degree C. Manufacturers use several techniques to compensate for temperature changes, such as using an oven to keep the oscillator at a constant temperature such as 70°C, building in a small amount of tuning that is either adjusted digitally or directly from a temperature sensor, and finally resonators can be built with temperature compensating capacitors or cavities [2]. Oscillator components also change with time, which causes a frequency drift due to aging. Aging is usually specified in ppm/year or some other time frame.

Power supply variation affects both the absolute accuracy of an oscillator frequency and the precision with which it maintains that frequency. The sensitivity of an oscillator to power supply variations is called "pushing" and is usually given in MHz/V. Drift in the supply voltage over temperature or with changes in the instrument state affect the accuracy of the frequency while noise on the power supply due to switching circuits will modulate the oscillation frequency through the same pushing mechanism. Time constants in the oscillator bias circuitry will cause the pushing factor to change as the modulation frequency increases, but this is rarely specified. Communication receivers and spectrum analyzers go to great lengths to filter oscillator power supplies.

Load variations also cause changes in oscillation frequency. Because the load on the oscillator output port has some finite coupling to the resonator, changes in the load reactance will change the resonator reactance and so change the oscillation frequency. Typically, a variable length of line is terminated in a standard return loss, such as 12 dB, and the oscillation frequency is measured as the line length is changed. Changing the line length creates a variable load that traces a circle on the Smith Chart. The maximum frequency change is quoted as the "pulling" for the oscillator at the given return loss. Precision oscillators will go through an isolator or a buffer stage to minimize pulling.

Oscillators with cavities and even suspended crystals can be affected by vibration. The vibration sensitivity specification depends on oscillator construction and mounting. Communication systems have been taken down by raindrops hitting the enclosure of an outdoor cavity oscillator. A vibration sensitivity in MHz/g can show sensitivity to vibration, but usually the frequency of the vibration is important as well.

Probably the most common specification of oscillator precision is phase noise or jitter [3]. Phase noise is the frequency domain equivalent of jitter in the time domain. Phase noise will be described in the theory section. Phase noise, FM noise, and jitter are all the same problem with different names. Because an oscillator contains a saturated gain element and a positive feedback loop, it will have very little gain for amplitude noise and a near infinite gain for phase noise. The amplitude and phase variations are with respect to the average oscillation frequency. If an oscillator is measured with a spectrum analyzer with sufficient resolution, the narrow line of the oscillator will appear broadened by noise which falls off at $1/f^3$ or $1/f^2$. The loop gain in an oscillator feedback loop reduces as $1/f^2$ for frequencies other than resonance. The additional $1/f$ factor comes from low frequency modulation within the device or resonator. The phase noise specification is usually given as script $L(f_m) = P_{SSB}(f_m)/Hz/P_C$, which is easily measured with a spectrum analyzer. The original script L definition noted that $P_{SSB}(f_m)/Hz$ was to be the phase noise power in one Hertz of bandwidth, but often people take the spectrum analyzer measurement and call it phase noise because phase noise dominates oscillator noise close to the carrier [4]. In reality, the spectrum analyzer cannot tell the difference between amplitude and phase noise. Whenever the phase noise approaches the noise floor or even a flat noise pedestal, it is likely that a significant amount of amplitude noise is present. In all of the above, f_m denotes the offset frequency from the carrier and corresponds to the frequency modulating the carrier. These same variations are called FM noise when measured with a frequency discriminator.

When digital systems are characterized, the time domain specification of jitter is more common than phase noise. Deviations in expected zero crossing times are measured and accumulated to give peak-to-peak and rms values. These values are given in picoseconds or in UI (unit intervals). UIs are just a fraction of the clock period, so UI = (jitter in picoseconds)/(clock period). Because phase noise shows the phase deviation at each frequency modulating the carrier, we can sum up all of the phase deviations and reach a total phase deviation which, when divided by 360°, also gives the jitter in UIs. This is the same as being able to compute the total power of a signal in frequency or in time. Often communication systems are more sensitive to jitter at certain modulation frequencies than others, so a tolerance plot of phase noise

versus offset frequency is given as a jitter specification [5]. The frequency domain view of jitter also makes it clear why PLLs work as "jitter attenuators." The narrow bandwidth of the PLL feedback loop filters higher modulation frequencies and so reduces the total phase deviations, but only if a significant amount of the jitter is due to frequencies above the loop bandwidth.

21.1.3 Tuning Bandwidth Specifications

For tunable oscillators there is an additional set of specifications. Typically broadband tunable oscillators have their bandwidth specified in terms of minimum and maximum frequency (e.g., f_{Max} and f_{Min}) with the center frequency not mentioned. Narrowband tunable oscillators, those with tuning bandwidths of 10% or less, have their center frequency and bandwidth specified. The tuning range is the voltage range of the tuning port for varactor-tuned oscillators (e.g., $V_{Max} - V_{Min}$) and the current range of the tuning ports for YIG tuned oscillators. Usually the minimum varactor voltage is greater than zero because the varactor diode needs reverse bias to maintain a high Q under the swing of the oscillator signal. The modulation sensitivity is the MHz change per volt at the tuning port. Often the modulation sensitivity is not equal to $(f_{Max} - f_{Min})/(V_{Max} - V_{Min})$ because the tuning sensitivity changes over the tuning range. The modulation sensitivity is usually measured at the center of the tuning range with a small voltage deviation. The modulation sensitivity ratio gives the ratio of maximum to minimum modulation sensitivity over the tuning range. This is especially important for varactor-tuned oscillators used in PLLs because the loop gain will vary by the modulation sensitivity ratio. At low voltages varactors have their maximum capacitance, as shown in Figure 21.2 [6]. The capacitance rapidly decreases as the tuning voltage increases until a minimum capacitance plateau is reached. The large capacitance change at low voltages means that the oscillator will have a more rapid frequency change at lower tuning voltages than at higher tuning voltages. This also means that the modulation sensitivity is higher at the minimum output frequency than it is at the maximum output frequency. The doping profile of a varactor affects its capacitance versus voltage curve. The simplest doping profile is an abrupt junction that gives the curve shown in Figure 21.2. A hyper-abrupt junction C–V curve is also shown in Figure 21.2. The hyper-abrupt curve will give the oscillator a more linear tuning characteristic, or modulation sensitivity ratio closer to unity. Another approach to linearizing a varactor oscillator is to shape the tuning voltage with an analog diode shaping network or with a digital lookup table [7,8]. YIG oscillators have an inherently linear tuning characteristic as given in Equation 21.1 [9,56]. In Equation 21.1, H_O is the magnetic bias field strength in Oersteds (Oe), H_a is the internal anisotropy field, and γ is 2.8 MHz/Oe.

$$f_{YIG} = \gamma\left(H_O \pm H_a\right) \tag{21.1}$$

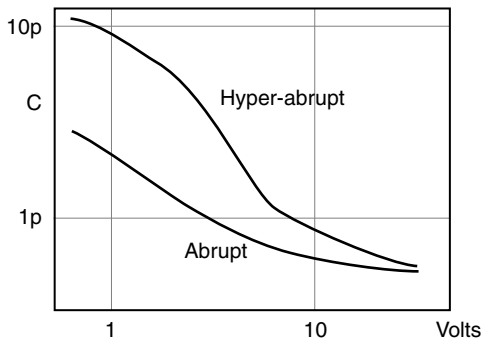


FIGURE 21.2 Varactor C–V plots.

The last specification is the tuning speed in MHz/second. This parameter determines the maximum modulation rate of an oscillator, or its agility in a frequency hopping application. Varactor-based oscillators can tune much faster than YIG based oscillators. For example, for 10 GHz oscillators, the tuning port of a VCO will typically have a bandwidth of 100 MHz, while the FM coil of a YIG oscillator will have a bandwidth of only 500 kHz.

21.2 Technologies and Capabilities

All of the characteristics discussed above depend on three aspects of oscillator technology: (1) active device; (2) resonator; and (3) packaging. Packaging mainly affects the oscillator's cost, size, temperature stability, susceptability to mechanical vibration, and susceptability to interference. Resonator technology mainly affects the oscillators cost, phase noise (jitter), vibration sensitivity, temperature sensitivity, and tuning speed. Device technology mainly affects the oscillator maximum operating frequency, output power, and phase noise (jitter).

Figure 21.3 shows various device technologies and their power versus frequency capability [10–12,57–63]. While Figure 21.3 shows fundamental frequency power, in many cases it is most cost effective to use a frequency multiplier to move a lower frequency oscillator up to a higher frequency [13]. While there is always a power loss from frequency multiplication, there is usually very little noise penalty because for equal resonator Qs, oscillator phase noise is proportional to the operating frequency [14]. Frequency multipliers can be simple resistive diode nonlinearities, tuned varactor diode circuits, or PLLs. Resistive nonlinearities are the simplest and broadest band, but have the most loss. Varactor multipliers can be extremely efficient and are used in the highest frequency multipliers [15]. At very high frequencies PLLs are implemented via subharmonic injection locking so that high frequency external frequency dividers and phase comparators are not needed [16,63]. The practical problem with injection locking is that without an external phase detector it is difficult to verify that the oscillator is in lock.

Most small signal oscillators are designed to source 0 to 20 dbm of power. Power oscillators are made to deliver watts of power, but frequency stability suffers due to extracting more power from the resonator and lowering the loaded Q. When stability is a concern, small signal oscillators are built and followed by a carefully designed chain of amplifiers.

21.2.1 Device Technologies

21.2.1.1 Bipolar Transistors

Silicon bipolar transistors are used in most low noise oscillators below 5 GHz. Hetero-junction bipolar transistors (HBTs) are common today and extend the bipolar range to 100 GHz as shown in Figure 21.3.

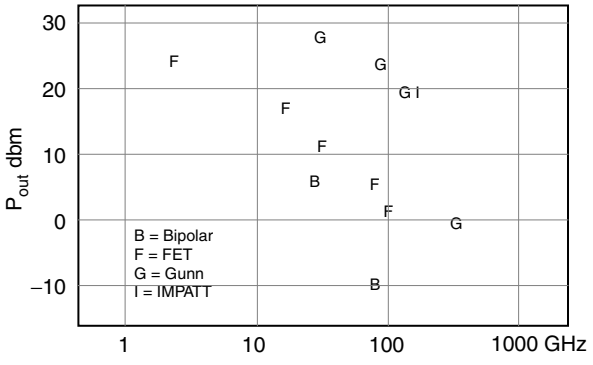


FIGURE 21.3 Device technology and power capability [10–12,57–63].

Bipolar transistors have high gains of over 20 dB at frequencies below 1 GHz and typically have $1/f$ corners in the kHz region. The $1/f$ corner of the oscillator phase noise is less than or equal to the $1/f$ corner of the active device because although the device low frequency noise modulates the device bias and causes phase modulation, it may not be enough to overcome the high frequency noise modulations. In crystal oscillators typically the resonator $1/f$ noise dominates [17,46]. The device $1/f$ noise corner scales with f_{Max} within a given technology, so smaller devices with higher f_{Max} will have higher $1/f$ noise corners. A 1 kHz $1/f$ corner is typical for a 10 GHz transistor while a 100 MHz transistor will have a $1/f$ corner in the tens of Hertz. Device technology such as ion implantation will raise the $1/f$ noise corner to 100 kHz or higher by introducing traps in to the device. $1/f$ noise is very sensitive to device construction [18]. The $1/f$ noise in a bipolar transistor is concentrated in the base current, as shown in Figure 21.4. The $1/f$ noise is due to traps at the base-emitter edge.

21.2.1.2 MOSFETs

CMOS integrated oscillators are becoming more common, although they are limited to the low GHz region [19,20]. Typically two transistors are used in a free running flip-flop configuration. CMOS transistors have lower gain, higher $1/f$ noise corners, and less output power than bipolar transistors, but they offer high integration density and low cost. The higher $1/f$ noise corner of CMOS is mitigated by the reduced modulation sensitivity of the device and by using balanced configurations to reduce noise modulation by symmetry [21].

21.2.1.3 JFETs

JFETs have excellent low frequency noise and limited gain relative to bipolar transistors. While JFETs are found in some extremely low noise discrete oscillators, they are not as common as other devices, especially above 200 MHz [22,54].

21.2.1.4 MESFETs and HEMTs

Above 5 GHz MESFETs and HEMTs are the most common 3-terminal oscillator engine. MESFETs and HEMTs have less gain than bipolar transistors at low frequencies, but have a much higher maximum frequency of operation, or f_{Max}, as shown in Figure 21.3. HEMT f_{Max} is higher than that of MESFETs due to transistor construction that maximizes mobility and provides better channel confinement and control [6]. Both MESFETs and HEMTs have much higher low frequency $1/f$ noise than bipolar transistors, with corner frequencies in the 10 to 100 MHz range being typical for a 600 μm device. As with bipolar transistors, the $1/f$ corner scales with f_{Max} of the device within a given process. Specifically, the $1/f$ level scales with the channel volume beneath the gate structure. GaAs MESFETs and HEMTs have problems with surface traps because of the lack of a native oxide, and there are problems with substrate and channel traps due to the material layering inside the FET. These traps are thermally spread into an approximately continuous $1/f$ distribution at room temperature [23,24]. As shown in Figure 21.4, the $1/f$ noise for a MESFET or HEMT is mostly in the equivalent gate voltage noise source. The reduced low frequency gain of FETs gives them less modulation sensitivity, so they typically have phase noise levels only 10 dB worse than bipolar oscillators even though the low frequency noise is often 30 dB worse.

21.2.1.5 Diodes

Diodes have the highest maximum useable frequencies for solid state devices, as shown in Figure 21.3. There are many different types of diodes for generating negative resistances and negative conductances. The diode negative immittance cancels the positive loss of the resonator and allows an oscillation to build up from the noise within the device. Traditionally microwave oscillators have been designed as negative immittance devices because it is much easier to measure reflections than to set up feedback loops at microwave frequencies. The distinction between negative resistance devices, such as IMPATT diodes, and negative conductance devices, such as Gunn diodes, is important because the device IV characteristic determines how the device saturates and whether it is stable with a series resonator or shunt resonator.

IMPATT diodes generate negative resistances and so are used at series resonant points in waveguides and planar circuits [25]. The avalanche mode of IMPATT operation creates high power but at the cost

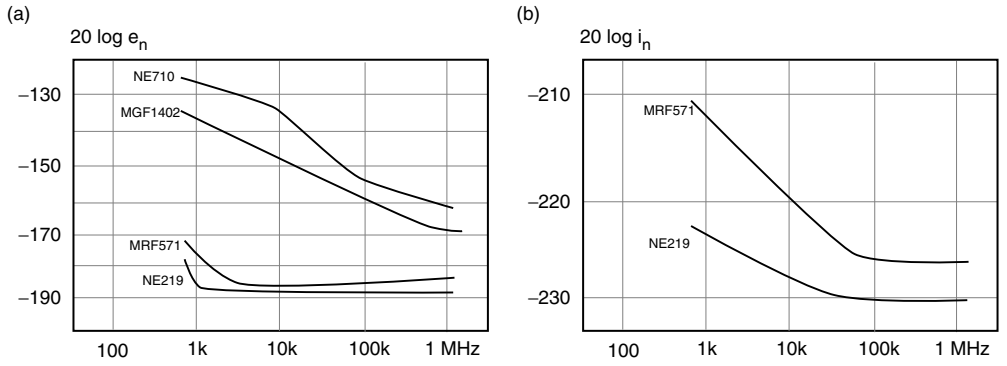


FIGURE 21.4 Device $1/f$ noise comparison for equivalent voltage input noise (a), and equivalent current input noise (b) [31]. FET input current noise is not shown because it is so small.

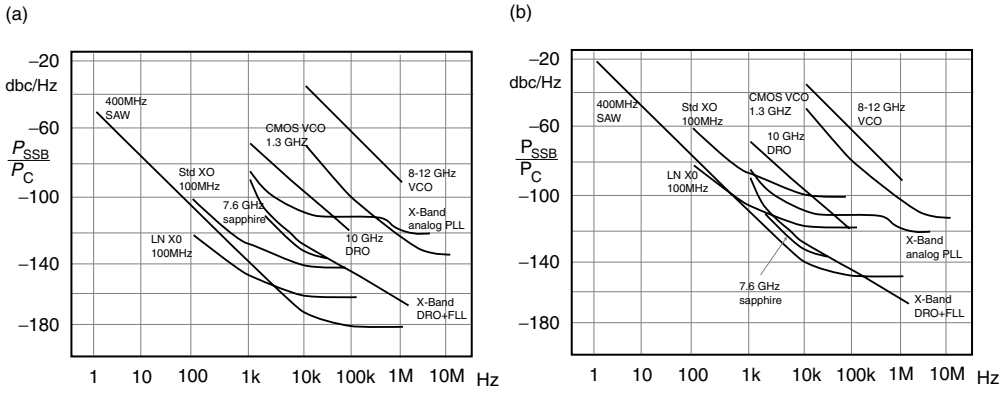


FIGURE 21.5 Oscillator noise performance of some microwave sources: (a) actual; and (b) referred to 10 GHz for comparison (scaled by 20 log[10 GHz/f_{OSC}]) [64–67].

of high noise levels. Gunn diodes have an inherently quiet Gunn domain negative conductance that requires a parallel resonant circuit [26]. Gunn diodes are among the quietest high frequency oscillators and exhibit excellent power into the 100s of GHz, as shown in Figure 21.3. The bulk nature of the Gunn device means that no third terminal metallization is required. This lack of a third terminal and bulk mode of operation reduces the device $1/f$ noise. As with FETs, more advanced material structures, such as using InP rather than GaAs, maximize the high frequency performance of Gunn diodes [12].

21.2.1.6 Multipliers

Frequency multipliers will always be a way of generating the highest frequencies. Reactive multipliers using varactor diodes offer low noise and efficient power generation almost to 1 THz [27]. Resistive multipliers are simple, broadband alternatives to tuned reactive multipliers. Resistive multipliers also suffer significant conversion losses, but are commonly used in broadband instrumentation. All frequency multipliers will increase the phase noise by the same factor that they multiply frequency because frequency

and phase are both multiplied, as shown in Equation 21.2. In dB this would be 20 log *N*. For example, if the oscillator signal is $V_O(t) = A\,\mathrm{Cos}(\omega_O\, t + \phi(t))$ then a times two multiplier would generate:

$$V_O(t)^2 = A^2\left(\tfrac{1}{2} + \tfrac{1}{2}\mathrm{Cos}\left(2\omega_O t + 2\phi(t)\right)\right) \tag{21.2}$$

21.2.2 Resonators

Table 21.2 shows an overview of resonator technologies for oscillators. Various abbreviations are used in the above table, with YIG being Yttrium-Iron-Garnet, TL being transmission line, DR being dielectric resonator, and SAW being surface acoustic wave. Resonator choice is a compromise of stability, cost, and size. Generally, *Q* is proportional to volume, so cost and size tend to increase with *Q*. Technologies such as quartz, SAW, YIG, and DR allow great reductions in size while achieving high *Q* by using acoustic, magnetic, and dielectric materials, respectively. Most materials change size with temperature, so temperature-stable cavities have to be made of special materials such as Invar or carbon fiber. Transmission line, dielectric resonator, and quartz resonators can easily have temperature coefficients below 10 ppm. *Q* changes with frequency for most resonators. Capacitors and dielectric resonators have *Q*s that decrease

TABLE 21.2 Resonators

Type	*Q* Range	Range (GHz)	Limitation	Benefit
LC	0.5–200	Hz–100 GHz	*Q*, lithography	cost
Varactor	0.5–100	Hz–100 GHz	*Q*, nonlinear, noise	tunable
Stripline	100–1000	MHz–100 GHz	size, lithography	cost, *Q*
Waveguide	1000–10,000	1–600 GHz	size, cost	*Q*
YIG	1000	1–50 GHz	cost, magnet, tuning speed	*Q*, tunable, linear
TL	200–1500	500 MHz–3 GHz	cost	*Q*, temperature stable
DR	5000–30,000	1–30 GHz	cost, size	*Q*, temperature stable
Sapphire	50 k	1–10 GHz	cost, size	*Q*
Quartz	100 k–2.5 M	kHz–500 MHz	frequency	*Q*, temperature stable
SAW	500 k	1 MHz–2 GHz	frequency, cost	*Q*

FIGURE 21.6 A picture of various resonators. From the left are three transmission line resonators for 500 MHz to 2 GHz operation, two dielectric resonators for 7 and 20 GHz operation, a 10 MHz crystal resonator, and a 300 MHz SAW resonator. The resonators are sitting on top of a 2.5 inch diameter dielectric cylindrical resonator for 850 MHz.

with frequency, while inductors and transmission line resonators have *Q*s that increase with frequency. Quartz resonators are an extremely mature technology with excellent *Q*, temperature stability, and low cost. Most precision microwave sources use a quartz crystal to control a high frequency tunable oscillator via a PLL. Oscillator noise power, and jitter, is inversely proportional to Q^2, making high resonator *Q* the most direct way to achieve a low noise oscillator.

Tunable resonators are very important because they offer the ability to transfer a reference frequency, with or without modulation, through a PLL. Tunable resonators also offer direct modulation and frequency agility for communication and test purposes. Varactor diodes are the most common device for tuning an oscillator. These devices are inexpensive, available in a variety of packages, and can be used at almost any frequency of interest. Varactors also offer rapid tuning for frequency hopping and high speed direct modulation. The only disadvantages of a varactor diode are low *Q* at high frequencies, low frequency noise, and a nonlinear tuning characteristic [6].

YIG resonators offer the advantages of tuning linearity and high *Q*. These resonators are excellent for instrumentation and special applications, but suffer from the needing a magnetic bias circuit, which increases the size and cost of the oscillator. Typically YIG resonators have both a broadband and a narrowband tuning port [9,56]. The narrowband tuning port requires much less inductance and so can be tuned faster than the broadband port, making it more useful for modulation.

21.3 Theory

21.3.1 Introduction

A brief review of oscillator theory will aid in understanding the oscillator specifications mentioned in the first section. First, an overview of oscillator topologies is shown in Figure 21.7. Traditionally, microwave oscillators have been viewed as one-port circuits with the active device presenting a negative immittance to the resonator, as shown in Figure 21.7a [28]. The one-port philosophy is easy to measure with a slotted

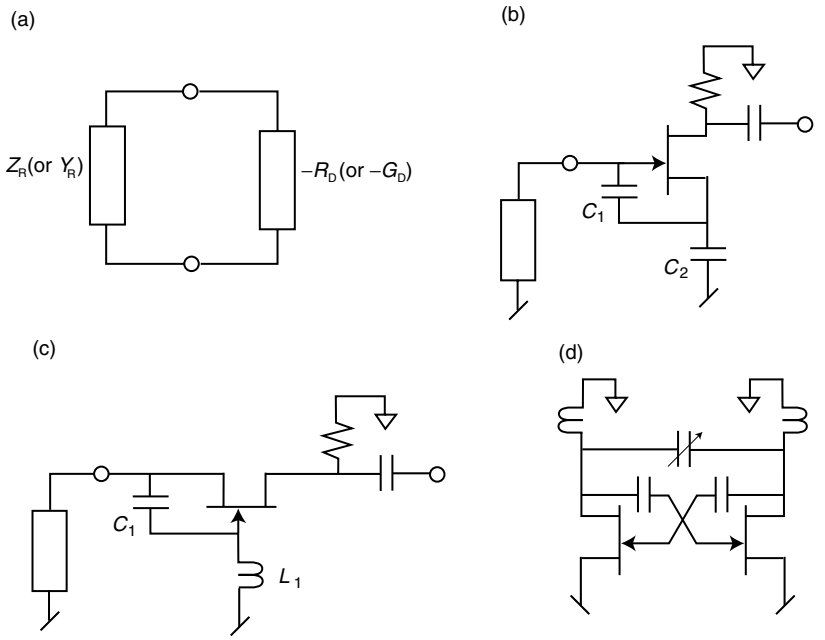


FIGURE 21.7 Oscillator configurations: (a) diode; (b) source feedback; (c) gate feedback; and (d) cross-coupled.

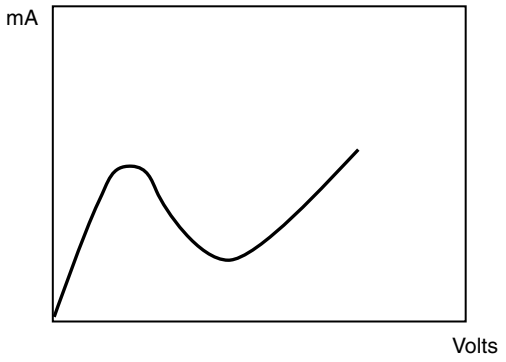


FIGURE 21.8 Gunn diode IV characteristic.

line or a network analyzer. Two-port oscillators are much more common at low frequencies, but probing voltages and currents in a feedback loop will always be difficult at the highest frequencies. Dielectric resonators have made feedback oscillators more common at microwave frequencies [29]. Integrated circuits have made the cross-coupled oscillator configuration, as shown in Figure 21.7d, popular [30].

One-port analysis does allow confusion over the device acting as a negative impedance or admittance. Knowing the device type is essential for establishing a stable oscillation. For example, a negative conductance Gunn diode has the IV characteristic shown in Figure 21.8. As the oscillation signal grows about the bias point it eventually extends into the positive resistance region. During saturation the load line becomes more horizontal, reducing the negative conductance. The resonator needs to have its minimum conductance at the resonance frequency, so that moving off the resonant frequency would require an increase in the device negative conductance. If a Gunn diode is loaded with a series resonant circuit, which has maximum conductance at resonance, noise will move the oscillation off the series resonance and onto a nearby parasitic parallel resonance where it will stabilize.

$$Z_D + Z_R = 0 \tag{21.3}$$

$$Y_D + Y_R = 0 \tag{21.4}$$

$$G\,H = 1 \tag{21.5}$$

The various oscillation conditions can be defined by the preceding three equations. Equation 21.3 describes the active device impedance, Z_D, canceling the resonator impedance, Z_R, to support an oscillation. Equation 21.4 describes the active device admittance, Y_D, canceling the resonator admittance, Y_R, to support oscillation. Equation 21.3 can be split into real and imaginary parts to give $R_D + R_R = 0$, and $X_D + X_R = 0$. Often X_D is small, so the Equations reduce to the active device negative resistance canceling the resonator (and load) resistance while the resonator is just off center frequency enough to cancel the device reactance. From a one-port point of view, when the resonator reactance is zero, no net phase shifts occur from the oscillation signal as it reflects back and forth from the active device to the resonator. Equation 21.5 is the oscillation equation for a feedback circuit with gain element G and feedback resonator H. Equation 21.5 describes a positive feedback situation where the gain cancels the loss in the feedback while the net phase shift around the loop is zero. With zero phase shift around a loop and no loss, a signal will be sustained at the frequency of zero phase. The equations describe a linear approximation to a stable oscillation. In reality each circuit is nonlinear. Oscillations start from noise in a circuit with positive feedback and grow until the circuit gain element saturates and the above equations are satisfied.

Typically oscillators are set up so that the negative immittance, or gain, is 1.5 to 2 times greater than the circuit loss so that the device saturates into a stable oscillation without being driven so hard that its operating point changes excessively [2]. Equations 21.1–21.5 can be brought into a single oscillator theory [31].

At lower frequencies the oscillator output power can be predicted analytically [32]. At microwave frequencies the accuracy of the oscillator frequency and output power is very dependent on the CAD model used. Both harmonic balance and SPICE simulators can be used to predict oscillator output power and frequency, but component and circuit parasitics can make exact frequency predictions difficult. Linear simulators have a long history in oscillator design, as might be predicted by looking at Equations 21.3–21.5. The operating frequency, tuning range, as well as sensitivities to load, bias, and power supply variations can all be obtained from a linear simulator with bias dependent S-parameters for the active device.

Although many oscillator circuits are used, Figures 21.7b and 21.7c show two of the most common discrete configurations. Figure 21.7b is used mostly with varactor-tuned inductors and dielectric resonators. Figure 21.7b is also used at lower frequencies where it is known as the Seiler oscillator [33]. Figure 21.7c is used mostly with YIGs, transmission line resonators, and dielectric resonators [34]. Simple analysis using an ideal transconductance for the device shows a negative resistance with capacitance at the gate of Figure 21.7b and a negative conductance with shunt inductance at the source for Figure 21.7c. Therefore, Figure 21.7b operates in the inductive region of a series resonator, and Figure 21.7c operates in the capacitive region of a parallel resonator. At microwave frequencies C_1 is simply the device capacitance. Close examination of Figure 21.7b,c shows that they are both the same circuit. The only real difference is where the tuning takes place. Each circuit uses the device drain to provide load isolation and expects to have a relatively low impedance in the device drain to take power out. Equation 21.6 shows how the negative input resistance of the source feedback oscillator changes with frequency, and that the residual reactance is capacitive. The input impedance in Figure 21.7b is that from the resonator looking into the device node. Equation 21.7 shows how the negative conductance of the gate feedback oscillator varies with frequency and that its residual susceptance is inductive. Note that the conductance is negative above the L_1C_1 resonance frequency and rapidly decreases with frequency. Both of these equations can be used for estimating the tuning bandwidth and oscillation frequency of an oscillator. The resonator used in either case will have its own immittance which can be compared to the device immittance using Equations 21.3 or 21.4 to determine if the oscillation condition is satisfied. CAD programs with S-parameter models will provide more accurate device characterization and tuning bandwidth analysis. For example, the configurations of Figure 21.7 can be analyzed as one-ports for S11, $S11_D$. The resonator S11, $S11_R$, forms a reflection loop so that Equation 21.5 is valid if $G = S11_D$ and $H = S11_R$.

$$Z_{IN} = -g_m / C_1C_2\omega^2 - j\left(C_1 + C_2\right) / C_1C_2\omega^2 \tag{21.6}$$

$$Y_{IN} = g_m / \left(1 - \omega^2 L_1C_1\right) + j\omega C_1 / \left(1 - \omega^2 L_1C_1\right) \tag{21.7}$$

21.3.2 Modulation, Noise, and Temperature

Many things perturb the oscillator frequency. To better study these perturbations consider the idealized oscillator shown in Figure 21.9. Note that this oscillator is connected in a positive feedback configuration. The active device could be two FETs connected to create positive feedback, or it could be a single negative conductance shunting the resonator. The feedback configuration is chosen because it clarifies that the loaded Q of the oscillator is:

$$Q_L = 1 / \left(\omega_O L_O G_O\right). \tag{21.8}$$

The loaded Q is only determined by circuit losses and external loading. The loaded Q is unaffected by the device g_m or negative conductance. The total losses, G_O, will be made up of internal losses, G_I, and external loading, G_{EXT}, so $G_O = G_I + G_{EXT}$. A little math will show that the total oscillator Q is a parallel combination of the internal Q_I and the external Q_{EXT}, as shown in Equation 21.8.

$$Q_L = \left(1/Q_I + 1/Q_{EXT}\right)^{-1} \tag{21.9}$$

The total oscillator Q is important because it determines the sensitivity to perturbations in oscillator frequency. Unfortunately, all we can measure directly is the external Q of an oscillator. Load pulling, as given in the pulling sensitivity specification, can be used to compute the external Q of an oscillator [22]. Load variations affect the oscillation frequency by changing the values of C_O or L_O in the same way G_{EXT} changes the value of G_O. Typically load pulling is measured by placing a 12 dB return loss load on a line stretcher and rotating the load through all phases while recording the maximum frequency deviation. The C_O and L_O of Figure 21.9 are made up of device capacitance, resonator elements, and load reactance. In low noise oscillators the resonator reactances dominate, but all oscillators have some influence from the other sources of reactance.

Power supply voltage variations change the oscillation frequency by changing the active device reactances and bias. The pushing specification in MHz/V defines the sensitivity to low frequency supply variations. Knowing the spectrum of the power supply variations allows the direct calculation of oscillator FM noise due to power supply variations through the pushing sensitivity. As the frequency of the supply variation is increased, the supply and bias filtering circuitry will reduce its influence on the oscillation frequency, so some measurements may have to be made if the power supply has large variations at high frequencies. Pushing can also be used as a technique to evaluate oscillator Q and minimize phase noise [69].

As discussed in the section on noise, the oscillator phase noise spectrum is the FM noise spectrum divided by f_m^2, where f_m is the modulating, or carrier offset frequency. Oscillator frequency variations are easily measured with a frequency discriminator or other FM detector such as a transmission line discriminator [35]. Spectrum analyzers are also very useful for measuring oscillator phase noise as long as the analyzer noise is much less than that of the measured device. Spectrum analyzers can measure script $L(f_m) = P_{SSB}(f_m)/\text{Hz}/P_C$ where the $P_{SSB}(f_m)/\text{Hz}$ is the phase noise power per Hertz within the $1/f^3$ region close to the carrier. Typically if the noise measurement is very near the carrier frequency, that is, within the f^{-3} region, the noise is dominated by phase noise. For modulating noise such as power supply pushing the actual phase noise spectrum, $S_\phi(f_m)$, is twice script $L(f_m)$ because the sidebands are correlated [31].

All active devices contain internal noise sources due to resistance (thermal noise), charge crossing an energy barrier (shot noise), and traps (G-R noise) [18]. All of these noise sources perturb the device bias point no matter how well filtered the bias circuit. The most significant source of bias noise is due to

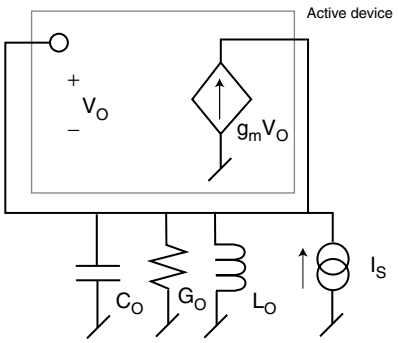


FIGURE 21.9 Idealized oscillator.

traps [31]. These traps are spread in energy distribution by thermal and mechanical processes [23,36,37]. The spreading of the traps in bipolar transistors, FETs, and MOSFETs results in a $1/f$ low frequency noise spectrum in each device. FET based devices have $1/f$ variations in the drain current which converts to an equivalent input noise via the device transconductance. Bipolar devices have $1/f$ variations in the base current due to traps at the edge of the base-emitter junction [38]. The low frequency bias variations change the device reactances and so change the oscillation frequency [31,39–41]. The oscillator sensitivity to bias current or voltage variation is easily simulated and combined with measured device noise to provide a prediction of oscillator noise. Although simulators have estimates of device $1/f$ sources, low frequency noise sources are so dependent on device processing that measuring low frequency noise is the best way to verify noise performance. Once the bias sensitivity and device noise are determined, the oscillator internal phase noise can be computed in a similar manner to power supply pushing. Various schemes using symmetry, low frequency device loading, and even low frequency feedback exist for reducing total oscillator phase noise [42–45]. Even if device noise can be eliminated, resonators such as quartz crystals have their own $1/f$ noise sources that modulate the oscillator spectrum [46].

For voltage- or current-controlled oscillators the noise spectrum contains an additional term due to noise at the modulation port. Even modulating devices, such as varactor diodes, have internal noise that affects the oscillator spectrum. These noise sources can be accounted for just like the pushing and internal contributions. While there are many contributors to oscillator noise, each term is easily accounted for once the modulation sensitivity is known. Some noise contributions, such as power supply switching noise, can come through several paths so the total noise cannot just be a power summation of the individual contributions, but must include any correlation between contributions. In most oscillators one noise source will dominate [31]. For example, in broadband varactor-controlled oscillators, VCOs, the modulation noise usually dominates.

Figure 21.9 also shows a current source, I_S, which adds to the oscillation. This current source is useful for analyzing the oscillator response to high frequency noise and injection locking signals. The modulative low frequency noise sources discussed earlier operate in a multiplicative or nonlinear way, whereas noise sources at the oscillator frequency simply add to the oscillator noise in a linear manner [47]. For additive noise sources, increasing the oscillator power decreases the oscillator phase noise because the additive noise sources are fixed. However, for modulative noise sources increasing the oscillator power does not change the relative noise power because the modulation affects a fraction of the total oscillator power. This seemingly "linear" behavior is the result of any second order nonlinearity, and has caused confusion for many people analyzing oscillator noise. This same fact is why reducing even order nonlinearities through symmetry is effective in reducing oscillator noise.

If I_S only consists of high frequency noise due to thermal and shot noise in the device and circuit, we can produce a simple noise analysis of the oscillator. A more rigorous analysis is given in References 31 and 47. The positive feedback of the oscillator loop will cause the noise at the center of the resonator frequency to be amplified until the amplifier nonlinearity causes device saturation. The amplitude saturation does two things: (1) it reduces the device gain until Equation 21.5 is satisfied; and (2) it reduces the loop gain to amplitude perturbations. I_S will cause both amplitude and phase variations in the oscillator carrier buildup. Amplitude saturation in the device will effectively strip the amplitude noise off the carrier as it passes through the active device, which means the amplitude noise of V_O will be only I_S passed through the resonator, or

$$S_{V\_AM}(f_m) = \left(S_{Is}(f_O + f_m) + S_{Is}(f_O - f_m)\right) \Big/ \left(2V_O^2/2\right) 1 \Big/ \left|G_O + j\left(\omega C_O - 1/\left(\omega L_O\right)\right)\right|^2, \quad (21.10)$$

which, near to the carrier, simplifies to

$$S_{V\_AM}(f_m) \approx S_{Is}(f_O + f_m) \Big/ \left(G_O^2 V_O^2/2\left(1 + 4Q^2 f_m^2/f_O^2\right)\right), \quad (21.11)$$

where $V_O^2/2$ represents the carrier mean square level and $S_{Is}(f_O + f_m)$ is approximately equal to $S_{Is}(f_O - f_m)$.

The noise source, I_S, will perturb the carrier with an equal amount of amplitude and phase noise, which can be seen by envisioning each noise sideband as a phasor rotating about the carrier phasor. The limiting action of the active device does not attenuate the phase modulation of the carrier, so the full loop gain acts on the phase variations causing phase noise to be greatly amplified [48]. The phase noise amplification causes the phase noise to dominate amplitude noise near the carrier. Just as in the amplitude noise case, a low pass equivalent analysis can be performed on the circuit [49]. Solving the low pass equivalent form for phase noise variations is Leeson's model for phase noise [50]. Equation 21.12 gives the phase noise spectral density for the circuit of Figure 21.9.

$$S_{V\_\phi}(f_m) \approx S_{Is}(f_O + f_m) \Big/ \left(G_O^2 V_O^2 / 2\ 4Q^2 f_m^2 / f_O^2\right), \tag{21.12}$$

The term $S_{Is}(f_O + f_m)/(G_O^2 V_O^2)$ is equal to $kTF/P_C\ (1 + f_k/f_m)$ in Leeson's analysis, where P_C is the carrier power and F is approximately the device noise figure. The $1 + f_k/f_m$ term accounts for $1/f$ noise modulation with f_k being less than or equal to the device $1/f$ noise corner. Although the preceding analysis assumed uncorrelated high frequency noise sidebands, the inclusion of a $1/f$ term in Leeson's result implies correlated sidebands and a factor of 2 between script L and phase noise spectral density, S_ϕ. Leeson's model for phase noise contains most of the important aspects of oscillator noise in a simple equation. More rigorous oscillator analysis appears with regularity, but Leeson's model continues to be useful and relevant [28,31,39]. Naturally, it is also possible to analyze oscillator noise in the time domain [51,52]. Several important aspects of oscillator noise predicted by Leeson's model are that the noise power decreases as Q^2, that the noise falls off as f_m^{-3} near the carrier, and that the noise power increases as f_O^2. The problems with Leeson's model are the approximations involved. Because oscillators are not linear, noise-matched amplifiers F cannot be used. Because the low frequency noise modulates the carrier, script L does not decrease in proportion to increases in P_C in well-designed, low-noise oscillators. And finally, f_k in the final oscillator depends on all the low frequency noise sources and modulation sensitivities and is usually less than or equal to the device $1/f$ noise corner for the dominant $1/f$ noise source.

$$script\ L(f_m) \approx kTF/P_C\left(1 + f_k/f_m\right) \Big/ \left(4Q^2 f_m^2 / f_O^2\right) \tag{21.13}$$

21.3.3 Injection Locking

Very high frequency oscillators are often phase locked by coupling a reference signal directly into the oscillator [28,53,55]. Understanding injection locking can help in understanding oscillator operation in general. In Figure 21.9 we can let I_S be an injection locking signal, $I_S e^{j\omega t}$. Then $V_O = V_O e^{j\omega t + \phi}$, where the phase shift ϕ occurs if ω is different from the oscillator loop center frequency of ω_O. From linear circuit analysis we can derive Equation 21.14, and Equation 21.15 follows from Euler's identity:

$$V_O e^{j\omega t + \phi}\left[1 - \omega^2 L_O C_O + j\omega L_O\left(G_O - g_m\right)\right] = j\omega L_O I_S e^{j\omega t} \tag{21.14}$$

$$V_O\left[1 - \omega^2 L_O C_O + j\omega L_O\left(G_O - g_m\right)\right] = j\omega L_O I_S\left(\cos(\phi) - j\sin(\phi)\right) \tag{21.15}$$

By grouping real and imaginary parts, Equation 21.15 can be expanded into Equations 21.16 and 21.17 which define the locking gain and the bandwidth of locking.

$$I_S = V_O\left(G_O - g_m\right) \Big/ \cos(\phi) \tag{21.16}$$

$$\sin(\phi) = \left(1/\omega L_O - \omega C_O\right) V_O / I_S \tag{21.17}$$

Equation 21.15 shows that as the device saturates and g_m approaches G_O, very little current, I_S, is required to maintain lock. However, as the injection frequency shifts away from the resonator frequency, ω_O, ϕ moves away from zero as shown in Equation 21.17. As ϕ increases from zero to ±90°, the $\cos(\phi)$ term in Equation 21.15 decreases to zero and an infinite injection current is required to maintain lock, so ϕ equal to ±90° defines the locking bandwidth. Equation 21.17 can be used to translate the ±90° limits into frequency limits that define the locking bandwidth. Equation 21.15 shows that very little injection locking signal is required at the band center while more signal is required as the frequency approaches the band edges. Another way of interpreting this is to say there is near infinite injection locking gain at the band center and zero locking gain at the band edges. When I_S is broadband noise this injection gain works to provide the commonly observed high levels of noise that decrease away from the carrier [48]. The relationship between the gain and offset frequency is linked through the loop phase shift. An injection locked oscillator is in fact a first order PLL.

21.4 Summary

Oscillators consist of an active device, a resonator, and a package. These three things determine the frequency, accuracy, available power, and cost of the source. For a simple task such as providing a sinusoid, oscillators require an inordinate number of specifications. The sections on specifications and theory try to provide a background for understanding oscillator requirements, while the technologies and capabilities section tries to show how modern devices and resonators are combined to meet oscillator specifications.

Acknowledgments

Thanks to Mark Shiman of Disman Bakner and Ron Korber of Stellex for providing various samples for the photographs.

References

1. Leier, R.M., and Patston, R.W., Voltage-Controlled Oscillator Evaluation for System Design, *MSN*, 102–125, Nov. 1985.
2. Rogers, R.G., *Low Phase Noise Microwave Oscillator Design*, Artech House, Boston, 1991.
3. Robins, W.P., *Phase Noise in Signal Sources*, Peter Peregrinus Ltd., London, U.K., 1982.
4. Blair, B.E., *Time and Frequency: Theory and Fundamentals*, U.S. Department of Commerce, NBS Monograph 140, 1974.
5. ANSI, *Telecommunications—Synchronous Optical Network (SONET)—Jitter at Network Interfaces*, T1.105.03-1994, 1994.
6. Bahl, I.J., and Bhartia, P.B., *Microwave Solid State Circuit Design*, John Wiley & Sons, New York, 1988.
7. Engineering Staff, *Nonlinear Circuits Handbook*, Analog Devices, Norwood, MA, 1976.
8. Huckleberry, B.E., Design Considerations for a Modern DTO, *Microwave Journal*, 291–295, May 1986.
9. Osbrink, N.K., YIG-Tuned Oscillator Fundamentals, *Microwave System Designer's Handbook*, 207–225, 1983.
10. Heins, M.S., Juneja, T., Fendrich, J.A., Mu, J., Scott, D., Yang, Q., Hattendorf, M., Stillman, G.E., and Feng, M., W-band InGaP/GaAs HBT MMIC Frequency Sources, *IEEE MTT-S Digest*, 239–242, 1999.
11. Dieudonne, J.-M., Adelseck, B., Narozny, P., and Dambkes, H., Advanced MMIC Components for Ka-Band Communication Systems: A Survey, *IEEE MTT-S Digest*, 409–415, 1995.
12. Eisele, H., and Haddad, G.I., Potential and Capabilities of Two-Terminal Devices as Millimeter- and Submillimeter-Wave Fundamental Sources, *IEEE MTT-S Digest*, 933–936, 1999.
13. Faber, M.T., Chramiec, J., and Adamski, M.E., *Microwave and Millimeterwave Diode Frequency Multipliers*, Artech House, Boston, 1995.

14. Scherer, D., Generation of Low Phase Noise Microwave Signals, *RF & Microwave Measurement Symposium*, Hewlett-Packard, Palo Alto, CA, 1983.
15. Bruston, J., Smith, R.P., Martin, S.C., Humphrey, D., Pease, A., and Siegel, P.H., Progress Towards the Realization of MMIC Technology at Submillimeter Wavelengths: A Frequency Multiplier to 320 GHz, *IEEE MTT-S Digest*, 399–402, 1998.
16. Roberts, M.J., Iezekiel, S., and Snowden, C.M., A Compact Subharmonically Pumped MMIC Self-Oscillating Mixer for 77 GHz Applications, *IEEE MTT-S Digest*, 1435–1438, 1998.
17. Bates, P.C., Measure Residual Noise in Quartz Crystals, *Microwaves & RF*, 95–106, Nov. 1999.
18. Ambrozy, A., *Electronic Noise*, McGraw-Hill, New York, 1982.
19. Banu, M., MOS Oscillators with Multi-Decade Tuning Range and Gigahertz Maximum Speed, *IEEE JSSC*, 1386–1393, Dec. 1988.
20. Svelto, F., Deantoni, S., and Castello, R., A 1.3 GHz Low-Phase Noise Fully Tunable CMOS LC VCO, *IEEE JSSC*, 356–361, Mar. 2000.
21. Aoki, H., and Shimasue, M., Noise Characterization of MOSFET's for RF Oscillator Design, *IEEE MTT-S Digest*, 423–426, 1999.
22. Vendelin, G., Pavio, A.M., and Rohde, U.L., *Microwave Circuit Design*, John Wiley & Sons, 1990.
23. Sodini, D., Touboul, A., Lecoy, G., and Savelli, M., Generation-Recombination Noise in the Channel of GaAs Schottky gate FET, *Electronics Letters*, 42–43, Jan. 22, 1976.
24. Hughes, B., Fernandez, N.G., and Gladstone, J.M., GaAs FETs with a Flicker Noise Corner Below 1 MHz, *WOCSEMMAD*, 20–23, 1986.
25. Goedbloed, J.J., Noise in IMPATT Diode Oscillators, *Philips Research Reports Supplement*, 1–115, 1973.
26. Sze, S.M., *Physics of Semiconductor Devices*, John Wiley & Sons, New York, 1981.
27. Crowe, T.W., Weikle, R.M., and Hesler, J.L., GaAs Devices and Circuits for Terahertz Applications, *IEEE MTT-S Digest*, 929–932, 1999.
28. Kurokawa, K., Noise in Synchronized Oscillators, *IEEE Trans MTT*, 234–240, Apr. 1968.
29. Popovic, N., Review of Some Types of Varactor Tuned DROs, *Applied Microwave and Wireless*, 62–70, Aug. 1999.
30. Abidi, A.A., Radiofrequency CMOS Circuits, *IEEE SCV-MTT Short Course*, Apr. 1997.
31. Riddle, A.N., Oscillator Noise: Theory and Characterization, N.C. State University, Raleigh, NC, PhD Dissertation, 1986.
32. Clarke, K.K., and Hess, D.T., *Communication Circuits: Analysis and Design*, Addison-Wesley, Reading, 1978.
33. Clapp, J.K., Frequency Stable LC Oscillators, *Proc. IRE*, 1295–1300, Aug. 1954.
34. Schiebold, C.F., An Approach to Realizing Multi-Octave Performance in GaAs-FET YIG-Tuned Oscillators, *IEEE MTT-S Digest*, 261–263, 1985.
35. Ondria, J.G., A Microwave System for Measurement of AM and FM Noise Spectra, *IEEE Trans. MTT*, 767–781, Sept. 1968.
36. Rohdin, H., Su, C.-Y., and Stolte, C., A Study of the Relationship Between Low Frequency Noise and Oscillator Phase Noise for GaAs MESFETs, *IEEE MTT-S Digest*, 267–269, 1984.
37. Christenssen, S., Lundstrum, I., and Svensson, C., Low Frequency Noise in MOS Transistors, *Solid-State Electronics*, 797–812, 1968.
38. van der Ziel, A., Noise in Solid State Devices and Lasers, *Proc. IEEE*, 1178–1206, Aug. 1970.
39. Siweris, H.V., and Schiek, B., Analysis of Nosie Upconversion in Microwave FET Oscillators, *IEEE Trans. MTT*, 233–242, Mar. 1985.
40. Pucel, R.A, and Curtis, J., Near-Carrier Noise in FET Oscillators, *IEEE MTT-S Digest*, 282–284, 1983.
41. Dallas, P.A., and Everard, J.K.A., Characterization of Flicker Noise in GaAs MESFETs for Oscillator Applications, *IEEE Trans. MTT*, 245–257, Feb. 2000.
42. Chen, H.B., van der Ziel, A., and Amberiadis, K., Oscillators with Odd-Symmetry Characteristics Eliminate Low-Frequency Noise Sidebands, *IEEE Trans. CAS*, 807–809, Sept. 1984.

43. Riddle, A.N., and Trew, R.J., A Novel GaAs FET Oscillator with Low Phase Noise, *IEEE MTT-S Digest*, 257–260, 1985.
44. Tutt, M.N., Pavlidis, D., Khatibzadeh, A., and Bayraktaroglu, B., The Role of Baseband Noise and its Unconversion in HBT Oscillator Phase Noise, *IEEE Trans. MTT*, 1461–1471, July 1995.
45. Prigent, M., and Obregon, J., Phase Noise Reduction in FET Oscillators by Low-Frequency Loading and Feedback Circuitry Optimization, *IEEE Trans. MTT*, 349–352, Mar. 1987.
46. Gagnepain, J.J., Olivier, M., and Walls, F.L., Excess Noise in Quartz Crystal Resonators, *Proc. 36th Symp. Frequency Control*, 218–225, 1983.
47. Edson, W.A., Noise in Oscillators, *Proc. IRE*, 1454–1466, Aug. 1960.
48. Spaelti, A., Der Einfluβ des Thermischen Widerstandrauschens und des Schroteffektes auf die Stoermodulation von Oscillatoren, *Bull. Schweiz Elektrotech. Verein*, 419–427, June 1948.
49. Egan, W.F., The Effects of Small Contaminating Signals in Nonlinear Elements used in Frequency Synthesis and Conversion, *Proc. IEEE*, 797–811, July 1981.
50. Leeson, D.B., A Simple Model of Feedback Oscillator Noise Spectrum, *Proc. IEEE*, 329–330, Feb. 1966.
51. Abidi, A.A., and Meyer, R.G., Noise in Relaxation Oscillators, *IEEE JSSC*, 794–802, Dec. 1983.
52. Lee, T.H., and Hajimir, A., Oscillator Phase Noise: A Tutorial, *IEEE JSSC*, 326–336, Mar. 2000.
53. Paciorek, L.J., Injection Locking of Oscillators, *Proc. IEEE*, 1723–1727, Nov. 1965.
54. Rohde, U.L., *Digital PLL Frequency Synthesizers*, Prentice-Hall, Englewood Cliffs, NJ, 1983.
55. Khanna, A.P.S., and Gane, E., A Fast-Locking X-Band Transmission Injection-Locked DRO, *IEEE MTT-S Digest*, 601–604, 1988.
56. Trew, R.J., Design Theory for Broad-Band YIG-Tuned FET Oscillators, *IEEE Trans MTT*, 8–14, Jan. 1979.
57. Prigent, M., Camiiade, M., Dataut, G., Raffet, D., Nebus, J.M., and Obregon, J., High Efficiency Free Running Class F Oscillator, *IEEE MTT-S Digest*, 1317–1324, 1995.
58. Maruhashi, K., Madihian, M., Desclos, L., Onda, K., and Kuzuhama, M., A K-Band monolithic CPW Oscillator Co-Integrated with a Buffer Amplifier, *IEEE MTT-S Digest*, 1321–1324, 1995.
59. Heins, M.S., Barbage, D.W., Fresina, M.T., Ahmari, D.A., Hartmann, Q.J., Stillman, G.E., and Feng, M., Low Phase Noise Ka-Band VCOs Using InGaP/GaAs HBTs and Coplanar Waveguide, *IEEE MTT-S Digest*, 255–258, 1997.
60. Eisele, H., Manns, G.O., and Haddad, G.I., RF Performance Characteristics of InP Millimeter-Wave N^+-N^--N^+ Gunn Devices, *IEEE MTT-S Digest*, 451–454, 1997.
61. Wollitzer, M., Buecher, J., and Luy, J.-F., High Efficiency Planar Oscillator with RF Power of 100 mW Near 40 GHz, *IEEE MTT-S Digest*, 1205–1208, 1997.
62. Siweris, H.J., Werthoff, A., Tischer, H., Schaper, U., Schaefer, A., Verweyen, L., Grave, T., Boeck, G., Schleichtweg, M., and Kellner, W., Low Cost GaAs PHEMT MMICs for Millimeter-Wave Sensor Applications, *IEEE MTT-S Digest*, 227–230, 1998.
63. Kadszus, S., Haydl, W.H., Neumann, M., Bangert, A., and Huelsmann, A., Subharmonically Injection Locked 94 GHz MMIC HEMT Oscillator using Coplanar Technology, *IEEE MTT-S Digest*, 1585–1588, 1998.
64. Galani, Z., Low Noise Microwave Sources for Radar and Missile Systems, Ultra Low Noise Microwave Sources Workshop, 1994.
65. Everard, J.K.A., and Page-Jones, M., Ultra Low Noise Microwave Oscillator with Low Residual Flicker Noise, *IEEE MTT-S Digest*, 693–696, 1995.
66. Vectron, *Frequency Control Products*, Vectron International, 1997.
67. Avantek, *Modular and Oscillator Components*, Avantek, 1989.
68. Feng, Z., Zhang, W., Su, B., Harch, K.F., Gupta, K.C., Bright, V., and Lee, Y.C., Design and Modeling of RF MEMS Tunable Capacitors Using Electro-Thermal Actuators, *IEEE MTT-S Digest*, 1507–1510, 1999.
69. Trans-Tech, Optimize DROs for Low Phase Noise, Application Note No. 1030, *Trans-Tech Temperature Stable Microwave Ceramics*, 1998.

22

Phase Locked Loop Design

Robert Newgard
Rockwell Collins

22.1 Introduction

Frequency synthesis through the use of phase locked loops (PLLs) is a topic that has received enormous amount of attention in the technical media. The objective of this chapter is to present the fundamental considerations that go into PLL design, without causing the reader to lose hope. The design of PLLs can be challenging and can even be daunting to those who have not stepped into the realm of frequency synthesis. With a basic understanding of the control theory and the building blocks of a PLL, the implementation of the PLL can be demystified. The PLL has been utilized in various systems for many years, but it was not until the development of integrated circuits in the 1970s that widespread use as a frequency synthesizer came about. With the expansion in the wireless industry and the ever-increasing demand for higher frequency systems, PLL design has now moved into the microwave realm. The frequency synthesizer plays a critical role in system performance. This chapter is by no means a complete discourse on PLL design, but rather a synopsis of PLL fundamental characteristics and design considerations. Only the single loop PLL will be addressed, but the reality is that in order for the frequency synthesizer to meet the system requirements, multiple loop architectures are typically necessary.

The architecture of the frequency synthesizer is often dependent upon the design of the transceiver, and the PLL designer must take the system requirements into account. In addition, many trade-offs must be performed in order to implement a successful frequency synthesizer design; the appropriate injection frequencies must be provided, but these should be generated with consideration given to tuning speed, phase noise performance, spurious requirements, and channel spacing to name a few. Many details are left to the reader, but it should be clear that successful PLL design is no accident.

22.2 Attributes of PLLs

In modern wireless communications systems the PLL plays a key role in the performance of the system. The primary function of the PLL is to generate a band of periodic signals with accurate frequencies that allow the transceiver to resolve the required channel. The injections may be synthesized by a combination of PLLs depending upon the system requirements. The spectral purity of the injection signal is a determining factor in the communication systems' performance. The PLL generates the injection frequencies from a stable reference source that is typically a crystal based oscillator (e.g., oven standard, temperature compensated crystal oscillator, etc). The frequency accuracy and temperature stability of the output signal of the PLL is proportional to the frequency accuracy and temperature stability of the crystal oscillator.

The PLL is a negative feedback control system utilizing a phase detector (PD), lowpass filter (LPF), voltage controlled oscillator (VCO), and frequency divider (FD). The basic block diagram of a PLL is shown in Figure 22.1. The VCO generates an output signal that is dependent upon a DC control voltage at its input. The PD compares the phase of the reference signal to the phase of the divided down VCO signal and generates a correction signal, which is proportional to the phase difference. The LPF's function is to (a) attenuate the reference sidebands, (b) shape the phase noise, and (c) tailor the PLL's dynamics. Frequency selection of the output of the PLL is accomplished by varying the value of the FD. The FD is typically programmable and has enough range to cover the desired amount of frequency tuning bandwidth.

The phase of the PLL's output signal is given by

$$\phi_o = \phi_{ref} N \tag{22.1}$$

where N is the division ratio of the FD and is stepped in integer values. Because frequency is the time derivative of phase, the output frequency of the PLL is given by

$$F_o = \frac{d\phi_o}{dt} = \frac{d\phi_{ref}}{dt} N = F_{ref} N \tag{22.2}$$

which shows that the output frequency is an integer multiple of the reference frequency. The reference frequency is chosen to attain the desired channel spacing since incrementing N increases the output frequency in multiplies of the reference frequency.

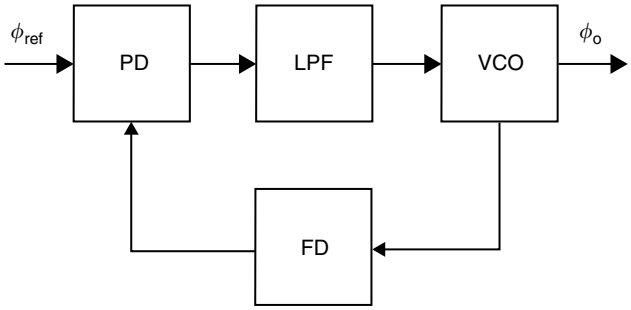


FIGURE 22.1 Phase locked loop.

The PLL makes a relatively unstable VCO to track the phase of the reference signal, which is derived from a stable oscillator. Free running VCOs drift with variation in temperature and power supply, as well as noise on the voltage control line. The action of the feedback loop is to keep the VCO phase-locked to the reference oscillator signal.

There are three primary problems that challenge the PLL designer. One is improving the settling time (i.e., when a command occurs to change channels, the PLL takes time to move from the old frequency to the new one and acquire lock). The second is reducing spurious signals from appearing on the PLL's output. Any discrete frequency components appearing on the VCO control line will modulate the VCO and appear as spurious sidebands on the output of the PLL. The primary spurious frequency source is modulation of the VCO by the error signal, at the comparison frequency, which is coming from the output of the PD. These spurs are referred to as reference sidebands. Other sources of spurious signals are conducted signals on power supplies (e.g., VCO power supply, PD power supply, etc.), radiated signals (e.g., induced on the VCO tank coil and loop filter coils), isolation from other signal sources (e.g., reverse isolation from the programmable divider to the PLL output). In addition to these spurious signal sources, mechanical vibration of the synthesizer assembly, which has the effect of inducing unwanted sidebands on the VCO by physically modulating the printed wiring board and/or the VCO's tank coil. The spurious signals are reduced through good design of the PLL, using good radio frequency (RF) shielding techniques, and providing mechanical support to the assemblies. The third problem is phase noise performance. As mentioned before, the long-term frequency stability of the output signal of the PLL is determined by the frequency standard used, but the phase noise performance, and thus the short-term stability, is dependent upon the design of the PLL. In the receive path, in order to down-convert the modulated RF signal, the output signal of the PLL (i.e., local oscillator [LO]) is mixed with the RF signal. The phase noise of the LO is superimposed onto the intermediate frequency (IF) or baseband signal and thereby affects the receiver's selectivity. In the transmit path, because the LO is mixed with the IF signal or baseband signal to generate the modulated RF signal, the transmit noise floor or signal-to-noise ratio (SNR) is a function of the LO's phase noise. Also, the performance of the receiver in the presence of a strong adjacent-channel signal is affected by the phase noise performance of the LO. The adjacent-channel signals mix with the LO's phase noise (i.e., reciprocal mixing) and produce signals at the IF, thereby decreasing the receiver's selectivity. In order to better understand these design considerations, the designer needs to develop a clear understanding of the mathematical models, which are used to characterize PLL behavior.

22.3 Modeling through Classical Control Theory

By making the assumption that the PLL is continuous in time, basic feedback control theory, utilizing Laplace transforms, can be utilized to determine the loop's behavior; provided that the loop bandwidth is much, much less than the reference frequency. While in practice, it is true that the PDs and FDs are not continuous in time, it is necessary to make this assumption in order to model the stability of the PLL using the Laplace transform. When wide loop bandwidth synthesizers are designed, the sampling nature of the FD and PD cannot be ignored. The time delay of these devices will introduce phase shift (i.e., reduction in phase margin), thereby affecting the dynamic performance of the PLL. Another assumption is that the PLL has reached steady state (i.e., it has reached a phase locked condition). The characteristics described in this section do not address the acquisition of phase lock.

The block diagram of a PLL and the gain of each of the functional blocks is shown in Figure 22.2. The PD is shown as an adder and gain block in order to clarify the understanding of the functionality of the PD. The forward gain, $G(s)$, is used to represent the product of the transfer function of each individual block within the forward path of the PLL. Likewise, the feedback gain, $H(s)$, represents the product of each individual transfer function within the feedback path of the PLL. The equations describing the PLL

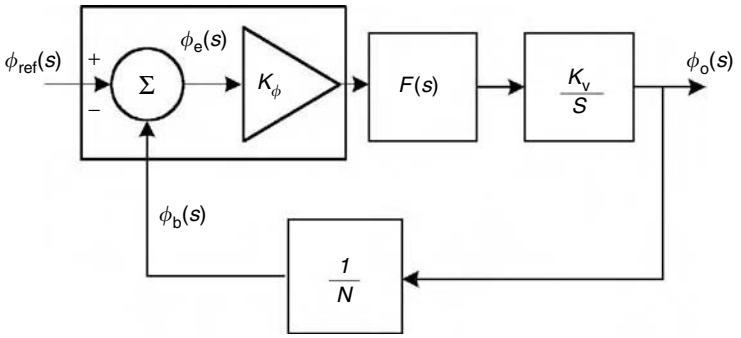


FIGURE 22.2 PLL gain block diagram.

shown in Figure 22.2, in terms of the transform variables, are

$$\phi_o(s) = \phi_e(s)G(s) \tag{22.3}$$

$$\phi_b(s) = H(s)\phi_o(s) \tag{22.4}$$

$$\phi_e(s) = \phi_{ref}(s) - \phi_b(s) \tag{22.5}$$

The overall closed-loop transfer function is found by solving Equations 22.3 through 22.5.

$$A_{CL}(s) = \frac{\phi_o(s)}{\phi_{ref}(s)} = \frac{G(s)}{1 + G(s)H(s)} \tag{22.6}$$

The denominator of the closed-loop response is defined as the characteristic equation. The forward and feedback gain of the PLL are

$$G(s) = K_\phi F(s)\frac{K_v}{s} \tag{22.7}$$

$$H(s) = \frac{1}{N} \tag{22.8}$$

Therefore, the transfer function of the PLL is

$$A_{CL}(s) = \frac{K_\phi F(s)(K_v/s)}{1 + K_\phi F(s)(K_v/s)(1/N)} \tag{22.9}$$

The transfer function given in Equation 22.9 is referred to as the closed-loop response. The open-loop transfer function is defined as the ratio of the output of the feedback path $\phi_b(s)$ to the system error signal $\phi_e(s)$. The open-loop transfer function is used in the analysis of the PLL's stability.

$$A_{OL}(s) = \frac{\phi_b}{\phi_e} = G(s)H(s) = K_\phi F(s)\frac{K_v}{s}\frac{1}{N} = M\angle\alpha \tag{22.10}$$

The closed- and open-loop responses, Equations 22.7 and 22.10, respectively, yield a phasor quantity for each unique complex parameter s. For the open-loop response, the magnitude is M and the phase angle is α. As can be seen from Equation 22.10, the open-loop response appears in the denominator of the closed-loop response. The frequency at which the magnitude of the open-loop response equals one is used to determine the stability of the PLL. As described in the following section, the phase of the open-loop response at this point is critical in determining the loop stability.

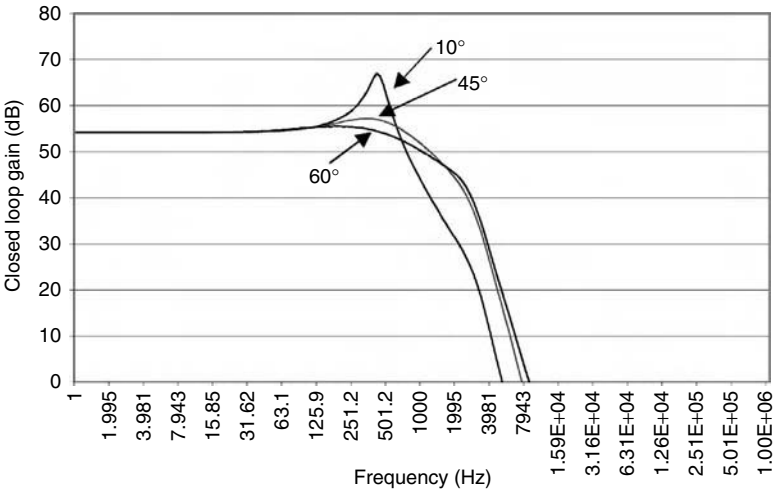


FIGURE 22.3 Closed-loop magnitude response of a PLL with 10°, 45°, and 60° of phase margin.

22.4 Stability

It is common practice to plot the log magnitude and phase of the open-loop transfer function to analyze stability. Phase margin is defined as 180° minus the phase measured at the point were the gain is 1. The frequency at which this occurs is referred to as the open-loop bandwidth. The phase margin is the amount of phase shift at the open-loop bandwidth that would produce instability. For a stable PLL, the phase margin should be greater than 30°, but needs to be greater than 60° not to impact the closed-loop phase noise performance. In designing the PLL, it is imperative that enough phase margin is allowed such that the loop's closed-loop response will not have peaking. The plot shown in Figure 22.3 shows the closed-loop gain of a PLL with three values of phase margin (10°, 45°, and 60°). When adequate phase margin is not provided, the loop will be unstable. When the PLL output signal is observed on a spectrum analyzer and there are spurs separated from the desired output signal by the loop bandwidth, this is an indication that the loop does not have adequate phase margin. The designer should make sure that variations in loop bandwidth, which occur as the PLL output signal is tuned, do not cause a loss in phase margin and thereby have an adverse effect on loop stability. Loop bandwidth variations are caused by changes in VCO gain, PD gain, component temperature coefficients, and loop division ratio.

22.5 Type and Order

It is imperative in modeling the transient and steady-state response of PLLs to develop an understanding of how the PLL will respond to various inputs. Most common PLL designs fall into two categories, Type I and Type II, albeit the PLL type is not limited. The type of system refers to the number of poles in the open-loop gain located at the origin (i.e., the number of perfect integrators in the PLL). The order of the system refers to the degree of the characteristic equation or the denominator of the closed-loop transfer function. As shown in Figure 22.2, there are two blocks that are a function of frequency, the loop filter, and VCO. Therefore, the filter block, $F(s)$, is the factor that determines the type and order of the PLL. The control system examples, which follow, will help the reader understand the basic principles of PLL design.

22.5.1 Type I First-Order Loop

The first PLL introduced is a Type I first-order PLL; although it is not practical owing to the fact that the sidebands caused by the error signal, $\phi_e(s)$, are in most cases too high without a loop filter. This

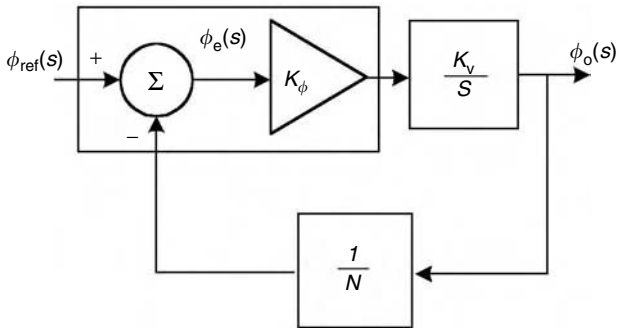


FIGURE 22.4 Type I first-order PLL.

is dependent upon many of the system parameters (e.g., reference signal frequency, VCO gain, division ratio, etc.). It is presented here as a basis for furthering the reader's understanding of the analysis of PLLs. A simplified block diagram is shown in Figure 22.4. As can be seen from the block diagram, the only integrator is the VCO. The assumption is made at this point that the PD does not have a pole at the origin, but such is not always the case. The closed-loop transfer function is given by

$$A_{CL}(s) = \frac{K_\phi K_v}{s + (K_\phi K_v / N)} \tag{22.11}$$

The uncompensated loop bandwidth is commonly defined as

$$\omega_n = \frac{K_\phi K_v}{N} \tag{22.12}$$

which is, in this case Type I first-order, the loop's bandwidth, since there is no compensation by a loop filter. Substituting Equation 22.12 into Equation 22.11, the closed-loop transfer function becomes

$$A_{CL}(s) = \frac{N\omega_n}{s + \omega_n} \tag{22.13}$$

The open-loop transfer function is given by

$$A_{OL}(s) = \frac{K_\phi K_v}{sN} \tag{22.14}$$

The plot in Figure 22.5 shows the open-loop gain and margin phase along with the closed-loop gain. It can be seen from the closed-loop gain that this example has 40 dB of gain inside the loop bandwidth. The loop bandwidth is defined, with respect to frequency, as the point where the open-loop gain equals 1. The phase margin is equal to 90°, which is more than adequate for stability. The problem is the PLL will provide little or no attenuation of the reference sideband spurs. Hence, it is imperative that the PLL designer adds a loop filter to the design and thereby the order of the PLL is increased. By examining the closed-loop transfer function, Equation 22.13, it can be seen that it is a LPF response with gain inside of the loop bandwidth. If the bandwidth of loop is narrow enough and the reference frequency is high enough, the loop will provide attenuation. The slope of the attenuation is 20 dB/decade. The limitations on the design owing to system requirements of vibration and settling time typically force the designer to add additional filtering to the forward path of the PLL.

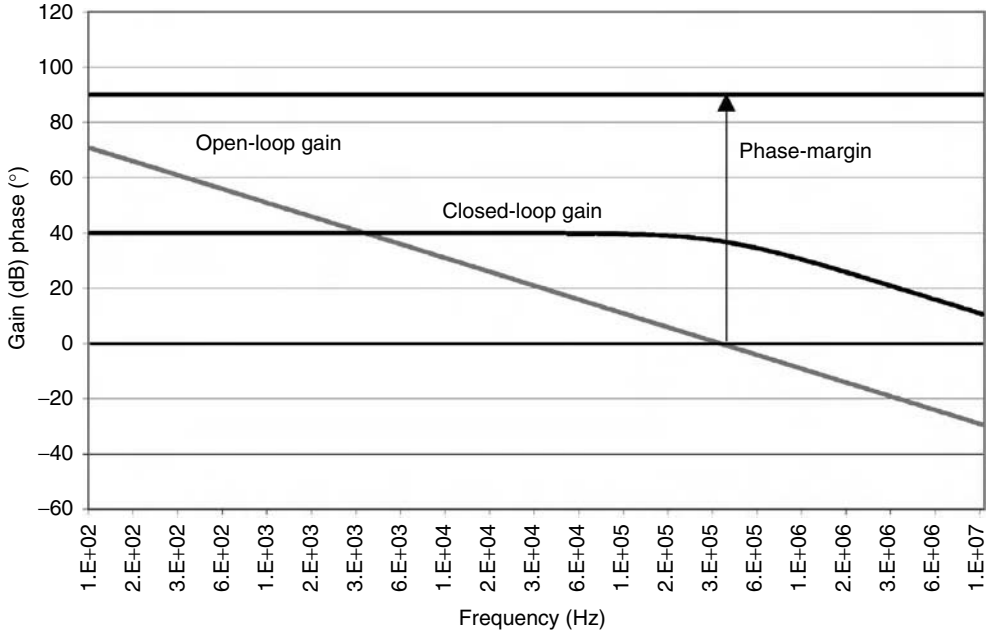


FIGURE 22.5 Type I first-order transfer functions.

22.5.2 Type I Second-Order Loop

A LPF is added in the forward path of the PLL in order to attenuate the reference sideband spurs. In this example, a single pole filter has been added and hence the PLL becomes a Type I second-order loop. A simplified block diagram is shown in Figure 22.6. The transfer function of the loop filter, $F(s)$, is given by

$$F(s) = \frac{1}{RCs + 1} \tag{22.15}$$

The closed-loop transfer function of the PLL is given by

$$A_{\mathrm{CL}}(s) = \frac{K_\phi K_\mathrm{v} F(s)}{s + (K_\phi K_\mathrm{v} F(s)/N)} \tag{22.16}$$

By substituting Equation 22.15 into Equation 22.16 and simplifying, the order of the PLL is plainly seen

$$A_{\mathrm{CL}}(s) = \frac{K_\phi K_\mathrm{v}(1/(RCs + 1))}{s + (K_\phi K_\mathrm{v}(1/(RCs + 1))/N)} = \frac{K_\phi K_\mathrm{v}}{RCs^2 + s + (K_\phi K_\mathrm{v}/N)} \tag{22.17}$$

The open-loop transfer function is given by

$$A_{\mathrm{OL}}(s) = K_\phi \frac{K_\mathrm{v}}{s} \frac{1}{RCs + 1} \frac{1}{N} = \frac{K_\phi K_\mathrm{v}}{s(NRCs + N)} \tag{22.18}$$

Figure 22.7 shows an example of the closed-loop gain, the open-loop gain, and phase margin of a Type I second-order loop. The addition of the filter has added phase shift to the open-loop response, but at 75°, the phase margin is adequate for stability. The loop filter, however, still does not offer much filtering for reference signal spurs. Higher-order filters are typically added to the PLL in order to provide the appropriate attenuation, but as can be seen in Figure 22.7, the additional filtering would add phase shift.

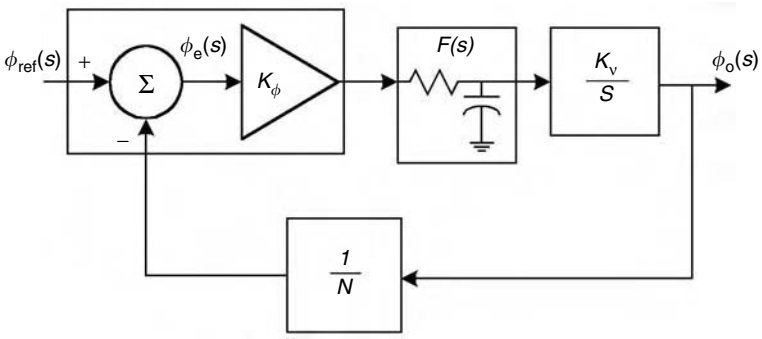


FIGURE 22.6 Type I second-order PLL.

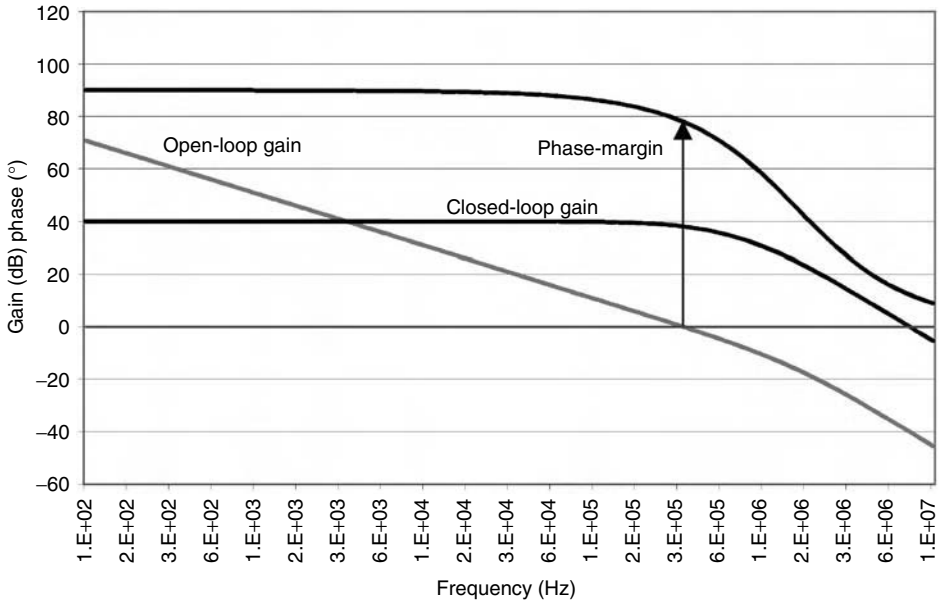


FIGURE 22.7 Type I second-order transfer functions.

The goal here is to maximize the filter's attenuation while realizing minimum phase shift. An elliptic filter is often used owing to the fact that they have higher selectivity (i.e., the passband is closest to the stopband) with minimum phase shift compared to other filters.

22.5.3 Phase Errors for Type I and Type II PLL

Dependent upon the system requirements, the PLL will have to respond to various kinds of inputs (i.e., phase of reference signal, change in division ratio, etc.). The designer is required to know how the PLL will respond to these inputs when the loop has reached steady state. In classical control theory, a system is characterized by its response to step changes in position, velocity, and acceleration. In PLL design, these changes correspond to step changes in phase, frequency, and time varying frequency. The steady-state response is determined by using the Laplace final value theorem, which is

$$\lim_{t\to\infty}[\phi_e(t)] = \lim_{s\to 0}[s\phi_e(s)] \tag{22.19}$$

For the Type I PLL, shown in Figure 22.4, the function $\phi_e(s)$ is the phase error signal generated within the PD and is referred to as the system error.

$$\phi_e(s) = \frac{1}{1 + (K_\phi K_v / sN)} \phi_{ref}(s) \tag{22.20}$$

A phase unit step function, $u(t)$, is applied to the input and the Laplace transform gives

$$\phi_{ref}(s) = \frac{A}{s} \tag{22.21}$$

where A is the magnitude of the phase step in radians. This would represent the input signal shifting phase of A radians. By substituting Equations 22.20 and 22.21 into the Laplace final value theorem gives

$$\lim_{t \to \infty} [\phi_e(t)] = \lim_{s \to 0} \left[s \frac{1}{1 + (K_\phi K_v / sN)} \frac{A}{s} \right] = 0 \tag{22.22}$$

Thus, when a step phase change is applied to the Type I PLL, the final value of the system error is zero, which better be the case or we do not have a PLL. Next, a unit step function of frequency is applied to the PLL. Phase is the integral of frequency, therefore, the reference signal becomes

$$\phi_{ref}(s) = \frac{A}{s^2} \tag{22.23}$$

Once again by substituting Equations 22.20 and 22.23 into the Laplace final value theorem gives

$$\lim_{t \to \infty} [\phi_e(t)] = \lim_{s \to 0} \left[s \frac{1}{1 + (K_\phi K_v / sN)} \frac{A}{s^2} \right] = \frac{AN}{K_\phi K_v} \tag{22.24}$$

Thus, when a step frequency change is applied to the Type I PLL, the final value of the system error is a constant, but, as can be seen, this constant is dependent upon the magnitude of the change, along with the division ratio of the loop. What this means to the PLL designer is that for any given N or output frequency, there will be a phase error between the reference signal and the PLL output. If the system cannot tolerate this error and needs the PLL output to be phase coherent with the reference signal, the designer will have to use a Type II loop. Next we will examine the case of a Type I loop with a time varying frequency input. The reference signal is given by

$$\phi_{ref}(s) = \frac{A}{s^3} \tag{22.25}$$

Substituting into the Laplace final value theorem gives

$$\lim_{t \to \infty} [\phi_e(t)] = \lim_{s \to 0} \left[s \frac{1}{1 + (K_\phi K_v / sN)} \frac{A}{s^3} \right] = \infty \tag{22.26}$$

What this indicates is that the phase error signal is continually increasing. The Laplace final value theorem can be applied to any Type of PLL and Table 22.1 is a quick reference to the system phase error within Type I and Type II loops.

TABLE 22.1 System Phase Error

Input Signal ϕ_{ref}	Type I	Type II
Phase	0	0
Frequency	Constant	0
Time varying frequency	Continually increasing	Constant

22.5.4 Type II Third-Order Loop

With the introduction of an integrator and a single pole RC as the LPF, $F(s)$, the PLL becomes a Type II third-order system. A popular configuration of the loop filter is shown in Figure 22.8. In order to prevent slew rate limiting in the amplifier, capacitor C_1 is commonly added to realize a single pole in front of the amplifier.

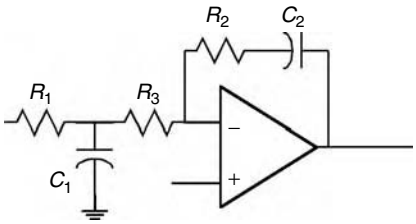


FIGURE 22.8 Integrator loop filter design.

The transfer function for the loop filter given in Figure 22.8 is given by

$$F(s) = \left[\frac{sR_2C_2 + 1}{s(R_1 + R_3)C_2}\right]\left[\frac{1}{s(R_1R_3/(R_1 + R_3))C_1 + 1}\right] \tag{22.27}$$

The zero is added to the filter transfer function to pull the phase margin up towards 90°. By substituting Equation 22.27 into Equation 22.16, the closed-loop transfer function becomes

$$A_{CL}(s) = \frac{K_\phi K_v(R_2C_2s + 1)}{(R_1 + R_3)C_1C_2s^3 + (R_1 + R_3)C_2s^2 + (K_\phi K_v R_2C_2/N)s + (K_\phi K_v/N)} \tag{22.28}$$

Figure 22.9 shows the closed-loop gain, the open-loop gain, and phase margin of a Type II third-order loop. The additional integrator causes the phase response to start from 0°. As mentioned earlier, the zero was added to the filter to move the phase response toward 90°. Owing to the additional integrator, the useable loop bandwidth is narrower than the Type I PLL. Typically, to achieve the necessary attenuation of the reference sidebands, the PLL needs to be designed with a higher order.

22.5.5 Higher-Order Loops

While the above examples serve well to further the understanding of PLLs, practical requirements often drive the designer to higher-order loops. To make the proper trade-off between settling time and spurious signals at the PLL output, higher-order filters are often necessary to minimize the amount of phase shift and maximize the amount of reference spur attenuation. Higher-order filters have a steeper attenuation characteristic, thereby, achieving less phase shift. Figure 22.10 illustrates this by plotting filter attenuation versus the phase bandwidth (i.e., filter's frequency at which the phase response is equal to 45° divided by the frequency of the stopband attenuation).

The system specifications, in some applications, require that the noise from the PLL meet a certain shape factor. The noise sources from within the loop can be tailored by the design of the loop's LPF

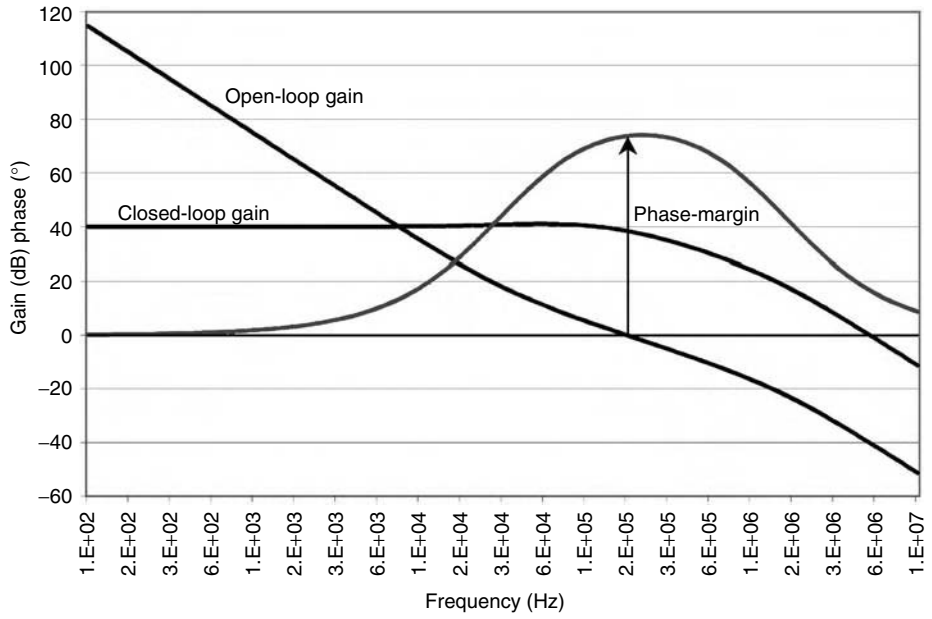


FIGURE 22.9 Type II second-order transfer functions.

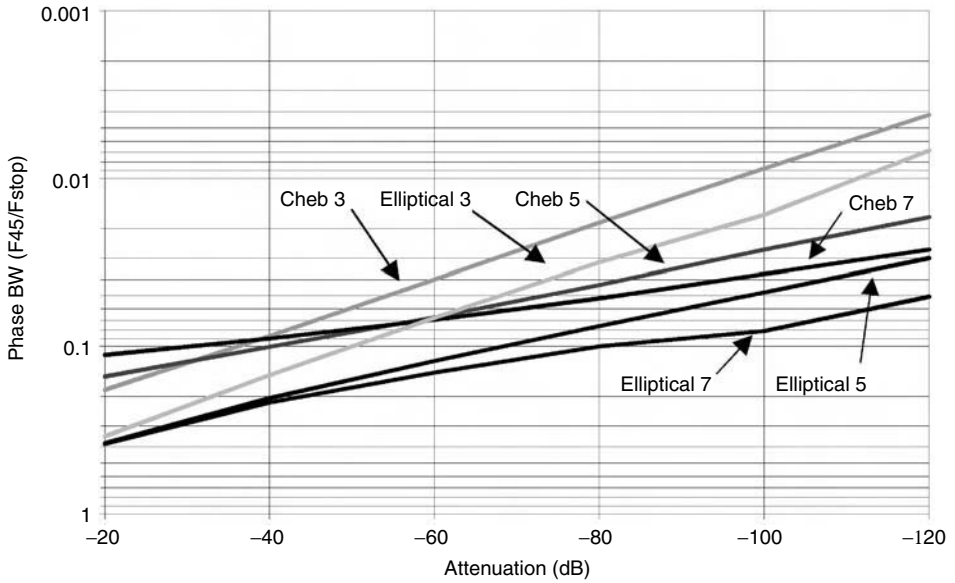


FIGURE 22.10 Lowpass filter phase shift comparisons.

(e.g., dual stopband filter). The noise shaping requirements typically forces the design to use high-order filters.

22.6 Phase Noise

The phase noise performance of a PLL is a critical parameter in the design of communication systems. The two main systems impacts of LO phase noise. The first is reciprocal mixing of adjacent channels,

which causes a decrease in the receiver's selectivity and the ability to receive weak signals. The second is a decrease in the systems SNR. The phase noise model developed in this chapter is directed toward the identification of the major contributors to the overall phase noise in a PLL, and evaluating the relative contributions of the significant sources that contribute to the output power spectral density. In certain cases, the source of noise in the loop can be pinpointed, but often it is difficult to characterize the noise precisely enough to make the necessary trade-offs. Most PLL designers are familiar with the different sources of noise that exist; in particular these include the frequency standard, VCO, FD, PD, and integrator/LPF (i.e., active filter).

The spectral characteristics of the oscillators (i.e., VCO and frequency standard) have been modeled in the past and are relatively well understood. Mathematically, an ideal sinewave can be described by the following equation:

$$V(t) = V_{\mathrm{o}} \sin(\omega t) \tag{22.29}$$

where V_{o} is the nominal amplitude and ω is the carrier frequency expressed in radians/second. In the real world, the sinewaves have error components related to both the phase and amplitude. A real sinewave signal is better modeled by

$$V(t) = [V_{\mathrm{o}} + \varepsilon(t)] \sin[\omega t + \Delta\phi(t)] \tag{22.30}$$

where $\varepsilon(t)$ is the amplitude fluctuation and $\Delta\phi(t)$ is the randomly fluctuating phase noise term. Both of these terms, $\varepsilon(t)$ and $\Delta\phi(t)$, are stationary random processes and are narrowband with respect to ω. For the purpose of this discussion, the amplitude spectral density will be ignored since it is of negligible significance compared to the phase perturbations.

There are two types of fluctuating phase terms. The first is the discrete signal components, which appears in the spectral density plot. These are commonly referred to as spurious signals. The second type of phase instability is random in nature and is commonly called phase noise. There are many sources of random phase perturbations in any electronic system, such as thermal, shot, and flicker noise.

One description of phase noise is the spectral density of phase fluctuations on a per-Hertz basis. The term spectral density describes the energy distribution as a continuous function, expressed in units of phase variance per unit bandwidth. The spectral density is described by the following equation:

$$s_{\phi}(f_{\mathrm{m}}) = \frac{\Delta\phi^2_{\mathrm{rms}}(f_{\mathrm{m}})}{\text{measurement BW}} \tag{22.31}$$

The units of spectral density are rad^2/Hz. The U.S. National Bureau of Standards has defined the single sideband spectral density as

$$\mathcal{L}(f_{\mathrm{m}}) = \frac{P_{\mathrm{ssb}}}{P_s} \tag{22.32}$$

where P_{ssb} is the power in 1 Hz of bandwidth at one phase modulation sideband and P_s is the total signal power. The single sideband spectral density, $\lambda(f_{\mathrm{m}})$, is directly related to the spectral density, $S_{\phi}(f_{\mathrm{m}})$, by

$$\mathcal{L}(f_{\mathrm{m}}) \cong \frac{1}{2} S_{\phi}(f_{\mathrm{m}}) \tag{22.33}$$

This holds true only if the modulation sideband, P_{ssb}, is such that the total phase deviation is much less than 1 radian. $\mathcal{L}(f_{\mathrm{m}})$ is expressed in dBc/Hz or dB relative to the carrier on a per hertz basis.

For the purpose of evaluating the noise performance of the PLL, each of the functional blocks is considered noiseless and a noise signal is added into the PLL at a summing junction in from of each of the functional blocks. In Figure 22.11, the noise sources within the PLL are shown along with the gains

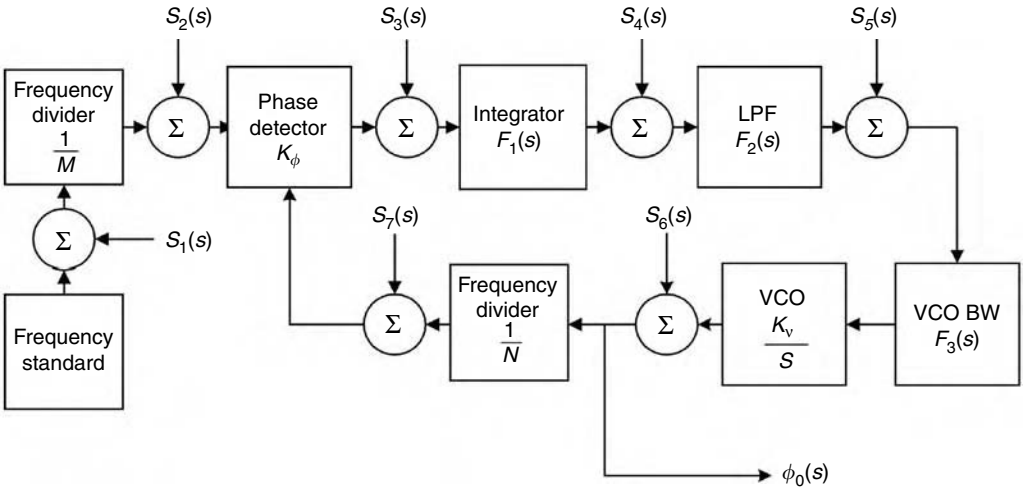


FIGURE 22.11 Phase noise sources in a PLL.

of the various blocks. In evaluating the contribution of each of the noise sources to the overall noise at the output of the PLL, each one will be considered alone. Since these noise sources are independent, superposition may be used to determine the phase noise at the output of the PLL. The transfer function for each of the noise sources is easily written. Once again, the transfer functions are derived using classic control theory. Two additional gain blocks have been added to the basic block diagram. The first one is the additional filter after the loop integrator as discussed in the section on higher-order PLLs. The second is the inclusion of the VCO's modulation bandwidth. Dependent upon the design of the VCO's input circuitry, the VCO's tune voltage input will have a finite bandwidth within which it will respond to an input signal. The 3 dB point of this response is defined as the modulation bandwidth. If the forward gain is defined as $G(s)$ and the feedback is defined as $H(s)$, then the closed-loop gain is given by

$$A_{\mathrm{CL}}(s) = \frac{G(s)}{1 + H(s)G(s)} \tag{22.34}$$

By applying Equation 22.34 to the PLL, the closed-loop gain for each of the noise sources is determined. This is done to characterize the loop's overall phase noise performance. By plotting the PLL's response to the individual noise sources, the proper trade-off for the optimization of the loop's performance can be made. The transfer function for each of the noise source and $K(\mathrm{s})$ are given in Table 22.2.

The overall transfer function of the PLL to each of the noise sources is shown below:

$$\Phi(s) = \frac{K(s)}{s + K(s)}\left[\frac{N}{M}S_1 + N(S_2 + S_7) + \frac{N}{K_\phi}S_3 + \frac{N}{K_\phi F_1}S_4 + \frac{N}{K_\phi F_1 F_2}S_5\right] + \frac{s}{s + K(s)}S_6 \tag{22.35}$$

First, the frequency standard noise S_1, reference FD S_2, PD noise S_3, integrator noise S_4, loop LPF noise S_5, and feedback divider noise S_7 are all acted upon the transfer function of a LPF. The actual cutoff frequency is determined by the designer while choosing the various parameters that establish the PLL open-bandwidth. The next noise component in Equation 22.35 to be considered is that of the VCO. The loop acts upon the VCO phase noise as if it was passed through a high pass filter. At offset frequencies that are much less than the open-loop bandwidth, the dominant noise sources are the digital noise and frequency standard noise. At offset frequencies that are much greater than the open-loop bandwidth, the dominant noise source is the VCO noise. Care needs to be taken in the designing of the loop parameters, such that peaking of the noise at the open-loop bandwidth does not occur. A general rule of thumb is that the designer would like to set the loop bandwidth at the point were the VCO phase noise crosses the digital

TABLE 22.2 Phase Noise Sources Transfer Functions

Source	Transfer Function	Simplification $K(s) = \frac{K_\phi F_1(s)F_2(s)F_3(s)K_v}{N}$
Frequency standard	$A_1(s) = \left[\frac{K_\phi F_1(s)F_2(s)F_3(s)K_v}{s + (K_\phi F_1(s)F_2(s)F_3(s)K_v/N)}\right]\frac{1}{M}$	$A_1(s) = \left[\frac{K(s)}{s+K(s)}\right]\frac{N}{M}$
Reference divider	$A_2(s) = \frac{K_\phi F_1(s)F_2(s)F_3(s)K_v}{s + (K_\phi F_1(s)F_2(s)F_3(s)K_v/N)}$	$A_2(s) = \left[\frac{K(s)}{s+K(s)}\right]N$
PD	$A_3(s) = \frac{F_1(s)F_2(s)F_3(s)K_v}{s + (K_\phi F_1(s)F_2(s)F_3(s)K_v/N)}$	$A_3(s) = \left[\frac{K(s)}{s+K(s)}\right]\frac{N}{K_\phi}$
Integrator	$A_4(s) = \frac{F_2(s)F_3(s)K_v}{s + (K_\phi F_1(s)F_2(s)F_3(s)K_v/N)}$	$A_4(s) = \left[\frac{K(s)}{s+K(s)}\right]\frac{N}{K_\phi F_1(s)}$
LPF	$A_5(s) = \frac{F_3(s)K_v}{s + (K_\phi F_1(s)F_2(s)F_3(s)K_v/N)}$	$A_5(s) = \left[\frac{K(s)}{s+K(s)}\right]\frac{N}{K_\phi F_1(s)F_2(s)}$
VCO	$A_6(s) = \frac{s}{s + (K_\phi F_1(s)F_2(s)F_3(s)K_v/N)}$	$A_6(s) = \left[\frac{s}{s+K(s)}\right]$
Feedback divider	$A_7(s) = \frac{K_\phi F_1(s)F_2(s)F_3(s)K_v}{s + (K_\phi F_1(s)F_2(s)F_3(s)K_v/N)}$	$A_7(s) = \left[\frac{K(s)}{s+K(s)}\right]N$

noise. In doing so, the optimum noise performance of the overall PLL can be achieved, but this is not always possible owing to settling time requirements. Next, it looks very advantageous to increase M and thereby decrease the contribution of the frequency standard phase noise. But if increasing M lowers the reference frequency, then N must increase if the overall multiplication factor to the output of the PLL is to remain the same. This actually decreases the frequency standard noise, but increases the multiplication of the phase noise contribution of the integrator, PD, loop filter, and loop divider. Therefore, the designer would want to keep N as low as possible in order to have the noise below open-loop bandwidth, ω_n, as low as possible.

The evaluation of the phase noise performance of a PLL can be an arduous task. Writing the transfer function for each of the noise sources within the loop will help to clarify how the loop acts upon each noise source as well as what trade-off can be made in the design for phase noise performance and settling time. The complete equation for the output phase noise of the PLL as a function of frequency is given by

$$S_\phi(f_m) = |A_1|^2 S_1 + |A_2|^2 S_2 + |A_3|^2 S_3 + |A_4|^2 S_4 + |A_5|^2 S_5 + |A_6|^2 S_6 + |A_7|^2 S_7 \tag{22.36}$$

where the $S_x s$ are power spectral densities that are a function of the offset frequency, f_m, from the carrier and the $A_x s$ are a function of the complex variable s. Also, by looking at the individual contributions of each of the phase noise sources, the designer can determine where to focus their energy in reducing the overall phase noise performance of the PLL.

22.7 Components of a PLL

22.7.1 Phase Detector

The PD produces an output signal that is proportional to the phase difference between the reference input, ϕ_{ref}, and the phase of the divided down VCO signal, ϕ_o/N. The most commonly used PD is a phase-frequency detector. In an out of lock condition, the output of the phase-frequency detector latches (i.e., the AC component is removed), thereby the error signal goes to the low or high rail, dependent upon the direction of phase error. The PD having the ability to perform frequency discrimination has greatly simplified the complexity of the PLL circuitry.

There are many factors to consider when determining which kind of PD the designer should use (e.g., PLL's tuning bandwidth, settling time, power consumption, phase error, etc.), but the most prevalent phase-frequency detector has a charge pump output.

The charge pump phase-frequency detector is used predominately within the commercial PLL ASIC industry. The charge pump PD output is a current that has an average value equal to the system phase error [2]. In Figure 22.12 the configuration of the PLL utilizing a charge pump phase-frequency detector is shown. One advantage of the charge pump phase-frequency detector is the reduction in complexity of the PLL. A second advantage is the ability to program the current, thereby being able to adjust the gain of the PD for optimum loop performance. A third advantage is that for narrow tuning bandwidth or fixed injection PLLs, only a passive filter on the charge pump output is required, thereby reducing the cost and size of the PLL. If the tuning bandwidth of the PLL is wide (e.g., octave) an op amp will usually have to be added to increase the range of the tuning voltage supplied to the VCO. A disadvantage of the charge pump phase-frequency detector is the leakage current. All attempts must be made to reduce the leakage current on the charge pump's output. As the level of the leakage current increases so will the amount of loop filtering needed to suppress the phase error spurious signals on the VCO's output.

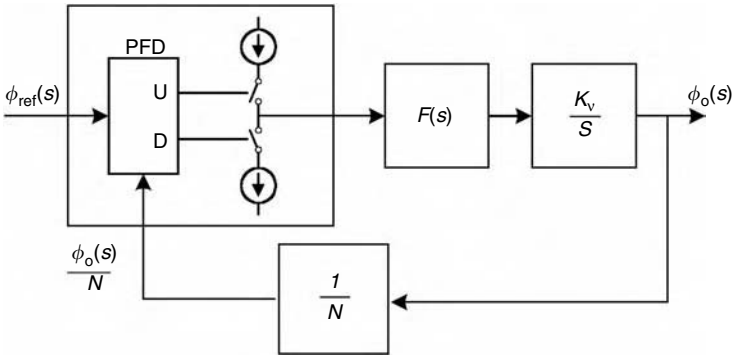


FIGURE 22.12 Charge pump PD PLL.

The charge pump phase-frequency detector has three states: (a) sourcing current, (b) sinking current, and (c) high impedance or tri-state. The amount of current being sourced or sunk is defined to be I_ϕ. Since the PD operates over a 2π range, the gain of the PD is, therefore, $I_\phi/2\pi$. If ϕ_o/N is leading ϕ_{ref} in phase, then the charge pump PD is sinking I_ϕ current, which is defined as the pull down current. If ϕ_o/N is lagging ϕ_{ref} in phase, then the charge pump PD is sourcing I_ϕ current, which is defined as the pull up current. When the two input waveforms have nearly identical phase, the charge pump PD is tri-stated. This is illustrated in Figure 22.13. The pull up and pull down currents must be equal for the gain of the PD to be linear. The charge pump current needs to be constant over the operating temperature of the

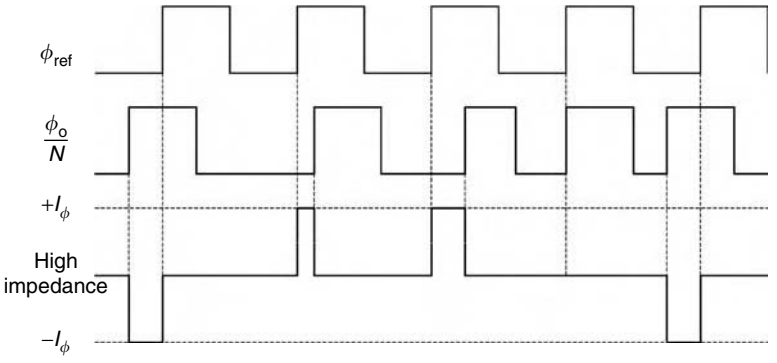


FIGURE 22.13 Charge pump PD output waveform.

system and operating voltage of the PD, owing to the fact that the PD gain changes, and thereby the loop gain changes, are proportional to the charge pump current.

22.7.2 Dividers

The output frequency of a PLL is determined by the value of the feedback divider as shown in Equation 22.2. Fundamentally, the feedback divider multiplies the reference frequency up to the output frequency. Hence, the controllability of the output frequency is dependent upon the programmability of the feedback divider. Typically, the digital circuitry of a PLL is implemented with CMOS technology, which limits the upper frequency range of the feedback divider. Therefore, a prescaler is added to the feedback path in order to maximize the input frequency to the divider while minimizing the overall power consumption.

The simplest form of a prescaler is a fixed modulus divider (i.e., divide by 2, 4, 8, etc.). The use of a fixed modulus prescaler will limit the step size at the output. For example, if you had a reference frequency of 1 MHz and an output frequency of 10 GHz, with a programmable divider you would be able to do 1 MHz channel resolution, but if you add a fixed divider by 2 prescaler, the channel resolution would be limited to 2 MHz. The designer could increase the division ratio of the reference divider, thereby decreasing the phase comparison frequency, but this would result in a 6 dB increase in the phase noise inside of the loop bandwidth. It would also force the open-loop bandwidth to be reduced and thereby potentially impacting the settling time of the PLL. With that said, it is not untypical in very high frequency (i.e., >6 GHz) synthesizer design to utilize fixed modulus prescalers.

The alternatives to fixed modulus prescalers are to implement a mixer in the feedback path or utilize a dual modulus prescaler. Albeit, the use of a mixer in the feedback path is very attractive for minimizing the phase noise multiplication (i.e., the mixer translates the frequency it does not divide down), it does require a considerable amount of hardware and is prone to causing spurious signals in the output. Great care must be used in the implementation and physical layout.

The dual modulus prescaler works in conjunction with low speed programmable dividers as shown in Figure 22.14. The core building block of this scheme is the dual modulus divider that can divide by one of two numbers, P or $P + 1$ (e.g., 10/11, 32/33, 64/65, etc.). The other divider is a counter that is programmable and controls the modulus of the prescaler; it is referred to as the A counter. The other divider is a programmable counter too, but its output is feed to the PD; it is referred to as the N counter. Its output also is used to reset the A counter. The two counters are clocked in parallel, therefore, the system will divide by $P + 1$ for A counts and by P for the $N - A$ counts. The operation of the dividers is shown by the following equation:

$$N_{\text{total}} = A \cdot (P + 1) + (N - A) \cdot P = N \cdot P + A \tag{22.37}$$

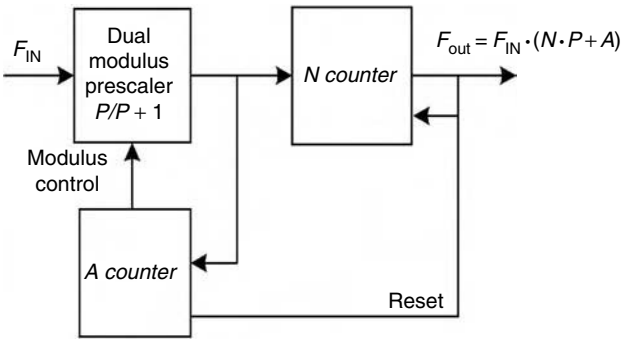


FIGURE 22.14 Dual modulus prescaler programmable divider.

For this system to work, A must always be less than or equal to the N counter; in order to be sure that no channels are missed the minimum total division ratio within the loop is given by

$$N_{\text{total}}(\min) = P \cdot (P - 1) \tag{22.38}$$

The other critical parameter that must be considered when designing a PLL with a dual modulus prescaler is propagation delay. The period of the VCO output frequency divided by the prescaler base division ratio must be greater than the sum of the times including (a) the propagation delay through the prescaler, (b) the prescaler setup time relative to the modulus control line, and (c) the propagation time from the input of the A counter to the modulus control output.

$$\frac{P}{F_{\text{in}}} > \text{Total delay} \tag{22.39}$$

The other consideration that must be given to all dividers is input power level. The divider will miscount if too much or too little RF power is applied to the input. The input sensitivity is a function of frequency. At high frequencies, the input sensitivity degrades owing to the transistor limitations; at lower frequency, the counter has problems making threshold decisions. When a divider miscounts the loop may appear to be in lock (i.e., the center frequency is correct), but the output signal is extremely noisy (i.e., it appears to have wideband modulation).

Another type of prescaler is the multimodulus that is used in integer and fractional-N synthesizers. A multimodulus prescaler is design by using cascaded divide-by 2/3 blocks, that are connected like a ripple counter [3]. It is not uncommon to utilize a 4 bit high speed prescaler, followed by a 6 bit low speed counter. The entire prescaler can be fabricated in a BiCMOS process. The multi-modulus prescaler has a much wider operating range and does not have the division ratio limitation seen in the dual modulus prescaler.

22.7.3 Voltage Controlled Oscillators

A great deal of research and development has been done on the design and fabrication of VCOs. The VCO is a critical building block of the any synthesizer design. Some of the issues that arise in the design of a VCO are the phase noise performance, spurious signals, power dissipation, tuning bandwidth, and tuning linearity. The predominate parameter is the phase noise performance. A semiempirical model was developed for the phase noise performance of an oscillator in 1966, Leeson–Culter:

$$\mathcal{L}(\omega_{\text{m}}) = 10 \log \left[\frac{2FkT}{P_{\text{s}}} \left[1 + \left(\frac{\omega_{\text{o}}}{2Q_{\text{L}}\omega_{\text{m}}} \right)^2 \right] \right] \tag{22.40}$$

where F is an empirical parameter called "device excess noise number," k is Boltzmann's constant, T is the absolute temperature, P_{s} is the average power dissipated in the resistive part of the tank, ω_{o} is the oscillator frequency, Q_{L} is the effective quality factor of the tank with the loading in place called "loaded Q," and ω_{m} is the offset frequency from the carrier [4,5]. In examining this model, it is apparent that the VCO designer should focus on three attributes in order to improve phase noise performance. The first is to utilize high Q resonators and minimize the loading on the resonator. The second is to minimize the number of active components in the oscillation path; this would also include the elimination of any lossy passive parts as well. The third is to maximize the oscillator swing. This, of course, can only be done in consideration of the overall system design and power budget. There are numerous circuit topologies that can be utilized to achieve these goals and there are many articles written on the design and implementation of VCOs.

22.7.4 Loop Filters

Once the designer has determined the type of PLL needed from the system requirements, the next step is to determine the configuration of the forward path of the PLL. In the design of a PLL, the level of the reference sideband spur on the VCO output is determined by a number of factors within the PLL. The amount of filtering needed for a desired reference sideband spur level can be calculated by utilizing the formulas presented in this section. Though these equations are empirical and by no means exact, they are a good starting point for determining the performance (i.e., filter attenuation) that is needed. In the basic form, the loop filter can take on three forms: (a) a lead-lag filter, (b) an active filter integrator, and (c) a charge pump passive filter. These three filters are shown in Figure 22.15. It is common practice to add an elliptic filter to the output of these basic filters, in order to achieve the needed attenuation of the reference sidebands. The configuration of the loop filter is dependent upon the type of phase-frequency detector circuitry used.

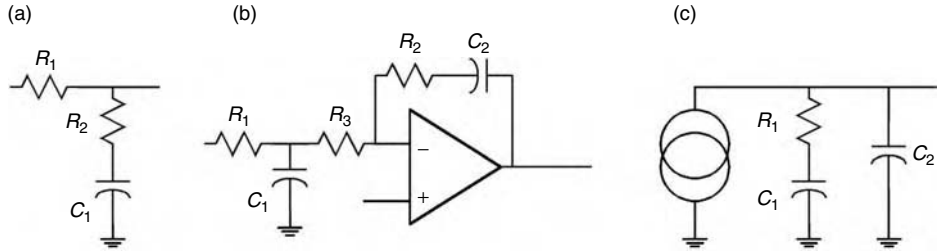


FIGURE 22.15 Loop filters: (a) lead-lag, (b) active filter integrator with zero, and (c) charge pump.

As mentioned before, the charge pump PD has become very popular in commercial PLL ASIC applications. A typical filter used with a charge pump PD is shown in Figure 22.16. In any application where wide tuning is needed, a higher tuning voltage must be supplied to the VCO. The higher tune voltage is supplied by adding an amplifier stage. A closer examination of the transfer function, $F_1(s)$, results in the determination of the component values. The transfer function of the filter, $F_1(s)$, is

$$F_1(s) = \frac{R_1 C_1 s + 1}{s(C_1 + C_2)(sR_1(C_1 C_2/(C_1 + C_2)) + 1)} = K\frac{\tau_z s + 1}{s(\tau_p s + 1)} \tag{22.41}$$

where

$$K = \frac{1}{C_1 + C_2} \tag{22.42}$$

$$\tau_z = R_1 C_1 \tag{22.43}$$

$$\tau_p = R_1 \frac{C_1 C_2}{C_1 + C_2} \tag{22.44}$$

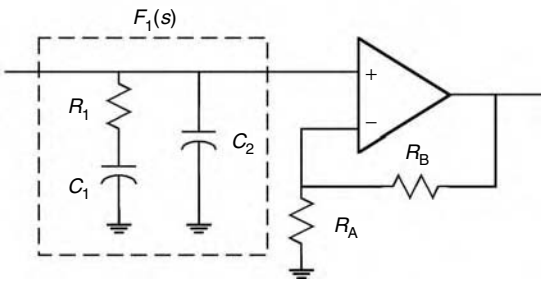


FIGURE 22.16 Basic forward path of a charge pump PLL.

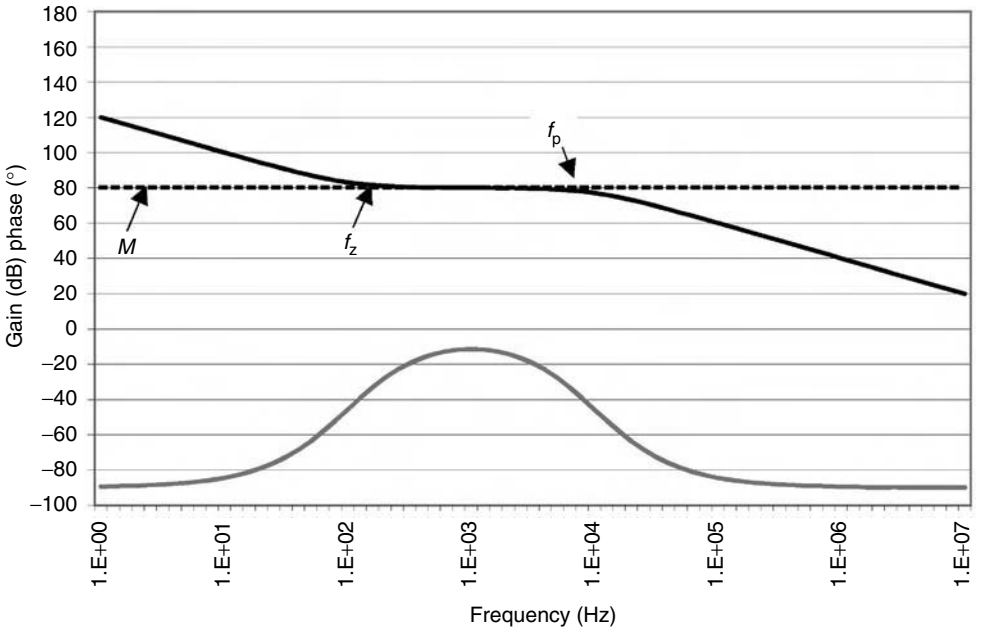


FIGURE 22.17 Transfer function of a charge pump filter.

By plotting the transfer function, $F_1(s)$, as a function of frequency, the pole and zero break frequencies can easily be seen, as shown in Figure 22.17.

In designing the loop filter for a specified loop bandwidth, the PLL's natural frequency will have to be adjusted by the filter gain. This will have to been done with the appropriate amount of phase margin. The pole and zero location and the filter gain determine the component values for the filter [6]. The magnitude of $F_1(s)$ at the geometric mean of the pole and zero frequencies is defined as

$$M \equiv \left| F_1 \left(j2\pi \sqrt{f_z f_p} \right) \right| \tag{22.45}$$

The component values for R_1, C_1, and C_2 may be calculated by the following formulas:

$$R_1 = M \frac{f_p}{f_p - f_z} \tag{22.46}$$

$$C_1 = \frac{1}{2\pi f_p M} \tag{22.47}$$

$$C_2 = \frac{f_p - f_z}{2\pi f_p f_z M} \tag{22.48}$$

Once the component values of the filter, $F_1(s)$, are determined, the amount of additional filtering for attenuation of the reference sidebands will need to be calculated. It is important to note that the closed-loop transfer function will not predict the amount of filtering needed for attenuation of the reference side bands. The closed-loop response does not take into account all of the parameters that effect the level of the reference sideband signal on the output of the VCO. An approximate calculation of the needed filtering for a given reference sideband level may be calculated from the following formula:

$$F(\text{dB}) = 20 \log \left[\frac{\sin(n\pi (V_{\text{DCmax}} / R_{\text{leak}} I_{\text{max}})}{n\pi F_{\text{ref}}} (m R_1 K_v I_{\text{max}}) \right] + \text{RSB(dB)} \tag{22.49}$$

where n is the harmonic of the reference frequency to be filtered, V_{DCmax} is the maximum voltage of the charge pump, R_{leak} is the leakage resistance across the charge pump, $I_{\max}$ is the maximum charge pump current, F_{ref} is the reference frequency, m is either the gain or loss in the forward path (i.e., op amp/lead-lag network, etc.), and R_1 is the value of the resistor in the filter, $F_1(s)$ [7]. An example of a charge pump PD utilizing the filter, $F_1(s)$, is given below.

22.7.5 Example

A charge pump PD utilizing the filter shown in Figure 22.16 with the following parameters: maximum voltage of the charge pump of 5 V, maximum VCO gain of 40 MHz/V, a reference frequency of 1 MHz, a filter gain of 1, a leakage resistance of 100 kΩ, a value of 200 Ω for R1 and the fundamental harmonic at a level of −70 dBc.

$$F(\mathrm{dB}) = 20\log\left[\frac{\sin(\pi(5/1\mathrm{E}5 \times 0.005))}{\pi \times 1\mathrm{E}6}(200 \times 40\mathrm{E}6 \times 0.005)\right] + 70 = 62.04\,\mathrm{dB} \tag{22.50}$$

Approximately, 62 dB of additional attenuation will be needed at the reference frequency of 1 MHz. If a simple single pole filter is used, the designer will have to take care that it does not introduce excessive phase shift at the loop bandwidth.

22.8 Fractional-*N* PLL

In the quest to improve performance of synthesizer, the fractional-N PLL has generated a lot of interest, owing to the fact that it can have small channel spacing with wide loop bandwidth, potentially all in one PLL. As in all things, this comes with a price. A fractional-N synthesizer differs from a traditional integer-N in which the division ratio, N, is a fixed integer, forcing a trade-off between loop bandwidth and frequency step size. In a fractional-N synthesizer, the division ratio is dynamically varied or dithered to achieve intermediate division ratios. The inherent problem with fractional-N PLLs is the undesired spurious signals that are generated. For this reason, designers have implemented a fractional-N scheme that utilizes a sigma-delta ($\Sigma\Delta$) modulator. The feedback divider is controlled by a form of $\Sigma\Delta$ modulator known as a multistage noise shaping (MASH). The division ratio (ΔN) is not confined to N and $N + 1$, but it depends on the number of stages (order) of the MASH given by

$$N^{th}\ Order:\ -2^{(n-1)} - 1{:}2^{(n-1)} \tag{22.51}$$

Therefore with a fourth-order, the integer value of the divider would range from $-7 \leq N \leq +8$. The MASH output, ΔN, is a pseudo-random sequence. The VCO frequency is governed by the long-term mean of ΔN. The MASH structure is a series of first-order sigma delta modulators, each fed by the quantization error of the previous stage shown in Figure 22.18. The block diagram is represented as a discrete-time system since the modulator samples the frequency control signal to be consistent with a MASH structure of $\Sigma\Delta$ modulators, the quantized outputs are shown to pass through a cascade of parallel input differencers. At the final summation stage, the modulation signal is added to the integer frequency control signal.

The $\Sigma\Delta$ modulator generates a sequence of integers based on a fractional frequency control word. The frequency control word is modulated by the sequence and the sum is directed to the FD. As the modulator shown in Figure 22.18 includes four accumulators, the divider hardware must be able to realize the resulting peak deviation and must accommodate a divisor that is sampled at the same rate as the PD. Consideration of the spectral characteristics of the modulator sequence needs to take place since the LPF must be designed to attenuate the added quantization noise. The quantization noise for

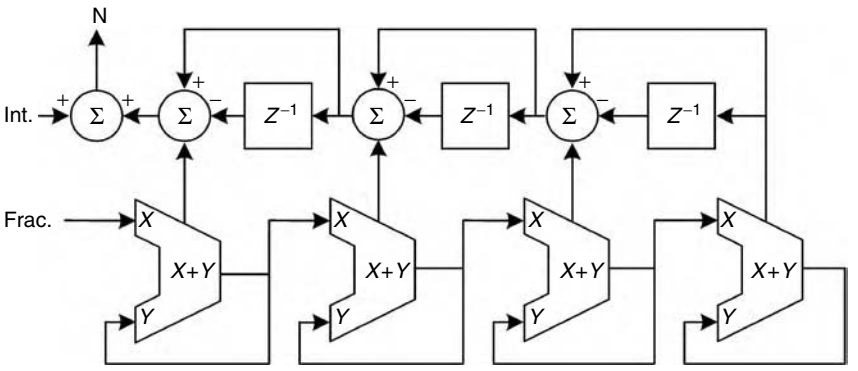


FIGURE 22.18 Block diagram of a MASH ΣΔ based modulator.

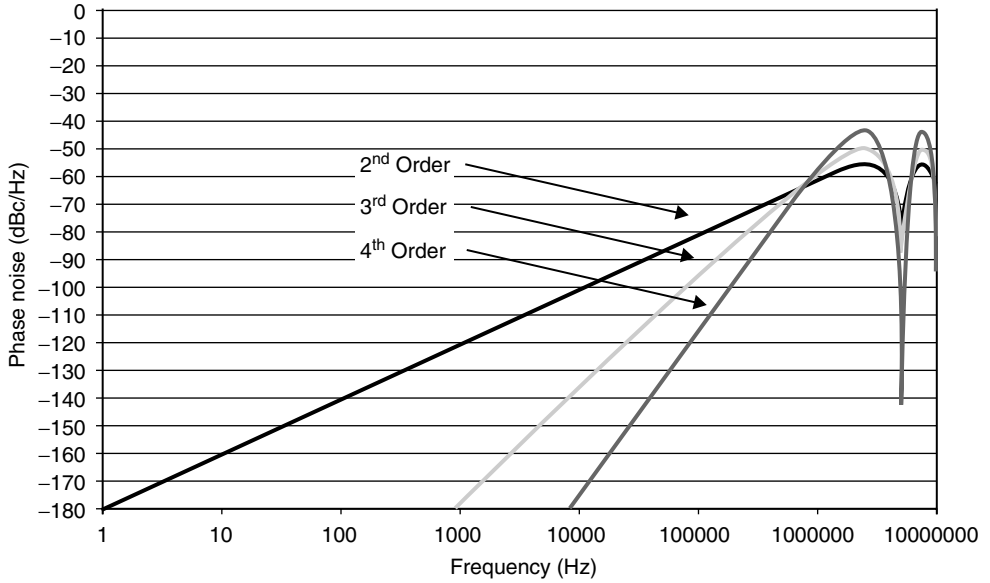


FIGURE 22.19 Spectral shaping of various order modulators.

the multimodulator scheme is given by

$$\mathcal{L}(f) = \frac{(2\pi)^2}{12 f_{\text{ref}}} \left[2 \sin\left(\frac{\pi f_{\text{m}}}{f_{\text{ref}}} \right) \right]^{2(m-1)} \tag{22.52}$$

where m is the number of modulators, f_{ref} is the comparison frequency at the PD and f_{m} is the offset frequency from the carrier [8]. The frequency spectrum shown in Figure 22.19 is consistent with the expected characteristic of ΣΔ noise shaping. The use of an ΣΔ modulator in a PLL has its advantages, in that small channel spacing can be realized. The limitations of this technique are the impact of nonlinear effects that the PLL has on phase noise performance. In order to realize the phase noise performance, the PD needs to be highly linear. Also, care must be taken in the layout of the circuitry owing to the fact that the VCO and PD comparison frequency are noncoherent (i.e., potential spurious signals getting through the loop filter).

22.9 Transient Response

So far, the assumption has been made that the PLL is operating in a steady state. What is the PLL's response when a disturbance is introduced or what is the transient response? In most of the articles written on PLL design, a second-order approximation is used to model the PLL's transient response to a change in phase or frequency. The most common is a change in the loop's division ratio to bring the PLL to a new frequency output. This change has two effects on the loop. One is, because the closed-loop response is dependent upon the value of the feedback divider, N, the loop bandwidth changes. The PLL has to acquire phase lock to the new VCO output frequency, which is known as the transient response. Of course, dependent upon how large the change in frequency, the gain of the VCO will also have effect on the loop bandwidth. With the mathematical modeling software tools available to the designer today, the second-order approximation is unnecessary, but is still useful in understanding the basic loop transient response.

In most systems, the PLL's transient phase error is the response of interest. Albeit, many system specifications define the frequency error, Phillips [9] has shown that the phase settling characteristic corresponds to the system performance better than the frequency settling characteristics. The transient phase response is the phase difference between the final value of the VCO and a steady-state signal, which has the same phase as the final value of the VCO. In the laboratory, the transient phase response is measured by mixing a signal generator and the VCO's output signal, both of which need to be phase locked to the same frequency standard. The output of the mixer shows the phase difference and needs to be lowpass filtered in order to remove the summed output. As an example, Figure 22.20 shows the result of a simulation of a PLL hopping between two frequencies. During the first hop, the loop has too little phase margin and there is excessive ringing. The second hop shows the phase response with a phase margin of 70°.

There are many factors within the loop that can affect the transient response. The nonlinearity of the VCO gain, which typically is less at the high end of its tuning range, will effect the loop bandwidth and, thereby, the phase margin. The frequency range and modulation bandwidth of the VCO will impact the transient response. Prepositioning of the VCO's control voltage, either with a digital control word into a DAC or by summing another PLL's tune voltage, can be done to reduce the amount of overshoot and thereby reduce the settling time. To prevent additional noise from being introduced into the PLL, careful

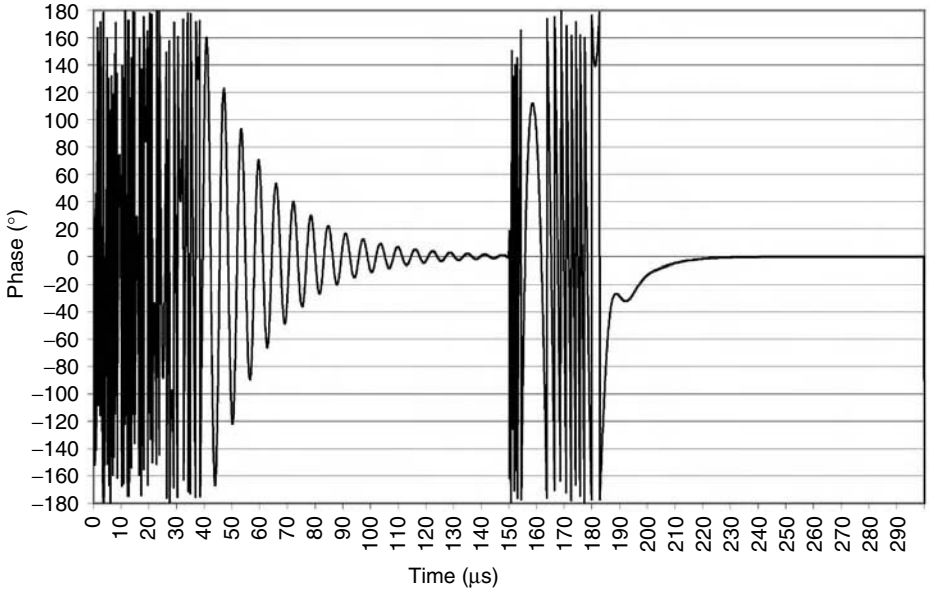


FIGURE 22.20 Phase settling response of a Type I second-order PLL.

design of the prepositioning circuitry must be done. The discrete or sampling nature of the PD and divider need to be considered in wide loop bandwidth designs. The continuous time approximation is typically used in transient modeling, but a number of papers have been written that use z-transforms to model the loop's response [10]. Once again, if the loop bandwidth is kept low relative to the reference signal frequency, the discrete nature of the PD can be ignored. When a charge pump PD is used, the mismatch between current sources, I_ϕ and $-I_\phi$, the leakage current of the charge pump as well as the leakage current of the components used in the loop LPF and board parasites all have an effect on the transient response. If an operational amplifier is used in the loop filter design, the slew rate and voltage limits will impact the transient response.

22.10 Conclusion

In this chapter, we have considered some of the fundamental design considerations that go into the design of a PLL. The design process is made up of a series of trade-offs (e.g., wide loop bandwidth for improved settling time, but narrow loop bandwidth for improved noise performance). There cannot be enough emphasis placed on the robustness of the PLL design. The designer needs to ensure that there is enough margin in the design parameters such that the loop works well over its operating temperature and other environmental conditions, as well as component tolerance. Modeling plays a key role in the development of the PLL and it is imperative that the designer have an understanding of those models and the limitations inherent in any mathematical model. Communication systems place severe requirements on the frequency synthesizer and the design of the PLL is critical in achieving the necessary system performance. The reader is encouraged to tap into the vast amount of literature on the design of PLLs to further their understanding of PLL design.

References

1. Egan, W.F., *Frequency Synthesis by Phase Lock*, Robert E. Krieger Publishing, Malabar, Florida, 1990.
2. Gardner, F.M., Charge-Pump Phase-Lock Loops, *IEEE Transactions on Communications* Vol. COM-28, pp. 1849–1858, Nov. 1980.
3. Wafe, A. and Ahmed, A., *High-Speed Multi-Modulus Prescaler Architecture for* $\Sigma\Delta$ *Fractional-N Frequency Synthesizer*, ISCAS 2004, pp. 241–244.
4. Leeson, D.B., A Simple Model of Feedback Oscillator Noise Spectrum, *IEEE Proceeding Letters*, vol. 54, no. 2, pp. 329–330, Feb. 1966.
5. Culter, L.S. and Searle, C.L., Some Aspects of the Theory and Measurement of Frequency Fluctuations in Frequency Standards, *IEEE Proceedings*, vol. 54, no. 2, pp. 136–154, Feb. 1966.
6. Opsahl, P.L., *Charge-Pump Filter Design*, Rockwell Collins, Inc., Frequency Control Team Internal Paper, 1999.
7. Mroch, A.B., *Charge-Pump Filter Attenuation*, Rockwell Collins, Inc., Frequency Control Team Internal Paper, 1999.
8. Miller, B. and Conley, R., A Multiple Modulator Fractional Divider, *IEEE Transactions on Instrumentation and Measurement*, vol. 40, No. 3, pp. 578–583, June 1991.
9. Phillips, D.E., Settling Time Specifications: Phase or Frequency? *IEEE Military Communications Conference*, vol. 3, pp. 806–810, Oct. 1994.
10. Crawford, J.A., *Frequency Synthesizer Design Handbook*, Artech House, Inc., Norwood, MA, 1994.

23

Filters and Multiplexers

Richard V. Snyder
RS Microwave

23.1 Introduction

"Filter: a device or material for suppressing or minimizing waves or oscillations of certain frequencies"…per Webster. This definition, while accurate, is insufficient for microwave engineers. Microwave systems and components enhance and direct, as well as suppress, waves and oscillations. Components such as circulators, mixers, amplifiers, oscillators, switches (in common with most complex systems) are, in fact, filters in which inherent physical properties are represented as smaller, constituent networks embedded within larger "filtering" (response-determining) structures or systems. Inclusion of the concept of embedding is, thus, central to understanding microwave filters. To clarify "embedding": the terminals of a well-defined subnetwork are provided with a known interface to the rest of the system, thus, providing selective suppression or enhancement of some oscillatory effect within the device or system. The subnetworks can be linear or nonlinear, passive or active, lumped or distributed, time dependent or not, reciprocal or non-reciprocal, chiral (handed) or non-chiral, or any combination of these or other properties of the basic constituent elements. The subnetworks are carefully defined (or characterized) so that the cascade response of a series of such subnetworks can be predicted (or analyzed) using software

simulation tools employing a variety of methods, such as linear, harmonic balance, Volterra series, finite-element, method of moments, finite-difference time domain, and so on. The careful definition normally involves the process called "synthesis," in which the desired response to a particular stimulus suggests a topological form for the subnetwork, followed by extraction of the specific elements of the subnetwork. The computed response of the synthesized network is compared with the desired response, with iteration as necessary using repeated synthesis or an perhaps optimization loop within the simulation tool. Hybrid combinations of these two iterative approaches are possible.

23.2 Analysis and Synthesis

The difference between prediction or "analysis" and definition or "synthesis" can be summarized as follows: the word analysis comes from the Greek *lysis*, a loosening, and *ana*, up; hence a loosening up of a complex. Synthesis, on the other hand, means the building up of a complex from parts or elements to meet prescribed excitation–response characteristics. Figure 23.1 illustrates an example of the excitation, network, and response.

Another difference between analysis and synthesis must be considered. There is always a unique solution for an analysis, although it might be hard to find. Synthesis, on the other hand, might result in several networks with the specified response, or possibly no solution whatsoever. In general, solutions are not unique but some might be more realizable than others. Figure 23.2 presents the general problem of synthesis. In the following sections, ω is the radian frequency, that is, $2\pi f$, and s represents the Laplace variable. What combination of elements in Figure 23.2 will give the prescribed response? It is important to realize that with a *finite* number of elements, in general, the required response *cannot* be realized at all. Functions having a required variation over some band of frequencies and zero value for all other frequencies cannot be represented by a rational function of the form of a quotient of polynomials. Thus, it is necessary to modify the response requirements to include some *tolerance.*

Figure 23.3 illustrates the imposition of a tolerance, or acceptable difference, between the desired response and the resultant response, for a particular synthesized characteristic. The approximation can take many possible forms. The approximations might require the magnitude squared of the voltage ratio

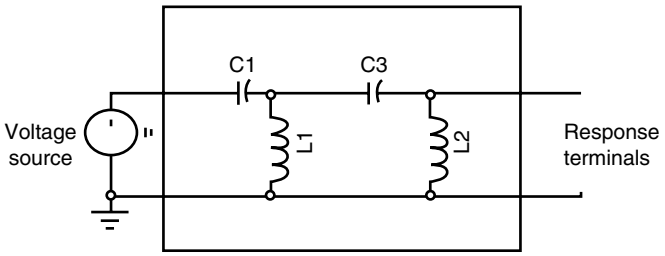


FIGURE 23.1 Example of a network.

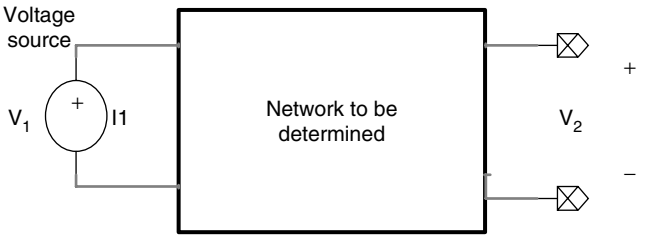


FIGURE 23.2 Transfer function synthesis. $|G_{12}(j\omega)| = |V_2(j\omega)/V_1(j\omega)|$

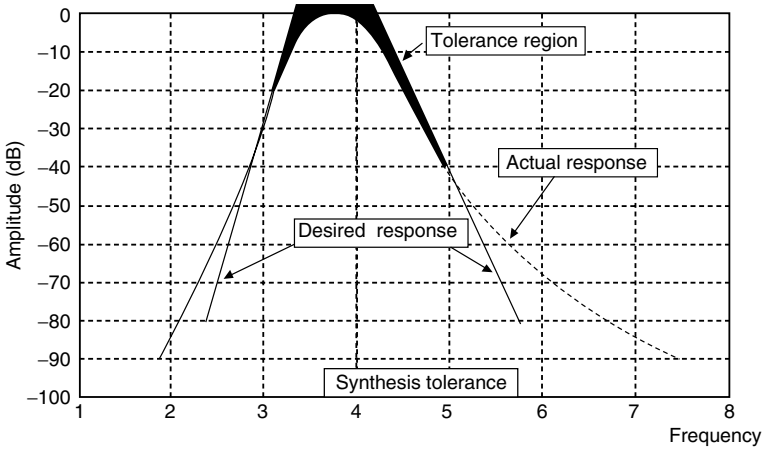


FIGURE 23.3 Polynomial approximation.

to be the quotient of rational and even polynomials in ω. A typical quotient of polynomials might be as given in Equation 23.1.

$$|G(\omega)|^2 = \frac{a_0\omega^6 + a_2\omega^4 + a_4\omega^2 + a_0}{b_0\omega^6 + b_2\omega^4 + b_4\omega^2 + b_0} \tag{23.1}$$

This is not a unique polynomial for realization. There are any number of other polynomial ratios of lower or higher degrees that may be used. The higher the degree of assumed polynomials, the better the approximation to the desired response, the smaller the tolerance region. The coefficients of Equation 23.1 are determined by the solution of a set of simultaneous linear equations, as many simultaneous equations as there are unknown coefficients. In general, we cannot match the desired response characteristic at all points in the spectrum; rather, we must choose to match exactly at certain points and approximately over the remainder of the tolerance region. We can choose to match the points, the derivatives, or use other criteria that will be discussed herein. The response shown, in Figure 23.3, is amplitude versus frequency. Typically, a network also has a desired time versus frequency response. Generally, these two requirements are inter-related and may not be specified independently. In very important classes of approximation, the amplitude and time responses are connected by the Hilbert transform, and ,thus, to know either the amplitude or the time response is sufficient to enable determination of the other. This will be discussed in the section on Approximations. The polynomial in Equation 23.1 has a simple dependence upon frequency ω, (representing lumped elements in the network). Generally, microwave filters include distributed elements such as quarter wave resonators, which display response characteristics dependent upon transcendental functions, such as $\tan(\omega)$. The resultant synthesis polynomials require more specialized techniques for element extraction. It is possible that real networks will include both lumped and distributed elements, with very complex transfer function polynomials. The synthesis process can, thus, be quite complex, and a comprehensive coverage is beyond the scope of the present article.

23.3 Types of Transfer Function

Before approximating a particular transfer function, one must determine the type of transfer function desired. These functions can be defined in terms of amplitude or time, expressing either of them function of frequencyIt is convenient to initially concentrate on amplitude versus frequency transfer characteristics. Table 23.1 presents the four available types:

It should be understood that the above transfer functions are defined between any pair of input–output ports of a potentially multi-port network. Transfer functions so defined are known as "two-port" transfer functions. We will initially restrict our efforts to such two-port circuits.

TABLE 23.1 Transfer Function Types

Transfer Function Type	Characteristics
Lowpass	Low loss region approximated over some bandwidth, prescribed rejection achieved at frequency some distance from the *highest* frequency in the low loss approximation region (Figure 23.4a)
Highpass	Low loss region approximated over some bandwidth, prescribed rejection achieved at frequency some distance from the *lowest* frequency in the low loss approximation region (Figure 23.4b)
Bandpass	Low loss region approximated over some bandwidth, prescribed rejection achieved at frequencies some distance from *both* the *highest and lowest* frequencies in the low loss approximation region (Figure 23.4c)
Bandstop	Low loss approximated over two regions, extending downward towards DC and upward towards infinity. Prescribed rejection achieved at frequencies some distance *above* the lower low loss region and *below* the upper low loss region (Figure 23.4d)

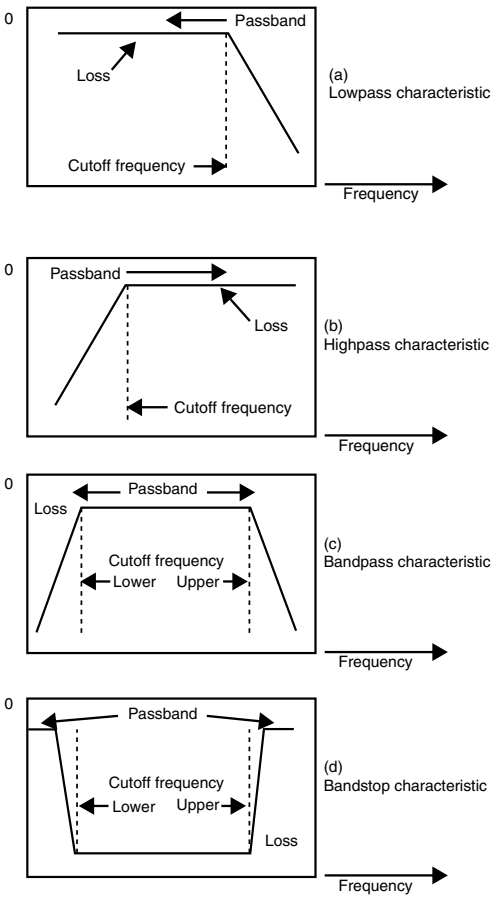


FIGURE 23.4 Basic transfer functions.

23.4 Approximations to Transfer Functions—Traditional

It is not possible to achieve the flat passbands and abrupt transitions illustrated in Figure 23.4 without using an infinite number of elements, each with zero resistance. We will discuss the properties of elements used to realize filters in a later section, but certainly the "Q" of available elements is less than infinity.

Thus, some approximation to the idealized transfer functions must be made in order to implement a filter network falling within the allowable tolerance shown in Figure 23.3. Essentially, the approximation procedure is directed towards writing mathematical expressions that approximate the ideal forms shown in Figure 23.4. These expressions include polynomial functions that are substituted into the left side of Equation 23.1 prior to element extraction. Some of the most common approximations will now be discussed. We will treat approximations to the amplitude response in some detail, and will briefly touch on approximations to phase or time delay.

23.4.1 Butterworth

The response function given by Equation 23.2 is known as the nth order Butterworth or maximally flat form.

$$|G_{12}(j\omega)| = \frac{1}{\sqrt{1+\omega^{2n}}} \tag{23.2}$$

From binomial series expansion

$$(1 \pm x)^{-n} = nx + \frac{(n+1)\,x^2}{2!} \mp \frac{n(n+1)(n+2)x^3}{3!} + \cdots, \quad x^2 \leq 1 \tag{23.3}$$

We see that near $\omega = 0$

$$(1+\omega^{2n})^{-1/2} = 1 - 0.5\omega^{2n} + 0.375\omega^{4n} - 0.313\omega^{6n} + \cdots \tag{23.4}$$

and from this expression, the first $2n-1$ derivatives are zero at $\omega = 0$. Thus, the magnitude

$$|G_{12}(j\omega)| = 0.707 \quad \text{for all values of } n. \tag{23.5}$$

The pole locations corresponding to the Butterworth response may be determined using analytic continuation of the binomial series expansion above. The poles of this function are defined by the equation (s is the Laplace variable)

$$1 + \left(-s^2\right)^n = 0 \tag{23.6}$$

The poles so defined are located on a unit circle in the s-plane and have symmetry, with respect to, both the real and the imaginary axes. Only the left half plane poles are used to form, what is known as, the all-pole response function which will yield the response required in Equation 23.2.

The form of the Butterworth response is shown in Figure 23.5a for several values of n. The $2n-1$ zero derivatives ensure the "maximally flat" passband characteristic.

23.4.2 Chebychev

If a rippled approximation to the passband region of the ideal transfer function is acceptable, one can use the expression

$$|G_{12}|^2 = \frac{1}{1+\varepsilon^2 C_n^{2\omega} 2} \tag{23.7}$$

where $C_n(\omega)$ is the nth order Chebychev polynomial and $\varepsilon < 1$ is a real constant. These polynomials are defined in terms of a real variable ω as follows:

$$C_n(\omega) = \cos\left(n \cos^{-1} \omega\right) \tag{23.8}$$

The response of a Chebychev-approximated transfer function is shown is Figure 23.5b. Analytic continuation can again be used to locate the poles, which will be found to be distributed on an ellipse, major

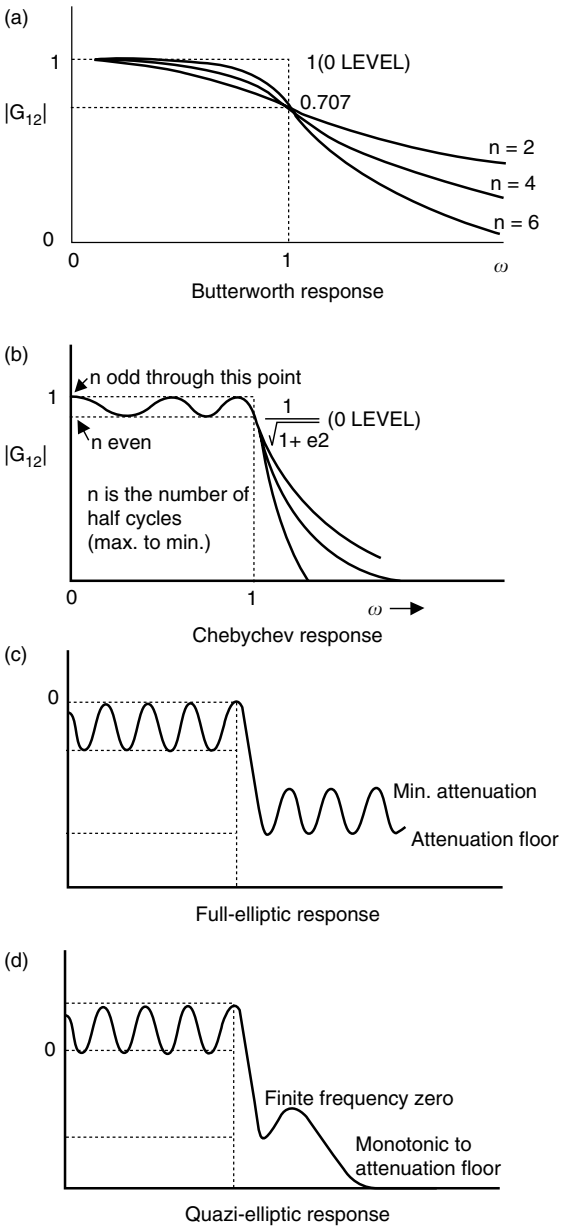


FIGURE 23.5 (a) Butterworth response, (b) Chebychev response, (c) full-elliptic response, and (d) quasi-elliptic response.

and minor axes, respectively, the imaginary and real axes of the *s*-plane, *s* the normal Laplace transform variable $\sigma + j\omega$. (Remember that the Butterworth poles were distributed on a circle, same axes). The reader is referred to many standard reference works for the details of element extraction, but the response of the Chebychev approximation will generally provide less passband performance but will achieve specified levels of stopband attenuation more quickly than the Butterworth approximation discussed previously. Butterworth and Chebychev transfer functions can be realized using single or resonant elements, with coupling only between adjacent elements. As such, the physical form for the network has the appearance of a ladder (see Figure 23.1) and these circuits are known as "ladder" networks.

23.4.3 Elliptic Approximation

If one can utilize rippled approximations to both the amplitude of passband and stopband regions, the resultant filter characteristics will display, somewhat, better passband performance coupled with steeper attenuation slopes, as compared to the Chebychev, but with attenuation slope characteristics with a level set on the minimum value of stopband. The response is shown in Figure 23.5c. The stopband region contains finite-frequency transmission zeros. Filters designed with elliptic responses are derived from expressions containing elliptic functions. Such filters are sometimes termed "full-elliptic" or "Cauer parameter." They can be derived from many starting points, but it is important to note that for bandpass cases, narrow passbands are hard to achieve and for bandstop, narrow stopbands are difficult. Designs of this type require extra resonant elements and sometimes coupling between nonadjacent resonators. Typically, such designs are used to achieve specified minimum stopband levels in close proximity to the passband (5–10% away in frequency).

23.4.4 Quasi-Elliptic Approximation

This class of function is achieved with rippled approximation to the passband amplitude and a limited number of ripples (finite frequency transmission zeros) in the stopband. Typically, the filter stopband slope displays monotonicity beyond the few ripples. Filters, in this category, can achieve the improved passband response of elliptic designs, with stopband performance almost as steep as an elliptic, and also with maximum stopband levels considerably improved as compared to the full-elliptic approach. If the extra zeros are real-axis, rather than real-frequency, the filter will display improved passband flatness and more constant group delay. Filters, in this category, can be what is known as "nonminimum-phase," as the extra zeros can be located in the right-half plane. The general response type is illustrated in Figure 23.5d. These filters usually require coupling between nonadjacent resonators (sometimes this is called "cross-coupling"), but do not need the extra resonant elements nor do they display the realization difficulties of full-elliptic designs. Although this design approach has been known since the 1970s, recent advances in simulation tools and synthesis techniques have resulted in the emergence of this category as the "cutting edge" in filter design. Such filters are also known as "pseudo-elliptic."

23.4.5 Other Approximations

The aforementioned approaches have all started with approximating the amplitude portion of the transfer function. In some cases, it is desirable to approach the delay or phase as a function to be approximated. In the ladder-derived filters above, to know the amplitude is to have determined the phase/delay (and contrawise). In more complex structures (cross-coupled), it is possible to have some degree of control over both amplitude and phase/delay. It is also possible to adjust the amplitude or phase/delay properties with the cascade of an additional circuit, known as an "equalizer." When the adjustments are internal, as in the cross-coupled cases, the equalization is known as "self-equalized." In this case, the extra transmission zeros are located on the s-plane real axis (imaginary frequency). If the additional circuit is used for equalization of either amplitude or phase/delay, the descriptive term is "externally equalized." Some of the more common approximations are summarized in Table 23.2.

23.5 Approximations to Transfer Functions—Modern Synthesis

In the last few years, a number of new topologies and concepts have arisen. Much attention has been paid to direct coupling from source or load to internal resonators, and also to the use of what have been termed "non-resonant nodes." Resonators have been grouped in clusters of three (called Trisections), groups of four (called Quadruplets or Quads) and even higher order groupings of resonators. These blocks are, then, cascaded to form higher order networks. For example, trisections and quads may be cascaded and are,

TABLE 23.2 Summary of Common Approximations

Approximation Type	Salient Characteristics
Butterworth	Lowest loss at center frequency, bandwidth is defined as −3 dB relative to center, maximally flat (no ripples) in passband, stopband slope depends on order, good group delay performance, generally easy to realize as ladder structure. Poles distributed on a unit of the *s*-plane (Laplace variable space). This is an all-pole filter
Chebychev	Rippled approximation to passband with reflections proportional to ripple level and higher than Butterworth, for the same order. Stopband slope depends on order, but steeper than Butterworth for the same order, poor group delay in passband (not constant), generally easy to realize as ladder structure. Poles distributed on an ellipse, major axis being the imaginary axis of the s-plane, with all poles within the unit circle of the *s*-plane. The ellipse intersects the Butterworth pole circle at two points on the imaginary axis. This is an all-pole filter
Elliptic	Rippled approximation to both passband and stopband. Passband reflection proportional to ripple level, stopband slope proportional to stopband ripple level. Finite-frequency transmission zeros supplement the amplitude response attributable to the poles and affect the phase response. Ratio between stopband and passband widths can be smaller than for all-pole designs
Quasi-Elliptic	Rippled approximation to passband, small number of ripples in stopband. Finite frequency or real-axis (imaginary frequency) zeros can be achieved, to emphasize either attenuation slopes, passband flatness, and delay, or both to some extent. Easier to realize for narrow bandwidths than full-elliptic designs. Physical structure requires coupling between non-adjacent resonators and thus folding of the structure, for microwave implementations
Max flat time delay (Bessel)	Analogous to Butterworth, but with the time-delay (group delay) function versus frequency containing $2n - 1$ zero derivatives. Poles lie on an ellipse-like path outside the *s*-plane unit circle (as contrasted to the Chebychev locations inside the unit circle). Provides flat, constant group delay within the passband but poor attenuation slopes
Gaussian	A Taylor-series approximation to a Gaussian magnitude function $\|G_{12}(j\omega)\| = e^{-\omega_2/2}$ is used to extract filters which have optimum transient overshoot characteristics and thus display minimum ringing when excited by a pulsed input signal. The group delays is not as flat as that of the Bessel design nor are the stopband slopes as steep. In common with Bessel designs, the filter will display high reflections at frequencies away from center frequency
Transitional	These are filters with transfer functions between those defined by the classical polynomials such as Chebychev, Butterworth, and so on An example is Gaussian-Chebychev.

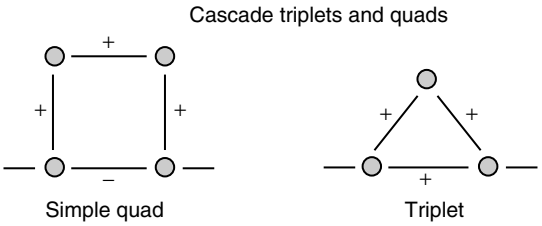


FIGURE 23.6 Cross-coupled blocks called quadruplets and triplets.

then, termed cascaded trisections (CT) or cascaded quadruplets (CQ). Figure 23.6 illustrates a trisection connection and a quadruplet connection.

The potential of such circuit connections is illustrated in the development of the complex quad, shown in Figure 23.7a, and the triplet, in Figure 23.7b. A comparison of some available circuits capable of obtaining the same transfer function, using either the traditional cross-coupled approach or the use of cascaded blocks (quadruplets) is shown in Figure 23.8. Many combinations of trisections and quadruplets are possible. Rather recently, inclusion of non-resonant circuit nodes and significant coupling from both source and load directly into the interior of a network, have added additional degrees of freedom to the circuit design art. While the full implication of these very new approaches are not yet fully understood, clearly innovative configurations are still being evolved, and in conjunction with synthesis employing

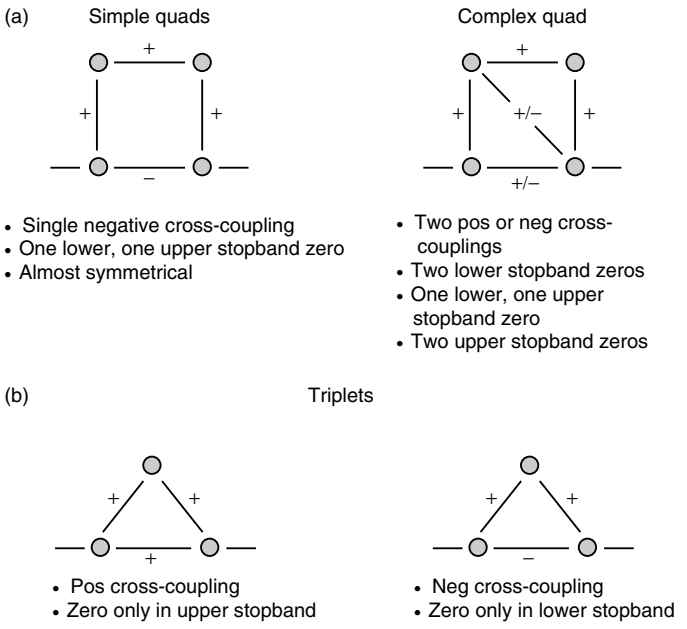


FIGURE 23.7 (a) Quadruplet and (b) Triplet possibilities.

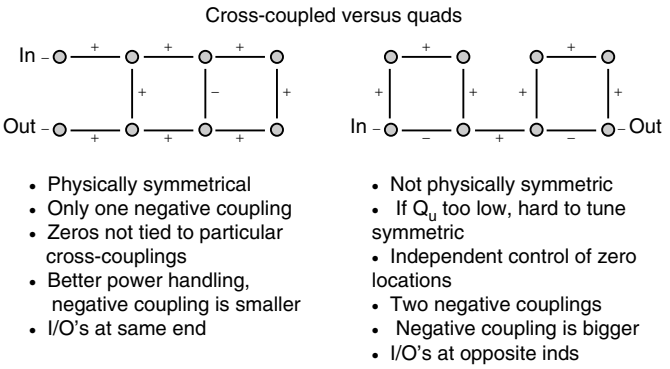


FIGURE 23.8 Comparison of cross-coupled and cascaded-quad configurations.

the response of lumped, distributed, and evanescent variables, design will probably outpace available technology and tuning capability for some time.

Examples of filters with both traditional and advanced network configurations will be presented later in this section.

23.5.1 Element Types and Properties

There are many ways to classify the available elements. Perhaps the most basic is to characterize the element as passive (no DC required) or active (DC required). Within the passive regime, classification includes lumped, distributed, non-reciprocal (and reciprocal), and combinations in which a particular element can display more than one of these characteristics. Lumped elements are those which present capacitive, inductive, resistive, or gyrator responses. The element impedance is essentially a function of ω. Typically, the enumerated elements are predominantly capacitive, inductive, and so on but will also display bits of the other possibilities. For example, at low frequencies, the lead inductance of a capacitor is not important, but as frequency increases, the inductive reactance becomes a significant fraction of the

element impedance, until at some frequency the capacitor behaves as a resonant circuit. It is possible to design networks with lumped element concepts all the way up to 100 GHz or so, but the usual limitation is below 10 GHz. Lumped circuitry usually displays an intrinsically lowpass behavior certainsome frequency. Distributed elements have impedance properties which are functions of $\tan(\omega)$ or $\tanh(\omega)$. These include quarter wavelength (or non-quarter wavelength) TEM mode resonators, waveguide resonators, cavities, dielectric resonators, and any structure built using essentially length-dependent techniques (as contrasted to length-independent but position-dependent lumped element circuits). The frequency range for distributed elements ranges from a few megahertz to the terahertz range. Waveguide elements have the property that internal wavelength is not linearly related to actual free-space wavelength, and are, thus, termed dispersive. Such elements display an intrinsically highpass response *below* a frequency known as the cutoff frequency (energy cannot freely propagate through the section below this frequency). Nonreciprocal elements contain ferrimagnetic structure (circulators, isolators, various gyrators) and are used in conjunction with other elements. An additional hybrid element of interest is formed by a resonated short length of below-cutoff waveguide or other dispersive structure, and is termed an "evanescent" section. Such a below-cutoff section presents an essentially inductive equivalent circuit and can be resonated with a capacitor. The result is formation of a high-Q resonant circuit which can be embedded into a variety of filter circuits. These elements have impedance characteristics similar to lumped at frequencies well below cutoff, and similar to distributed, near cutoff. Unloaded Q is an important property of any circuit element, and is a measure of the ability of the element to store energy without dissipation. High-Q means low loss and is a desirable property. These elements can be fabricated using superconductive material to obtain remarkably high unloaded Q values. Table 23.3 summarizes the properties of many common circuit elements. Application depends on various factors, including basic electrical specification, ambient environment with concomitant difficulties associated with temperature, humidity, vibration, and shock. In general, lumped elements must be potted in place. Distributed and lumped elements have natural changes in impedance or resonant frequency as functions of temperature, which must be compensated using elements with opposite drift properties. Filters can be built that will be stable to no worse than

TABLE 23.3 Common Filter Implementations

Element Type	Frequency Range	Unloaded Q	Implementation
Inductor, lumped	Almost DC to 100 GHz	50–300 at room temperature, 1,000's if superconductive	Coils (air and ferrite-loaded), helices, printed, shorted stubs, evanescent waveguide
Capacitor, lumped	To 100 GHz	50–1,000 at room temperature, 1,000's if superconductive	Multilayer, single layer, open stubs, coaxial
Resistor, lumped	DC-5 GHz	N/A (parasitic capacitance and inductance can be problems)	Metal, composition, chips
Stub or line, printed, TEM, suspended substrate, coplanar stripline, coplanar waveguide, finline, coaxial, other TEM or almost TEM lines	DC-100 GHz	100–500 at room, 1,000's if superconductive	Microstrip, stripline, finline, CPS, CSS, SSS
Evanescent	200 MHz–90 GHz	300–10,000 at room temperature, no data on superconductive application	Below cutoff waveguide of various aspect ratios, machined sections resonated using various capacitive schemes
Dispersive (guided but non-TEM modes)	100 MHz to terahertz	1,000–20,000 at room temperature, 100,000 or more if superconductive	Waveguide, air or dielectric filled cavities with metal walls, dielectric resonators, multimode

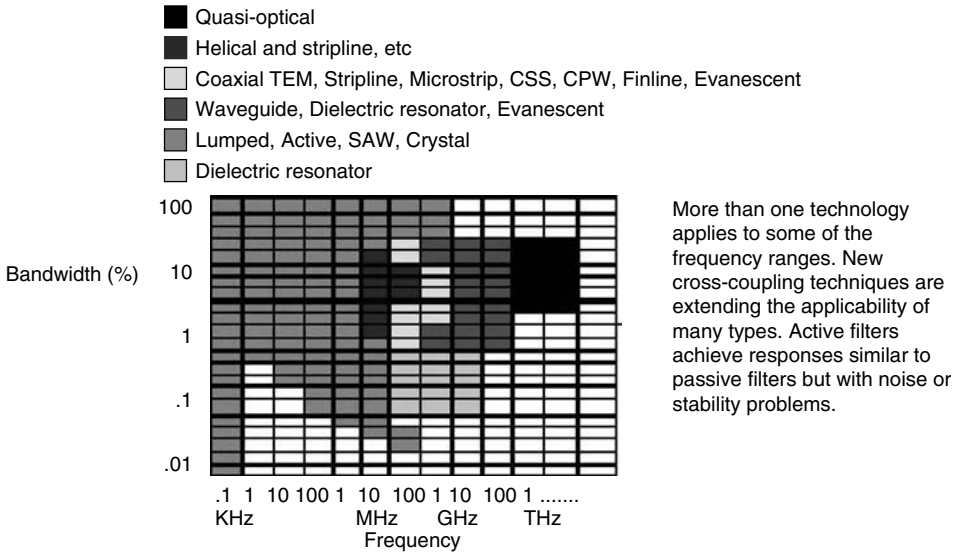


FIGURE 23.9 Filter selection guide.

1 ppm, without the need for external stabilization. Vibration and shock must be damped or isolated from the circuitry. Humidity will affect resonant frequency as well as degrade performance over time. Typically, filter circuits are sealed to eliminate the presence of moisture and to prevent the intrusion of moisture as temperature changes, in a humid environment. Salt will degrade performance and must be eliminated through sealing and special plating systems. In general, filters can be built that combine the properties of the various lumped and distributed element types. This is a difficult, if not impossible, synthesis problem but with the modern simulation and optimization tools, such globally-designed networks are practical and offer the optimum in electrical and environmental performance with associated production cost reduction. The cost reduction stems from the fact that performance over the full range of ambient environment can be predicted, with sensitivity to production tolerances easily taken into account prior to cutting metal. Tolerances are, thus, fit to the problem at hand, with proper care taken and waste minimized.

Figure 23.9 presents a filter selection guide applicable to current technology.

23.6 Physical Examples

1. GPS Multiplexer, using cross-over coupling:

This is a high power GPS application, bandpass/bandstop diplexer, fully militarized. The channel responses are quasi-eliptic, using crossover coupling to implement real frequency transmission zeros for close-in rejection while still achieving maximum passband width.

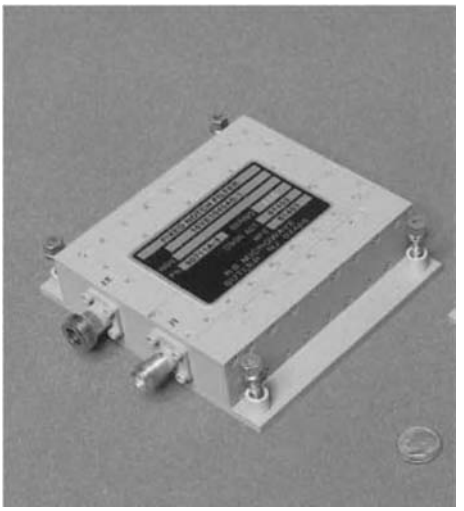
SPECIFICATIONS

Channel 1	
Frequency range:	800-1208 MHz
Insertion lOSS (1 to 3 or 3 to 1)	2.2 dB maximum; 0.6 dB average
VSWR:	1.5:1 maximum, 50 ohms, across the band
Power Handling:	1 Kw Peak, 50 W avg 1030 ± 7 MHz, 1090 ± 7 MHz
	200 W peak, 50 W avg across the rest of the band
Channel 2	
Frequency range:	1217–1237 MHz; 1565–1585 MHz; 1610–1630 MHz
Insertion Loss (1 to 3 or 3 to 1)	4 dB maximum across above ranges
VSWR:	1.5 maximum, 50 ohms, across above frequency ranges
Power Handling:	At least 1 Watt across above frequency ranges
Channel 3	
Frequency range:	800–1700 MHz
Power Handling:	100 W Peak, 5 W average
Isolation, Ports 1 to 2 and 2 to 1:	20 dB minimum, 1207.5–1227.5 MHz

2. Dual mode waveguide bandpass filter

Typically used in satellites as part of communication packages.

3. Notch filter for cosite interference



Designed for rejection of both 1030 and 1090 MHz bands in MIDS/JTIDS applications, the new 90711A-3 dual-notch filter provides the current state-of-the-art performance in fixed notch design. Employing the RS Microwave proprietary quasi-elliptic design, this filter is suitable for the most demanding fighter aircraft environments. The filter can also be combined with a high-power switch for high-speed lobing between MIDS and navigation subsystems.

SPECIFICATIONS

Minimum passbands:	962–1012, 1048–1070, 1108–1215 MHz
Maximum −3 dBc stopband points:	1012, 1048,1070, 1108 MHz
Average passband loss:	1.5 dB maximum
Maximum passband VSWR:	1.5:1
Average passband group delay:	15 nS maximum
Stopbands:	1023–1037, 1083–1097
Minimum stopband rejection:	−58 dBc, 1023–1037
	−41 dBc, 1083–1097
	−50 dBc, 1900–18000
Minimum TDMA power handling:	250 W peak, 50 W AV.
Minimum TACAN power handling:	400 W peak, 4 W AV.
Temperature:	At least −30° C TO +85° C
Altitude:	To 80,000 *(Good for at least 30 minutes)*

4. Advanced design notch filter with wide stopband, wide passband

Low-loss high rejection bandstop filter

This filter is for rejection of MIDS/JTIDS.

SPECIFICATIONS

Minimum −50 dB stopband:	969–1206 MHz
Maximum 1.3B insertion loss:	DC-920, 1286–5000 MHz
RF Power:	At least 80 W peak, 20 W Avg
Operating temperature:	−10 to +50° C

This is a bandstop design of a new type, offering very wide stopbands combined with passbands significantly wider than any previously available. Similar broadband designs are available for rejection of

other radar and communication bands. These filters can be combined with switches to form receiver protection systems, reducing or eliminating broadband receiver desensitization issues in large-signal cosite scenarios.

23.7 Multiplexers

The interconnection of more than one filter at a common junction results in a network termed a multiplexer. With one common port and two individual ports, we have a diplexer. With three individual ports, a triplexer, and so forth through quadruplexer, quintaplexer, sextaplexer, and so on. The individual networks can be lowpass, highpass, bandpass, or bandstop. The common connection presents significant difficulty, as without proper precaution, the interaction between the individual filters causes severe degradation of the desired path transfer function. Many techniques have evolved for performing the interconnection. A multiplexer is normally used if a wide spectrum must be accessed equally and instantaneously. Conventionally, multiplexers have had the disadvantage of requiring at least 3 dB excess loss (crossover loss) at frequencies common to two channels. Thus, the passband characteristics for contiguous structures always showed an insertion loss variation over the passband of, at least, 3 dB. To construct any multiplexer, it is necessary to connect networks to the constituent filters such that each filter appears as an open circuit to each other filter (see Figure 23.10). While this is simple for narrow band channels, it is difficult for broadband or contiguous filters. Normally, the filters and the multiplexing network are synthesized as a set, with computer optimization being used to simulate the results before construction begins. Some of the more common multiplexing techniques include line lengths, circulators, hybrids, and transformers. More recently, the multiplexer filter channels have been combined using power dividers (Figure 23.11). This recent adaptation of always-available technology is due to newly-available cheap and compact amplifier stages. Such gain blocks provide flat gain and low noise over wide bandwidths. In the case of two-way combining, conservation of energy means that the 3 dB insertion loss is still experienced …but on a flat-loss basis. Although each channel is subject to the additional 3 dB loss, it is essentially constant loss over each channel and, thus, the excess passband loss variation is less than 1 dB. Excess loss is defined as that loss not attributable to the individual channel filter roll-off. This power divider based combining can be extended to triplexers (4.7 dB flat loss), quadruplexers (6 dB flat loss), and so on. Because the loss variation is minimized, the overall insertion loss can frequently be made up using amplifiers, which display

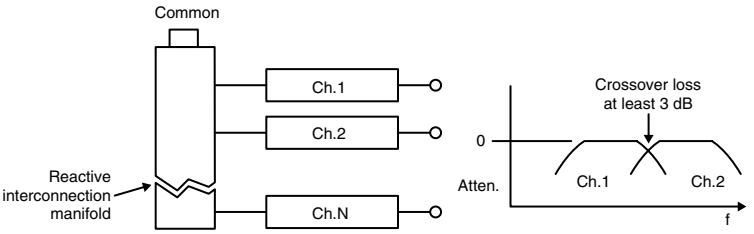


FIGURE 23.10 *N*-way reactively connected multiplexer.

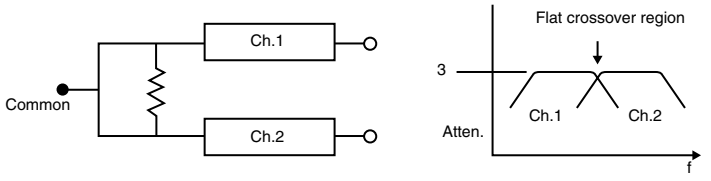


FIGURE 23.11 Power divider combined diplexer.

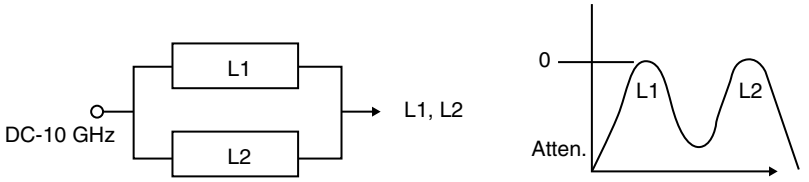


FIGURE 23.12 Double-diplexed two channel filter for GPS.

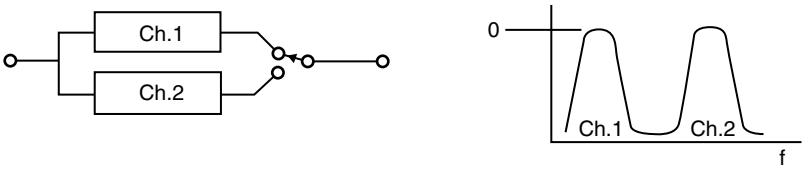


FIGURE 23.13 CH.1 widely separated from CH.2 input diplexed, output switched.

flat gain versus frequency. Filters can be multiplexed by parallel combination at both ends. For example, if two bandpass filters are multiplexed at both input and output, a network results, which provides one input and one output, with two passbands, essentially attenuating everything else. Such assemblies are useful in systems such as GPS which have two or more operating frequencies, with the requirement for isolation between the operating channels and adjacent cluttered regions of the spectrum (Figure 23.12). Another approach employs switched selection of filters. Hybrid combinations using multiplexers with power dividers, switches, and amplifiers are now possible (see Figure 23.13). The interactions of these essentially reactive components can cause undesirable degradation of stopbands or passbands, if precautions are not taken. Available computer simulation techniques are sufficiently sophisticated that accurate prediction of performance and dimensions minimizes the time required to develop and deliver such complex assemblies. Interconnection of sub-components or sub-modules within multiplexers is sometimes difficult, with parasitic lengths causing degradation of performance. Although the computer can predict these problems, sometimes the parasitics reach levels for which compensation cannot be effected. It is possible to use blind-mate interconnection of sub-modules, which is used to minimize both parasitic interconnections and spurious crosstalk. Thus, the physical structure, including all interactions, can be predicted accurately and the unacceptable interactions and crosstalk are eliminated using the mechanical elegance and electrical isolation of blind-mate internal connections. Multiplexer development is impacted heavily by network synthesis and computer simulation techniques. As it becomes possible to synthesize combinations of lumped, distributed, and evanescent elements as well as predict and compensate their interactions, multiplexers will shrink in size, increase in order (number of channels), and display improved performance in insertion loss, isolation, and bandwidth.

23.8 Simulation and Synthesis Software

The process of simulation involves four separate, but related steps:

1. Synthesis and analysis of a theoretical network compliant to specification, under idealized terminating conditions and with idealized construction.
2. Representation of the synthesized network by an appropriate set of very accurate lumped elements. For any circuit, this involves modeling the physical structure and computing the lumped elements which best represent the actual, electromagnetic structure (i.e., solving Maxwell's equations inside the proposed filter structure).

3. Optimizing the filter response with the stipulated terminating impedances (i.e., the complex source and load impedance), using the representation of the circuit as computed in Step 2.
4. Revising the physical structure, if required, by iterating the analysis portion of Step 1.

The solutions to Maxwell's equations which allow for derivation of the lumped equivalents requires the comparison of a set of scattering parameters which will describe the physical structure (computed using *E–M*) to a set which will describe the characteristics of an assumed lumped element topology (computed using linear simulation). The difference between the two sets is reduced using optimization [7]. The data set is stored, and is used in an iterative manner as described in Step 4. All physical structures can be described by a set of lumped elements of arbitrary complexity. Unfortunately, not every set of lumped parameters describes a physically realizable structure, so care must be taken to assume a realizable lumped circuit topology.

Traditional filter designs proceed from the basis of network synthesis. Over the last 90 years or so, the application of matrix, transform, complex variable theory, and advanced algebra have led to many clever network topologies. Numerical methods have also advanced the design process, not only simplifying the calculation process but enabling determination of the design suitability through the use of linear simulators which essentially compute the response of the synthesized structure so that the computed response may be compared to the desired response. If it is found that the synthesis is inadequate, the design can be iterated without the necessity for actual laboratory experimentation. Synthesis techniques have been developed to a very high degree for networks consisting of linear lumped elements or linear distributed elements, but to a much lesser extent for combinations of lumped and distributed elements. This is because the natural frequency variation for a lumped element is in terms of ω, while the variation of a distributed element is in terms of $\tan(\omega)$. Thus, it is difficult to perform a synthesis which requires extraction of elements based upon the location of poles and zeros in a complex plane, when the coordinates of the complex plane are different for lumped and distributed structures.

23.9 Linear Simulators

availability of linear simulators, combined with mathematical optimization, has reduced the need for advanced synthesis development (probably to the detriment of our profession and certainly to the dismay of many). The various elements can be readily combined and calculated in the simulator, as long as the elements can be described in transfer matrix (*s*-parameter) format. However, most physical elements have complex matrix descriptions because the elements are embedded into the surrounding structure in such a way as to respond to more than one mode of excitation. For example, a simple waveguide resonant cavity is analogous to an *L–C* tank circuit, but the waveguide cavity will resonate at more than one frequency based on field distribution. Thus, computation of the analogous (or equivalent) *L–C* values for the waveguide cavity requires knowledge of the excitation field. Combining microwave elements, such as cavities, probes, irises, and so on with each other (or for that matter with *R–L–C* elements), thus, requires inclusion of the effects of the excitation field and the effects upon the field of each of the microwave elements encountered within the composite structure. Accomplishing this requires solutions to Maxwell's equations within the structure.

23.10 Electromagnetic (*E–M*) Simulators

Fortuitously, numerical methods have been applied to the solution of Maxwell's equations resulting in the development of what have come-to-be-known as *E–M* simulators. These programs employ techniques such as finite elements, elements in frequency or time domains, method of moments, spectral domain, and so on, combined with advanced gridding methods and various structure generation software. Although quite advanced, most of these simulators are far too slow to use in conjunction with the mathematical optimization techniques that originally reduced the need for developing new and elegant synthesis

techniques. It is well known [7–10] that frequency-dependent equivalent circuits can be derived, which are adequate lumped representations of distributed structures, to some degree of accuracy. When these structures are so represented, the equivalent circuits depend on the aforementioned mutual interaction of excitation and element. When the response modes are widely separated in the frequency or space domains, a single-mode computation provides a sufficiently accurate representation to enable the resultant lumped circuit to be used for computation of the approximate response of the distributed element or some combination of elements.

23.11 Synthesis Software

There have been a few software packages created that automate the design process to a large extent. However, most practitioners elect to create custom software to facilitate the transition from theory to practical filter networks. Some of the currently available most notable packages include Filter (Eagleware–Agilent), Filpro (Middle Eastern Technical University in Turkey) [11]. Packages that integrate linear and electromagnetic simulation are available from several sources, but inclusion of filter synthesis as an integrated package is rarely available (Eagleware–Agilent has such an integrated package). Recently, researchers have made available design packages that enable synthesis of some of the advanced topologies such as Trisections, Quads, extended box, multiple cross couplings, and so on For example, the Dedale-HF [12] package provides the ability to write Matlab routines and design both symmetric and asymmetric filter networks based on the work of Cameron et al. [13,14]. Another new package is called a FilSynth (from FilResearch) [15], also providing automated matrix rotations and other synthesis operations for a wide variety of filters.

23.12 Active Filters

Since about 1970, it has been possible to simulate a high-Q inductance using a bipolar or FET transistor to convert the output capacitance into equivalent input inductance. The introduction of DC as an external power source acts to compensate for the loss properties of the inductor and make available the inductive element for inclusion into filter circuits. Over the years, other techniques have been developed for using active elements to realize high-Q filter circuits. These filters differ from the better known low frequency op amp-based filters, in that, the synthesis generally is identical to that used for conventional passive RF filters, in which there is no requirement for constant voltage or constant current sources—typical impedances are 20–150 Ω. Such active filters and multiplexers have been built from 100 MHz to over 10 GHz. Stability, noise figure, thermal stability, and power consumption are problems yet to be fully overcome. Miniature passive filters suffer from poor performance due to the low-Q of the available elements. In principle, miniature active filters can be constructed that will provide great selectivity, low loss, and so on with the price being the need for DC power. Applications to handheld cell phones are now to be found, with considerable progress reported. Combinations of active and passive devices are also possible, with the passive elements being used in such a way as to provide stability for the active, high-Q components.

References

1. Blinchikoff and Zverev, *Filtering in the Time and Frequency Domains*, John Wiley & Sons, New York, 1976.
2. Matthaei, Young, and Jones, *Microwave Filters, Impedance-Matching Networks and Coupling Structures*, McGraw-Hill (now Artech House). This is the so-called "black-book" of filters and should be purchased by any serious student.
3. Sydney Frankel, *Multiconductor Transmission Line Analysis*, Artech House.

4. Craven and Skedd, *Evanescent Mode Microwave Components*, Artech House.
5. M.E. Van Valkenburg, *Introduction to Modern Network Synthesis*, John Wiley and Sons, New York, 1960.
6. J. Malherbe, *Microwave Transmission Line Filters*, Artech House, 1979.
7. R. V. Snyder, "Embedded-Resonator Filters," *Proceedings of the ESA-ESTEC Conference on Filter CAD, ESA*, The Netherlands, Nov. 6–8, 1995.
8. R. V. Snyder, "Inverted Resonator Evanescent Mode Filters," *IEEE-MTT-S Symposium Proceedings*, San Francisco MTT IMS, 1996.
9. N. Marcuvitz, *Waveguide Handbook*, Vol 10, MIT RadLab Series.
10. R. V. Snyder, "Filter Design Using Multimode Lumped Equivalents Extracted From E-M Simulations," *MTT/ED Workshop on Global Simulators*, La Rochelle, France, May 27, 1998.
11. N. Yildirim, FILPRO Manual, November, 1996, METU, Ankara, Turkey.
12. Dedale-HF, http://www-sop.inria.fr/apics/Dedale
13. R. J. Cameron. General coupling matrix synthesis methods for chebyshev filtering functions. *IEEE Transaction on Microwave Theory and Techniques*, 47(4):433–442, 1999.
14. R. J. Cameron, J. C. Faugere, and F. Seyfert. Coupling matrix synthesis for a new class of microwave filters. *Proceedings of the International Microwave Symposium 2005*, Long Beach, 2005.
15. FilSynth 1.0, from FilResearch.

24

RF Switches

Robert J. Trew
North Carolina State University

Microwave switches are control elements required in a variety of systems applications. They are used to control and direct, under stimulus from externally applied signals, the flow of RF energy from one part of a circuit to another. For example, all radars that use a common send and receive antenna require an RF switch to separate the send and receive signals, which often differ in amplitude by orders of magnitude. The large difference between the send and receive signals places severe demands upon the switching device, which must be able to sustain the high power of the transmitted signal, as well as have low loss to the returning signal. Isolation is very important in this application since the switch must be able to protect the sensitive receive circuit from the large RF transmitted power. The isolation requirement places severe restrictions upon the switch, and high power radars generally use gas discharge tubes to implement the switch function. Phased-array radars generally use semiconductor transmit/receive modules and use large numbers of switches. A phased-array radar, for example, may require thousands or tens of thousands of switches to permit precise electronic control of the radiated beam. The distributed nature of a phased-array permits the switches to operate at lower power, but the devices still need to operate at power levels on the order of 1 to 10 watts. In general, switches can be manually or electronically switched from one position to the next. However, most microwave integrated circuit applications require switching times that cannot be achieved manually, and electronic control is desirable. Integrated circuit implementation is ideal for switching applications since a large number of components can easily be accommodated in a relatively small area.

Electronically controlled switches can be fabricated using pin diodes [1,2] or transistors, generally GaAs MESFETs [3]. Both types of switches are commonly employed. Switches fabricated using pin diodes have often been used in radar applications [4], achieving insertion slightly over 1 db in L-band, with isolation greater than 35 db. Broadband operation can also be obtained [5] and 6 to 18 GHz bandwidth with insertion loss less than 2 db, isolation greater than 32 db, and CW power handling in excess of 6 watts has been reported using pin diodes connected in a shunt circuit configuration. Such switches have also demonstrated the ability to be optically controlled [6]. GaAs MESFETs are commonly used to fabricate RF switches suitable for use in integrated circuit applications [3,7]. High performance is achieved and a 1 watt SPDT switch with insertion loss of 0.6 db and isolation greater than 20 db has been reported [8]. These switches often use multi-gate GaAs MESFETs specifically designed for switching applications [9] that permit switching control of high RF power with low gate voltage. Such IC switches have demonstrated the ability to handle large power levels [10] and RF power on the order of 38 dbm can be effectively controlled. Switching at extremely high frequency is also possible by replacing the GaAs MESFET with a HEMT, and high performance Q-band [11] and W-band [12] operation has been achieved. A comparison

of the RF performance of MESFET and HEMT switches [13] indicates that HEMT devices generate more distortion than MESFET devices, but are useful at high millimeter-wave frequency. All semiconductor switches, whether fabricated using pin diodes or transistors, can be considered as two-state, one-port devices. Recently, a unified method for characterizing these networks has been presented [14].

24.1 PIN Diode Switches

A pin diode is a nonlinear device fabricated from a p^+nn^+ structure, as shown in Figure 24.1. These devices are widely used in switch applications such as phase shifters [2] and have properties that result in low loss and high frequency performance. A pin diode can also be optically controlled [6], which is desirable for certain applications. The diode is a pn junction device with a lightly doped or undoped (intrinsic) region located between two highly doped contact regions. The presence of the intrinsic region yields operational characteristics very desirable for switching applications. That is, under reverse bias the intrinsic region produces very high values for breakdown voltage and resistance, thereby providing a good approximation to an "open" switching state. Both the breakdown voltage and off-state resistance are dependent upon the length of the intrinsic region, which is limited in design length only by transit-time considerations associated with the frequency of operation. Under forward bias, the conductivity of the intrinsic region is controlled or modulated by the injection of charge from the end regions and the diode conducts current, thereby providing the "on" switching state. The "on" resistance of the diode is controlled by the bias current and in forward bias, the diode has excellent linearity and low distortion.

An equivalent circuit for the PIN diode is shown in Figure 24.2, and in operation the diode functions as a single-pole, double-throw (SPDT) switch, depending upon the bias state. Under reverse bias, the equivalent circuit reduces to that shown in Figure 24.3, and under forward bias it reduces to the forward resistance R_f. The reverse bias resistance can be expressed as [3]

$$R_r = R_c + R_i + R_m \tag{24.1}$$

where R_c is the contact resistance of the metal semiconductor interfaces, R_i is the channel resistance of the intrinsic region, and R_m is the resistance of the contact metals. The resistance of the intrinsic region

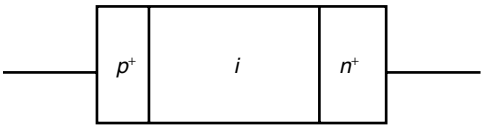


FIGURE 24.1 Schematic diagram for a PIN diode.

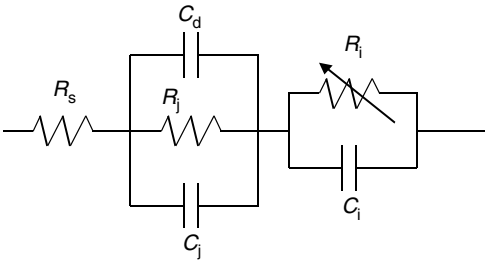


FIGURE 24.2 PIN diode equivalent circuit.

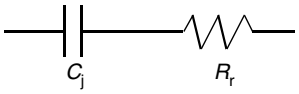


FIGURE 24.3 Reverse biased PIN diode equivalent circuit.

dominates and the reverse resistance becomes essentially that of the intrinsic region, which in terms of physical parameters can be expressed as

$$R_i \cong \frac{3(kT)L^2}{8qI_0L_a^2} \tag{24.2}$$

where L is the length of the intrinsic region, typically in the range of 1 to 100 μm. Depending upon design frequency, I_0 is the bias current, and L_a is the ambipolar diffusion length, which is a constant of the material [3]. The other parameters have their usual meanings. Note that the reverse resistance, which can be in the kΩ range, is inversely proportional to bias current, and decreases with the magnitude of the applied bias current. The greatest off-state resistance, therefore, occurs under low reverse bias voltage. Under reverse bias the intrinsic region is essentially depleted of free charge, so the series capacitance is simply the capacitance of the intrinsic region, and can be expressed as

$$C_i = \frac{\varepsilon A}{L} \tag{24.3}$$

where A is the cross-sectional area of the diode. Note that the capacitance is constant under reverse bias.

Under forward bias the diode is dominated by the forward charge injection characteristics of the pn junction, and the diode can be represented as a resistance, with magnitude determined by the forward current. The on-state resistance can be expressed as

$$R_i = \frac{nkTA}{qI_0} \tag{24.4}$$

where n is the ideality factor for the diode (given in the diode specifications). The resistance of the diode in forward bias is inversely proportional to bias current, and the lowest resistance is obtained at high currents. The impedance of the diode can be tuned for RF circuit matching by adjustment of the bias current.

The rate at which the pin diode can be switched from a low-impedance, forward biased condition to a high-impedance, reverse biased condition, is determined by the speed at which the free charge can be extracted from the diode. Diodes with longer intrinsic regions and larger cross-sectional areas will store more charge, and require, therefore, longer times to switch. The actual switching time has two components: the time required to remove most of the charge (called the delay time) from the intrinsic region, and the time during which the diode is changing from a low- to a high-impedance state (called the transition time). The transition time depends upon diode geometry and details of the diode doping profile, but is not sensitive to the magnitude of the forward or reverse current. The delay time is inversely proportional to the charge carrier lifetime. Diodes with short carrier lifetime have short delay times, but suffer from high values of forward bias resistance. High forward bias resistance increases the insertion loss for the diode, and this will produce attenuation of the signal through the device in the on-state.

24.2 MESFET Switches

A schematic diagram for a GaAs MESFET is shown in Figure 24.4, and these devices are often used in switching applications. In general, a MESFET can be used in two different modes as passive or active elements. In the active mode the transistor is used as a three-terminal switch where the transistor is configured similar to an amplifier circuit. Either single- or dual-gate FETs can be used. The transistor is biased with a positive drain and a negative gate voltage, which are set so that the transistor is active. Switching action is accomplished by control of the transistor gain, which can be varied over several orders of magnitude. Dual-gate devices are particularly attractive for this application since the second gate can be used as a control port for efficient control of the gain.

In the passive mode of operation, the MESFET is configured to function as a passive two-terminal device, with the gate terminal acting as a port for only the control signal. That is, the RF signal is not applied to the gate and only travels between the drain and source terminals. The magnitude of the RF impedance between the drain and source terminals is controlled by a DC signal applied to the gate terminal. The drain-to-source impedance can be varied from a low value, obtained under open channel conditions when a zero potential is applied to the gate, to a high value, obtained when the gate is biased with a negative potential of sufficient amplitude to prevent current from flowing through the transistor. This occurs when the gate voltage achieves the transistor pinch-off voltage, which has a magnitude that is a function of the particular MESFET used.

In the passive mode the low-impedance state of the MESFET switch is dominated by the fully open conducting channel, and the open-channel resistance for the device is low. The equivalent circuit is essentially the "on" resistance for the transistor. In the high-impedance state the MESFET is dominated by the depleted channel or "off" resistance, which is large, and the switch has an equivalent circuit as shown in Figure 24.5. The high-impedance state for the MESFET switch can be approximated with the simplified equivalent circuit shown in Figure 24.6, where the "off" state resistance and capacitance are

$$R_{off} = \frac{2r_d}{2 + r_d \omega^2 C_g^2 r_g} \tag{24.5}$$

where ω is the radian frequency, and

$$C_{off} = C_{ds} + \frac{C_g}{2} \tag{24.6}$$

Note that the "off" state resistance is an inverse function of frequency and, therefore, the magnitude of the blocking resistance decreases as frequency increases. The performance of the switch will degrade at high frequency and switch design becomes more difficult.

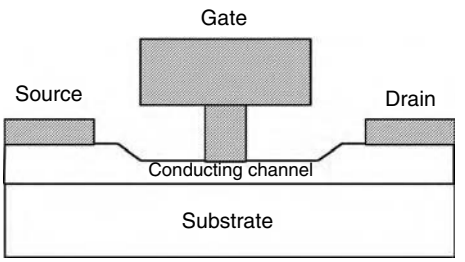


FIGURE 24.4 Schematic diagram for a MESFET switching element.

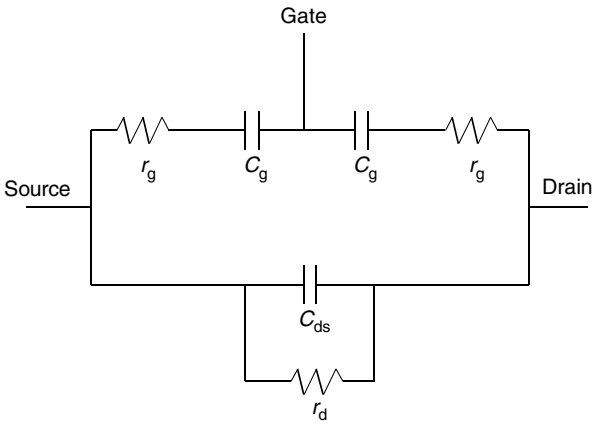


FIGURE 24.5 High-impedance, off-state equivalent circuit for a MESFET switch.

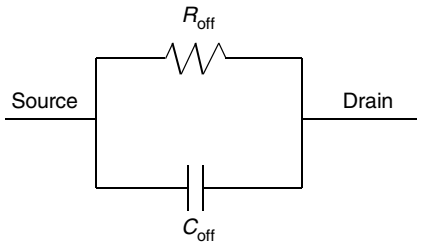


FIGURE 24.6 Simplified off-state equivalent circuit for a MESFET switch.

24.3 Switching Circuits

There are two basic configurations used for single-pole, double-throw (SPDT) switches that are commonly used to control the flow of microwave signals along a transmission line. The basic configurations can be fabricated using either pin diodes or MESFET switching elements, and are realized by utilization of the diode or transistor in a series or shunt connection to the transmission line. A simplified equivalent circuit for a series connected switch is shown in Figure 24.7, and a shunt connected switch is shown in Figure 24.8. The two configurations are complimentary in that the low-impedance state of the series switch permits signal flow, while the high-impedance state of the shunt switch permits signal flow. In the "off" state for both configurations, the microwave power incident upon the switching device is primarily reflected back toward the source. A small fraction of the incident power is dissipated in the switching element and transmitted through the device toward the load. It is this fraction of the incident power that accounts for the insertion loss and the finite and nonideal isolation of the device. The fraction of microwave power that is transmitted through the device increases with frequency due to parasitic paths due to mounting, bonding, packaging, etc. elements, and switch isolation tends to degrade as operating frequency increases. It is possible, however, to minimize the parasitic signal flow by RF tuning and impedance compensation techniques.

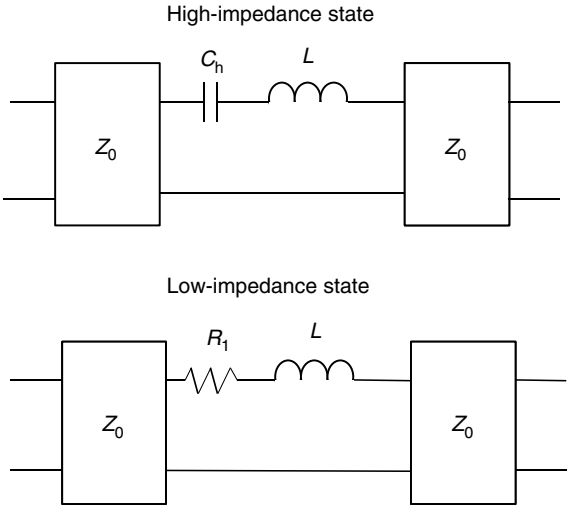


FIGURE 24.7 Simplified series connected switch circuit.

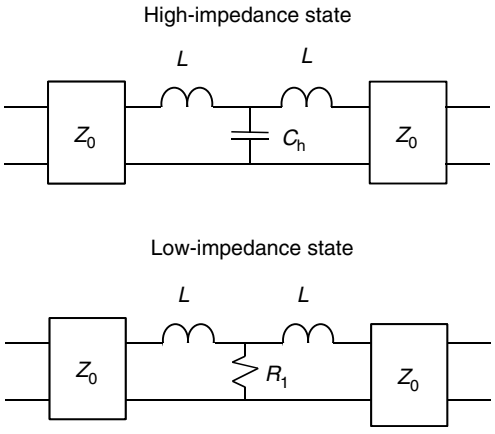


FIGURE 24.8 Simplified shunt connected switch circuit.

24.4 Insertion Loss and Isolation

Insertion loss (IL) and isolation are important parameters that are used to characterize the performance of microwave switches. Insertion loss is defined as the ratio, generally in decibels, of the power delivered to the load in the "on" state of an ideal switch to the actual power delivered by the switch. The insertion loss can be calculated from consideration of the series and shunt equivalent circuits shown in Figures 24.9 and 24.10. If V_L represents the voltage developed at the load for an ideal switch, the insertion loss can be written as

$$IL = \left| \frac{V_L}{V_{LD}} \right|^2 \tag{24.7}$$

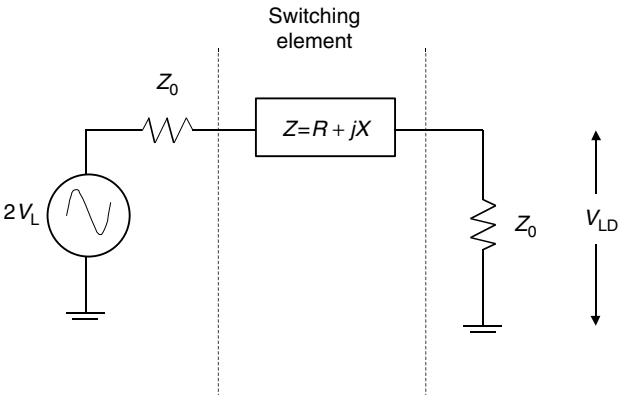


FIGURE 24.9 Equivalent circuit for a series connected switch.

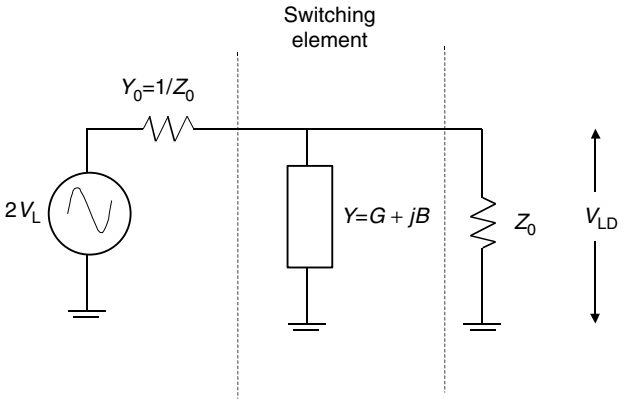


FIGURE 24.10 Equivalent circuit for a shunt connected switch.

where, for the series configuration

$$V_{LD} = \frac{2V_L}{2+Z/Z_0} \tag{24.8}$$

and

$$Z = R + jX \tag{24.9}$$

is the impedance of the switching device. The insertion loss is expressed as

$$IL = \left|\frac{2+Z/Z_0^2}{2Z_0}\right| = \left|1+\frac{R+jX}{2Z_0}\right|^2 = 1+\frac{R}{Z_0}+\frac{1}{4}\left(\frac{R}{Z_0}\right)^2+\frac{1}{4}\left(\frac{X}{Z_0}\right)^2 \tag{24.10}$$

where R and X are the resistance and reactance of the switching device in the low-impedance state.

For the shunt configuration, the voltage across the load is

$$V_{LD} = \frac{2V_L Y_0}{2Y_0+Y} \tag{24.11}$$

and the insertion loss is written as

$$IL = \left|\frac{2Y_0 + Y}{2Y_0}\right|^2 = \left|1 + \frac{G + jB}{2Y_0}\right|^2 = 1 + \frac{G}{Y_0} = \frac{1}{4}\left(\frac{G}{Y_0}\right)^2 + \frac{1}{4}\left(\frac{B}{Y_0}\right)^2 \tag{24.12}$$

where $Y_0 = 1/Z_0$, and $Y = G + jB$.

Isolation is a measure of the off-state performance of the switch. It is defined as the ratio of microwave power delivered to the load for an ideal switch in the "on" state, to the actual power delivered to the load when the switch is in the "off" state. In order to calculate isolation, the insertion loss expressions given above are used with the real and reactive terms for the device in the low-impedance state interchanged with the appropriate device parameters for the high-impedance state.

24.5 Switch Design

Switch design procedures are based upon the principle that the switching element in the "on" and "off" states can be considered as a reactance or susceptance that can be included in a filter configuration. Switch design, therefore, makes use of filter design procedures and all approaches to filter design can be used. The "on" and "off" state equivalent circuits are used to embed the switch element in the filter design. Generally, the "on" or low-insertion loss state is considered first, and the network is designed to yield the desired pass-band performance. The "off" state can be considered as a detuned network, and the impedances are adjusted to achieve the desired isolation. This approach to switch design may require several iterations until satisfactory performance in both the "on" and "off" states are achieved. Mounting and lead reactances are considered in the design and are absorbed and incorporated into the filter network. The actual filter element values may, therefore, differ in value from the design values. The performance of the insertion loss and isolation will vary with tuning and the lowest insertion loss and greatest isolation generally are obtained over narrow bandwidth. Increased bandwidth produces degradation in switch performance. Bias control circuits and thermal handling are accomplished in the same manner as for amplifier circuits.

References

1. H.A. Watson, *Microwave Semiconductor Devices and Their Circuit Applications*, McGraw-Hill, New York, 1969.
2. S.K. Koul and B. Bhat, Microwave and Millimeter Wave Phase Shifters, in *Semiconductor and Delay Line Phase Shifters,* Norwood, MA, Artech House, 1991.
3. I. Bahl and P. Bhartia, *Microwave Solid State Circuit Design*, Wiley Interscience, New York, 1988.
4. M.E. Knox, P.J. Sbuttoni, J.J. Stangel, M. Kumar, and P. Valentino, Solid State 6x6 Transfer Switch for Cylindrical Array Radar, *1993 IEEE International Microwave Symposium Digest*, 1225–1228.
5. P. Omno, N. Jain, C. Souchuns, and J. Goodrich, High Power 6-18 Transfer Switch Using HMIC, *1994 International Microwave Symposium Digest*, 79–82.
6. C.K. Sun, C.T. Chang, R. Nguyen, and D.J. Albares, Photovoltaic PIN Diodes for RF Control—Switching Applications, *IEEE Trans. Microwave Theory Tech.,* 47, 2034–2036, Oct. 1999.
7. M. Shifrin, P. Katzin, and Y. Ayasli, High Power Control Components Using a New Monolithic FET Structure, *1989 IEEE Monolithic and Millimeter-Wave Integrated Circuits Symposium Digest,* 51–56.
8. T. Yamaguchi, T. Sawai, M. Nishida, and M. Sawada, Ultra-Compact 1 W GaAs SPDT Switch IC, *1999 IEEE International Microwave Symposium Digest*, 315–318.
9. H. Uda, T. Yamada, T. Sawai, K. Nogawa, and Yu. Harada, A High-Performance GaAs Switch IC Fabricated Using MESFET's with Two Kinds of Pinch-Off Voltages, *GaAs IC Symp. Digest,* 139–142, 1993.

10. M. Masuda, N. Ohbata, H. Ishiuchi, K. Onda, and R. Yamamoto, High Power Heterojunction GaAs Switch IC with P-1db of More Than 38 dbm for GSM Application, *GaAs IC Symp. Digest*, 229–232, 1998.
11. D.L. Ingram, K. Cha, K. Hubbard, and R. Lai, Q-Band High Isolation GaAs HEMT Switches, *IEEE GaAs IC Symposium Digest*, 289–292, 1996.
12. H. Takasu, F. Sasaki, H. Kawasaki, H. Tokuda, and S. Kamihashi, W-Band SPST Transistor Switches, *IEEE Microwave Guided Wave Letters*, 315–316, Sept. 1996.
13. R.H. Caverly, and K.J. Heissler, On-State Distortion in High Electron Mobility Transistor Microwave and RF Switch Control Circuits, *IEEE Trans. Microwave Theory Tech.*, 98–103, Jan. 2000.
14. I.B Vendik, O.G. Vendik, and E.L. Kollberg, Commutation Quality Factor of Two-State Switchable Devices, *IEEE Trans. Microwave Theory Tech.*, 802–808, May 2000.

25

Monolithic Microwave IC Technology

Lawrence P. Dunleavy
Modelithics, Inc.

25.1 Monolithic Microwave Integrated Circuit Technology

25.1.1 MMIC Definition and Concepts

An excellent introduction to monolithic microwave integrated circuit (MMIC) technology is presented by Robert Pucel in 1981 [1]. Pucel went on to assemble a collection of papers on the subject in which he states in the introduction [2]:

> "the monolithic approach is an approach wherein all active and passive circuit elements and interconnections are formed, in situ on or within a semi-insulating semi-conductor substrate by a combination of deposition schemes such as epitaxy, ion implantation, sputtering, and evaporation."

Figure 25.1 [3] is a conceptual MMIC chip illustrating most of the major constituents. These include field effect transistor (FET) active devices, metal–insulator–metal (MIM) capacitors, thin film resistors, spiral strip inductors via hole grounding, and airbridges.

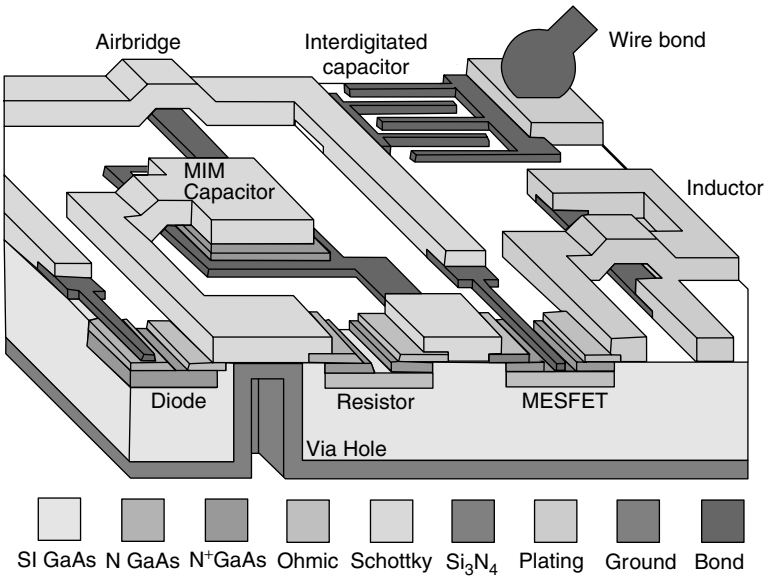


FIGURE 25.1 Three dimensional conceptual illustration of MMIC technology. (From Ladbrooke, P., *MMIC Design GaAs FETs and HEMTs*, Artech House, 1989, p. 46. With permission.)

As implied by the above quote and Figure 25.1, in a MMIC all of the circuit components, including transistors, resistors, capacitors, and interconnecting transmission lines are integrated onto a single semi-insulating/semi-conducting (usually gallium arsenide (GaAs)) substrate. Use of a mask set and a corresponding series of processing steps achieve the integrated circuit fabrication. The mask set can be thought of as a mold. Once the mold has been cast the process can be repeated in a turn-the-crank fashion to batch process tens, hundreds, or thousands of essentially identical circuits on each wafer.

25.1.2 A Brief History of GaAs MMICs

The origin of MMICs may be traced to a 1964 government program at Texas Instruments [4]. A few key milestones are summarized in the following:

1. 1964—U.S. Government funded research program at TI based on Silicon integrated circuit technology [5]:
   - The objective was a transmit/receive module for a phased array radar antenna.
   - The results were disappointing due to the poor semi-insulating properties of silicon.
2. 1968—Mehal and Wacker [6] used semi-insulating GaAs as the substrate with Schottky diodes and Gunn devices as active devices to fabricate an integrated circuit comprising a 94 GHz receiver front-end.
3. 1976—Pengelly and Turner [7] used MESFET devices on GaAs to fabricate an X-band (∼10 GHz) amplifier and sparked an intense activity in GaAs MMICs.
4. 1988 (approximately)—U.S. Government's Defense Advanced Research Projects Agency (DARPA, today called ARPA) launched a massive research and development program called the MIMIC program (included Phase I, Phase II, and Phase III efforts) that involved most of the major MMIC manufacturing companies.

In the early 1980s a good deal of excitement was generated and several optimistic projections were made predicting the rapid adoption of GaAs MMIC technology by microwave system designers, with correspondingly large profits for MMIC manufacturers. The reality is that there was a much slower rate of progress to widespread use of MMIC technology, with the majority of the early thrust being provided

by the government for defense applications. Still, steady progress was made through the 1980s, and the government's MIMIC program was very successful in allowing companies to develop lower cost design and fabrication techniques to make commercial application of the technology viable. The 1990s have seen good progress towards commercial use with applications ranging from direct broadcast satellite (DBS) TV receivers, to automotive collision avoidance radar, and many wireless communication applications (Cell phones, WLANs, etc.).

In recent years GaAs MMICs are giving ground to SiGe HBT- and Si CMOS-based silicon radio frequency integrated circuits (RFICs), especially at Wireless frequencies below 6 GHz. Since a review of Silicon RFIC technology is outside our scope here, the interested reader is referred to one of the many treatments on these topics [8–10]. As will be discussed later on in this section, GaAs and related III–V compounds like GaN, InP remain strong contenders for switching, power amplifier, and millimeter-wave *IC* applications.

As a point of terminology clarification, according to Dr. Gailon Brehm of Triquint "Many recent publications have adopted the usage of 'RFIC' to refer to monolithic integrated circuits performing radio frequency functions for high volume applications up to 5 GHz, and have applied this term to such circuits independent of the semiconductor—GaAs or silicon. The term 'MMIC' can be used interchangeably as well [11]."

25.1.3 Hybrid versus MMIC Microwave IC Technologies

The conventional approach to microwave circuit design that MMIC technology competes with, or is used in combination with is called "hybrid microwave integrated circuit," "discrete microwave integrated circuit" technology, or simply MIC technology. In a hybrid MIC the circuit pattern is formed using photolithography. Discrete components are then assembled onto the substrate (e.g., using solder or silver epoxy) then connected using bondwires. In contrast to the batch processing afforded the MMIC approach, MICs has to be assembled with discrete components attached using relatively labor intensive manufacturing methods. Table 25.1 summarizes some of the contrasting features of hybrid and monolithic approaches.

The choice of MMICs versus hybrid approach is mainly a matter of volume requirements. The batch processing of MMICs gives this approach advantages for high volume applications. Significant cost savings can be reaped in reduced assembly labor; however, for MMIC the initial design and mask preparation costs are considerable. The cost of maintaining a MMIC manufacturing facility is also extremely high and this has forced several companies out of the business. A couple of examples are Harris who sold their GaAs operation to Samsung and put their resources on Silicon. Another is AT& T, who also is relying on Silicon for its anticipated microwave IC needs.

The high cost of maintaining a facility can only be offset by high volume production of MMICs. Still this does not prevent companies without MMIC foundries from using the technology, as there are several commercial foundries that offset the costs of maintaining their facilities by manufacturing MMIC chip products for third party companies through a foundry design working relationship.

TABLE 25.1 Features of Hybrid and Monolithic Approaches

Feature	Hybrid	Monolithic
Type of substrate	Insulator	Semiconductor
Passive components	Discrete/deposited	Deposited
Active components	Discrete	Deposited
Interconnects	Deposited and wire-bonded	Deposited
Batch-processed?	No	Yes
Labor intensive/chip	Yes	No

Source: Pucel, R.A., ed., *Monolithic Microwave Integrated Circuits*, IEEE Press, 1985, p. 1. With permission.

A key advantage of MMICs is small size. To give an example, hybrid MIC the size of a business card can easily be reduced to a small chip one or two millimeters on a side. An associated advantage is the ease of integration that allows several functions to be integrated onto a single chip. For example, Anadigics and Raytheon have both manufactured DBS related MMICs, wherein, the functions of amplification (LNA and IF amplifiers), signal generation (VCO), and signal conversion (mixer) and filtering are all accomplished on a 1×2 mm, or smaller chip.

In contrast to MMIC, MIC lithography is quite inexpensive and a much smaller scale investment is required to maintain a MIC manufacturing capability. There are also some performance advantages of the hybrid approach. For example, it is much easier to tune or repair a hybrid circuit after fabrication than it is for a MMIC. For this reason, for applications where the lowest noise figure is required, such as in a satellite TV receiver, an individually tuned hybrid LNA may be preferred as the first stage.

Ultimately, there is no such thing as a truly all MMIC system. Monolithic technology can be used to integrate single functions, or several system functions, but cannot sustain a system function in isolation. Usually, a MMIC is packaged along with other circuitry to make a practically useful component, or system. Figure 25.2, a radar module made by Raytheon, exemplifies how the advantages of both MMIC and hybrid approaches are realized in a hybrid connection of MMIC chips. Another example of combined hybrid/MMIC technology is shown in Figure 25.3, in the form of a low noise block downconverter for DBS Applications. An example of combined MMIC and hybrid MIC technologies, this radar module includes several MMIC chips interconnected using microstrip lines. The hybrid substrate (white areas) is an alumina insulating substrate; the traces on the hybrid substrate are microstrip lines (courtesy Raytheon Systems Company).

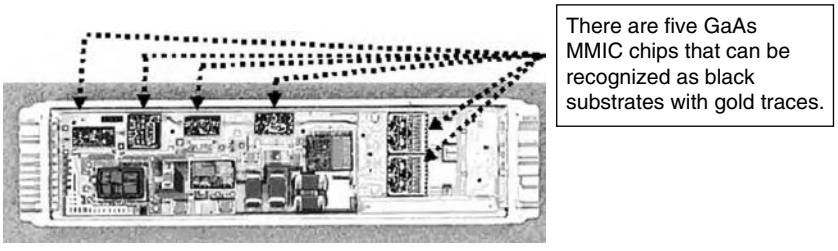


FIGURE 25.2 An example of combined MMIC and hybrid MIC technologies, this radar module includes several MMIC chips interconnected using microstrip lines. The hybrid substrate (white areas) is an Alumina insulating substrate; the traces on the hybrid substrate are microstrip lines. (Courtesy Raytheon Systems Company.)

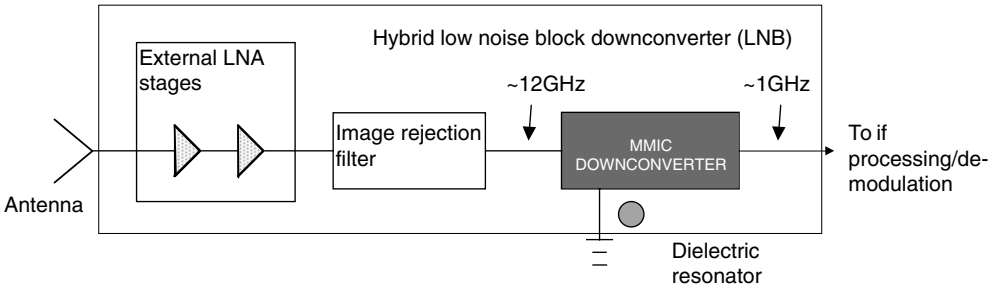


FIGURE 25.3 Application of MMICs in Ku-Band Direct Broadcast Satellite downconverters. The antenna is typically a small parabolic dish.

25.1.4 GaAs MMICs in Comparison to Silicon VLSI Computer Chips

Everyone is familiar with silicon digital IC chips or, at least, the enormous impact silicon-based very large scale integrated (VLSI) circuits have had on the computer industry. Silicon computer chips are digital circuits that contain hundreds to thousands of transistors on each chip. In a digital circuit the transistors are used as switches that are in one of two possible states depending on the "logic" voltage across a pair of terminals. The information processed by a digital circuit consists of a sequence of "1's" and "0's", which translate into logic voltages as the signal passes through the digital IC. Noise distorts the logic waveform in much the same way that it distorts a sinusoidal signal, however, as long as the signal distortion due to noise is not severe, the digital circuitry can assign the correct (discrete) logic levels to the signal as it is processed. Signal interpretation errors that occur due to noise are measured in terms of a bit error rate (BER).

The speed of the digital processing is related to how fast the transistors can switch between one state and another, among other factors. Because of certain material factors, such as electron mobility, digital circuits made on GaAs have been demonstrated to have speed advantages over silicon digital ICs, however, the speed advantages have not been considered by the majority of companies to outweigh the significant economic advantages of well established, lower cost, silicon processing technology.

Because of the large volumes of Silicon chips that have been produced over the last 20 years silicon processing techniques are significantly more established and in many cases standardized as compared to GaAs processing techniques which still vary widely from foundry to foundry. The digital nature of the signals and operation modes of transistors in digital ICs makes uniformity between digital ICs, and even similar ICs made by different manufacturers, much easier to achieve than achieving uniformity with analog GaAs MMICs.

In contrast, GaAs MMICs are analog circuits that usually contain less than 10 transistors on a typical chip. The analog signals (can take on any value between certain limits) processed may, generally, be thought of as combinations of noisy sinusoidal signals. Bias voltages are applied to the transistors in such a way that each transistor will respond in one of several predetermined ways to an applied input signal. One common use for microwave transistors is amplification, whereby the result of a signal passing through the transistor is for it to be boosted by an amount determined by the gain of the transistor.

A complication that arises is that no two transistors are identical in the analog sense. That is, taking gain for example, there will be a statistical distribution of gain for a set of amplifier chips measured on the same GaAs MMIC wafer, a different (wider) set of statistics applies to variations in gain from wafer-to-wafer for the same design. These variations are caused primarily by variations in transistors, but also by variations in other components that make up the MMIC, including MIM capacitors, spiral inductors, film resistors, and transmission line interconnects. Successful foundries are able to control the variations within acceptable limits, in order to achieve a satisfactory yield of chips meeting a customer's requirements; however, translating a MMIC design mask set to a different manufacturing foundry, generally, requires considerable effort on the part of the designer and the foundry.

25.1.5 MMIC Yield and Cost Considerations

Yield is an important concept for MMICs and refers to the percentage of circuits on a given wafer with acceptable performance relative to the total number of circuits fabricated. Since yield may be defined at several points in the MMIC process, it must be interpreted carefully.

- DC yield is the number of circuits whose voltage and currents measured at DC (zero frequency) conditions meets fall within acceptable limits.
- Radio frequency (RF) yield is generally defined as the number of circuits that have acceptable RF/microwave performance when measured "on-wafer," before circuit dicing.
- Packaged RF yield is the final determination of the number of acceptable MMIC products that have been assembled using the fabricated MMIC chips.

If measured in terms of the total number of circuits fabricated each of these yields will be successively lower numbers. For typical foundries DC yields exceed 90% , while packaged yields may be around 50%. Final packaged RF yield depends heavily on the difficulty of the RF specifications, the uniformity of the process (achieved by statistical process control), as well as how sensitive the RF performance of the circuit design is to fabrication variations.

The costs involved with MMICs include

- Material
- Design
- Mask set preparation
- Wafer processing
- Capital equipment
- Testing
- Packaging
- Inspection

A typical wafer run may cost $20,000–50,000 with $5,000–10,000 for the mask set alone, and design costs excluded! per-chip MMIC costs are determined by

- Difficulty of design specifications
- Yield
- Material (wafer size and quality)
- Production volume
- Degree of automation

Some 1989–1990 example prices for MMIC chips are as follows [13]:

- 1–5 GHz wideband amplifier—$30.00
- 2–8 GHz wideband amplifier—$45.00
- 6–18 GHz wideband amplifier—$100.00
- DC-12 GHz attenuator—$60.00
- DBS downconverter chip—$10.00

In comparison, example prices for 1999 MMIC chips [14]

- DBS downconverter chip $<$ $2.50
- DC-8 GHz HBT MMIC amp. $\sim$ $3
- Packaged power FET MMIC $1 for $\sim$2 GHz
- 4–7 and 8–11 GHz high power/high eff. Power amps. $85 (2 W)–$330 (12 W)

As can be seen above, prices vary widely and the key driver that dictates the lower end of the chip prices is the highly competitive price pressure of high volume commercial applications. Prices for low volume specials at high frequency have not moved much since 1989, however, 2006 prices for higher volume chips are about a factor of 2 lower than the above prices [12].

25.1.6 Silicon versus GaAs for Microwave Integrated Circuits

The subject of silicon versus GaAs has been a hotly debated subject since the beginning of the MMIC concept in around 1965 (see Section 25.1.2). Although a comprehensive treatment of this issue is outside of the main scope of interest, for the present treatment, a brief discussion will be given. Two of the main discriminating issues between the technologies are microwave transistor performance and the loss of the semi-conductor when used as a semi-insulating substrate for passive components.

A comparison of relevant physical parameters for silicon and GaAs materials is given in Table 25.2. The dielectric constants of the materials mainly affects the velocity of propagation down transmission line

TABLE 25.2 Properties of GaAs, Silicon, and Common Insulating Substrates

Property	GaAs	Silicon	Semi-Insulating GaAs	Semi-Insulating Silicon	Sapphire	Alumina
Dielectric constant	12.9	11.7	12.9	11.7	11.6	9.7
Thermal conductor (W/cm-K)	0.46 (Fair)	1.45 (Good)	0.46 (Fair)	1.45 (Good)	0.46 (Fair)	0.37 (Fair)
Resistivity (Ω-cm)	—	—	10^7–10^9 (Fair)	10^3–10^5 (Poor)	$>10^{14}$ (Good)	10^{11}–10^{14} (Good)
Elec. mobility	4300* (Best)	700* (Good)	—	—	—	—
Sat. elec. velocity	1.3×10^7 (Best)	9×106 (Good)	—	—	—	—

* At a doping concentration of $10^{17}/cm^3$.

Source: Adapted from Pucel, R., ed., *Monolithic Microwave Integrated Circuits*, IEEE Press, 1985, p. 2. With permission.

interconnects, and for the materials compared this parameter is on the same order. For the other factors considered significant differences are observed. The thermal conductivity, a measure of how efficiently the substrate conducts heat (generated by DC currents) away from the transistors, is best for Silicon and one key advantage of the silicon approach. This advantage is offset by silicon's lower mobility and lower resistivity. Mobility in a semiconductor is a measure of how easily electrons can move through the "doped" region of the semiconductor (see discussion of FET operation in the following section). The mobility, as well as the saturated velocity, have a strong influence on the maximum frequency that a microwave transistor can have useful gain.

Turning our attention to passive component operation, GaAs has much better properties for lower loss passive circuit realization. With the exception of a resistor, the ideal passive component is a transmission line, inductor, or capacitor that causes no signal loss. Resistivity is a measure of how "resistant" the substrate is to leakage currents that could flow, for example, from the top conductor of a microstrip line and the ground plane below. Looking first at the properties of the insulators Sapphire and Alumina, the resistivity is seen to be quite high. Semi-insulating GaAs is almost as high, and silicon has the lowest resistivity (highest leakage currents for a given voltage). GaAs has a radiation tolerance advantage as well, although silicon-on-insulator (SOI) has been used with some success in this area at Honeywell and Peregrine [15].

These considerations have led many companies to historically invest heavily in GaAs technology for microwave applications. However, the use of silicon RFICs is now dominant in wireless applications like GPS, Bluetooth, Zig Bee, 802.11a-b-g, WiFi, UWB, and automotive radar, which according to one RFIC design expert are almost 100% Silicon [15].

Also, atleast one company has pushed silicon-based microwave ICs into Ku-band frequencies to compete in the DBS satellite receiver market [16]. Silicon–Germanium Heterojunction Bipolar Transistor Technology is also paving the way for increasing the applicability of silicon technology to even higher frequencies.

To close this section, let us summarize an industry viewpoint on a few areas where GaAs is still better than silicon [17]:

1. RF high power/linearity switches. You can make CMOS switches, but it just is not as good as you can do with GaAs due to the lower breakdown voltage and poor substrate (loss).
2. High power output—once you get above about 1 W, silicon, or SiGe runs out of steam, particularly linear power. Silicon-based PAs have made it into cell phones, but only so far on a limited basis, for example, in Asia where connection over distance is not as much of a concern as in the United States, and hence lower power output requirements.

3. High frequency—GaAs and III–V materials (e.g., AlGas, InP, and related composites) still have an advantage. To be sure, papers reporting high frequency integrated transceiver circuits at 60 GHz and above are starting to appear [18], including an interesting development of a power amplifier at 77 GHz [19].

Generally beyond that, GaAs has higher breakdown, higher substrate isolation (hence better *Q* inductors), better caps, and higher fTs/fMAXs (for the moment). Silicon has better integration potential, complementary devices (hole mobility in GaAs is very bad) and is available in better mass production capability than GaAs.

It is also worth mentioning that progress continues to be made in both silicon and GaAs (as well as related II–V) technologies and that today's silicon limits tend to be pushed further and further out with each passing year. While, it is likely that GaAs material advantages will always translate into advantages for certain MMIC applications, the limits will continue to be challenged and moved toward higher performance. The other factor that will affect this list in the future is the evolution in radio systems. Today's handset-to-basestation requirements give GaAs an advantage in handset PAs. If system architectures of the future move to shorter distances between basestations, much of the GaAs advantage goes away. In contrast, if direct-to-satellite radios become pervasive, GaAs gains significant advantage. While RF system architectures of the future are not known, they will have an impact on the GaAs–silicon technology use.

25.2 Basic Principles of GaAs MESFETs and HEMTs

25.2.1 Basic MESFET Structure

One of the primary active devices in a GaAs MMIC is the metal-electrode-semiconductor FET, or MESFET. In the past 10 years or so, the heterojunction bipolar transistor (HBT) has emerged as an important alternative, especially for power amplifiers and oscillator applications. The reader is referred to one of the many excellent treatments of HBT device technology for coverage of that topic [21].

The basic construction of a MESFET is shown in Figures 25.4 and 25.5. An "active layer" is first formed on top of a semi-insulating GaAs substrate by intentionally introducing an *n*-type impurity onto the surface of the GaAs, and isolating specific channel regions. These channel regions are semi-conducting in

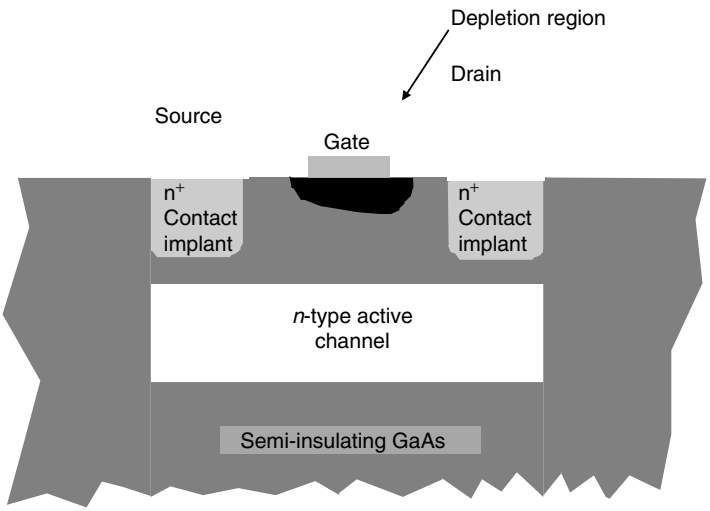


FIGURE 25.4 Cross section of a MESFET transistor. In operation, current I_{ds} flows between the gate and drain terminals through the doped *n*-type active channel. An AC voltage applied to the gate modulates the size of the depletion region causing the I_{ds} current to be modulated as well. Notice that in a MMIC the doped *n*-type region is restricted to the transistor region leaving semi-insulating GaAs outside to serve as a passive device substrate.

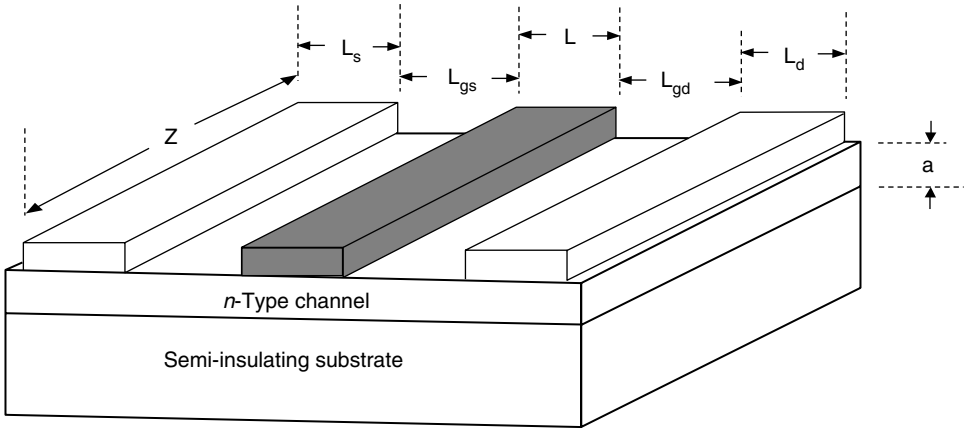


FIGURE 25.5 Aspect view of MESFET. Important dimensions to note are the gate length (L_g) and the gate width (Z). (From J. M. Golio, Ed., Microwave MESFETs and HEMTs, Artech House 1991, p. 28. With permission.)

that they contain free electrons that are available for current flow. When a metal is placed in direct contact with a semi-conductor, as in the case of the gate, a "Schottky diode" is formed. One of the consequences of this is that a natural "depletion region," a region depleted of available electrons, is formed under the gate. A diode allows current flow easily in one direction, while impeding current flow in the other direction. In the case of a MESFET gate, a positive bias voltage between the gate and the source "turns on" the diode and allows current to flow between the gate and the source through the substrate. A negative bias between the gate and the drain "turns off" the diode and blocks current flow, it also increases the depth of the depletion region under the gate.

In contrast to the gate contact, the drain and source contacts are made using what are called "ohmic contacts." In an ohmic contact, current can flow freely in both directions. Whether an ohmic contact or Schottky diode is formed at the metal/semi-conductor interface is determined by the composition of the metal placed on the interface and the doping of the semi-conductor region directly under the metal. The introduction of "pocket $n+$ implants" help form the ohmic contacts in the FET structure illustrated in Figure 25.4. In the absence of the gate, the structure formed by the active channel in combination with the drain and source contacts essentially behave as a resistor obeying ohms law. In fact, this is exactly how one type of GaAs-based resistor commonly used in MMICs is made.

25.2.2 FETs in Microwave Applications

The most common way to operate a MESFET, for example, in an amplifier application, is to ground the source (also called "common source" mode), introduce a positive bias voltage between the drain and source and a negative bias voltage between the gate and source. The positive voltage between the drain and source, V_{ds}, causes current, I_{ds}, to flow in the channel.

As negative bias is applied between the gate and source, V_{gs}, the current, I_{ds}, is reduced as the depletion region extends farther and farther into the channel region. The value of current that flows with zero gate to source voltage is called the saturation current, I_{dss}. Eventually, at a sufficiently large negative voltage the channel is completely depleted of free electrons and the current, I_{ds}, will be reduced to essentially zero. This condition is called "pinch-off." In most amplifier applications the negative gate voltage is set to a "bias condition" between 0 V and the pinch-off voltage, V_{po}.

Figure 25.6 gives a simplified view of a FET configured in an amplifier application. An input sinusoidal signal, $V_{gs}(t)$, is shown offset by a negative DC bias voltage. The sinusoidal variation in, V_{gs}, causes a likewise sinusoidal variation in depth of the depletion region that, in turn, creates a sinusoidal variation (or modulation) in the output current. Amplification occurs because small variations in the, V_{gs}, voltage cause relatively large variations in the output current. By passing the output current through a resistance

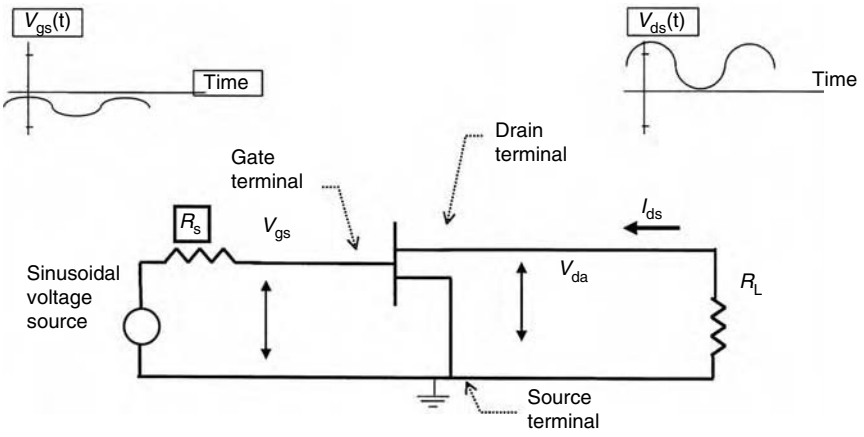


FIGURE 25.6 Simplified schematic of a FET in an amplifier application. Not shown are matching networks needed to match between the source resistance R_s and the FET input impedance, and the load resistance R_L and the FET output impedance. Also omitted are the networks needed to properly apply the DC bias voltages to the device and provide isolation between the RF and DC networks.

RL the voltage waveform, $V_{ds}(t)$, is formed. The, V_{ds}, waveform is shown to have higher amplitude than, V_{gs}, to illustrate the amplification process.

Other common uses, which involve different configurations and biasing arrangements, include use of FETs as the basis for mixers, oscillators (or VCOs).

25.2.3 FET Fabrication Variations and Layout Approaches

Figure 25.7 illustrates a MESFET fabricated with a recessed gate, along with a related type of FET device called a high electron mobility transistor (HEMT) [22]. A recessed gate is used for a number of reasons. First, in processing, it can aid in assuring that the gate stripe is placed in the proper position between the drain and source, and it can also result in better control and uniformity in I_{dss} and V_{po}. A HEMT is a variation of the MESFET structure that, generally, produces a higher performing device, this translates, for example, into a higher gain and a lower noise figure at a given frequency.

In the light of the above cursory understanding of microwave FET structures, some qualitative comments can be made about some of the main factors that cause intended and unintended variations in FET performance. The first factor is the doping profile in the active layer. The doping profile refers to the density of the charge carriers (i.e., electrons) as a function of depth into the substrate. For the simplest type, uniform doping, the density of dopants (intended impurities introduced in the active region) is the same throughout the active region. In practice, there is a natural "tail" of the doping profile that refers to a gradual decrease in doping density as the interface between the active layer and the semi-insulating substrate is approached. One approach to create a more abrupt junction is the so-called "buried *p*-layer" technique. The buried *p*-layer influences the distribution of electrons versus depth from the surface of the chip in the "active area" of the chip where the FET devices are made. The idea is to create better definition between where the conducting channel stops and where the non-conducting substrate begins. (More specifically the buried *p*-layer counteracts the n-type dopants in the "tail" of the doping profile.) The doping profile and density determine the number of charge carriers available for current flow in a given cross section of the active channel. This has a strong influence on the saturation current, I_{dss} and pinch-off voltage, V_{po}. This is also influenced by the depth of the active layer (dimension "*a*" of Figure 25.5).

Other variables that influence MESFET performance are the gate length and gate width ("*L*" and "*Z*" of Figure 25.5). The names for these two parameters are counterintuitive since the gate length refers to

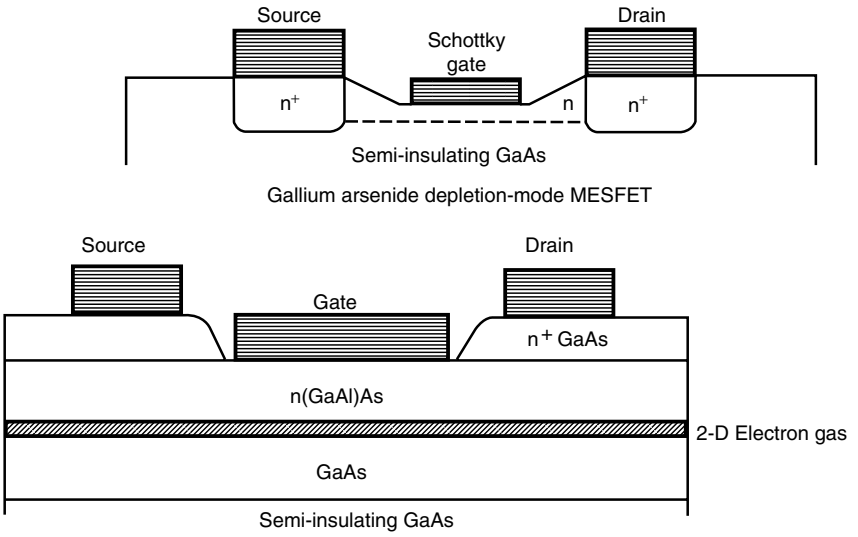


FIGURE 25.7 MESFET (a) and HEMT (b) structures showing "gate recess" structure whose advantages include better control of drain-to-source saturation current I_{dss} and pinch-off voltage V_{po}. (From Goyal, R., Ed., *Monolithic Microwave Integrated Circuits: Technology and Design*, Artech House 1989, p. 113. With permission.)

the shorter of the two dimensions. The gate width and channel depth determine the cross sectional area available for current flow. An increase in gate width increases the value of the saturation current, which translates into the ability to operate the device at higher RF power levels (or AC voltage amplitudes). Typical values for gate widths are in the range of 100 μm for low noise devices to over 2000 μm for high power devices. The gate length is usually the minimum feature size of a device and is the most significant factor in determining the maximum frequency where useful gain can be obtained from a FET; generally, the smaller the gate width, the higher the gain for a given frequency. However, the fabrication difficulty increases, and processing yield decreases, as the gate length is reduced. The difficulty arises from the intricacy in controlling the exact position and length (small dimension) of the gate.

The way a FET is "laid out" can also influence performance. Figure 25.8 shows plan views of common ways by which FETs are actually "laid out" for fabrication [23]. The layout of the FET affects what are called "external parasitics" which are undesired effects that can be modeled as a combination of capacitors, inductors, and resistors added to the basic FET electrical model.

In MMIC fabrication, variations in the most of the above-mentioned parameters are a natural consequence of a real process. These variations cause variations in observed FET performance even for identical microwave FETs made using the same layout geometry on the same wafer. Certainly there are many more subtle factors that influence performance, but the factors considered here should give some intuitive understanding of how unavoidable variations in the physical structure of fabricated FETs cause variations in microwave performance. As previously mentioned, successful GaAs MMIC foundries use statistical process control methods to produce FET devices within acceptable limits of uniformity between devices.

25.3 MMIC Lumped Elements: Resistors, Capacitors, and Inductors

25.3.1 MMIC Resistors

Figure 25.9 shows three common resistor types used in GaAs MMICs. For MMIC resistors the type of resistor material, and the length and width of the resistor determine the value of the resistance [24].

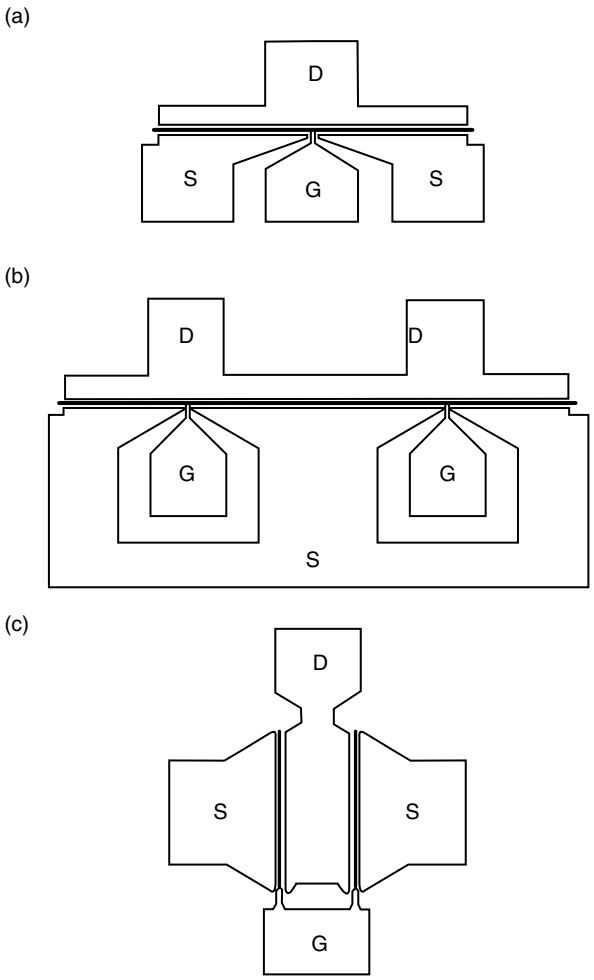


FIGURE 25.8 Three common layout approaches for MESFETs and HEMTs. (From Ladbrooke, P., *MMIC Design: GaAs FETs and HEMTs*, Artech House, 1989, p. 92. With permission.)

In practice, there are also unwanted "parasitic" effects associated with MMIC resistors which can be modeled, generally, as a combination of series inductance, and capacitance to ground, in addition, to the basic resistance of the component.

25.3.2 MMIC Capacitors

The most commonly used type of MMIC capacitor is the metal-insulator-metal capacitor shown in Figure 25.10 [25]. In a MIM capacitor, the value of capacitance is determined from the area of the overlapping metal (smaller dimension of two overlapping plates), the dielectric constant ε of the insulator material, typically silicon nitride, and the thickness of the insulator.

For small value of capacitance, less than 0.5 pF, various arrangements of coupled lines can be used as illustrated in Figure 25.11 [26]. For these capacitors the capacitance is determined from the width and spacing of strips on the surface of the wafer.

At microwave frequencies, "parasitic" effects limit the performance of these capacitors. The two main effects are signal loss due to leakage currents, as measured by the quality factor Q of the capacitor, and a self-resonance frequency, beyond which the component no longer behaves as a capacitor. The final wafer

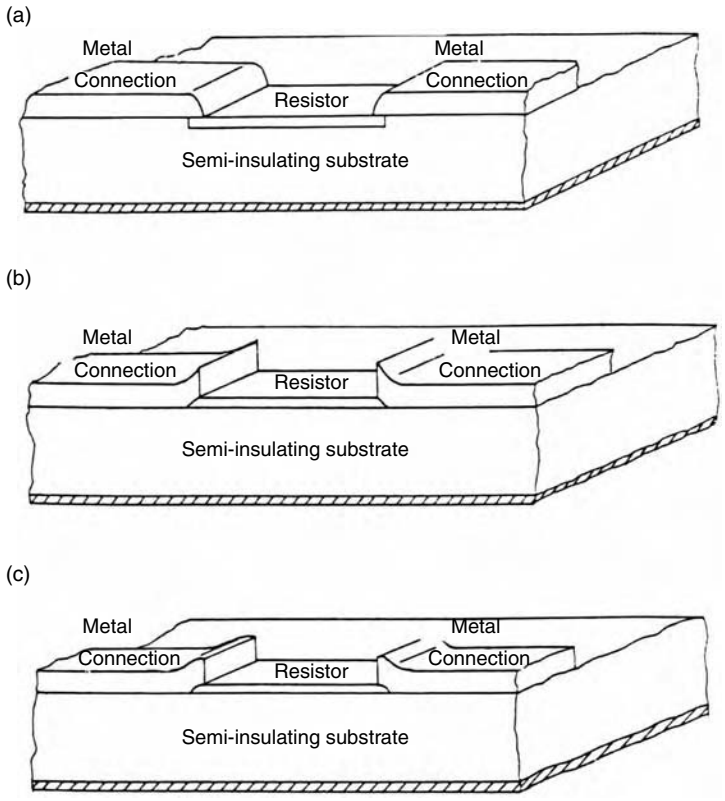


FIGURE 25.9 Common MMIC resistor fabrication approaches: (a) implanted GaAs, (b) mesa-etched/epitaxialy grown GaAs, and (c) thin film (e.g., TaN). (From Goyal, R., Ed., *Monolithic Microwave Integrated Circuits: Technology and Design,* Artech House 1989, p. 342. With permission.)

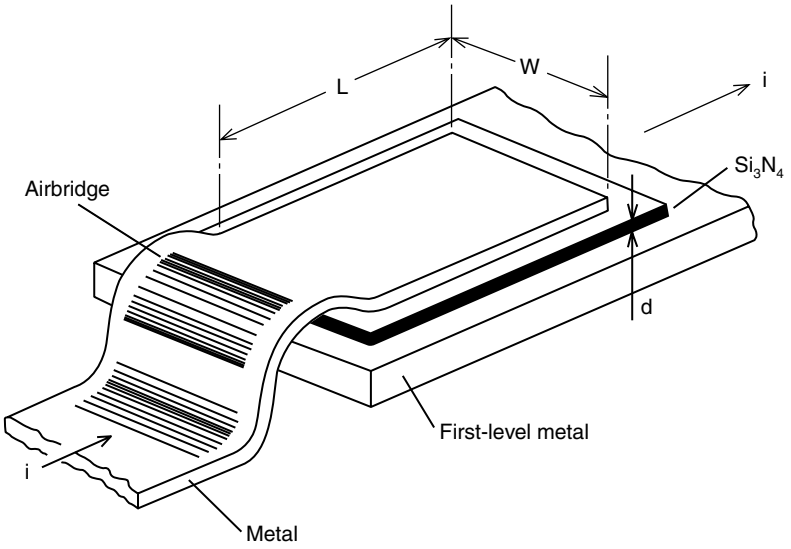


FIGURE 25.10 Metal-insulator-metal conceptual diagram. (From Ladbrooke, P., *MMIC Design GaAs FETs and HEMTs,* Artech House, 1989, p. 46. With permission.)

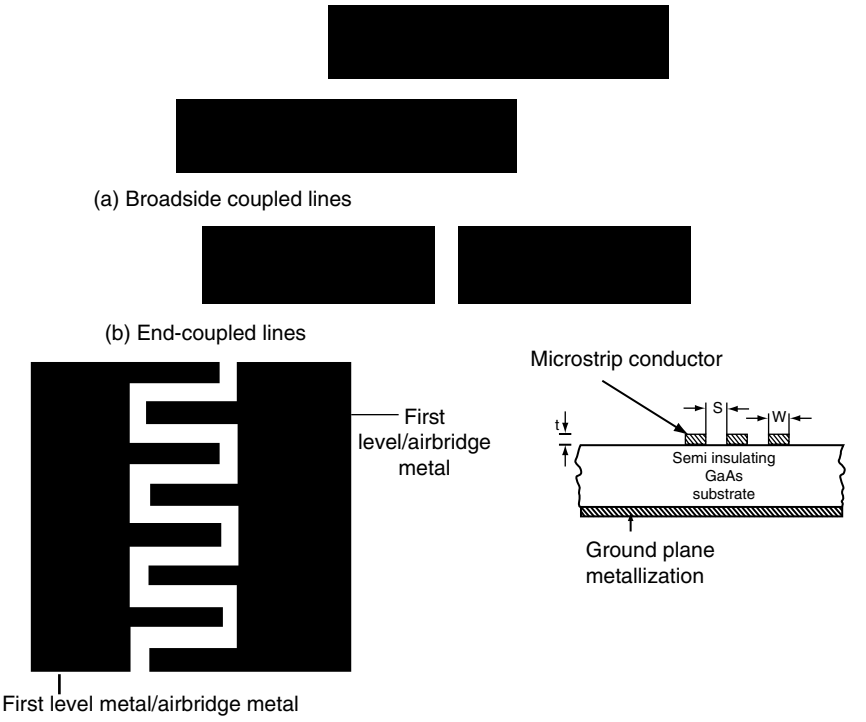


FIGURE 25.11 MMIC approaches for small valued capacitors. (From Goyal, R., Ed. *Monolithic Microwave Integrated Circuits: Technology and Design*, Artech House, 1989, p. 331. With permission.).

or chip thickness can have a strong influence on these parasitic effects and the associated performance of the capacitors in the circuit. Parasitic effects must be accurately modeled for successful MMIC design usage.

25.3.3 MMIC Inductors

MMIC Inductors are realized with narrow strips of metal on the surface of the chip. Various layout geometries are possible as illustrated in Figure 25.12 [27]. The choice of layout is dictated mainly by the available space and the amount of inductance L that is required in the circuit application, with the spiral inductors providing the highest values.

The nominal value of inductance achievable from strip inductors is determined from the total length, for the simpler, layouts, and by the number of turns, spacing and line width for the spiral inductors.

At microwave frequencies, "parasitic" effects limit the performance of these inductors. The two main effects are signal loss due to leakage currents, as measured by the quality factor Q of the capacitor, and a self-resonance frequency, beyond which the component no longer behaves as an inductor. The final thickness of the substrate influences not only the nominal value of inductance, but also the quality factor and self-resonance frequency. Inductor parasitic effects must be accurately modeled for successful MMIC design usage.

25.3.4 Airbridge Spiral Inductors

Airbridge spiral inductors are distinguished from conventional spiral inductors by the presence of metal traces that make up the inductor suspended from the top of the substrate using MMIC airbridge technology. The MMIC airbridges are, generally, used to allow crossing lines to jump over one another

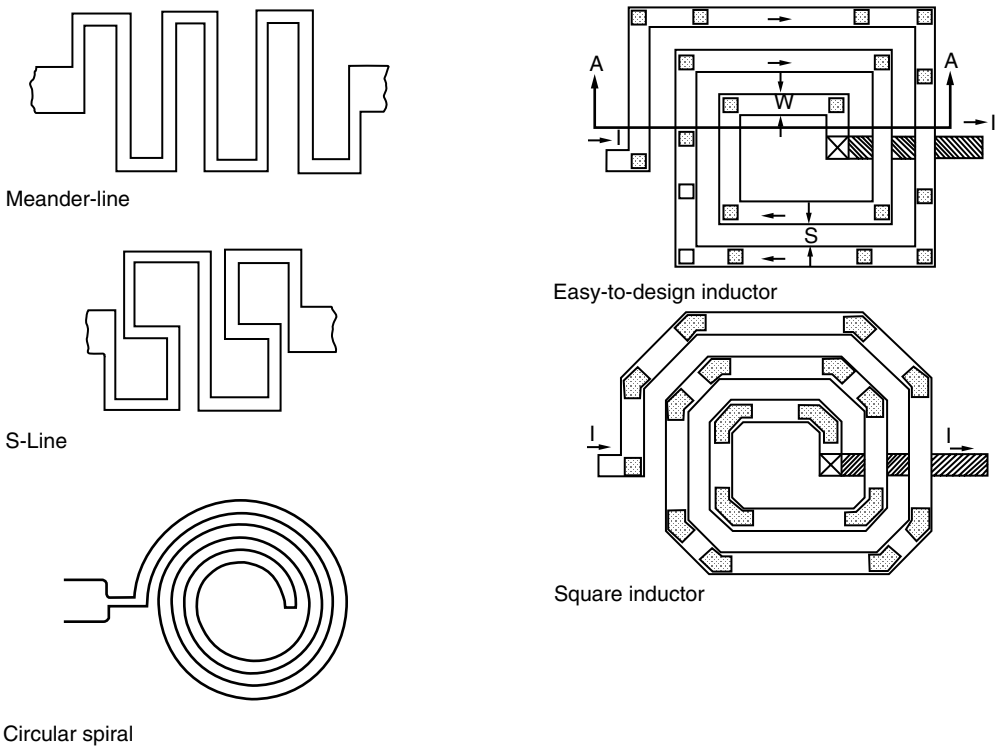


FIGURE 25.12 Various MMIC inductor layouts. (From Goyal, R., Ed. *Monolithic Microwave Integrated Circuits: Technology and Design,* Artech House, 1989, pp. 320–325. With permission.).

without touching and are, almost invariably, used in conventional spiral inductors to allow the center of the spiral to be brought through the turns of the spiral inductor for connection to the circuit outside of the spiral. In an airbridge spiral inductor all the turns are suspended off the substrate using a series of airbridges supported by metalized posts. The reason for doing this is to improve inductor performance by reducing loss as well as the effective dielectric constant of the lines that make up the spiral. The latter can have the effect of reducing inter-turn capacitance and increasing the resonant frequency of the inductor. Whether or not airbridge inductors are "worth the effort" is a debatable subject as the airbridge process is an important yield-limiting factor. This means circuit failure due to collapsed airbridges, for example, occur at an increasing rate the more airbridges that are used.

25.3.5 Typical Values for MMIC Lumped Elements

Each MMIC fabrication foundry sets its limits on the geometrical dimensions and range of materials available to the designer in constructing the MMIC lumped elements discussed in the previous section. Accordingly, the range of different resistor, capacitor and inductor values available for design will vary from foundry to foundry. With this understanding a "typical" set of element values associated with MMIC lumped elements are presented in Table 25.3 [28].

25.4 MMIC Processing and Mask Sets

Most common MMIC processes in the industry can be characterized as having 0.25 μm to 0.5 μm gate lengths fabricated on a GaAs wafer whose final thickness is 100 μm, or 4 mils. The backside of the wafer has plated gold; via holes are used to connect from the backside of the wafer to the topside of the wafer.

TABLE 25.3 Ranges of MMIC Lumped Element Values Available to Designer for a "Typical" Foundry Process

Type	Value	*Q*-Factor (10 GHz)	Dielectric or Metal	Application
INDUCTOR: Single loop, meander line, and so on	0.01–0.5 nH	30–60	Plated gold	Matching
INDUCTOR: Spiral	0.5–10 nH	20–40	Plated gold	Matching, DC power (bias) supply choke
CAPACITOR: Coupled lines	0.001–0.05 pF		Plated or un-plated gold	Matching, RF/DC signal separation
RESISTOR: Thin film	5 Ω to 1 kΩ		NiCr, TaN	DC bias ckts, feedback, matching, stabilization
RESISTOR: GaAs monolithic	10 Ω to 10 kΩ		Implanted or epitaxial GaAs	DC biasing, feedback, matching, stabilization

Source: Adapted from Goyal, R., *Monolithic Microwave Integrated Circuits: Technology and Design,* Artech House, 1989, p. 320. With permission.

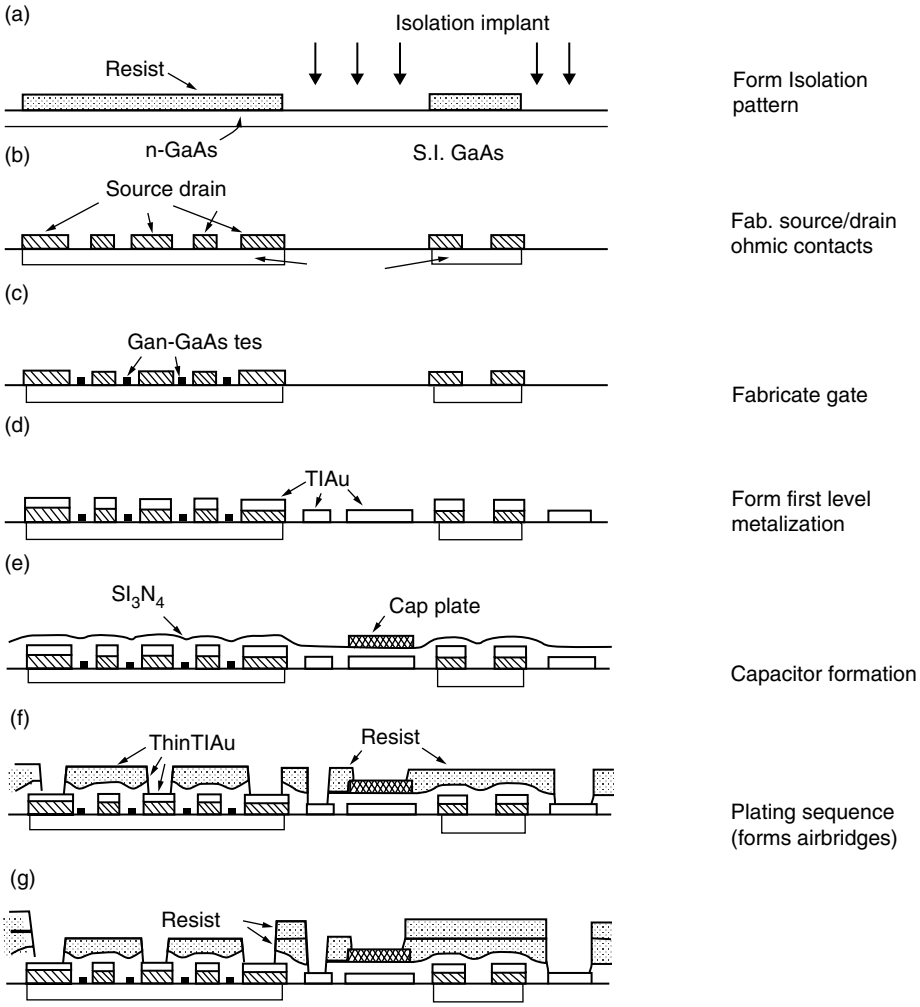


FIGURE 25.13 Conceptual diagrams illustrating flow for typical MMIC process. (From Williams, R., *Modern GaAs Processing Methods*, Artech House, 1990, pp. 11–15. With permission.).

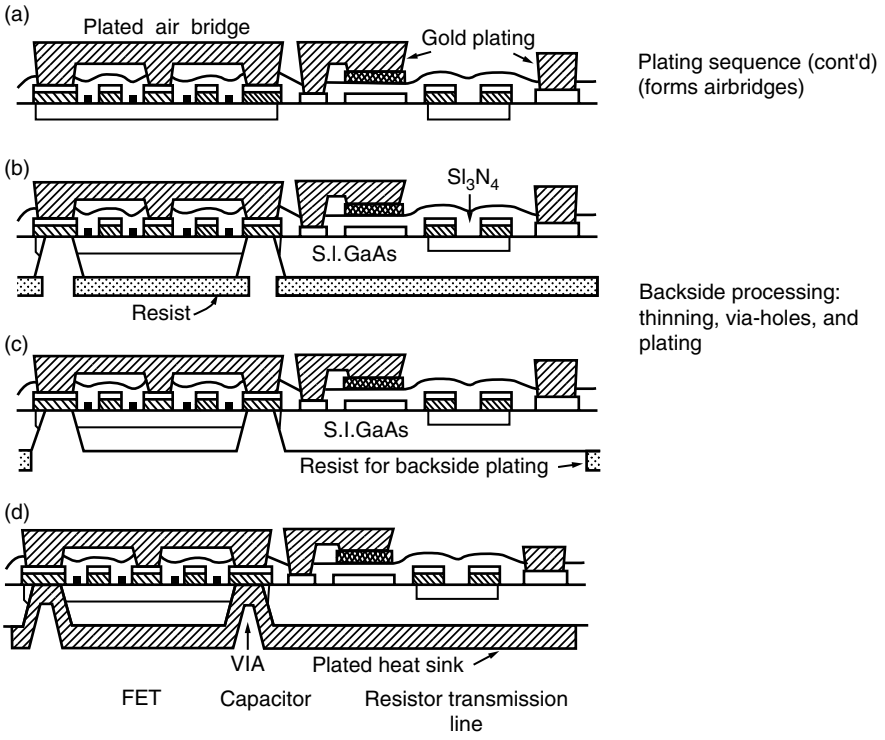


FIGURE 25.13 Continued.

Although specific procedures and steps vary from foundry to foundry, Figure 25.13 [29] illustrates a typical process.

References

1. Pucel, R. A., "Design considerations for monolithic microwave circuits," *IEEE Transactions on Microwave Theory and Techniques*, vol. MTT-29, pp. 513–534, June 1981.
2. Pucel, R.A. ed., *Monolithic Microwave Integrated Circuits*, IEEE Press, 1985, p. 1.
3. Ladbrooke, P., *MMIC Design GaAs FETs and HEMTs*, Artech House, 1989, p. 29.
4. Pucel, R.A. ed., *Monolithic Microwave Integrated Circuits*, IEEE Press, 1985, p. 1–2.
5. Hyltin, T.M. "Microstrip transmission on semiconductor substrates," *IEEE Transactions Microwave Theory and Techniques,* vol. MTT-13, pp. 777–781. Nov. 1965.
6. Mehal, E. and Wacker, R. "GaAs integrated microwave circuits," *IEEE Transactions Microwave Theory and Techniques,* vol. MTT-16, pp. 451–454, July 1968.
7. Pengelly, R. S. and Turner, J.A. "Monolithic broadband GaAs FET amplifiers," *Electronics Letters*, vol. 12, pp. 251–252, May 13, 1976.
8. Cressler, J.D. and Niu, G., *Silicon-Germanium Heterojunction Bipolar Transistors,* Artech House, 2003.
9. T. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits,* Cambridge University Press, 1998.
10. Joseph, A.J., Dunn, J., Freeman, G., Harame, D.L., Coolbaugh, D., Groves, R., Stein, K.J., et al. "Product applications and technology directions with SiGe BiCMOS**",** *IEEE Journal of Solid-State Circuits,* vol. 38 (9), pp. 1471–1478, Sept. 2003.
11. Brehm, G., Triquint Semiconductor, *personal communication*, Mar., 2006.
12. Pucel, R.A., ed., *Monolithic Microwave Integrated Circuits*, IEEE Press 1985, p. 1.
13. Goyal, R., *Monolithic Microwave Integrated Circuits: Technology and Design,* Artech House, 1989, p. 17.
14. Pengelly R., Cree, and Schmitz, Bert M/A-Com, *personal communication*, Aug., 1999.

15. Wyatt, M., Insyte Corporation, *personal communication*, Feb. 2006.
16. Frank, Mike, Hewlett Packard Company, *personal communication*, Dec. 22, 1995.
17. Kane, B., Insyte Corporation, *personal communication*, Feb., 2006.
18. Floyd, B.A., Reynolds, S.K., Pfeiffer, U.R., Zwick, T., Beukema, T., and Gaucher, B. "SiGe bipolar transceiver circuits operating at 60 GHz,"
19. Komijani, A. and Hajimiri, A., "A wideband 77GHz, 17.5 dBm power amplifier in silicon," *Proceedings of the Custom Integrated Circuits Conference, 2005. IEEE 2005,* 18–21, pp. 566–569, Sept. 2005.
20. Pucel, R., ed., *Monolithic Microwave Integrated Circuits*, IEEE Press, 1985, p. 2.
21. Liu, W. *Fundamentals of III-V Devices: HBTs, MESFETs, and HFETs/HEMTs,* Wiley-Interscience, 1999.
22. Goyal, R., Ed., *Monolithic Microwave Integrated Circuits: Technology and Design*, Artech House 1989, p. 113.
23. Ladbrooke, P., *MMIC Design: GaAs FETs and HEMTs*, Artech House, 1989, p. 92.
24. Goyal, R., Ed., *Monolithic Microwave Integrated Circuits: Technology and Design,* Artech House 1989, p. 342.
25. Ladbrooke, P., *MMIC Design GaAs FETs and HEMTs*, Artech House, 1989, p. 46.
26. Goyal, R., Ed. *Monolithic Microwave Integrated Circuits: Technology and Design,* Artech House, 1989, p. 331.
27. Goyal, R., Ed. *Monolithic Microwave Integrated Circuits: Technology and Design,* Artech House, 1989, pp. 320–325.
28. Goyal, R., *Monolithic Microwave Integrated Circuits: Technology and Design,* Artech House, 1989, p. 320.
29. Williams, R., *Modern GaAs Processing Methods*, Artech House, 1990, pp. 11–15.

26

Bringing RFICs to the Market

John C. Cowles
Analog Devices—Northwest Labs

26.1 Introduction: A Brief History of Radios

The art and science of RFIC design developed over a period of at least 150 years. The fundamental work of the philosopher–scientists of the nineteenth century paved the way with invaluable empirical discoveries related to electricity and magnetism. The identities of these early players are known to us today through the basic units that electrical engineers have been using over the past century: Volta, Ohm, Ampere, Henry, and Faraday. In 1873, Maxwell laid the theoretical foundation for the propagation of electromagnetic energy through space in his famous treatise. Maxwell's equations demonstrated that electric and magnetic fields are inexorably coupled and fueled by charges and currents.

The nineteenth century also brought a flurry of activity as electricity was used in practice to transmit information over a distance, first through wires and then through air. Samuel Morse's invention of the telegraph in 1835, accompanied by his digital, self-named code, and Alexander Graham Bell's invention of the telephone in 1876 demonstrated to the world the potential of electricity as an effective means for communications. In the late 1800s, Hertz proved the existence of electromagnetic waves and was shortly followed by Marconi, who in 1896 demonstrated the first wireless transmission by sending telegraph signals over 2 km from ship to shore. Thus, began the age of wireless transmission.

In the first quarter of the twentieth century, broadcasts began in earnest with Reginald Fesseden getting first credit by broadcasting Christmas carols with a crude form of amplitude modulation in 1906. Shortly thereafter, in the 1920s the first commercial radio and television transmissions began operation. Such broadcasts required radio transmitters and receivers along with the means to carry the information: modulation. The great Edward Armstrong is credited with the invention of frequency modulation, the regenerative receiver and the super-heterodyne receiver, which had been the work-horse until the advent of digital direct-conversion radios. In a matter of 25 years, the world was brought closer together than

ever before by radio and television, although only few could afford the equipment for technology since manufacturing had not yet come of age.

The next step involved a confluence of disparate developments that propelled the era of electronics to the next level. Coincident with the advancements in radios and broadcasting was the observation of effects not predicted by the classical sciences, which led to the birth of quantum mechanics. The unraveling of the underlying physical processes in materials and the harnessing of their properties launched the era of vacuum tubes and the seeds for future solid-state electronic devices. As early as 1883, Thomas Edison stumbled on a crude thermionic diode while in 1906 DeForest invented the Audion, the first triode tube amplifier. World War II gave impetus to the development of the RADAR, the magnetron, the klystron and the traveling-wave-tube, further advancing the generation and detection of signals. There was also the appearance of the digital computer in the form of the ENIAC developed by Eckert and Mauchly in 1946, although there was no thought as yet of combining it with a radio. This period of discovery reached a pinnacle with the discovery of the point-contact transistor in 1947 by Shockley, Brattain, and Bardeen. Ironically, their "discovery" of bipolar action was allegedly a fortuitous by-product of a failed experiment in search of a field-effect device. In time, the first true JFETs and MOSFETs followed suit and launched the age of solid-state integrated circuits.

By the mid-1960s there was a growing semiconductor industry led by Jack Kilby of Texas Instruments and Robert Noyce of Fairchild among others. To supplement the technology, Prof. Robert Pederson of Berkeley led the development of the simulation tool SPICE which enabled solid-state circuits to grow in complexity by replacing laborious bread-boarding with computer simulations. As levels of integration increased towards VLSI, the microprocessor and then the personal computer appeared in the marketplace, followed by numerous consumer items for entertainment that also relied on digital technology such as video camcorders, DVD/CD/MP3 players, and digital cameras. The regular release of consumer products containing complex digital and analog ICs at affordable prices reflected a mature solid-state electronics industry consisting of robust device technologies, accurate models, sophisticated simulation tools, organized design teams, and a strong sense of entrepreneurship.

Coincident with the growth of consumer-oriented electronics was the development of microwave electronics, primarily driven by defense and satellite industries. These niche applications drove the development of materials, design philosophies, and test techniques required to deliver performance above 1 GHz. A typical early microwave amplifier might consist of a discrete GaAs MESFET or silicon BJT mounted on FR4 board material along with passive matching networks either as printed transmission lines or as mounted components. In the 1980s and 1990s, the fabrication technologies for III–V materials, their modeling and simulation tools had reached the level where multitransistor, multifunction, monolithic ICs could be designed with confidence, the era of microwave (MICs) and millimeter-wave ICs (MMICs). The devices, circuit topologies, spawning testing methods and even the nature of the figures of merit were foreign to traditional analog and digital designers. It would take yet one more link to bring these disparate worlds together to forge the concept of the RFIC.

The confluence of radio architectures, semiconductor technologies, low-cost manufacturing, and a strong consumer electronics market required a key catalyst before launching the age of wireless technology. The spark was simply the latent need for wireless connectivity by the public at large. Prosaic devices as the pager, the garage door opener, and the television remote control were the primitive ancestors of the cellular phone that appeared in the late 1980s along with the Internet. The expectation today is wireless access to digital information anywhere, at any time, with anyone, using whichever access method is available: GSM, UMTS, WiMax, WiFi, and so on. As the public has adapted quickly to the new opportunities afforded by such access to information, the wireless industry has also adapted to create new markets and services. The wireless industry, in cooperation and in competition with wired access methods such as optical-fiber, CATV, and Ethernet, are vying for the delivery of information to the public and enterprise. The distinction between telephony, television, internet, satellite, and computer is quickly vanishing, opening opportunities for service providers to compete for customer access. There is a "productivity cycle" at work, shown in Figure 26.1, in which radio techniques, semiconductor technologies, and bandwidth constraints are in constant flux, one driving the other.

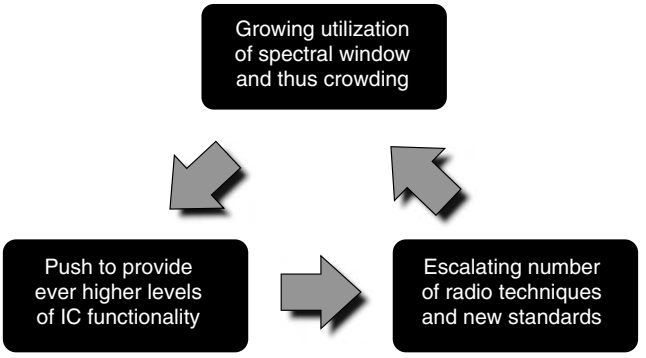


FIGURE 26.1 The productivity cycle drives innovation and progress.

The on-going challenges are perfectly clear:

1. Low cost to reach the widest consumer market
2. Re-configurability and flexibility to adapt to changing channels and standards
3. More efficient use of bandwidth to increase data capacity
4. Higher frequency allocations to increase available bandwidth

Addressing these challenges has required the marriage of classical high-speed analog and microwave designs to deliver timely, robust products at the appropriate integration level, at a competitive price while operating at microwave frequencies. The possibilities are endless as the public demands more access and services which in turn keep the cycle in motion.

26.2 Preliminaries

The RFIC is the building block that links the digital engines that perform high-level mediation of information transfer and processing and the antennas that form the air-interface to the radio. Before the product design is to begin there are many questions, both technical and strategic, to be asked. First of all, one must identify the basic technical function and articulate, as completely as possible at this point, the specifications. In addition to the key metrics, this might include the package with pin descriptions, the proposed technical approach and even a rough die layout, and any applications' details indicating how the part might be used. This information is nothing more than a concept data sheet that describes the product even before committing heavy design and manufacturing resources. It is surprising how many critical decisions can be made in the process of creating such a data sheet.

Underlying the technical requirements captured in the concept data sheet is a set of strategic requirements. As shown in Figure 26.2, there are market forces and corporate realities which must be considered. First of all, is the proposed product of such entrepreneurial vigor that it is likely to create new markets or is it an upgrade of an existing product or an answer to competition? The classic market questions of time-to-market and cost/profit must be clear and understood. If the product has no precedence, one must rely on experience with customers and an anticipation of what customers WILL need. In defining timing and costing, it is essential to understand the target market. The so-called "vertical markets" are often driven by consumer electronics and demand very specific functionality and specifications, low cost, and quick cycle times. The volumes tend to go to a few customers and typically skyrocket for a period of time and then disappear. The "horizontal markets" are driven by a slower-paced infrastructure market that requires high-performance parts but is not under the same time and cost pressure. The customer base tends to be broad with moderate volumes growing systematically over time and holding their value.

The other component required for product development is, of course, corporate resources. This covers the design expertise, the available semiconductor fabrication technologies, the product engineering

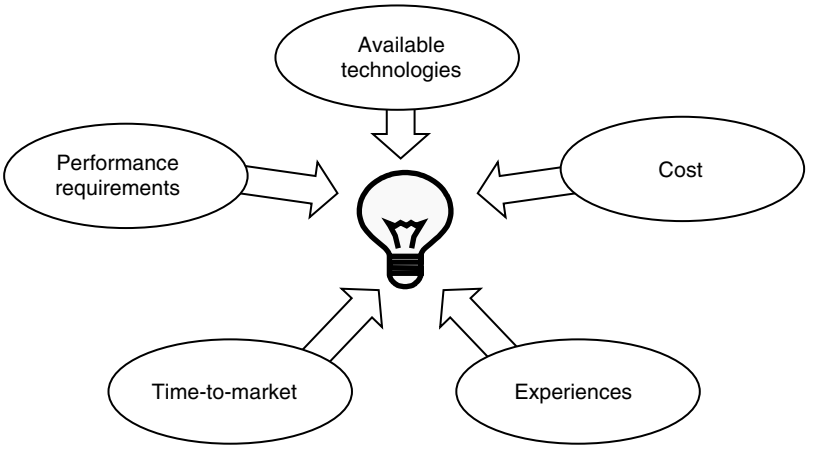


FIGURE 26.2 There are many variables to consider when proposing a product.

required to bring a product to a market release and customer applications' support once the part in released. If any of these components is miscalculated, the product development is likely to fail. It is unlikely that a junior digital designer with access to a low-cost CMOS technology will produce a breakthrough RF product in GaAs without adequate training in a short TTM scenario. It must be remembered that circuits, in themselves, are not products; products require a team of dedicated specialists to deal with the device modeling, the actual design, the fabrication, the packaging, the evaluation, the reliability, the data sheet, and so on. Any weak link in this chain is likely to be catastrophic.

26.3 The Design Process

The design of a product is built on the foundation established early on during the concept phase. At this point, there is a concept data sheet with preliminary information to guide the design phase and the necessary resources allocated to the task. As noted earlier, this includes the designer's capabilities, the fabrication technology, the development schedule with key milestones, and the greater team responsible for bringing the design into production. The design phase enters its first part in which detailed technical approaches are investigated to validate whether the proposed concept is feasible. Any building block can be implemented in a myriad of ways. Figure 26.3 illustrates a handful of amplifiers whose embodiments are quite distinct. This is where an understanding of the bandwidth, gain, dynamic range, and impedance levels would eliminate many of the candidate approaches. Similar examples for mixers, oscillators, variable gain amplifiers, and power detectors are shown in Figures 26.4–26.7. The designer should be encouraged to consider all options and investigate new topologies as well. The simulation tools allow for "experiments" that can quickly weed out pretenders and, sometimes, even identify an undiscovered star. When topologies and technologies seem to conflict, the option of integration at the package level can also be an option depending on the cost. Figure 26.8 illustrates a common example of a portable handset power amplifier in which a GaAs power stage and a silicon power detector are integrated on a substrate along with passive lumped and distributed components. There might also be extra functionality to improve the efficiency-distortion trade-off and to protect against dangerous loading conditions. This form on integration has become possible due to advances in packaging that have turned expensive, military module technology into low cost consumer-level components.

During this exploratory, feasibility period, the designer would focus on the areas of greatest technical risk and difficulty, leaving the more mundane details for later. Simultaneously, the market feasibility should also be explored further with customer visits, survey of the competition, and preliminary cost analysis. The team then reaches a critical gating point in which the overall feasibility of the product is

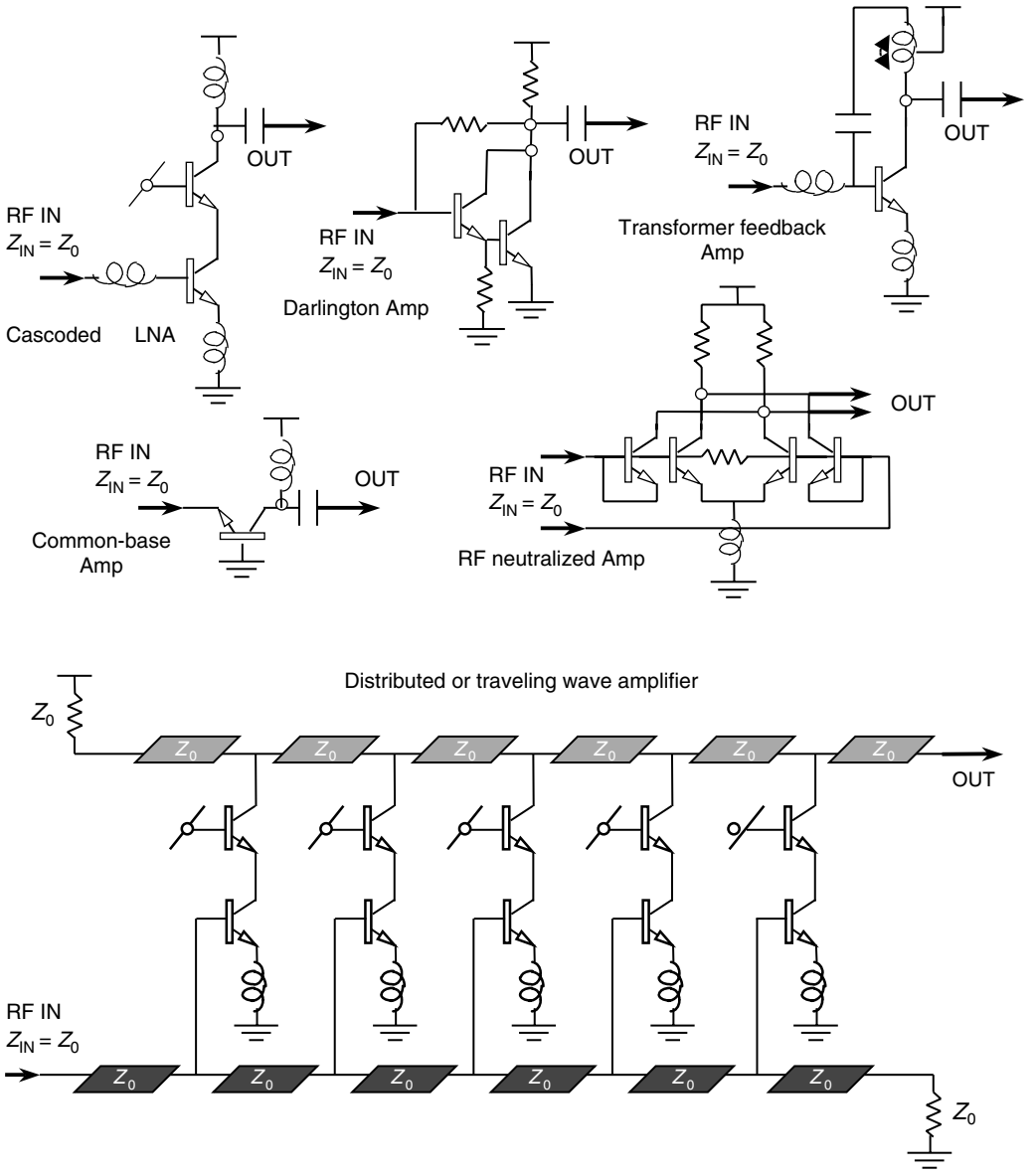


FIGURE 26.3 Which amplifier is appropriate for your application?

reviewed from technical and business points-of-view. From here, there should be an agreement to proceed towards a product, return to feasibility, or stop the work altogether. Note that, as painful as it may be to cancel a development, it is much less painful to do so earlier rather than later. In many cases, the feasibility of a product is not in question, and this phase can be mostly by-passed as long as the goals and requirements are clearly articulated.

A positive feasibility review moves the designer towards a full-blown development. The technical approach is cemented early in this phase with the balance of the time dedicated to fine tuning the design to address robustness and manufacturability. The time spent on these details is paramount to a successful product as it maximizes the production yield, the profits, and customer satisfaction. Simulation sweeps over temperature and over supply and over certain device model parameters quickly reveal weak points in the design. RFIC design has emerged as a symbiosis of high-speed analog with its ties to silicon technology

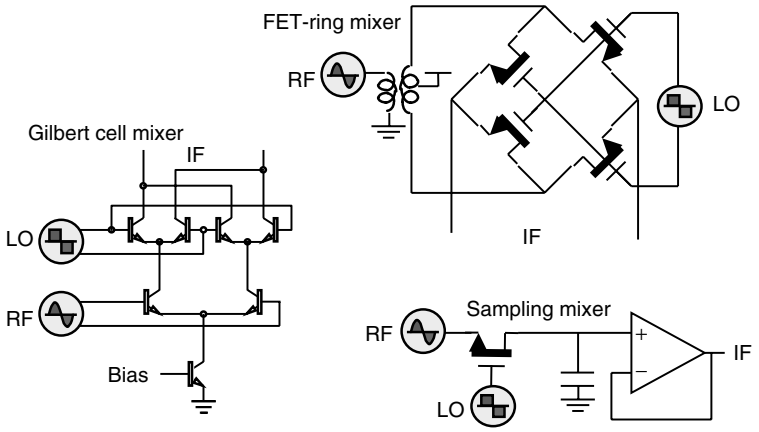


FIGURE 26.4 Mixers have not changed much in 20 years yet few understand the merits of passive and active topologies.

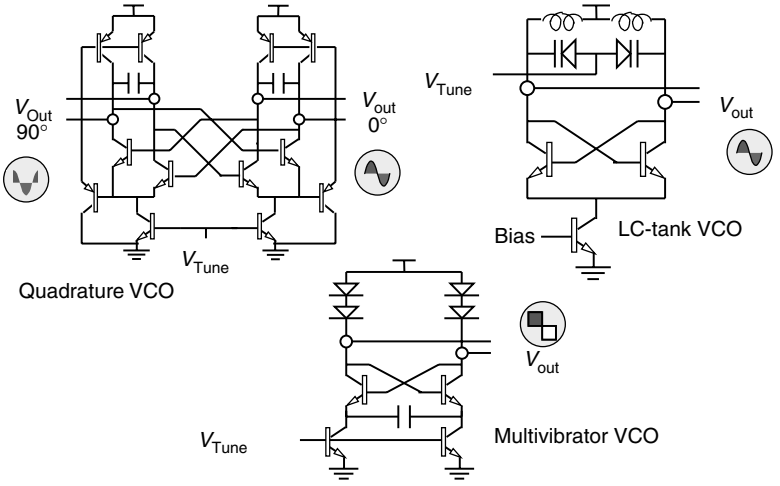


FIGURE 26.5 For a small circuit, oscillators generate more disagreement than any other.

and of more traditional microwave design with its ties to GaAs. Adopting the best aspects of both worlds in design, simulation and layout methodologies is the best toolbox.

Certain practices have long been established in the realm of analog design. Many parameters such as gain are dimensionless and should be defined by ratios of like quantities. The desensitization of such scalar quantities is addressed by use of unit elements of resistors, capacitors, and transistors placed in close proximity and in the same orientation to improve their matching. It is also prudent to use the largest permissible dimensions for these unit elements and surround them with dummies to create a common topography for all. The schematic should give the engineer responsible for the layout the visual clues as to the placement of the unit cells and the dummies. The more clearly the schematic can communicate the intent of the designer, the less is left up to chance. For critical matching, such as in the differential pairs shown in Figure 26.9, components can be connected in a cross-quad arrangement where any temperature or process gradients are spatially averaged. At high frequencies, matching of capacitances in differential circuits may be more important than minimizing them, particularly since minimum sized devices are rarely used due to noise or current handling reasons.

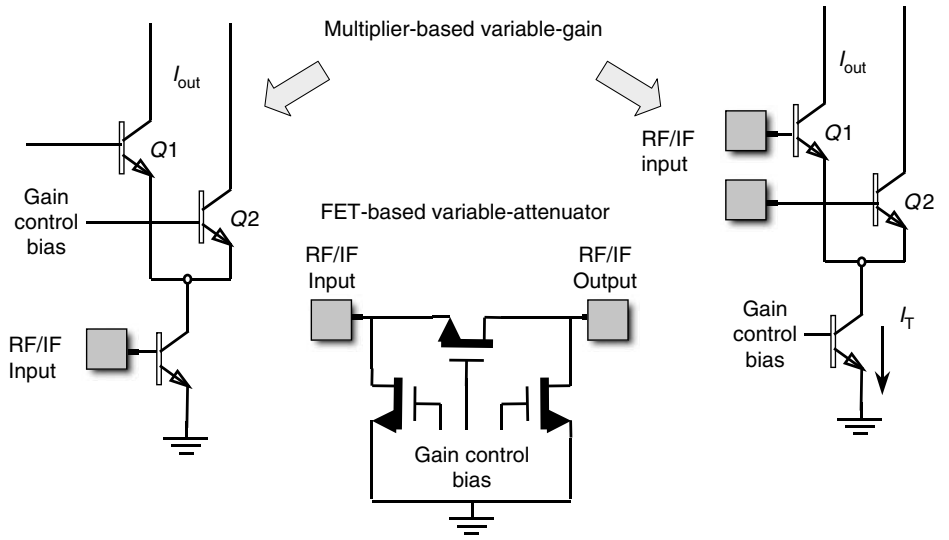


FIGURE 26.6 The classic analog and microwave approaches for gain control differ greatly.

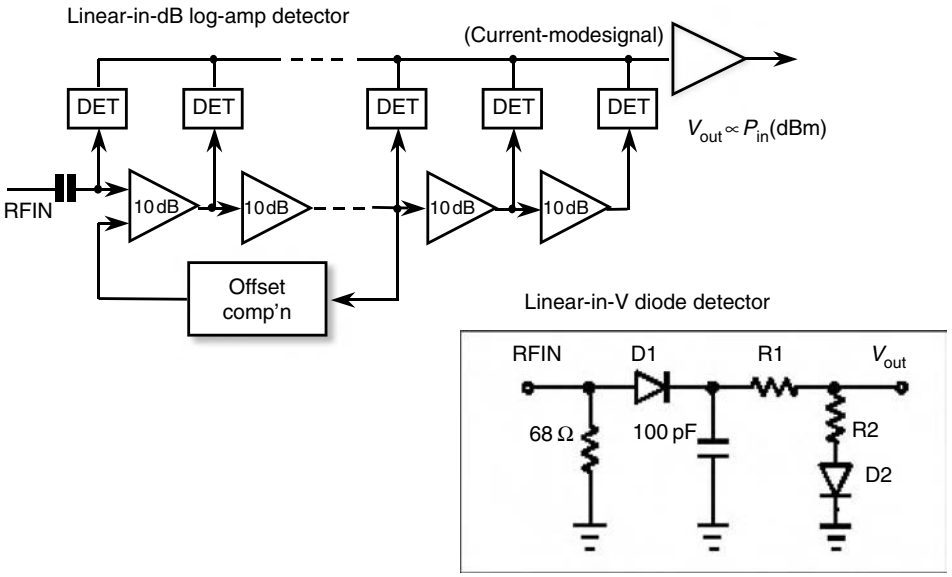


FIGURE 26.7 Power detection has evolved from magic to science in the age of RFICs.

In microwave design, the effect of package and layout parasitics has always had a dramatic impact on the high-frequency performance of circuits. With the increased complexity, integration level and requirements for RFICs, package models and layout parasitic extraction have become critical. Many inexperienced and experienced designers have been caught saying something akin to, "This is a low-frequency circuit so the package model is not important," only to realize later that the 90 GHz transistors in the process were not told about being a low frequency circuit. The accuracy of these parasitic models depends directly on the complexity of the networks that represent them. The iteration between design, package, and layout simulation can be protracted and, at some point, fruitless as many competing effects are present simultaneously. Proper judgment is required to trade-off model accuracy against time to tape-out. It is

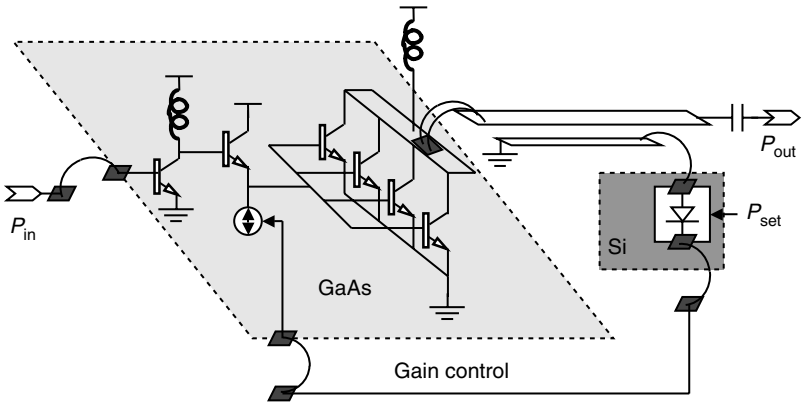


FIGURE 26.8 A cellular PA is often a complex multi-chip system implemented as a low-cost module.

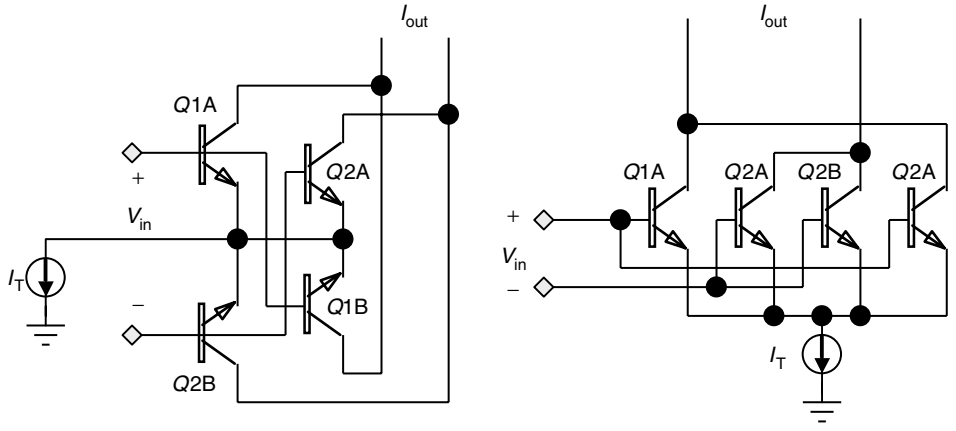


FIGURE 26.9 Cross-quads ubiquitous in analog design have added merit in RFICs. Two forms are shown here.

useful to be able to turn "on" and "off" elements in the package and parasitic network for debugging purposes. This might help identify the need for extra bond-wires or pins for a pesky ground oscillation or the need to balance a pair of differential lines for a lower offset voltage.

The greatest chasm between analog and RF design lies in the design of the bias circuits. While analog IC designers have been using carefully crafted bias currents to achieve a certain precision or stability, microwave designers were just happy to get any current through a device. The era of RFICs has brought the sophistication of biasing to the RF domain. Consider the differential pair in Figure 26.10a. In order to achieve a gain that is independent of temperature, the current must be shaped to be proportional to temperature so that the G_m becomes constant. Now consider a heavily degenerated differential pair in Figure 26.10b. What shape should the current take now that $G_m \sim 1/R$? The answer can be complicated. If constant input capacity or output swing (both related to $I_{Bias}R$) is the important factor, then the current should be made stable with temperature. If the output of a differential pair is resistively loaded to form a CML gate, shown in Figure 26.10c, then the current might be made complementary to temperature (decrease at higher temperatures) since the voltage swing needed to toggle a CML input drops at higher temperatures. There is much to be gained in thinking about the appropriate bias cell. A fair question to pose is whether a formal bias cell is needed or can a simpler bias scheme suffice. The incidental bias for the simple Darlington amplifier in Figure 26.3 might be derived from the supply while the current that defines the reference for a square law detector must be precise to ensure accuracy.

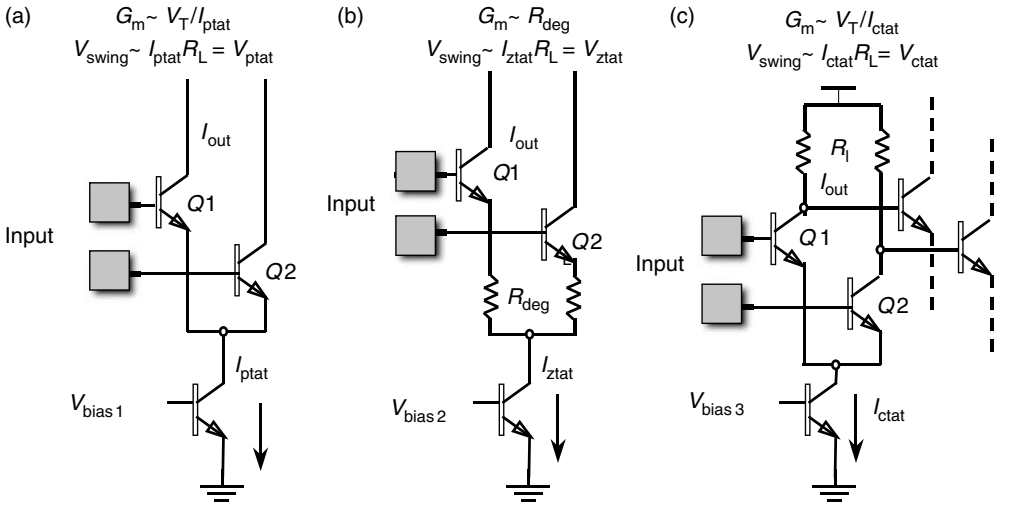


FIGURE 26.10 (a)–(c) Biasing differential pairs requires some forethought on the goals.

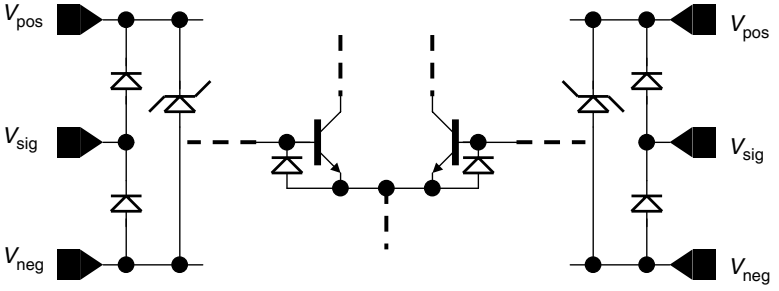


FIGURE 26.11 Typical ESD protection for differential pair signal pins.

An area of robustness that is often neglected is ESD protection. All die and packaged parts are subject to ESD events during handling. It is typical to test every product for a human body model (HBM), machine model (MM), and a charged device model (CDM) up to a certain discharge level. Failing to protect pins can be catastrophic to factory yield and to reliability in the field. Every technology tends to have suggested "good practice," but there is no substitute for thinking ahead. Figure 26.11 shows one example of protecting the input pins of a differential pair. Note that the technology recommendation is to place specially designed ESD diodes from each pin to the supply and ground along with a power ESD diode across the supply. However, it is clear that an ESD event across the input pins would destroy the emitter–base junction that was reverse biased. The solution is to place a diode antiparallel to each emitter–base junction. Under normal operation, these diodes are off; under a pin-to-pin ESD event, the diodes clamp the reverse bias and protect the inputs. It is difficult to anticipate all the possible permutations of ESD events in a complex circuit; here again experience and some forethought are essential.

Once the design is completed, there is usually an event called a design review. The goals of a design review can vary greatly according to the product and the requirements. It is intended to address a broad audience of interested folks ranging from colleagues who share the technical expertise, colleagues from different design disciplines, the marketing team and those who will inherit the silicon and carry it to release and support the customers. The content should address all the attendees' concerns such as achieved technical specification, package and fabrication technologies, circuit techniques of interest and concern, remaining risk factors, robustness, revised cost estimates, schedules, and so on. There will inevitably be

a list of actions to investigate and correct if needed. A successful design review hinges on a relevant and participatory audience asking probing questions and giving productive criticism. A passive, uninvolved crowd does the team no favor.

During the design phase, the layout activity should begin. Note that parasitic extraction and validation that the die will fit in the package are necessary to complete the simulation of the RFIC. The greater the realism captured in the schematic, the closer it will be represented in layout. A layout review with a team is just as crucial as the design review. Great patience is required to look for the needle in the haystack. Common topics for discussion are the routing grid of power supplies and ground, critical matching areas, current density along metal lines, ESD protection scheme, unintentional coupling of signals and the bonding diagram associated with the package. As in the design review, there are always issues raised that deserve investigation. The layout and design at this point are inextricably intertwined and tape-out of the mask set only takes place once both are deemed complete by the team. The tape-out authorization is another key gating point where the market and technical merits of the part are once again scrutinized before committing money and resources to mask-making and manufacturing.

26.4 From Silicon to Release

Once the wafers are in process, the team does not go dormant. The time schedules for when material returns must be outlined. The data sheet needs to be updated with a more comprehensive set of specifications, final package, details, key features, and basic connections to the part. In addition to advancing the status of the official data sheet, the update helps define the limits of the test and characterization plans needed to release the part into full production. The evaluation boards can now be designed, go to layout and be reviewed in much the same way the RFIC was scrutinized. The choice of board material, supporting passive components, board traces, proper grounding, supply decoupling and connectors all affect the measurements of the RFIC high-frequency performance.

Once the wafers are finished, the die must be separated, packaged, and mounted on the evaluation board. Characterization and the inevitable debugging phases begin. At this point a full look at the part is critical to help decide whether revisions are necessary. It is common to have parts from wafers that have had certain parameters, such as a resistor value or a doping level, purposefully skewed in order to force certain sensitivities. From nominal and skew material, the robustness of the part to process variations can be inferred. Full temperature and ESD testing complete the characterization. If all seems well at this point, stocking quantities can be generated and the official data sheet can be posted, indicating that the product has been released to the world.

More often, some revisions are necessary. The process of debugging is perhaps the most frustrating and murky of all. In the best of circumstances, the changes may involve only revisions to metallization or to resistors and capacitors. In the worst of cases, a whole new mask set needs to be ordered. It is prudent at this point to re-evaluate the future of this product in terms of time-to-market, consumption of resources, and continued risk. The debug process relies on ingenuity of testing and extreme familiarity with the internal structure of the part. It is absolutely essential for the designer to be present with the test engineer, if they are not one and the same. Quick, resourceful thinking is the only answer here. With some luck and forethought, de-capped ICs, in which the plastic has been etched away, can be probed to gather more information not accessible from the package pins alone.

26.5 Toward the Future

Once the product is released into production, the whole team shifts modes. The applications' support teams go to the field to aid customers and gather feedback to help drive the next development. In the meantime, the design team is already pondering the next concept, tracking the competition, and looking for new product ideas. Just as there is a macro-level productivity engine cycle where markets, technology,

and manufacturing chase each other, there is a micro-level cycle in which product definition, design, and release go roundabout. Both cycles yield technological progress, more competitive solutions, and continuous process improvement. This cycle has carried the industry from the first vacuum-tube radios that heated a small room to a video-enabled cellular phone of today which was, coincidentally, inspired by Dick Tracy's watch in the 1930s at the dawn of radios.

III

CAD, Simulation and Modeling

27

System Simulation

Joseph Staudinger
Freescale Semiconductor, Inc.

The concept of system simulation is an exceptionally broad topic. The term itself, system, does not have a rigid definition and in practice the term is applied to represent dramatically differing levels of circuit integration, complexity, and interaction. In a simple sense, the term is applied to represent the interconnection and interaction of several electrical circuits. In a broader sense, such as in communication systems, the term may be applied to represent a much higher level of complexity including part of, or the composite mobile radio, base unit, and the transmission medium (channel). Regardless of complexity of the level, the issue of simulation is of critical importance in the area of design and optimization.

As one might expect, the techniques and methods available in the engineering environment to simulate system level performance are quite diverse in technique and complexity [1–3]. Techniques include mathematically simple formula-based methods based upon simplified models of electrical elements. Such methods tend to be useful in the early design phase and are applied with the intent of providing insight into performance level and trade-off issues. While such methods tend to be computationally efficient allowing simulations to be performed rapidly, accuracy is limited in large part due to the use of simplified models representing electrical elements. Other techniques tend to be computationally intensive CAD-based where waveforms are generated and calculated throughout the system. The waveform level technique is more versatile and can accommodate describing electrical elements to almost any level of detail required, thereby exploring the system design and performance space in fine detail. The models may be simple or complex in mathematical form. In addition, it is possible to use measured component data (e.g., scattering parameters) or results from other simulators (e.g., small- and large-signal circuit simulations using harmonic balance where the active device is represented by a large-signal electrical model). The price for the improvement in accuracy is significantly longer, perhaps much longer simulation times and the requirement to very accurately describe the characteristics of the electrical components.

The intent of this section is to examine fundamental issues relative to simulating and evaluating performance of microwave/RF related system components. A number of terms describing system level performance characteristics will be examined and defined. In addition, first order methods for calculating the performance of systems consisting of cascaded electrical circuits will be examined. To begin, consider three parameters of interest in nearly all systems: gain, noise figure (NF), and intermodulation distortion (IMD).

27.1 Gain

A usual parameter of interest in all systems is the small signal (linear) gain relationship describing signal characteristics at the output port relative to the input port and/or source for a series of cascaded circuits.

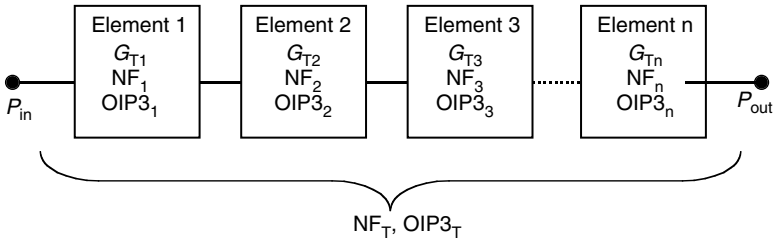


FIGURE 27.1 System formed by cascading three electrical elements.

Numerous definitions of gain have been defined in relationship to voltage, current, and power (e.g., power gain can be defined in terms of transducer, available, maximum stable, etc.) [1]. In general for system analysis, the concept of transducer power gain (G_T) is often applied to approximate the small signal gain response of a series of cascaded elements. Transducer power gain (G_T) is defined as the magnitude of the forward scattering parameter (S_{21}) squared (i.e., $G_T = |S_{21}|^2$, the ratio of power delivered to the load to that available from the source). This assumes the source (Γ_S) and load (Γ_L) voltage reflection coefficient are equal to zero, or alternatively defined as a terminating impedance equal to characteristic impedance Z_0 (typically 50 Ω).

Consider several two-port networks cascaded together as illustrated in Figure 27.1 where the transducer power gain of the *i*th element is represented as G_{Ti}. The transducer power gain of the cascaded network is:

$$G_{T_T} = G_{T1}(dB) + G_{T2}(dB) + G_{T3}(dB) + \ldots G_{Tn}(dB) \tag{27.1}$$

The accuracy of Equation 27.1 relies on the assumption that that *i*th two-port network is terminated by characteristic impedance Z_0 per the above definition. In practice, the source and load termination provided to the *i*th element is defined by the *i*th – 1 and *i*th + 1 elements, respectively. Even though in a well-designed subsystem, each two-port network is designed such that the input (S_{11}) and output (S_{22}) scattering parameters are near zero in magnitude, they cannot be exactly zero resulting in impedance mismatch effects. Hence, Equation 27.1 is approximate and its accuracy dependent on each element satisfying the above criteria. A more thorough analysis accounting for impedance mismatches can be performed at the expense of more complexity. In general this requires a more precise description of each element using perhaps some form of network parameters. For example, the *T* and scattering parameters (T_T and S_T, respectively) for two networks, A and B, cascaded together are given by [3,7]

$$T_T = \begin{bmatrix} T_{11}^A & T_{12}^A \\ T_{21}^A & T_{22}^A \end{bmatrix} \begin{bmatrix} T_{11}^B & T_{12}^B \\ T_{21}^B & T_{22}^B \end{bmatrix} \tag{27.2}$$

$$S_T = \begin{bmatrix} S_{11}^A & 0 \\ 0 & S_{22}^B \end{bmatrix} + \begin{bmatrix} S_{12}^A & 0 \\ 0 & S_{21}^B \end{bmatrix} \begin{bmatrix} -S_{22}^A & 1 \\ 1 & -S_{11}^B \end{bmatrix}^{-1} \begin{bmatrix} S_{21}^A & 0 \\ 0 & S_{12}^B \end{bmatrix} \tag{27.3}$$

While the above methods allow an exact analysis for cascaded linear circuits, it is often difficult to apply them to practical systems since the network parameters for each of the elements comprising the system are usually not known precisely. For example, in systems consisting of interconnected circuits, board layout effects (e.g., coupling between elements) and interconnecting structures (board traces) must also be included in applying the network analysis techniques shown in Equations 27.2 and 27.3. This, of course, requires accurate knowledge of the electrical nature of these structures, which is often unknown.

In critical situations, the network parameters for these elements can be determined by measurement or through the use of electromagnetic simulations assuming the geometrical and physical nature of these structures are known.

27.2 Noise

A second parameter of interest important to all systems is noise. In receivers, noise performance is often specified by noise figure, defined as

$$NF(dB) = \frac{S_i/N_i}{S_o/N_o} \tag{27.4}$$

where S_i/N_i and S_o/N_o are the signal-to-noise ratio at the input and output ports, respectively. Note that *NF* is always greater than or equal to unity (0 dB). When several circuits are cascaded together as illustrated in Figure 27.1, the cascaded noise figure (NF_T) is given by

$$NF_T = NF_1 + \frac{NF_2 - 1}{G_1} + \frac{NF_3 - 1}{G_1 G_2} + \frac{NF_4 - 1}{G_1 G_2 G_3} + \ldots \frac{NF_n - 1}{G_1 G_2 G_3 \ldots G_n} \tag{27.5}$$

where G_i and NF_i are the gain and noise figure of the i^{th} element, respectively. Note the importance of the contribution of the first element's noise figure to the total cascaded noise figure. Hence, the noise figure of the low noise amplifier contained in a receiver is a major contributor in setting the noise performance of the receiver.

27.3 Intermodulation Distortion

Intermodulation distortion (IMD) has been a traditional spectral measure (frequency domain) of linearity applied to both receiver and transmitter elements. The basis of IMD is formed around the concept that the input-output signal relationship of an electrical circuit can be expressed in terms of a series expansion taking the form:

$$E_o = a_1 E_{in} + a_2 E_{in}^2 + a_3 E_{in}^3 + a_4 E_{in}^4 + a_5 E_{in}^5 + \ldots \tag{27.6}$$

where E_{in} and E_o are instantaneous signal levels at the input and output ports of the electrical circuit, respectively. If the circuit is exactly linear, all terms in the expansion are zero except for a_1 (i.e., gain). In practice, all circuits exhibit some nonlinear behavior and hence higher order coefficients are nonzero. Of particular interest is the spectral content of the output signal when the circuit is driven by an input consisting of two sinusoids separated slightly in frequency taking the form

$$E_{in}(t) = \cos(\omega_1 t) + \cos(\omega_2 t) \tag{27.7}$$

where ω_1 and ω_2 are the angular frequencies of the input stimuli and where ω_2 is slightly greater than ω_1. The output signal will exhibit spectral components at frequencies $m\omega_1 \pm n\omega_2$, m = 0,1,2, … and n = 0,1,2, … as illustrated in Figure 27.2. Notice that the third series term ($a_3 E_{in}^3$) generates spectral tones at $2\omega_1 - \omega_2$ and $2\omega_2 - \omega_1$. These spectral components are termed third order intermodulation distortion (*IM*3). It can also be shown that series terms greater than three also produce spectral components at these frequencies and hence the total *IM*3 is the vector sum from E_{in}^3, E_{in}^4 …. In a similar manner, higher order IMD products exist (e.g., *IM*5 at $3\omega_1 - 2\omega_2$ and $3\omega_2 - 2\omega_1$) due to the higher order series terms. Notice however, that all IMD products are close in frequency to ω_1 and ω_2, fall within the desired frequency band of the circuit, and hence cannot be removed by external filtering. In practice, third order products are often the highest in magnitude and thus of greatest concern, although in some cases fifth and

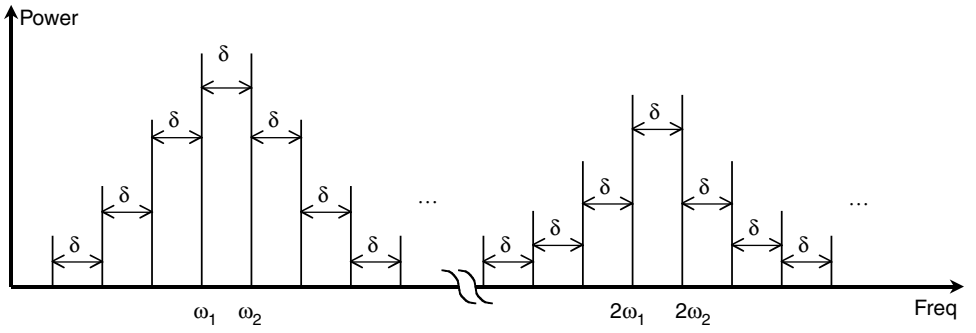


FIGURE 27.2 Resultant spectrum representing the nonlinear amplification of two equal amplitude sinusoidal stimuli at frequencies ω_1 and ω_2 ($\delta = \omega_2 - \omega_1$).

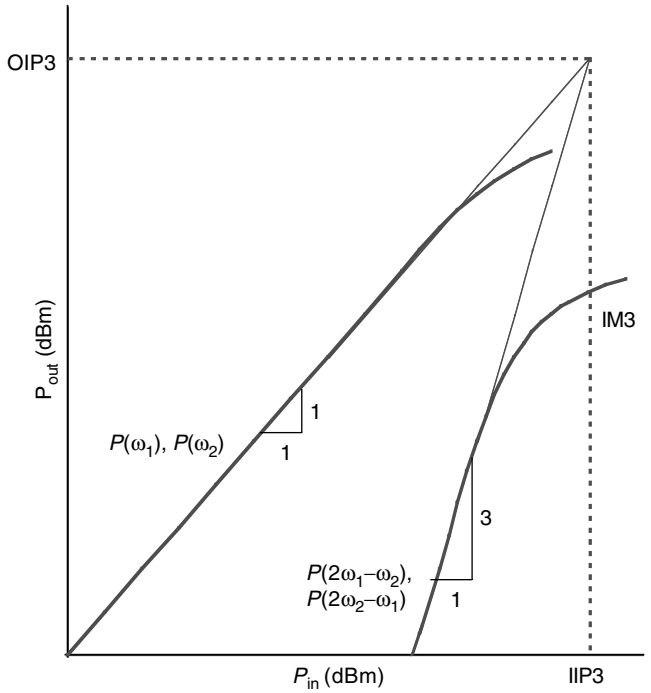


FIGURE 27.3 Expected relationship between fundamental frequency components and third order intermodulation distortion products neglecting effects due to higher order series coefficients.

higher order products may also be of interest. Spectral analysis of the circuit can be greatly simplified if the input signal is assumed small in magnitude such that the dominant contributor to *IM3* is from E_{in}^3.

Intermodulation distortion products (*IM3*, *IM5*,…) can be expressed in terms of power, either absolute (dBm) or relative to the carrier (dBc), or by a fictitious intercept point. For certain circuits where the input stimulus is small in magnitude (e.g., low noise amplifier and certain receiver components), an intercept point distortion specification is useful. Consider the nonlinear response of a circuit represented by Equation 27.6 and driven with an equal amplitude two-tone sinusoidal stimuli as given in Equation 27.7. Assume further than only a_1 and a_3 in Equation 27.6 are nonzero. A plot of the circuit's output power spectrum as a function of input power is illustrated in Figure 27.3. The output spectral tones at ω_1 and ω_2 increase on a 1:1 (dB) basis with input power. The *IM3* products ($2\omega_1 - \omega_2$, $2\omega_2 - \omega_1$) increase on a

3:1 (dB) basis with input power (due to E_{in}). The intersection of the fundamental tones with the third order products is defined as the third order intercept point. Note that the intercept point can be specified relative to input or output power of each tone, *IIP*3 and *OIP*3, respectively. Given this linear relationship, the output intercept point can easily be calculated based on the power of the fundamental and third order terms present at the output port [2,6]

$$OIP3(dBm) = P_{out}(\omega_1) + \frac{P_{out}(\omega_1) - P_{out}(2\omega_1 - \omega_2)}{2} \tag{27.8}$$

where $P_{out}(\omega_1)$ and $P_{out}(2\omega_2 - \omega_1)$ is the power (dBm) in the fundamental and third order products referenced to the output port. Notice that since the input stimuli is an equal amplitude two-tone sinusoidal stimulus, like order spectra products are assumed equal in magnitude (i.e., $P_{out}(\omega_1) = P_{out}(\omega_2)$, $P_{out}(2\omega_2 - \omega_1) = P_{out}(2\omega_1 - \omega_2)$, …).

The relationship between input and output intercept points is given by

$$IIP3(dBm) = OIP3(dBm) - G(dB) \tag{27.9}$$

In a similar manner, the fifth order output intercept point can be defined as [5]

$$OIP5(dBm) = P_{out}(\omega_1) + \frac{P_{out}(\omega_1) - P_{out}(3\omega_1 - 2\omega_2)}{4} \tag{27.10}$$

where $P_{out}(3\omega_2 - 2\omega_1)$ is power (dBm) of fifth order products referenced to the output port.

Similarly,

$$IIP5(dBm) = OIP5(dBm) - G(dB) \tag{27.11}$$

In system analysis, it is often desirable to consider the linearity performance for a series of cascaded two-port circuits and the contribution of each circuit's nonlinearity to the total. The IM3 distortion of the complete cascaded network can be approximated based on the third order intercept point of each element. Consider the two-tone sinusoidal stimulus (Equation 27.7) applied to the input of the cascaded circuits shown in Figure 27.1. The magnitude of the fundamental and IM3 products (i.e., ω_1, ω_2, $2\omega_1 - \omega_2$, and $2\omega_2 - \omega$) at the output port of the 1st element can be calculated based on knowledge of the input power level of tones ω_1 and ω_2, transducer gain, and the third order intercept point of element one using Equation 27.8. Next, consider the second element where the input stimulus now consists of spectral components at $2\omega_1 - \omega_2$ and $2\omega_2 - \omega_1$ in addition to those at ω_1 and ω_2. The IM3 spectral products at the second element's output port will be the result of contributions from two sources, (1) those due to intermodulation distortion of ω_1 and ω_2 in element 2 and (2) those due to amplifying spectral products $2\omega_1 - \omega_2$, and $2\omega_2 - \omega$ present at the input port of element 2. The IM3 products due to the former are, again, calculated from Equation 27.8. The IM3 products at the output of element 2 due to the latter will be the IM3 products at the input amplified by G_2. Hence, the total *IM*3 spectral products are the vector sum from each. Both a minimum and maximum value is possible depending on the vector relationship between the various signals. A worst case (lowest *OIP*3) results when they combine in phase and are given by [2]

$$\frac{1}{IIP3_T} = \frac{1}{IIP3_1} + \frac{G_1}{IIP3_2} + \frac{G_1 G_2}{IIP3_3} \cdots \tag{27.12}$$

with *IIP*3 expressed in Watts.

Or alternatively from [6]

$$OIP3_{T\min} = \left(\sum_{i=1}^{n} \frac{1}{OIP3_i g_i} \right)^{-1} \tag{27.13}$$

with *OIP*3 expressed in Watts and where g_i is the cascaded gain from the output of the i^{th} element to the system output, including impedance mismatch effects. A best case scenario (highest *OIP*3) results when they combine out of phase with the results given by [6]:

$$OIP3_{T\max} = \left(\sum_{i=1}^{n} \frac{1}{OIP3_i^2 g_i^2} - 2 \sum_{i=2}^{n} \sum_{\substack{j=1 \\ i>j}}^{n} \frac{1}{OIP3_i OIP3_j g_i g_j} \right)^{-1/2} \tag{27.14}$$

Hence, Equations 27.13 and 27.14 specify bounds for intercept performance of cascaded networks.

An illustration of the measured spectral content of an amplifier driven with a two-tone sinusoidal stimuli is shown in Figure 27.4. At low power levels, third and fifth order IM products closely follow a 3:1 and 5:1 (dB) relationship with input power. Hence, per the previous discussion, *OIP*3 and *OIP*5 can be calculated based on measurements of the output spectral products at a given input power level. In this example, the spectral content is $P_{out}(3\omega_2 - 2\omega_1) = -87.4$ dBm, $P_{out}(2\omega_2 - \omega_1) = -50.2$ dBm, $P_{out}(\omega_1) = +3.0$ dBm and G = 10.4 dB for the input level shown. Applying Equations 27.8 and 27.10 yield *OIP3* = 29.6 dBm and *OIP5* = 25.6 dBm, respectively. The input intercept points are determined from Equations 27.9 and 27.11.

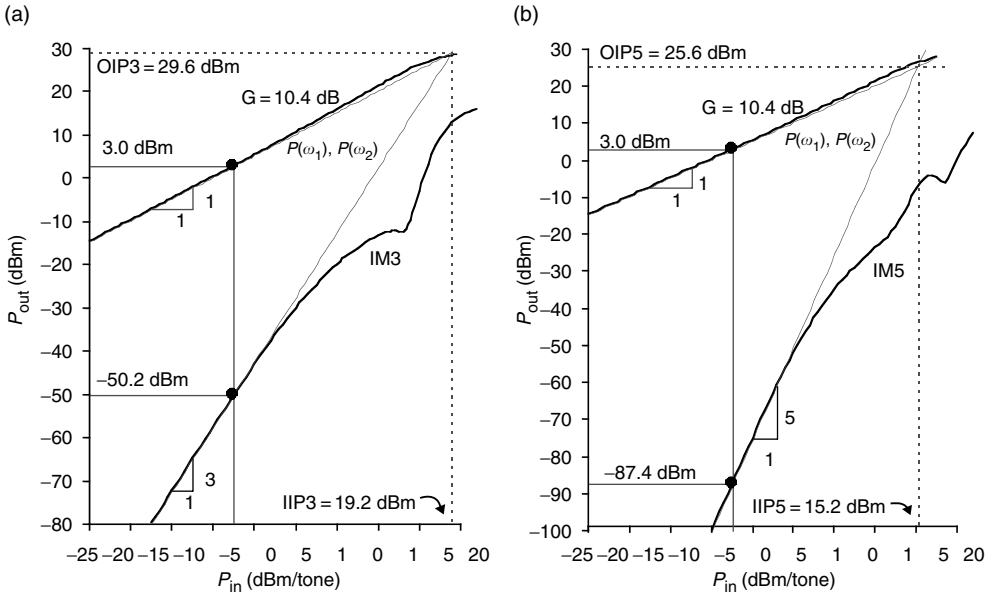


FIGURE 27.4 Typical relationship between fundamental frequency components and 3rd and 5th order intermodulation distortion products including effects due to higher order series coefficients. a) Third order IMD, and b) Fifth order IMD.

Limitations rooted in the approximations in deriving intercept points become more apparent at higher power levels where the relationship between input power and spectral products deviates dramatically from their assumed values. At some increased power level, the effects due to the higher order series coefficients in Equation 27.6 become significant and cannot be ignored. Hence, for certain circuits, such as power amplifiers for example, the concept of intercept point is meaningless. A more meaningful measure of nonlinearity is the relative power in the IMD products (dBc) referenced to the fundamental tones, with the reference generally made to output rather than input power.

27.4 System Simulation with a Digitally Modulated RF Stimuli

Many modern communications systems, including second (2G) and third (3G) generation cellular, satellite communications, and wireless local area networks (WLAN), to name but a few, utilize some form of digital modulation to encode information onto an RF carrier. As discussed earlier in this chapter, these signals are complex in that the RF carrier's phase, amplitude, or both are modulated in some manner to represent digital information. An extensive examination of the mathematical techniques and available methods to simulate the system response of passing such signals through various RF circuits is well beyond the scope of this section. The reader is referred to Reference 1 for a more detailed discussion of simulation techniques. Nevertheless, some of the fundamental RF-related aspects of system simulation will be examined in the context of a mobile wireless radio. Consider the architecture illustrated in Figure 27.5 which is intended to represent major elements of wireless radio such as presently utilized in 2G and 3G cellular systems. The radio utilizes frequency division multiplexing whereby a diplexer confines RX and TX signals to the respective receiver and transmitter paths.

To begin, consider the TX path where digital information is first generated by a DSP. This data is modulated onto an RF carrier whereby the information is encoded and modulated onto a carrier conforming to a particular modulation format. From this point, the signal is injected into a mixer where it is raised in frequency to coincide with the TX frequency band. The signal is then amplified by a power amplifier and passed to the antenna via a diplexer.

Simulation of the TX signal path begins by considering the digital information present at the modulator. In a cellular system, this information corresponds to a digital sequence representing voice information and/or data. The precise structure of the sequence (i.e., patterns of zeros and ones) is important in that it is a major contributor in defining the envelope characteristics of the RF signal to be transmitted. Also note for simulation purposes, the RF stimulus is generally formed by repeating this sequence and modulating it onto an RF carrier. Hence, the resultant RF signal is periodic per the digital bit sequence. The effect of a particular digital bit sequence in defining the RF signal is illustrated by considering the

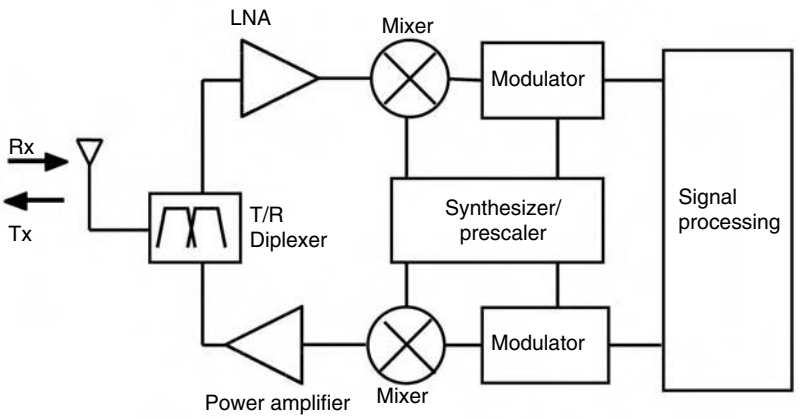


FIGURE 27.5 Block diagram representing major elements in mobile cellular radio.

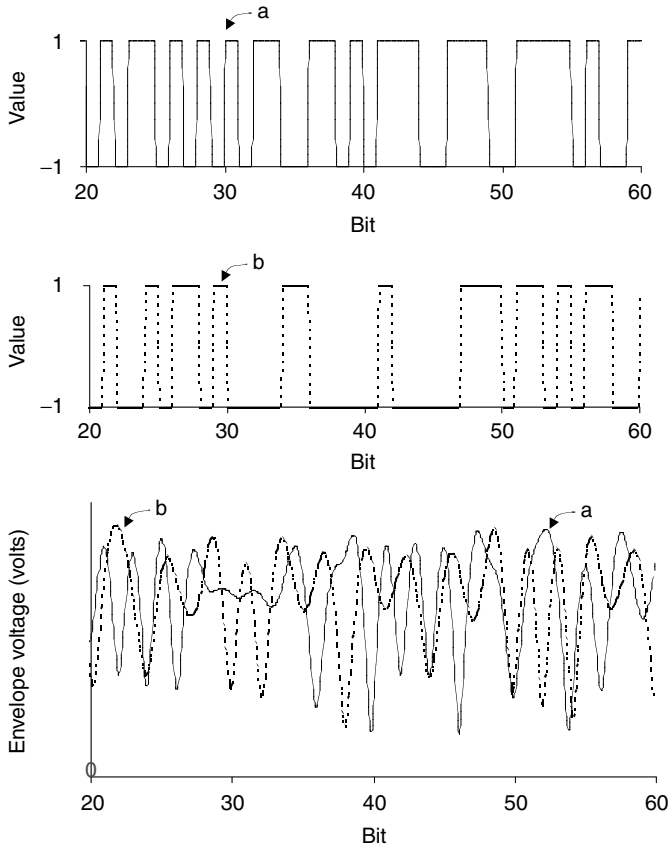


FIGURE 27.6 The RF modulated envelope formed by considering two randomly generated bit sequences.

two randomly generated NRZ bit patterns shown in Figure 27.6. The amplitude modulated RF envelope voltage developed by utilizing these sequences in a $\pi/4$ DQPSK modulator is also shown in Figure 27.6. While the two envelope signals are nearly identical in their average power levels, they are substantially different, especially in their peak voltage excursions. Hence, the digital bit sequence and the resultant modulated RF waveform can be particularly important when evaluating the performance of nonlinear circuits such as power amplifiers in that the bit sequence can affect spectral distortion. Nevertheless for simulation purposes, it is necessary to choose a suitable sequence to represent the digital information to be transmitted. In general, either a predefined binary NRZ sequence, a randomly generated one, or a pseudo-noise (PN) sequence is generally chosen. Often the latter is considered a more desirable choice due to its statistical properties. For example, a maximal length PN sequence of length 2^m-1 contains all but one m bit combinations of 1s and 0s. This property is particularly important in that it allows all possible bit patterns (except for the all zeros pattern) to be utilized in generating the RF modulated waveform. In contrast, a much longer random sequence would be needed with no guarantee of this property. Further, the autocorrelation function of a PN sequence is similar to a random one [1]. A potentially significant disadvantage of applying a random sequence in evaluating nonlinear circuit blocks is that the simulation results will change from evaluation to evaluation as the randomly generated sequence is not identical for each simulation.

Once a sequence is chosen, the performance of the modulator can be evaluated. The modulator can be modeled and simulated at the component level using time-based methods such as SPICE. However, in the early system design phase, such detail and the time required to perform a full circuit level simulation

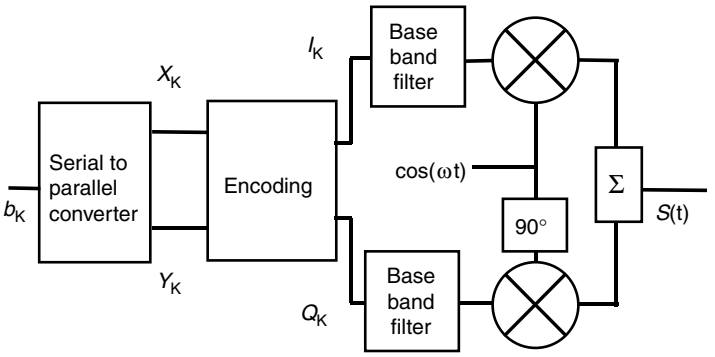


FIGURE 27.7 The process of generating modulated signal $S(t)$ in a QPSK system.

may be unattractive. On the other hand, it may be more appropriate to model the modulator at a higher level (e.g., behavioral model) and only consider selected first order effects. For example, the simplified diagram in Figure 27.7 depicts the functionality of a quadrature modulator (in this case to represent a QPSK modulator). Starting with a data steam, every two bits are grouped into separate binary streams representing even and odd bits as indicated by X_K and Y_K. These signals (X_K and Y_K) are encoded in some manner (e.g., as relative changes in phase for IS-136 cellular) and are now represented as I_K and Q_K with each symbol extending over two bit time intervals. These signals now pass though a baseband filter, often implemented as a finite impulse response filter with impulse response h(t). The filtered signal can be calculated based upon the convolution theorem for time sampled signals. These signals are then modulated onto a carrier (IF) with the output modulated signal taking the form

$$S(t) = \sum_n g(t-nT)\cos(\Phi_n)\cos(\omega_c t) - \sum_n g(t-nT)\sin(\Phi_n)\sin(\omega_c t) \tag{27.15}$$

where ω_c is the radian carrier frequency, Φ_n represents phase, $g(t)$ is a pulse shaping factor, and n = 0, 1, 2, … are discrete time samples.

At this point, some first order effects can be evaluated by considering the mathematical nature of Equation 27.15. For example, phase imbalance (Φ_{imb}) within the modulator can be modeled as:

$$S(t) = \sum_n g(t-nT)\cos(\Phi_n)\cos(\omega_c t) - \sum_n g(t-nT)\sin(\Phi_n)\sin(\omega_c t + \Phi_{imb}) \tag{27.16}$$

Given a higher level model, the modulated envelope can be simulated using time-based methods to determine $S(t)$.

The power amplifier represents an element in the transmitter chain where linearity is of concern, especially in those RF systems employing modulation methods resulting in a nonconstant amplitude envelope. These cases, which incidentally include a number of cellular and PCS systems, require a linear power amplifier to preserve the amplitude/phase characteristics of the signal. The nonlinear characteristics of the amplifier can be simulated at the circuit level using time- and/or frequency-based methods. However, circuit-based simulations require accurate and detailed knowledge of all circuit components within the amplifier, including an appropriate large signal model for all active devices as well as highly accurate models for all passive structures. Such knowledge is often unavailable at the system level with sufficient detail and accuracy to perform such a simulation. In addition, circuit level nonlinear simulations of the amplifier driven by digitally modulated RF stimulus are generally quite computationally intensive resulting in long simulation times, making this approach even more unattractive.

A more common approach to modeling the nonlinearity of a power amplifier at the system level is through the use of behavioral models [1,7,8]. While a number of behavioral models have been proposed with varying levels of complexity and sophistication, all of them rely to some extent on certain approximations regarding the circuit nonlinearity. A common assumption in many of the behavioral models is that the nonlinear circuit/RF-modulated stimulus can be represented in terms of a memoryless and bandpass nonlinearity [1]. Although the active device within the amplifier generally exhibits some memory behavior, and additional memory-like effects can be caused by bias networks, these assumptions are generally not too limiting for power amplifiers utilized in cellular communications systems. Further, when the above-noted assumptions are met or nearly met, the simulation results are very accurate and the needed simulation time is very short.

In general, an input–output complex envelope voltage relationship is assumed with the complex output envelope voltage $v_{out}(t)$ taking the form

$$v_{out}(t) = RE\left\{G(V(t))e^{j\left\{\Phi(t)+\varphi(V(t))+\omega_c t\right\}}\right\} \tag{27.17}$$

where $G(V(t))$ and $\varphi(V(t))$ describe the instantaneous input–output envelope voltage gain and phase. Note that functions $G(V)$ and $\varphi(V)$ represent the amplifier's AM–AM and AM–PM response, respectively. The term ω_c represents the carrier frequency.

An inspection of Equation 27.17 suggests the output envelope voltage can be calculated in a time-based method by selecting time samples with respect to the modulation rate rather than at the RF carrier frequency. This feature is particularly advantageous in digitally modulated systems where the bit rate and modulation bandwidth are small in comparison to the carrier frequency. Significantly long and arbitrary bit sequences can be simulated very quickly since the time steps are at the envelope rate. For example, consider an NRZ bit sequence on the order of several mS which is filtered at baseband and modulated onto an RF carrier with a 1nS period (i.e., 1 GHz). Time-based simulation at the RF frequency would likely require time samples significantly less than 0.1 nS and the overall number of sample points would easily exceed 10^7. Alternatively, simulating the output envelope voltage by choosing time steps relative to the modulation rate would result in several orders of magnitude fewer time samples.

A particular advantage of the above model is that the entire power amplifier nonlinearity is described in terms of its AM–AM (gain) and AM–PM (insertion phase) response. The AM–AM response is equivalent to RF gain and the AM–PM characteristics are equivalent to the insertion phase — both measured from input to output ports with each expressed as a function of input RF power. For simplicity, both characteristics are often determined using a single tone CW stimulus, although modulated RF signals can be used at the expense of more complexity. The nature of the behavioral model allows it to be characterized based on the results from another simulator (e.g., harmonic balance simulations of the amplifier circuit), an idealized or assumed response represented in terms of mathematical functions describing am-am and am-pm characteristics, or from measurements made on an actual amplifier.

Computation of the output envelope voltage from Equation 27.17 allows evaluation of many critical system characteristics, including adjacent and/or alternate channel power ratio (ACPR), error vector magnitude (EVM), and bit error rate (BER) to name but a few. For example, using Fourier methods, the spectral properties of the output signal expressed in Equation 27.17 can be calculated. Nonlinear distortion mechanisms are manifested in spectral regrowth or equivalently a widening of the spectral properties. This regrowth is quantified by figure of merit ACPR, which is defined as a ratio of power in the adjacent (or in some cases alternative) transmit channel to that in the main channel. In some systems (e.g., IS-136), the measurement is reference to the receiver (i.e., the transmitted signal must be down-converted in frequency and filtered at base band). An illustration of spectral distortion for a $\pi/4$ DQPSK stimulus compliant to the IS-136 cellular standard is shown in Figure 27.8. In this system the channel bandwidth is 30 KHz. Hence, using the specified root raised cosine baseband filter with excess bandwidth

ratio $\alpha = 0.35$, a 24.3 KS/s data sequence will result in a non-distorted modulated RF stimulus band limited to 32.805 KHz (i.e., 1.35 * 24.3 KS/s) as illustrated in Figure 27.8. Nonlinear distortion mechanisms cause regrowth as illustrated. Note the substantial increase in power in adjacent and alternate channels.

Another measure of nonlinear distortions involves examining the envelope characteristics of $S(t)$ in time as measured by the receiver. Consider an input $\pi/4$ DPSK modulated RF signal which is developed from a 256 length randomly selected NRZ symbol sequence and passed through both linear and nonlinear amplifiers. Figure 27.9 illustrates the constellation diagram for both signals as measured at the receiver. Note that both signals represent the I and Q components of the envelope where $S(t)$ has been passed through a root-raised-cosine receiver filter. The constellation plot only shows the values of the I and Q

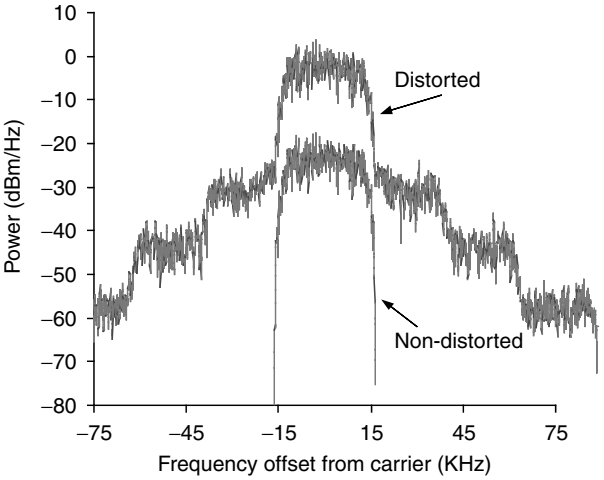


FIGURE 27.8 Nonlinear distortions due to a power amplifier result in a widening of the spectral properties of the RF signal.

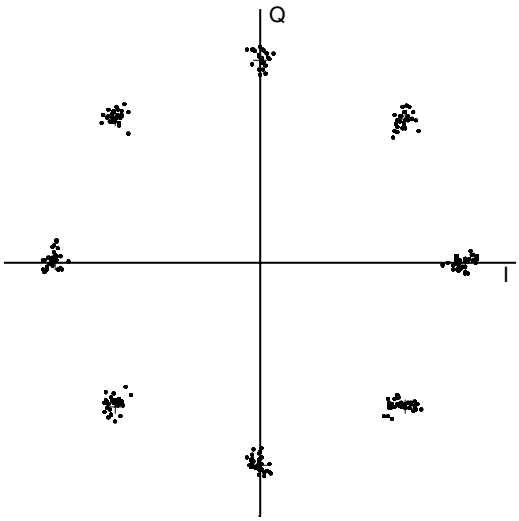


FIGURE 27.9 Constellation plot for both a non-distorted and distorted $\pi/4$ DPSK stimulus.

components at the appropriate time sample intervals. The non-distorted signal exhibits expected values of ± 1, $\pm\sqrt{1/2} \pm i\sqrt{1/2}$, and $\pm i$. The distorted signal is in error from these expected values in terms of both amplitude and phase.

References

1. M.C. Jeruchim, P. Balaban, and K.S. Shanmugan, *Simulation of Communication Systems,* Plenum Press, New York, 1992.
2. J. Tsui, *Microwave Receivers and Related Components,* Air Force Avionics Lab., Wright-Patterson AFB, OH, 1983.
3. G. Gonzalez, *Microwave Transistor Amplifiers,* Prentice-Hall, Inc., Englewood Cliffs, N.J., 1984.
4. J. Sevice, Nonlinear Analysis Methods for the Simulation of Digital Wireless Communication Systems, *International Journal of Microwave and Millimeter-Wave Computer-Aided Engineering,* 6, 1997, 197–216.
5. H. Xiao, Q. Wu, F. Li, Measure a Power Amplifier's Fifth-Order Intercept Point, *RF Design,* April 1999, 54–56.
6. N. Kanaglekar, R. Mcintosh, and W. Bryant, Analysis of Two-Tone, Third-Order Distortion in Cascaded Two-Ports, *IEEE Transactions on Microwave Theory and Techniques,* April 1988, 701–705.
7. System Theory of Operation, OmniSys Manual, Hewlett-Packard Co.
8. J. Staudinger, Applying the Quadrature Modeling Technique to Wireless Power Amplifiers, *Microwave Journal,* Nov. 1997, 66–86.

28

Numerical Techniques for the Analysis and Design of RF/Microwave Structures

Manos M. Tentzeris
Georgia Institute of Technology

28.1 Introduction

Recent advances in wireless and microwave communication systems (higher operation frequency bands, more compact topologies containing MMICs and MEMS) have increased the necessity of fast and accurate numerical simulation techniques [1–20]. Unlike hybrid microwave integrated circuits at low frequencies, it is extremely difficult and essentially impossible to adjust the circuit and radiation characteristics of communication modules once they are fabricated. The starting point for the development of efficient numerical algorithms is an accurate characterization of the passive and active structures involved in the topologies. Although most commercial CAD programs are based on curve-fitting formulas and lookup tables and not on accurate numerical characterization, the latter can be used if it is fast enough. In addition, it can be used to generate lookup tables and to check the accuracy of empirical formulas.

Any numerical method for characterization needs to be as efficient and economical as possible in both CPU time and temporary storage requirement, although recent rapid advances in computers impose less severe restrictions on the efficiency and economy of the method. Another important aspect in the development of numerical methods has been the versatility of the method. In reality, however, numerical methods are chosen on the basis of trade-offs between accuracy, speed, storage requirements, versatility, and so forth and are often structure dependent. Among these techniques, the most popular ones include the method of moments (MoM), integral equation based techniques, mode-matching (MM), the finite-difference time-domain (FDTD), the transmission line matrix method (TLM), and the finite element method (FEM).

28.2 Integral Equation-Based Techniques

28.2.1 Method of Moments—Integral Equation

The term "method of moments" was introduced in electromagnetics by R. Harrington [3] in 1968 in order to specify a certain general method for reducing linear operator equations to finite matrix solutions [4–7]. The MoM computations invariably reduce the physical problem, specified by the Maxwell's equations and the boundary conditions, into integral equations having finite, and preferably small, domains. In this small domain, the discretization is performed through the expansion of unknowns as a series of basis functions. An example is the magnetic field integral equation (MFIE) for the scattering of a perfectly conducting body illuminated by an incident field H^{i} [21],

$$\hat{n} \times H(r) = 2\hat{n} \times H^{\mathrm{i}}(r) + 2\hat{n} \times \int_s [\hat{n}^r \times H(r')] \times \nabla' G(r, r') \mathrm{d}s' \quad \text{on } S \tag{28.1}$$

where H^{i} is defined as the field due to the source in the absence of the scattering body S, and

$$G(r, r') = \frac{\mathrm{e}^{-jk|r-r'|}}{4\pi |r - r'|} \tag{28.2}$$

where r and r' are the position vectors for the field and source positions, respectively. A continuous integral, such as the one above, can be written in an abbreviated form as

$$L(f) = g \tag{28.3}$$

where f denotes the unknown, which is H above, and g denotes the given excitation, which is H^{i}. Also, L is a linear operator. Let f be expanded in a series of functions $f_1, f_2, f_3, \ldots$ in the domain S of L, as

$$f = \sum_n a_n f_n \tag{28.4}$$

where a_n are constants and f_n are called expansion (basis) functions. For exact solutions, the above summation is infinite and f_n forms a complete set of basis functions. For approximate solutions, this solution is truncated to

$$\sum_n a_n L(f_n) = g \tag{28.5}$$

Assume that a suitable inner product $\langle f, g \rangle = \int f(x) g(x) \mathrm{d}x$ has been determined for the problem. Defining a set of weighting (testing) functions, $w_1, w_2, w_3, \ldots$ in the range of L, and taking the inner

product of the above summation with each w_m leads to the result

$$\sum_n a_n \langle \omega_m, L(f_n) \rangle = \langle \omega_m, g \rangle, \quad m = 1, 2, 3, \ldots \tag{28.6}$$

which can be written in matrix form as

$$[l_{mn}][a_n] = [g_m] \tag{28.7}$$

where

$$[l_{mn}] = \begin{bmatrix} \langle \omega_1, L(f_1) \rangle & \langle \omega_1, L(f_2) \rangle & \cdots \\ \langle \omega_2, L(f_1) \rangle & \langle \omega_2, L(f_2) \rangle & \cdots \\ \cdots & \cdots & \cdots \end{bmatrix}, \quad [a_n] = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \end{bmatrix}, \quad [g_m] = \begin{bmatrix} \langle w_1, g \rangle \\ \langle w_2, g \rangle \\ \vdots \end{bmatrix} \tag{28.8}$$

If the matrix $[l]$ is nonsingular, its inverse $[l]^{-1}$ exists and the a_n are given by

$$[a_n] = [l_{mn}]^{-1}[g_m] \tag{28.9}$$

and the unknown f is given from the weighted summation Equation 28.4. Assuming that the finite expansion basis is defined by $[\tilde{f}_n] = [f_1 f_2 f_3 \cdots]$, the approximate solution for f is

$$f = [\tilde{f}_n][a_n] = [\tilde{f}_n][l_{mn}]^{-1}[g_m] \tag{28.10}$$

Depending upon the choice of f_n and w_n this solution could be exact or approximate [22]. The most important aspect of MoM is the choice of expansion and testing functions. The f_n should be linearly independent and chosen such that a finite-term superposition approximates f quite well. The w_n should also be linearly independent. In addition, this choice is affected by the size of the matrix that has to be inverted (should be minimal), the ease of evaluation of the matrix elements, the accuracy of the desired solution, and the realization of a well-conditioned matrix $[l]$. The special choice $w_n = f_n$ gives Galerkin's method. The two most popular subsectional bases are the pulse function (step approximation) and the triangle function (piecewise linear approximation), as shown in Figure 28.1. The numerical Gaussian quadrature rule [7] is used when the integrations involved in the evaluation of $l_{mn} = \langle w_m, Lf_n \rangle$ are

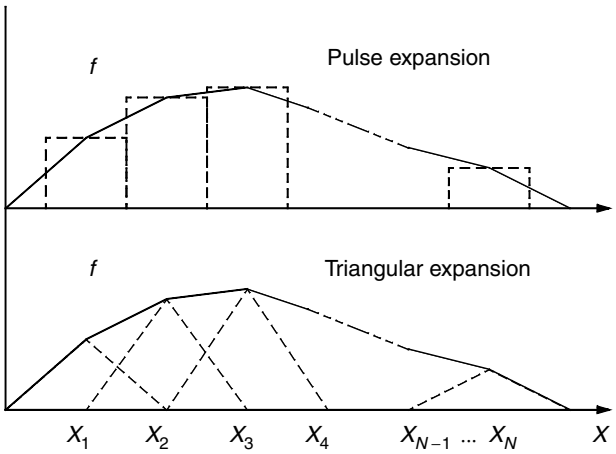


FIGURE 28.1 Moment method expansion in pulse and triangular basis.

difficult to perform for common w_m's for a specific problem or when a more complex expansion basis is used. An alternative approach makes use of the Dirac delta functions for testing. This technique is called point-matching and effectively satisfies Equation 28.2 at discrete points in the region of interest. When the basis functions f_n exist only over subsections of the domain of f, each affects the approximation of f only over a subsection of the region of interest (method of subsections). In this way, the evaluation of l_{mn} is simplified and the form of the matrix $[l]$ is stripped and easier to invert.

The MoM involves setting up and solving dense, full, complex-valued systems. For a one-dimensional (1D) structure (e.g., a wire) of length L wavelengths, the size of the matrix is typically on the value of 10 L; for a three-dimensional (3D) structure with surface area S square wavelengths, the size is, typically, of the order of 100 S. Consequently, there are applications such as radar scattering from aircraft when the system of equations has order in excess of one million. Though MoM has been proven to be a very robust technique, it is plagued by the significant computational burden of having to solve a dense matrix equation of large order. In addition, modeling of a new structure requires the reformulation of the integral equation, a task that may require the very difficult derivation of a geometry-specific Green's function. The MoM is used for the solution of various integral equations such as the electric field integral equation (EFIE) and the MFIE. The integral equation approach has the advantage that it rigorously takes into account the radiation conditions of any open structure and therefore it is not necessary to implement absorbing boundary conditions. The kernel of the integral equation is the Green's function that accurately describes all possible wave propagation effects, such as radiation, leakage, anisotropy, stratification, and so on.

The MoM has been used in many scattering problems [23–25], microstrip lines on multilayered dielectrics [26], microstrip antennas [27,28], integrated waveguides in microwave [29,30], and optical frequencies [31] even on anisotropical substrates [32]. In addition, results have been derived for the characterization of junctions [33], high-speed interconnects [34], viaholes [35], couplers [36], and infinite aperture arrays [37]. It should be emphasized that the discretization may be nonuniform, something that has been demonstrated in the application of MoM to planar circuits in Reference 38.

Chew [39] introduced a new method to precondition the matrix equation resulting from applying MoM to EFIE. This method leads to dramatic reductions in iteration count and allows for the use of fast solvers such as the low-frequency multilevel fast multipole algorithm (LF–MLFMA) [40]. In addition, Brown [41] suggested one computationally effective way to enable the efficient large-domain MoM solutions to electrically large practical EM problems. Johnson [42] used genetic algorithms to facilitate the MoM modeling of integrated antennas elements and arrays. Recently, the MoM has been used for the accurate modeling and design of smart antennas [43], 3D multilayer RF geometries [44,45] and electromagnetic band-gaps (EBG's) [46].

In Reference 47, novel radial basis function-neural network (RBF–NN) models were presented for the efficient filling of the majority of the coupling matrix of the MoM.The proposed neural network method of moments (NN–MoM) was applied to the analysis of a number of microstrip patch antenna arrays. References 48 and 49 present the multiple sweep method of moments (MSMM) analysis of the radiation and scattering from electrically large, smooth, perfectly conducting bodies, such as wideband 3D antennas. The MSMM is an $O(N^2)$ recursive method for solving the large matrix equations, which arise in the MoM analysis of electrically large bodies. In the MSMM, the body is split into P sections and the currents on these sections are found in a recursive fashion. The first MSMM recursion or sweep includes the dominant scattering mechanisms and each subsequent recursion or sweep includes higher order mechanisms. Thus, as the electrical size of the body increases, fewer and fewer sweeps are required to accurately compute the current on the body.

In Reference 50, the adaptive cross approximation (ACA) algorithm is presented as a generic effective approach to reduce memory and CPU time overhead in the MoM application to electromagnetic compatibility-related problems of moderate electrical size. Specifically, the ACA algorithm is used to study compact-range ground planes and electromagnetic interference and shielding in vehicles. Through numerical experiments, it is concluded that for moderate electrical size problems the memory and CPU time requirements for the ACA algorithm scale as N/sup 4/3/log N. The periodic method of moments

(PMMs) has been applied to frequency-selective surfaces (FSSs) in Reference 51 with results presented for single-layer FSS with asymmetrically split segmented rings. Last, the impedance boundary MoM method has been utilized for the analysis of quantum wells (QWs) and quantum wires with arbitrary potential profiles [52]. However, both EFIE and MFIE suffer from nonunique solutions at resonant frequencies. To alleviate this problem, the combined field integral equation (CFIE) has been utilized leading to the fast multipole method (FMM). The basic principle behind FMM is to decompose the computation of matrix vector products into two parts: one involving the interaction between nearby sources and the other involving those between well-separated ones. The technique employs a recursive subdivision of the spatial domain and a corresponding hierarchical tree structure. For this, the sources are first enclosed in a box. Then the box is divided into equal children. Each child is then recursively divided into smaller and smaller boxes and, thus, a tree structure is established. The FMM technique has been parallelized and used in large-scale problems [53], in time-domain PEEC analysis [54], in magnetic simulations [55], in far/near-field transformation problems [56], in scattering problems [57], while there have been publications on distributed-memory multilevel implementations of the algorithm [58].

28.2.2 Spectral Domain Approach

Electromagnetic analysis in the spectral (Fourier transform) domain is preferable to many spatial domain numerical techniques, especially for planar transmission lines, microstrip antennas, and other planar layered structures. If applied to an equation describing EM fields in planar media, a Fourier transform reduces a fully coupled, 3D equation to a 1D equation that depends on two parameters (the transform variables), but is uncoupled and can be solved independently at any value of those parameters. After solving the 1D equation, the inverse Fourier transform performs a superposition of the uncoupled 1D solutions to obtain the 3D solution. Thus, the process offers substantial savings since it effectively converts a 3D problem into a series of 1D problems.

Yamashita and Mittra [59] introduced the Fourier domain analysis for the calculation of the phase velocity and the characteristic impedance of a microstrip line. Using the quasi-TEM approximation (negligible longitudinal E- and H-fields), the line capacitance is determined by the assumed charge density through the application of the Fourier domain variational method. Denlinger [60] extended this approach to the full wave analysis of the same line. However, his solution depends strongly on the assumed current strip distributions. Itoh and Mittra [61] introduced a new technique, the spectral domain approach (SDA), that allows for the systematic improvement of the solution accuracy to a desired degree. In SDA, the transformed equations are discretized using MoM, yielding a homogeneous system of equations to determine the propagation constant and the amplitude of current distributions from which the characteristic impedance is derived.

For metallic strip problems, the Fourier transform is performed along the direction parallel to the substrate interface and perpendicular to the strip. The first step is the formulation of the integral equation that correlates the E-field and the current distribution J along the strip and the application of the boundary conditions for E- and H-fields. Then, the Fourier transform is applied over E and J and the MoM technique produces a system of algebraic equations that can be solved. Different choices of expansion and testing functions have been discussed in Reference 62. The SDA is applicable to most planar transmission lines (microstrips, finlines, CPWs) [63–68], microstrip antennas and arrays [69,70], interconnects [71], junctions [72], dielectric waveguides [73], resonators of planar configurations [74], and embedded passives [75] and micromachined devices [76] on single or multilayered [77] dielectric substrates including conductors with finite conductivity [78]. This method requires significant analytical preprocessing, something that improves its numerical efficiency, but also restricts its applicability, especially, for structures with finite conductivity strips and arbitrary dielectric discontinuities.

In addition, SDA has been applied to integrated lasers [79], cylindrical coplanar waveguides [80], microstrip leaky-wave antennas [81], planar and quasi-3D circuits in stratified media [82], and boxed planar lines when complex materials (anisotropic dielectrics, ferrites, magnetoplasmons, chiral media, and so on) are used as substrates [83]. A nonuniform fast Fourier transform (NUFFT) technique has

been incorporated into the SDA for the analysis of shielded single and multiple coupled microstrip lines and allows for the accurate investigation of the dispersion characteristics of coupled lines with finite metallization thickness and of coupled lines at different levels [84].

28.2.3 Mode Matching Technique

This technique is usually applied to the analysis of waveguiding/packaging discontinuities that involve numerous field modes. The fields on both sides of the discontinuity are expanded in summations of the modes in the respective regions with unknown coefficients [85] and appropriate continuity conditions are imposed at the interface to yield a system of equations. As an example, to analyze a waveguide step discontinuity with TE_{n0} excitation, E_y and H_x fields are written as the superposition of the modal functions $\phi_{an}(x)$ and $\phi_{bn}(x)$ for $n = 1, 2, \ldots$, respectively for the left waveguide (waveguide A) and the right waveguide (waveguide B), as it is displayed in Figure 28.2. Both of these fields should be continuous at the interface $z = 0$. Thus,

$$\sum_{n=1}^{\infty}(A_n^+ + A_n^-)\,\phi_{an} = \begin{cases} \sum_{n=1}^{\infty}(B_n^+ + B_n^-)\phi_{bn}, & 0 < x < b \\ 0, & b < x < a \end{cases} \tag{28.11}$$

$$\sum_{n=1}^{\infty}(A_n^+ - A_n^-)\,Y_{an}\phi_{an} = \begin{cases} \sum_{n=1}^{\infty}(B_n^+ - B_n^-)\,Y_{bn}\phi_{bn}, & 0 < x < b \\ 0, & b < x < a \end{cases} \tag{28.12}$$

where $(+)$ and $(-)$ indicate the modal waves propagating to the positive and negative z direction and Y_{an}, Y_{bn} are the mode impedances. Sampling these equations with ϕ_{bm} (m-mode in waveguide B) and making use of the mode orthogonality in waveguide B,

$$\sum_{n=1}^{\infty} H_{nm}(A_n^+ + A_n^-) = B_m^+ + B_m^- \tag{28.13}$$

$$\sum_{n=1}^{\infty} H_{nm}\,Y_{an}(A_n^+ - A_n^-) = Y_{bm}(B_m^+ - B_m^-) \tag{28.14}$$

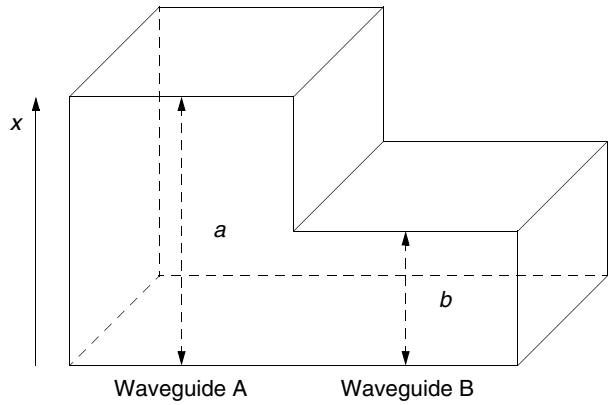


FIGURE 28.2 Rectangular waveguide discontinuity.

with $H_{nm} = \int_0^b \phi_{an}(x)\phi_{bm}(x)\mathrm{d}x$. Similarly, sampling Equations 28.11 and 28.12 with ϕ_{am} (m-mode in waveguide A) and making use of mode orthogonality in waveguide A, we have

$$A_m^+ + A_m^- = \sum_{n=1}^{\infty} H_{mn}(B_n^+ + B_n^-) \tag{28.15}$$

$$Y_{am}(A_m^+ - A_m^-) = \sum_{n=1}^{\infty} H_{mn} Y_{an}(B_n^+ - B_n^-) \tag{28.16}$$

Assuming that the structure is excited through A_n^+ terms, the calculation of A_n^- (reflected modes in A) and B_m^+ (transmitted modes in B) will provide the scattering parameters for the analyzed structure through a procedure that involves matrix inversions.

The foundation of this technique is the expansion of an electromagnetic field in terms of an infinite series of normal modes. Because a computer's capacity for numerical calculation is finite, these summations have to be truncated, something that could lead to incorrect solutions if not performed efficiently. The main criterion for this truncation is the convergence of the summation. A natural way to check it is to plot the numerical values of some desired parameters versus the number of retained terms. The truncation is considered sufficient when the change in the parameters is smaller than prespecified criteria. The procedure becomes more complicated where there is a need for the truncation of two or more infinite series (bifurcated waveguide, step junction). The numerical results appear to converge to different values depending on the manner of the truncation, a phenomenon that is called relative convergence. It has been found that the relative convergence is related to the violation of field distributions at the edge of a conductor at the boundary [9] and to the ill-conditioned situation of the linear system of the computation process [85,86]. Thus, either the edge condition or the condition number of the linear system can be used as a criterion to ensure the validity of modal analyses. Another common criterion is to plot the field distributions on both sides of the boundary and observe their matching conditions. The mode matching method has been applied to analyze various discontinuities in waveguides with rectangular or circular cross-sections [87–90], finlines [91,92], microstrip lines [93,94], and coplanar waveguides [95,96]. In addition, it has been used for closed-region scattering geometries involving a discrete set of modes, such as E-plane filters [97,98], waveguide impedance transformers [99], power dividers [100], and microstrip filters [101]. Moreover, this technique has been implemented for the solution of eigenvalue problems, such as the resonant frequency of a cavity [102] and the performance of evanescent mode filters [103], since it can efficiently model both evanescent and propagating modes.

Recently, reduced-order optimized mode matching was used for the design of microwave waveguide components [104]. In addition, the MM technique was used in the design of large aperture waveguide coupling [105], in the analysis of infinite array and waveguide problems [106], in the modeling of circular microresonators [107], in EMC problems [108], in the analysis of bent waveguides [109], EBG structures [110], photonic crystal microcavity filters [111], horn with superquadric apertures [112], ferrite-filled waveguides [113], and photonic crystal slabs [114]. Also, this technique has been employed for the characterization of the coupling between neighboring microwave components [115] and for the modeling of metafilms [116], of left-handed waveguides [117], of metamaterials [118], of FSS/EBG structures [119], of patch antennas on arbitrary shaped multilayer structures [120], as well as for the full-wave design and optimization of canonical waveguide filters [121], spatial power combiners/dividers with horns [122], terahertz antennas [123], micromachined gyroscopes [124] and for the calculation of the scattering from chiral periodic structures [125], as well as for the characterization of the human exposure to RF radiation [126].

28.3 PDE-Based Techniques FDTD, TLM, MRTD, FEM

In contrast to the previous techniques, numerical methods based on the partial differential equation (PDE) solutions of Maxwell's equations yield either sparse matrices (frequency domain, finite-element

methods) or no matrices at all (time domain, finite-difference, or finite-volume methods). In addition, specifying a new geometry is reduced to a problem of mesh generation only. Thus, PDE solvers could provide a framework for a space/time (frequency) microscope permitting the EM designer to visualize with submicron/sub-picosecond resolution the dynamics of electromagnetic wave phenomena propagating at light speed within proposed geometries.

28.3.1 Finite-Difference Time Domain Technique

The FDTD [10–13] is an explicit solution method for Maxwell's time-dependent curl equations. It is based upon volumetric sampling of the electric and magnetic field distribution within and around the structure of interest over a period of time. The sampling is set below the Nyquist limit and typically more than 10 samples per wavelength are required. The time step has to satisfy a stability condition. For simulations of open geometries, absorbing boundary conditions (ABCs) are employed at the outer grid truncation planes in order to reduce spurious numerical reflections from the grid termination.

In 1966, Yee [127] suggested the solution of the first-order Maxwell equations in time and space instead of solving the second-order wave equation. In this way, the solution is more robust and more accurate for a wider class of structures. In Yee's discretization cell, E- and H-fields are interlaced by half space and time gridding steps, as shown in Figure 28.3. The spatial displacement is very useful in specifying field boundary conditions and singularities and leads to finite-difference expressions for the space derivatives that are central in nature and second-order accurate. The time displacement (leapfrog) is fully explicit, completely avoiding the problems involved with simultaneous equations and matrix inversion. The resulting scheme is nondissipative; numerical wave modes propagating in the mesh do not spuriously decay due to a nonphysical artifact of the time-stepping algorithm. Denoting any function u of space and time evaluated at a discrete point in the grid and at a discrete point in time as $u(i\Delta x, j\Delta y, k\Delta z, l\Delta t) = {}_l u_{i,j,k}$, where Δt is the time step and Δx, Δy, Δz the cell size along the x-, y-, and z-directions, the first partial space derivative of u in the x-direction and the first time derivative of u are approximated with the following central differences, respectively

$$\frac{\partial u}{\partial x}(i\Delta x, j\Delta y, k\Delta z, l\Delta t) = \frac{{}_l u_{i+1/2,j,k} - {}_l u_{i-1/2,j,k}}{\Delta x} + O[(\Delta x)^2]$$
$$\frac{\partial u}{\partial x}(i\Delta x, j\Delta y, k\Delta z, l\Delta t) = \frac{{}_{l+1/2} u_{i,j,k} - {}_{l-1/2} u_{i,j,k}}{\Delta x} + O[(\Delta x)^2] \tag{28.17}$$

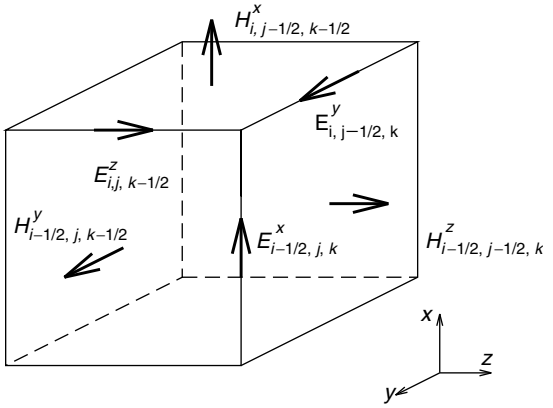


FIGURE 28.3 FDTD Yee cell.

Applying the above notation, the FDTD equations are derived for all field components. For example,

$$
{}^{l+0.5}H^{x}_{i,j-0.5,k-0.5} = \left(\frac{1-\left(\rho'_{i,j,k}\Delta t/2\mu_{i,j,k}\right)}{1+\left(\rho'_{i,j,k}\Delta t/2\mu_{i,j,k}\right)}\right) {}^{t-0.5}H^{x}_{i,j-0.5,k-0.5}
$$
$$
+\left(\frac{(\Delta t/\mu_{i,j,k})}{1+\left(\rho'_{i,j,k}\Delta t/2\mu_{i,j,k}\right)}\right)\left(\frac{{}^{l}E^{y}_{i,j-0.5,k}-{}^{l}E^{y}_{i,j-0.5,k-1}}{\Delta z}-\frac{{}^{l}E^{z}_{i,j,k-0.5}-{}^{l}E^{z}_{i,j-1,k-0.5}}{\Delta y}\right) \tag{28.18}
$$

where $\sigma_{i,j,k}$ and $\rho'_{i,j,k}$ are the electric and the magnetic loss coefficient for the (i, j, k)-cell. It can be observed that a new value of a field vector component at any space lattice point depends only on its previous value and the previous values of the components of the other field vectors at adjacent points. Therefore, at any given time step, the value of a field vector component at p different lattice points can be calculated, simultaneously, if p parallel processors are employed, demonstrating that the FDTD algorithm is highly parallelizable. Holland [128] suggested an exponential time stepping to model the exponential decay of propagating waves in certain highly lossy media that the standard Yee time-stepping algorithm fails to describe. Stability analysis [129] has shown that the upper bound for the FDTD time step for a homogeneous region of space (ε_r, μ_r) is given by

$$
\Delta t \le \frac{\sqrt{\varepsilon_r \mu_r}}{c\sqrt{\frac{1}{(\Delta x)^2}+\frac{1}{(\Delta y)^2}+\frac{1}{(\Delta x)^2}}} \quad \text{(3D simulations)}
$$

$$
\Delta t \le \frac{\sqrt{\varepsilon_r \mu_r}}{c\sqrt{\frac{1}{(\Delta x)^2}+\frac{1}{(\Delta y)^2}}} \quad \text{(2D simulations)}
$$

Lower values of upper bounds are used in case a highly lossy material or a variable grid is employed. Discretization with at least 10–20 cells/wavelength almost guarantees that the FDTD algorithm will have satisfactory dispersion characteristics (phase error smaller than $5°/\lambda$ for time step close to the upper bound value).

The computational domain must be large enough to enclose the structure of interest, and a suitable ABC on the outer perimeter of the domain must be used to simulate its extension to infinity to minimize numerical reflections for a wide range of incidence angles and frequencies. Central differences cannot be implemented at the outermost lattice planes, since, by definition, no information exists concerning the fields at points one-half space cell outside of these planes. The perfectly matched layer (PML) ABC, introduced in 2D by Berenger in 1994 [130] and extended to 3D by Katz et al. [131], provides numerical reflection comparable to the reflection of anechoic chambers with values −40 dB lower than the previous absorbers. A new ABC based on Green's functions that absorbs efficiently propagating and evanescent modes has also been demonstrated [132,133] for waveguide and RF packaging structures.

An incoming-plane wave source [134] is very useful in modeling radar scattering problems, since in most cases of this type the target of interest is in the near field of the radiating antenna, and the incident illumination can be considered to be a plane wave. The hard source [134] is another common FDTD source implementation. It is set up simply by superimposing a desired time function onto specific electric or magnetic field components in the FDTD space lattice, which are regularly updated by the FDTD equations. Collinear arrays of hard-source field vector components in 3D can be useful for exciting waveguides and strip lines. In the FDTD simulations of microstrip and stripline structures, the Gaussian pulse (nonzero DC content) is used as the excitation of the microstrip and stripline structures. The Gabor function

$$
s(t) = \mathrm{e}^{-((t-t_0)/(pw))^2}\sin(wt) \tag{28.19}
$$

where $pw = 2 \cdot (\sqrt{6}/\pi(f_{\max} - f_{\min}))$, $t_0 = 2pw$, $w = \pi(f_{\min} + f_{\max})$, is used as the excitation of the waveguide structures, since it has zero DC content. By modifying the parameters pw and w, the frequency spectrum of the Gabor function can be practically restricted to the interval $[f_{\min}, f_{\max}]$. As a result, the envelope of the Gabor function represents a Gaussian function in both time and frequency domain. Monochromatic simulations are performed through the use of continuous-wave (sinusoidal) excitations.

It is very common, especially for high-speed circuit structures, to use a cell size Δ that is dictated by the very fine dimensions of the circuit and is almost always much finer than needed to resolve the smallest spectral wavelength propagating in the circuit. As a result, with the time step Δt bound to Δ by numerical stability considerations, FDTD simulations have to run for tens of thousands of time steps in order to fully evolve the impulse responses needed for calculating impedances, *s*-parameters, or resonant frequencies. One popular way to avoid virtually prohibitive execution time has been to apply contemporary analysis techniques from the discipline of digital signal processing and spectrum estimation. The strategy is to extrapolate the electromagnetic field time waveform by 10:1 or more beyond the actual FDTD time window, allowing a very good estimate of the complete system response with 90% or greater reduction in computation time. This extrapolation can be performed using forward–backward predictors [135] or autoregressive (AR) models [136].

The FDTD technique has found numerous applications in modeling microwave devices such as waveguides, resonators, transmissions lines, vias, antennas, and active and passive elements. In 1985, DePourcq [137] used FDTD to analyze various 3D waveguide devices. Navarro et al. [138] investigated rectangular, circular, and T junctions in square coaxial waveguides and narrow-wall, multiple-slot couplers. Wang et al. [139] studied the *Q* factors of resonators using FDTD. Liang et al. [140] used FDTD to analyze coplanar waveguides and slotlines and Sheen et al. [141] presented FDTD results for various microstrip structures including a rectangular patch antenna, a low-pass filter, and a branch-line coupler. Cangellaris estimated the effect of the staircasing approximation of conductors of arbitrary orientation in Reference 142.

The characterization of interconnect transitions in multichip and microwave circuit modules has also been investigated using FDTD. Lam et al. [143] used a nonuniform mesh to model microstrip-to-via-to-stripline connections. Piket-May et al. [144] studied pulse propagation and crosstalk in a computer module with more than ten metal-dielectric-metal layers and numerous vias. Luebbers et al. [145] and Shlager and Smith [146] developed and described in detail efficient 3D, time domain, near-to-far-field transformations. In 1990, Maloney et al. [147] presented accurate results for the radiation from rotationally symmetric simple antennas such as cylindrical and conical monopoles, while Luebbers and coworkers [148,149] presented mutual coupling and gain computations for a pair of wire dipoles, and Tirkas and Balanis [150] modeled 3D horn antennas. Uehara and Kagoshima [151] analyzed microstrip phased-array antennas and Jensen and Rahmat-Samii [152] presented results for the input impedance and gain of monopoles, planar inverted-F antennas (PIFAs) and loop antennas on handheld transceivers. In addition, Taflove [153] used FDTD to model scattering and compute near/far fields and radar cross-section (RCS) for 2D and 3D structures. Britt [154] calculated the RCS of both 2D and 3D perfectly conducting and dielectric scatterers. In 2002, Bushyager and Tentzeris [155] proposed an adaptive FDTD-based modeling technique that couples Maxwell's, mechanical and solid-state equations for the simulation of MEMS devices with moving parts [156] and of submicron transistors in wireless and microwave transceivers.

Another area of FDTD applications is active and passive device modeling. Two different approaches are used. In the first, analytical device models are coupled directly with FDTD. In the second, lumped-element subgrid models are used with the device behavior being determined by other software, something that may be preferable in the modeling of active devices with complicated equivalent circuits. In 1992, Sui et al. [157] reported a 2D FDTD model with lumped circuit elements, including nonlinear devices, such as diodes and transistors and this approach was extended to 3D by Piket-May et al. [144] and Ciampollini et al. [158]. Kuo et al. [159] presented a large-signal analysis of packaged nonlinear active microwave circuits. Alsunaidi et al. [160] developed an active device model that couples the Yee update equations with the solution of the current continuity equation, the energy-conservation equation, and

the momentum-conservation equations. Thomas et al. [161] developed an approach for coupling SPICE lumped elements into the FDTD method.

Dey and Mittra [162,163] introduced a very simple, stable, and accurate contour-path technique to model curved metal surfaces in practical microwave and mixed-signal structures. Maloney and Kesler [164] introduced several novel means to analyze periodic (PBG, EBG) structures using FDTD, and Painter et al. [165] employed FDTD to design, construct and successfully test the world's smallest microcavity laser based upon a 2D photonic bandgap structure. In 2000 Zheng et al. [166] introduced the first 3D alternating-direction implicit (ADI) FDTD algorithm with proven unconditional stability regardless of the size of the time step, something that could be useful for large geometries, such as photonic structures [167,168]. It has to be noted that the ADI has to be applied in a careful way considering the effects of numerical dispersion for large cell sizes. Recently, the computationally improved envelope ADI–FDTD method has been extensively utilized [169,170].

The latest FDTD publications involve the microwave breast cancer detection [171], the analysis of ferrites and other lossy magnetic materials [172] and of uniaxial bianisotropic media [173], the modeling of UWB Ground Penetrating antennas [174], the design of optical waveguides [175,176], the simulation of 3D multilayer LTCC structures [177], the shielding effectiveness of whole planes [178], the design of biomedical devices [179], and the modeling of biological tissues [180]. In addition, FDTD has been used for the propagation of lightning pulses [181], the modeling of FSS [182] and of photonic bandgap structures [183], the modeling of WLAN switchable antennas [184], the simulation of biological effects due to cell phones [185]. FDTD has been also used to model power distribution networks [186], ELF propagation effects [187], lightning-generated EM fields [188], transient propagation in chiral metamaterials [189] and leaky-wave structures with negative-refractive-index materials [190].

It has to be noted that there have been various research efforts for an efficient multigridding/subgridding approach [191,192], the development of higher-order finite-difference subgridding schemes [193], and the hardware implementation of FDTD algorithms [194], as well as the algorithm parallelization using PC clusters [183]. Special attention has been paid to the analysis of fine metal/dielectric details involving the modeling of off-grid boundary conditions [195,196] and the implementation of sensitivity analysis of complicated structures [197].

28.3.2 Transmission Line Matrix Method

The transmission line matrix (TLM) method [198–200] is similar to FDTD. The main difference is that the electromagnetic problem is analyzed through the use of a 3D equivalent network problem [201]. It is a very versatile time-domain technique and discretizes the computational domain using cubic cells with a period (cell size) Δl. Boundaries corresponding to perfect electric (magnetic) conductors are represented by short-circuited (open-circuited) parallel nodes on the boundary. Variations of dielectric and diamagnetic constant [202] are introduced by adding short-circuited series stubs of length $\Delta l/2$ at the series (H-field) nodes and open-circuited $\Delta l/2$ stubs at the shunt (E-field) nodes. Losses can be introduced by resistively loading the shunt nodes. After the time-domain response is obtained, the frequency response is calculated using the Fourier transform. Due to the introduction of periodic lattice structures, a typical passband-stopband phenomenon appears in the frequency domain data. The frequency range must be below the upper bound of the lowest passband and is determined by the mesh size Δl. The TLM technique has been used in the analysis of waveguiding structures [203] making use of the Diakoptics Theorem [204] and has been extensively compared to the FDTD technique [205]. Effective numerical absorbers [206–208] including the FDTD-popular PML absorber have been derived and implemented in the modeling of radiating structures [209,210]. In addition, TLM has been employed in the analysis of bondwire packaging [211] and MEMS switches [212]. So and Hoefer [213] coupled TLM with SPICE and performed a coupled field and circuit time-domain simulation. Paul and Christopoulos presented ways to model with TLM materials with frequency-dependent [214], anisotropic [215], nonlinear [216] properties. In addition, TLM has been applied to the calculation of SAR properties of cellular phones [217] and to the modeling of longitudinally periodic waveguides [218]. Pierantoni applied TLM in

combination with integral equation to the analysis of photonic bandgap integrated optical components [219], while, So modeled metamaterials using the 2D SCN–TLM [220]. Lorenz proposed an efficient way to apply parallelized TLM to complex electromagnetic structures [221], and Abolghasem utilized TLM for sensitivity analysis [222,223]. Recently, there have been presented approaches for tetrahedral-based TLM [224] and unstructured triangular TLM meshes [225], and for the TLM modeling of reconfigurable multiband antennas [226]. Also, TLM has been applied to the analysis of ferrites [227], the modeling of thin wires [228], the analysis of lightning-induced EMC problems [229], the modeling of nonlinear optical effects [230] and the modeling of photodetectors using a drift-diffusion TLM implementation [231].

28.3.3 Finite Element Method

In the FEM [14–20], instead of partial differential equations with boundary conditions, corresponding functionals (e.g., power) are set up and variational expressions are applied to each cell (element) of the area of interest. Most of the time, the elements are rectangles or triangles for two-dimensional (2D) problems and parallelepiped (bricks) or tetrahedra for 3D problems, something that allows for the efficient representation of most arbitrary shapes.

Assume that the 2D $(x - y)$ Laplace equation is to be solved

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} = 0 \tag{28.20}$$

The solution is equivalent to the minimization of the functional

$$I(\phi) = \langle \phi, \nabla^2 \phi \rangle = \iint_s \phi \left(\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} \right) \mathrm{d}x\, \mathrm{d}y = - \iint_S \left[\left(\frac{\partial \phi}{\partial x} \right)^2 + \left(\frac{\partial \phi}{\partial y} \right)^2 \right] \mathrm{d}x\, \mathrm{d}y \tag{28.21}$$

The last integral is the superposition of each element's contribution. In each 2D element ϕ can be approximated as polynomial of variables x and y. For example, for a triangular element-p (Figure 28.4),

$$\phi = \alpha + \alpha_x x + \alpha_y y \tag{28.22}$$

where the constants $\alpha, \alpha_x, \alpha_y$ depend on the ϕ values at the three vertices of the triangle

$$\phi_p = \alpha + \alpha_x x_p + \alpha_y y_p \tag{28.23}$$

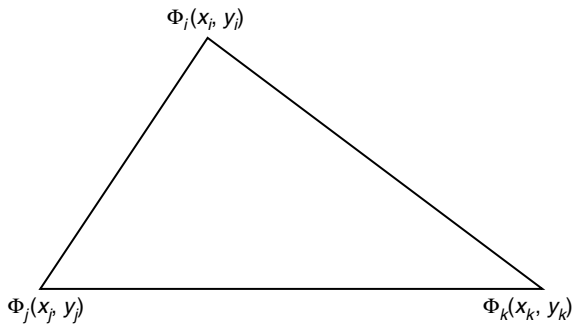


FIGURE 28.4 FEM triangular (2D) element.

where $p = i, j, k$ are the three triangle vertices. Due to the first derivatives of Equation 28.21, only α_x, α_y are needed for the calculation of $I(\phi)$, thus

$$\begin{bmatrix} a_x \\ a_y \end{bmatrix} = A \begin{bmatrix} \phi_i \\ \phi_j \\ \phi_k \end{bmatrix} \tag{28.24}$$

For each (i, j, k)-element, $I(\phi)$ gets the value

$$I_{i,j,k}(\phi) = [\phi_i, \phi_j, \phi_k] A^{\mathrm{T}} A \begin{bmatrix} \phi_i \\ \phi_j \\ \phi_k \end{bmatrix} |\Delta S| \tag{28.25}$$

where A^{T} is the transpose of A and ΔS is the area of the triangle (element) calculated by

$$\Delta S = \frac{1}{2} \begin{vmatrix} 1 & x_i & y_i \\ 1 & x_j & y_j \\ 1 & x_k & y_k \end{vmatrix} \tag{28.26}$$

The Rayleigh–Ritz technique is used for the minimization of $I_{i,j,k}$

$$\frac{\partial I_{i,j,k}}{\partial \phi_i} = \frac{\partial I_{i,j,k}}{\partial \phi_j} = \frac{\partial I_{i,j,k}}{\partial \phi_k} = 0 \tag{28.27}$$

As a result, for each (i, j, k)-element

$$A^{\mathrm{T}} A \begin{bmatrix} \phi_i \\ \phi_j \\ \phi_k \end{bmatrix} = 0 \tag{28.28}$$

After iterating this procedure to all elements of the computational domain S and using a connection matrix to account for points that are vertices common to more than one element, the following matrix equation is derived:

$$B \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_N \end{bmatrix} = 0 \tag{28.29}$$

After plugging in the known values of the ϕ_i's that are located on the boundaries, the rest of the ϕ's are calculated through the inversion of matrix B. In this way, the potentials of all interior points can be given by Equation 28.22.

Various element forms [232–236] have been used in order to minimize the memory requirements and facilitate the gridding procedure and the modeling of boundary conditions (PECs, dielectric interfaces). The effect of discretization error on the numerical dispersion has been extensively studied and gridding guidelines have been derived in References 237 and 238. In addition, the analysis of radiating (antennas) and scattering problems has led to the development of numerical absorbers [239–241] with very low numerical reflection coefficients. Due to the shape of the finite elements, the FEM technique can accurately represent very complex geometries and is one of the most popular techniques for scattering problems [242], discontinuities and transitions [243], packaging [244], interconnects [245], and MMIC modeling [246].

Li [247] optimized patch antennas on ferrite substrates and Zhu [248] presented a way to model lossy media. Hung applied FEM to MEMS [249] including mechanical motion equations and Ammus [250] modeled the thermal characteristics of power semiconductor devices. Also, Volakis [251] and Cangellaris [252] used hierarchical vector finite elements to perform an adaptive multiresolution modeling of antennas and microwave structures.

Recently, FEM has been used in electromagnetic scattering problems of targets with strong/weak coupling for randomly rough surfaces using an FFT accelerated boundary condition [253], in the modeling of eddy-currents [254], in the thermal assessment of RF integrated 3D LTCC modules [255], in the modeling of large-domain 3D multiport waveguiding structures [256], in 3D tomography problems [257], in the calculation of passive/active electromagnetic shielding [258], and in the modeling of photonic crystals [259]. The FEM has been also utilized in medical problems involving soft tissue response to acoustic impulses [260], in nanodevices through the use of 3D spectral element method [261], and in the analysis of MEMS problems [262]. In addition, high-order vectorial finite elements have been used for nonsymmetric transversely anisotropic waveguides [263]. Also, FEM has been used in the modeling of semiconductor traveling-wave devices [264]. One time-domain finite-element formulation has been used in the analysis of periodic structures combined with Floquet ABC [265]. A very good comparison between the capabilities of volume integral equation versus those of FEM for printed antennas on thin finite dielectric substrates is given in Reference 266. Recently, FEM has been coupled with neural networks to optimize coil geometries [267] and with particle swarm optimization techniques to optimize microwave filters [268]. Also, p-refinement has been used for the effective large-domain 2D modeling of arbitrary waveguides [269]. Hierarchical vector finite elements have been proposed for the analysis of complicated waveguiding structures [270] and have been evaluated in 2D curvilinear domains [271]. It has to be noted that various multigrid techniques have been used for the acceleration of edge-based FEM analysis [272].

The boundary element technique (BE) [273,274] is a combination of the boundary integral equation and a finite-element discretization applied to the boundary. In essence, it is a form of the integral equation–MoM approach discussed previously. The wave equation for the volume is converted to the surface integral equation through Green's identity. The surface integrals are discretized into N elements, and the evaluation is performed for each element after E- and H-fields are approximated by polynomials. Due to the reduction of the number of dimensions, there is a significant reduction in memory and CPU time requirements. The BE technique has been utilized in the analysis of cavities [275] and of planar layered media [276], and in the incorporation of lumped elements in the FEM analysis [277]. Recently, a novel symmetric FEM–BE approach has been proposed for the analysis of complex electromagnetic problems [278]. Also, an edge-based FEM–BE technique has been presented for wideband electromagnetic computations [279]. A very good discussion about the pros and cons of FEM and BE techniques has been published in Reference 280.

28.3.4 Hybrid Techniques

As has become clear from the previous discussion, all numerical techniques have advantages and disadvantages depending on the geometry to be modeled. Integral equation techniques allow for the quick and efficient modeling of radiation phenomena, but derivation of Green's function for complex structures is tedious. The MM method is more appropriate for waveguiding structures where modes are easily determined. The FDTD (and TLM) technique is quite general and requires no preprocessing, though it has often to be run for medium-to-large execution times. The FEM technique is adaptive due to the shape of the elements, but gridding and functional optimization demands significant computational effort. Thus, there have been numerous efforts for the development of hybrid simulation approaches that use different techniques for different sub-geometries and utilize connection relationships for the areas of numerical interfaces.

Jin [281], Gedney [282], and Musolino [283] proposed a hybrid FEM/MoM method for the modeling of wave scattering by 3D apertures, wave diffraction in gratings and nonlinear analysis of microwave devices.

Wu [284] and Monorchio [285] suggested an FEM/FDTD approach for the multi-frequency modeling of complex geometries. The MM technique has been coupled with integral equation [286], spectral domain [287], and FEM [288,289] to analyze complicated passive microwave circuits and waveguide problems including inductive loading and wave scattering. Lindenmeier [290,291] introduced a hybrid TLM/MIE for thin wire modeling and for the estimation of the electromagnetic interaction between complex objects (e.g., MMICs) separated by large free-space regions, and Pierantoni [292] analyzed numerical aspects of MoM–FDTD and TLM–Integral equation used in EMC modeling. Brandt [293] applied a hybrid surface integral-equation/MM technique to the analysis of dielectric waveguides of arbitrary shape. Wu [294] used a hybrid MM/FEM approach to perform parameter extraction for multilayered RF passives. Arndt [295] has proposed various fast CAD and optimization tools of waveguide structures and aperture antennas based on hybrid MM, FEM, MoM, and FDTD techniques. The FEM method has been coupled with the method of lines for the analysis of planar or quasi-planar transmission lines [296]. Also, FEM has been coupled with the MoM to calculate the human exposure effects to GSM base stations [297]. Lastly, FEM and spherical mode expansion have been combined for the analysis of finite arrays of coupled antennas [298].

To model effectively large structures with very fine details, the FDTD technique has been hybridized with time-domain MM [299], as well as with FEM/MoM [300] with significant economies in memory and execution time. In addition, FDTD has been hybridized with FEM for the hybrid electrical/mechanical analysis of composite antennas and 3D modules [301].

28.3.5 Wavelets: A Memory-Efficient Adaptive Approach?

The term "wavelet" [302–306] has a very broad meaning, ranging from singular integral operators in harmonic approach to sub-band coding algorithms in signal processing, from coherent states in quantum analysis to spline analysis in approximation theory, from multi-resolution transform in computer vision to a multilevel approach in the numerical solution of partial differential equations, and so on. Most of the time, wavelets could be considered mathematical tools for waveform representations and segmentations, time-frequency analysis, and fast and efficient algorithms for easy implementation in both time and frequency domains. One of the most important characteristics of expansion to scaling and wavelet functions is the time (domain)-frequency (Fourier-transformed domain) localization. Another very salient feature of these new expansions is that the entire set of basis functions is generated by the dilation and shifting of a single function, called the mother wavelet. The standard approach in ideal lowpass ("scaling," Figure 28.5) and bandpass ("wavelet," Figure 28.6) filtering for separating an analog signal into different frequency bands emphasizes the importance of time localization. Multiresolution analysis (MRA), introduced by Mallat and Meyer [307,308], provides a very powerful tool for the construction of wavelets and implementation of the wavelet decomposition/reconstruction algorithms. In the case of cardinal *B*-splines [309], an orthonormalization process is used to produce an orthonormal scaling function and, hence, its corresponding orthonormal wavelet by a suitable modification of the two-scale sequence. The orthonormalization process was introduced by Schweinler and Wigner [310] and the resulting wavelets are the Battle–Lemarie wavelets, obtained independently by Battle [311] and Lemarie [312] using different methods. The only orthonormal wavelet that is symmetric or antisymmetric and has compact support (to give finite decomposition and reconstruction series) is the Haar [313] wavelet. Nevertheless, these wavelets exhibit poor time-frequency localization. Another set of orthonormal basis is the Daubechies wavelets [314]. At present, MRA has been applied to alleviate the numerical disadvantages mainly of Integral Equation and FDTD methods, though preliminary efforts for FEM are currently being investigated.

It is well known that the integral equation method, described in previous sections, offers a straightforward and efficient numerical solution when applied to small- to medium-scale problems. Difficulties arise when the complexity of the geometry and, subsequently, the number of the unknowns increases, resulting in very large matrices. All conventional basis functions traditionally used in MoM generate full moment matrices. The computational problems associated with the storage and manipulation of large, densely populated matrices easily rules out the practicality of the integral equation techniques.

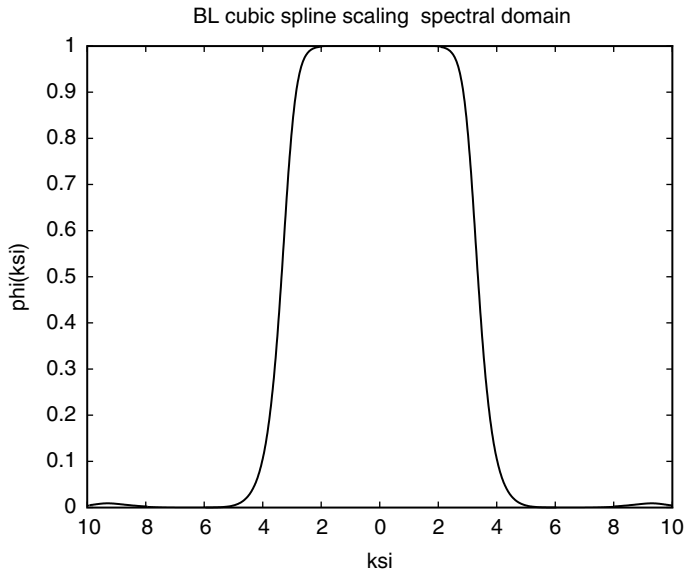


FIGURE 28.5 Battle–Lemarie scaling-spectral ("lowpass") domain.

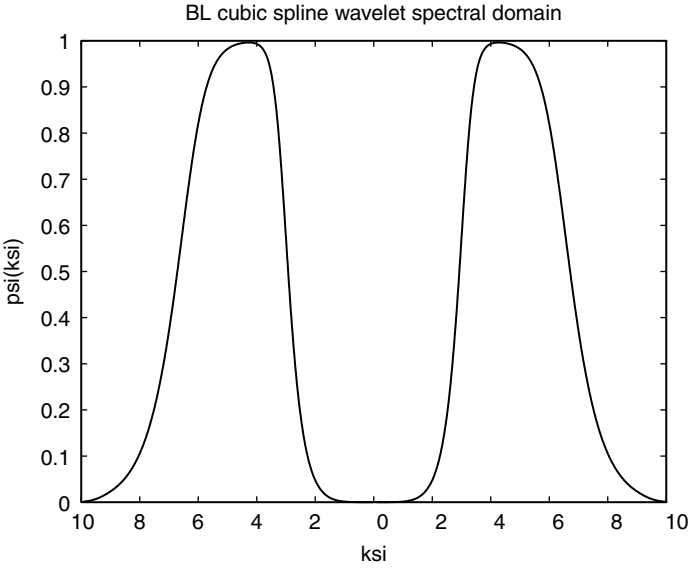


FIGURE 28.6 Battle–Lemarie wavelet-spectral ("bandpass") domain.

The potential application of wavelet theory in the numerical solution of integral equations led to the finding that wavelet expansion of certain types of integral operators generates highly sparse linear systems [315]. This proposition was used [316–318] in the MoM formulation of 1D electromagnetic scattering problems and in the analysis of integrated millimeter- and submillimeter-wave waveguides with a Battle–Lemarie orthonormal basis [319]. Sparsity results above 90% allowed for the accurate modeling of structures that could not be analyzed with IE using conventional expansions. Various expansion basis have been used [320] leading to significant computational economies in the modeling of 2D/3D path propagation/scattering problems from regular or rough surfaces [321] and in the modeling of high-speed interconnects [322].

As far as it concerns FDTD, despite its numerous applications, many practical geometries, especially in microwave and millimeter-wave integrated circuits (MMIC), packaging, interconnects, sub-nanosecond digital electronic circuits [such as multi-chip modules (MCM)], and antennas used in wireless and microwave communication systems, have been left untreated due to their complexity and the inability of the existing techniques to deal with requirements for large size and high resolution. Krumpholz has shown that Yee's FDTD scheme can be derived by applying the MoM for the discretization of Maxwell's equations using pulse basis functions for the expansion of the unknown fields. The use of scaling and wavelet functions as a complete expansion basis of the fields demonstrates that multiresolution time domain (MRTD) [323] schemes are generalizations of Yee's FDTD and can extend this technique's capabilities by improving computational efficiency and substantially reducing computer resources. In an MRTD scheme the fields are represented by a twofold expansion in scaling and wavelet functions with respect to time/space. Scaling functions ("lowpass") guarantee a correct modeling of smoothly varying fields. In regions characterized by strong field variations or field singularities, higher resolution is enhanced by incorporating wavelets in the field expansions. The major advantage of the use of MRTD in the time domain is the capability to develop time and space adaptive grids through the thresholding of the wavelet coefficients for each time step throughout the grid. The MRTD schemes based on cubic spline Battle–Lemarie scaling and wavelet functions have been used for the derivation of time/space-adaptive schemes [324,325] in real time. Various types of wavelets have been successfully applied to the simulation of nonlinear effects [326], 3D dielectric cavities [327], filters [328], mixers [329], RF packaging structures [330], optical waveguides [331], mine detection [332], and MMIC [333] offering economies in memory and execution time of order(s) of magnitude with respect to FDTD. Dispersion analysis by Tentzeris [334] shows the capability of excellent accuracy with up to 2 points/wavelength (Nyquist limit) for the Battle–Lemarie basis. Nevertheless, the functions of this family do not have compact support; thus, the MRTD schemes have to be truncated with respect to space, something that requires the use of image theory for the modeling of hard boundaries (e.g., PECs). Thus, specific problems may require the use of functions with compact support (e.g., Haar, Figure 28.7) especially for the approximation of time derivatives. In this way, the modeling of boundary conditions is most straightforward, though the computational economies are not so dramatic [335–337]. Sarris proposed a way for an improvement of the accuracy of MRTD algorithms modifying the offset of electric and magnetic fields [338] and Cao [339] presented an anisotropic PML for the modeling of open structures. Also, Sarris presented an efficient interface between FDTD and Haar-MRTD technique [340].

Bushyager presented [341] a very complete methodology *for the modeling of multiple PEC's and dielectric discontinuities within a single-cell*, an approach that could potentially revolutionize the simulation of very complex structures. S. Goasguen [342,343] and Bushyager [155] coupled the electromagnetic and the semiconductor device simulations using interpolating and Haar wavelets, respectively, and achieving very significant computational economies. In addition, bi-orthogonal wavelets have been used in electromagnetic scattering problems [344] and in the modeling of elastic-wave interactions with buried

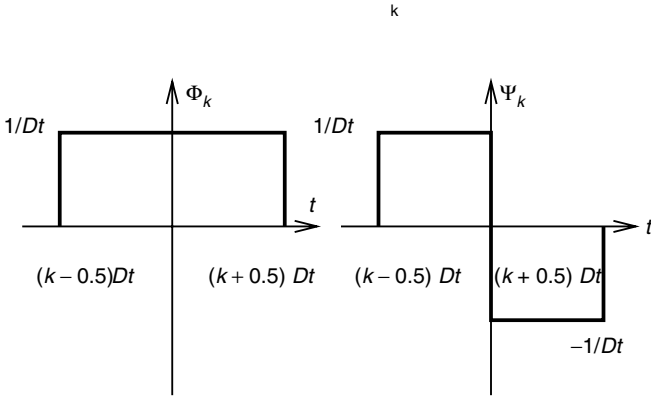


FIGURE 28.7 Haar expansion basis.

objects [345] and PEC targets [346]. Also, Zhaoxian has presented an adaptive time-domain integral equation for the transient analysis of wire scatterers [347].

28.4 Optimization Tools

Existing optimization packages included in commercial electromagnetic simulators, often do not take into account the specific effect of each of the factors involved in the design process and the degree of interaction between them, thus, leading to time-consuming "trial-and-error" approaches. For the complicated 3D architectures of RF/microwave multilayer modules, a comprehensive and sophisticated tool is necessary to account for complex phenomena such as coupling and fringing effects. Alternative optimization methods suited to this kind of complexity are neural networks [348] and genetic algorithms [349]. These are very precise methods which, unfortunately, require a hefty amount of prior knowledge and are computationally complex and time consuming. The statistical tools to be presented in this section [350] overcome these disadvantages. They provide a thorough understanding of all of the factors involved in the design process, and identify which are more significant and which are not significant at all, how they interact with each other, and if the design goals are achievable in the given conditions. Most importantly, these methods are very easy to implement with negligible computational overhead. They are suited as a first approach to understand the behavior of a system, to decide whether an optimization is possible under certain conditions and filter out all the insignificant factors from the beginning of the process. Also, due to their flexibility, the simultaneous optimization of the electrical, mechanical, and thermal parameters of the system is possible. Finally, Monte Carlo type analyses [351] model the performance capability of the system, when the models are used to predict both the nominal values and the variation expected for system performance.

The first and most simple tool to be presented is design of experiments (DOE). The DOE is a series of tests in which a set of input variables or factors is purposely changed so that the experimenter can observe and identify the reasons for changes in the output response. The factorial designs are used in experiments involving several factors where the goal is the study of the joint effects of the factors on a response. Prior knowledge of the analyzed system is required for choosing the factors, their studied ranges, and the figures of merit. The 2^k factorial design is the simplest one, with k factors at 2 levels each. It provides the smallest number of runs for studying k factors and is widely used in factor screening experiments [352]. The statistical analysis of the DOE shows which parameters are significant for each of the figures of merit and the ones that are not significant are eliminated from further analysis, as well as how factors are interacting with each other. Then, the model is checked for curvature and if detected, response surface methods (RSM) can account for curvature through second-order model development. The optimization is performed based on the developed first- or second-order models. A typical second-order model development is illustrated as a flow chart below.

For cases in which the optimal is not in the initial design space, path of ascent (POA) is the method of choice. The POA is applied to determine a path, which extends beyond the DOE design space, for further optimization of the figures of merit. Using this path, another design space is defined and the process repeats with another DOE, as illustrated in Figure 28.8. The process is complete when the performance goal or optimum performance is achieved. The POA comes in as the method of choice for systems with large curvature, when the initial design space cannot be modeled either with DOE or RSM, and the design space has to be reduced to a smaller size, where the curvature can be modeled. After that, the optimization is performed by using the POA to find the new design space where the optimal lies. Also, the performance of RF and microwave systems is severely affected by the fabrication process tolerances, especially for the small circuit dimensions at millimeter-wave frequencies. By using Monte Carlo analysis, the developed transfer functions predict both the nominal values and the variation expected for system performance at the beginning of the design process, thus, saving the time and the frustration of the trial-and-error approaches. The above reported methods are very easy to implement with the use of commercially available electromagnetic and statistical software packages. In addition, these methods can be easily extended to

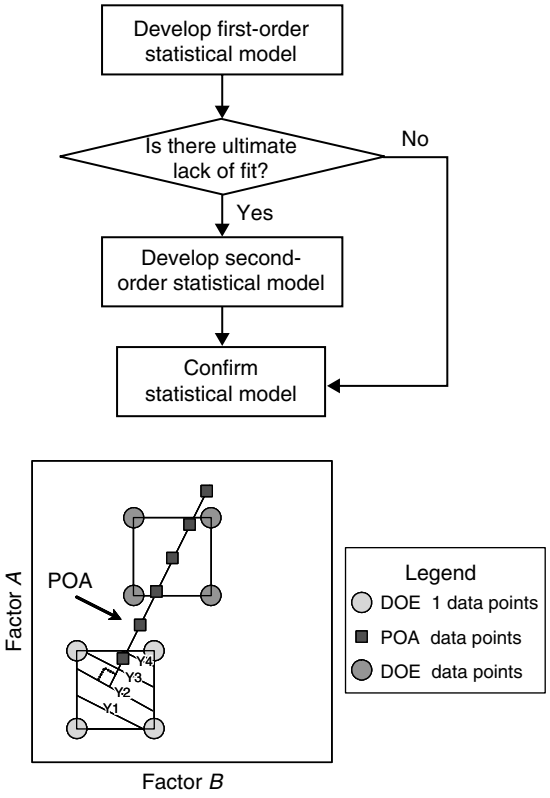


FIGURE 28.8 Procedures for statistical model development.

predict performance of multi-level systems, by separately analyzing each level and, then, combine them all together by using output variables for lower level as inputs for the next level [352,353].

28.5 Conclusions

This chapter has briefly presented various numerical techniques that are commonly used for the analysis and design of RF and microwave circuits. Their fundamental features as well as their advantages and disadvantages have been discussed and representative references have been reported. It must be emphasized that there is no numerical scheme that can achieve optimal performance for all types of structures. Thus, the hybridization of these techniques or the implementation of novel approaches (e.g., wavelets), successfully applied in other research areas, would be better candidates for the generalized, efficient, and accurate modeling of modern complex devices used in telecommunication, radar, space, biomedical, and computing applications.

References

1. T. Itoh, *Numerical Techniques for Microwave and Millimeter-Wave Passive Structures*, John Wiley & Sons, New York, 1989.
2. R.C. Booton, *Computational Methods for Electromagnetics and Microwaves*, John Wiley & Sons, New York, 1992.
3. R.F. Harrington, *Field Computation by Moment Methods*, The Macmillan Company, New York, 1968.
4. R.C. Hansen, *Moment Methods in Antennas and Scattering*, Artech House, Norwood, MA, 1990.

5. J.H. Wang, *Generalized Moment Methods in Electromagnetics*, John Wiley & Sons, New York, 1991.
6. R. Bancroft, *Understanding Electromagnetic Scattering Using the Moment Method: A Practical Approach*, Artech House, Norwood, MA, 1996.
7. A.F. Peterson, S.L. Ray, and R. Mittra, *Computational Methods for EM*, IEEE Press, Oxford University Press, USA, 1998.
8. C.R. Scott, *The Spectral Domain Method in Electromagnetics*, Artech House, Norwood, MA, 1989.
9. R. Mittra and W.W. Lee, *Analytical Techniques in the Theory of Guided Waves*, Macmillan, New York, 1971.
10. K.S. Kunz and R.J. Luebbers, *The Finite Difference Time-Domain Method for Electromagnetics*, CRC Press, Boca Raton, FL, 1993.
11. A. Taflove and S.C. Hagness, *Computational Electrodynamics: The Finite-Difference Time-Domain Technique*, 2nd Edn., Artech House, Norwood, MA, 2000.
12. A. Taflove, *Advances in Computational Electrodynamics*, Artech House, Norwood, MA, 1998.
13. T. Itoh and B. Housmand, *Time-Domain Methods for Microwave Structures: Analysis and Design*, Wiley IEEE Press, USA, 1998.
14. G. Strang and G.J. Fix, *An Analysis of the Finite Element Method*, Prentice-Hall, Englewood Cliffs, NJ, 1973.
15. P.P. Silvester and R.L. Ferrari, *Finite Elements for Electrical Engineers*, Cambridge University Press, Cambridge, 1983.
16. J.N. Reddy, *An Introduction to the Finite Element Method*, McGraw Hill, New York, 1984.
17. J.-C. Sabonnadiere and J.-L. Coulomb, *Finite Element Methods in CAD: Electric and Magnetic Fields*, Springer-Verlag, New York, 1987.
18. O.C. Zienkiewicz and R.L. Taylor, *The Finite Element Method*, McGraw Hill, London, 1988.
19. T. Itoh, G. Pelosi, and P.P. Silvester, *Finite Element Software for Microwave Engineering*, John Wiley & Sons, New York, 1996.
20. J.L. Volakis, A. Chaterjee, and L.C. Kempel, *Finite Elements Method for Electromagnetics*, IEEE Press, New York, 1998.
21. A.W. Glisson and D.R. Wilton, Simple and efficient methods for problems of radiation and scattering from surfaces, *IEEE Trans. Antennas and Propagation*, 28, 593–603, 1980.
22. T.K. Sarkar, A.R. Djordjevic, and E. Arvas, On the choice of expansion and weighting functions in the numerical solution of operator equations, *IEEE Trans. Antennas and Propagation*, 33, 988–996, 1985.
23. J.R. Mossig and F.E. Gardiol, Analytical and numerical techniques in the Green's function treatment of microstrip antennas and scatterers, *IEE Proc.*, Part H, 130, 175–182, 1983.
24. A.J. Poggio and E.K. Miller, Integral equation solutions of 3D scattering problems, in *Computer Techniques for Electromagnetics*, R. Mittra, Ed., Hemisphere, New York, 1987.
25. A.F. Peterson and P.W. Klock, An improved MFIE formulation for TE-wave scattering from lossy, in homogeneous dielectric cylinders, *IEEE Trans. Antennas and Propagation*, 25, 518–524, 1987.
26. F. Ling, D. Jiao, and J.M. Jin, Efficient EM modeling of microstrip structures in multilayer media, *IEEE Trans. Microwave Theory Techniques*, 47, 1810–1818, 1999.
27. D.M. Pozar, Input impedance and mutual coupling of rectangular microstrip antennas, *IEEE Trans. Antennas and Propagation*, 30, 1191–1196, 1982.
28. W.-T. Chen and H.-R. Chuang, Numerical computation of the EM coupling between a circular loop antenna and a full-scale human-body model, *IEEE Trans. Microwave Theory Techniques*, 46(10-1), 1516–1520, 1998.
29. G. Athanasoulias and N.K. Uzunoglu, An accurate and efficient entire-domain basis Galerkin's method for the integral equation analysis of integrated rectangular dielectric waveguides, *IEEE Trans. Microwave Theory Techniques*, 43, 2794–2804, 1995.
30. M. Swaminathan, T.K. Sarkar, and A.T. Adams, Computation of TM and TE modes in waveguides based on surface integral formulation, *IEEE Trans. Microwave Theory Techniques*, 40, 285–297, 1992.
31. S.J. Polychronopoulos and N.K. Uzunoglu, Propagation and coupling properties of integrated optical waveguides—an integral equation formulation, *IEEE Trans. Microwave Theory Techniques*, 44, 641–650, 1996.
32. P. Cottis and N. Uzunoglu, Integral equation approach for the analysis of anisotropic channel waveguides, *J. Opt. Soc. A.*, 8, 4, 608–614, 1991.

33. P. Guillot, P. Couffignal, H. Baudrand and B. Theron, Improvement in calculation of some surface integrals: application to junction characterization in cavity filter design, *IEEE Trans. Microwave Theory Techniques*, 41, 2156–2160, 1993.
34. T.E. van Deventer, L.P.B. Katehi and A.C. Cangellaris, Analysis of conductor losses in high-speed interconnects, *IEEE Trans. Microwave Theory Techniques*, 42, 78–83, 1994.
35. A.W. Mathis and A.F. Peterson, Efficient electromagnetic analysis of a doubly infinite array of rectangular apertures, *IEEE Trans. Microwave Theory Techniques*, 46(1), 46–54, 1998.
36. S-G. Hsu and R.-B. Wu, Full-wave characterization of a through hole via in multi-layered packaging, *IEEE Trans. Microwave Theory Techniques*, 1073–1081, 1995.
37. A.M. Rajeek and A. Chakraborty, Analysis of a wide compound slot-coupled parallel waveguide coupler and radiator, *IEEE Trans. Microwave Theory Techniques*, 43(4), 802–809, 1995.
38. G.V. Eleftheriades, J.R. Mosig and M. Guglielmi, A fast integral equation technique for shielded planar circuits defined on nonuniform meshes, *IEEE Trans. Microwave Theory Techniques*, 44, 2293–2296, 1996.
39. J.-S. Zhao and W.C. Chew, Integral equation solution of Maxwell's equations from zero frequency to microwave frequencies, *IEEE Trans. Antennas and Propagation*, 48(10), 1635–1645, 2000.
40. W.C. Chew, J.M. Jin, C.C. Lu, E. Michielssen, and J.M. Song, Fast solution methods in electromagnetics, *IEEE Trans. Antennas and Propagation*, 45(3), 533–543, 1997.
41. R.A. Brown, B.M. Notaros, B.D. Popovic, Z. Popovic, and J.P. Weem, Efficient large-domain MoM solutions to electrically large practical EM problems, *IEEE Trans. Microwave Theory Techniques*, 49(1), 151–159, 2001.
42. J.M. Johnson and Y. Rahmat-Samii, Genetic algorithms and method of moments (GA/MOM) for the design of integrated antennas, *IEEE Trans. Antennas and Propagation*, 47(10), 1606–1614, 1999.
43. K.R. Dandekar, H. Ling, and G. Xu, Experimental study of mutual coupling compensation in smart antenna applications, *IEEE Trans. on Wireless Communications*, 1(3), 480–487, 2002.
44. F. Ling, J. Liu, and J.-M. Jin, Efficient electromagnetic modeling of 3D multilayer microstrip antennas and circuits, *IEEE Trans. Microwave Theory Techniques*, 50(6), 1628–1635, 2002.
45. M.R. Abdul-Gafoor, H.K. Smith, A.A. Kishk, and A.W. Glisson, Simple and efficient full-wave modeling of electromagnetic coupling in realistic RF multilayer PCB layouts, *IEEE Trans. Microwave Theory Techniques*, 50(6), 1445–1457, 2002.
46. A.S. Barlevy and Y. Rahmat-Samii, Characterization of electromagnetic band-gaps composed of multiple periodic tripods with interconnecting vias: concept, analysis and design, *IEEE Trans. Antennas and Propagation*, 49(3), 343–353, 2001.
47. E.A. Soliman, M.H. Bakr, and N.K. Nikolova, Neural networks-method of moments (NN-MoM) for the efficient filling of the coupling matrix, *IEEE Trans. Antennas and Propagation*, 52(6), 1521–1529, 2004.
48. D. Colak and E.H. Newman, The multiple sweep method of moments (MSMM) design of wideband antennas, *IEEE Trans. Antennas and Propagation*, 46(9), 1365–1371, 1998.
49. D. Tonungrueng and E.H. Newman, The multiple sweep method of moments (MSMM) analysis of electrically large bodies, *IEEE Trans. Antennas and Propagation*, 45(8), 1252–1258, 1997.
50. K. Zhao, M.N. Vouvakis, and J.-F. Lee, The adaptive cross approximation algorithm for accelerated MoM computations of EMC problems, *IEEE Trans. Electromagnetic Compatibility*, 47(4), 763–773, 2005.
51. J. Blackburn and L.R. Arnaut, Numerical convergence in periodic method of moments analysis of frequency-selective surfaces based on wire elements, *IEEE Trans. Antennas and Propagation*, 53(10), 3308–3315, 2005.
52. K.A. Remlev and A. Weisshar, A versatile impedance boundary method of moments computational technique for solving the one-dimensional Schrodinger equation with application to quantum well and quantum wire problems, *IEEE Journal of Quantum Electronics*, 34(7), 1171–1179, 1998.
53. F. Wu, Y. Zhang, Z.Z. Oo, and E. Li, Parallel multilevel fast multipole method for solving large-scale problems, *IEEE Antennas and Propagation Magazine*, 47(4), 110–118, 2005.
54. G. Antonini, Fast multipole method for time domain PEEC analysis, *IEEE Trans. Mobile Computing*, 2(4), 275–287, 2003.
55. G. Brown, T.C. Schultness, D.M. Apalkov, and P.B. Visscher, Flexible fast multipole method for magnetic simulations, *IEEE Trans. Magnetics*, 40(4), Part 2, 2146–2148, 2004.

56. N. Bartoli, F. Collino, F. Dodu, and T. Koleck, A far-near field transformation using the fast multipole techniques, *IEEE Trans. Antennas and Propagation*, 52(12), 3329–3336, 2004.
57. S. Velamparambil and W.C. Chew, Analysis and performance of a distributed memory multilevel fast multipole algorithm, *IEEE Trans. Antennas and Propagation*, 53(8), Part 2, 2719–2727, 2005.
58. J.M. Song and W.C. Chew, Multilevel fast multipole algorithm for solving combined field integral equations of electromagnetic scattering, *Microw. Opt. Techniques Lett.*, 10, 14–19, 1995.
59. E. Yamashita and R. Mittra, Variational method for the analysis of microstrip line, *IEEE Trans. Microwave Theory Techniques*, 16, 251–256, 1968.
60. E.J. Denlinger, A frequency dependent solution for microstrip transmission lines, *IEEE Trans. Microwave Theory Techniques*, 19, 30–39, 1971.
61. T. Itoh and R. Mittra, Spectral-domain approach for calculating the dispersion characteristics of microstrip lines, *IEEE Trans. Microwave Theory Techniques*, 21, 496–499, 1973.
62. M.I. Aksun and R. Mittra, Choices of expansion and testing functions for MoM applied to a class of EM problems, *IEEE Trans. Microwave Theory Techniques*, 41, 503–509, 1993.
63. R.H. Jansen, Unified user-oriented computation of shielded, covered and open planar microwave and millimeter wave transmission-line characteristics, *IEE J. Microwave Opt. Acoust.*, 3, 14–22, 1979.
64. T. Itoh, Spectral domain immitance approach for dispersion characteristics of generalized printed transmission lines, *IEEE Trans. Microwave Theory Techniques*, 28, 733–736, 1980.
65. J. Sercu, N. Fache, F. Libbrecht, and D. De Zutter, Full-wave space-domain analysis of open microstrip discontinuities including the singular current-edge behavior, *IEEE Trans. Microwave Theory Techniques*, 41, 1581–1588, 1993.
66. F. Olyslager, D. De Zutter, and K. Blomme, Rigorous full-wave analysis of propagation characteristics of general lossless and lossy multiconductor transmission lines in multilayered media, *IEEE Trans. Microwave Theory Techniques*, 41, 79–88, 1993.
67. K.K.M. Cheng and J.K.A. Everard, A new technique for the quasi-TEM analysis of conductor-backed coplanar waveguide structures, *IEEE Trans. Microwave Theory Techniques*, 41, 1589–1592, 1993.
68. N. Gupta and M. Singh, Investigation of periodic structures in a fin line: a space-spectral domain approach, *IEEE Trans. Microwave Theory Techniques*, 43, 2708–2710, 1995.
69. Y. Imaizumi, M. Shinagawa, and H. Ogawa, Electric field distribution measurement of microstrip antennas and arrays using electro-optic sampling, *IEEE Trans. Microwave Theory Techniques*, 43, 2402–2407, 1995.
70. Y.-D. Lin, J.-W. Sheen, and C.-K.C. Tzuang, Analysis and design of feeding structures for microstrip leaky wave antenna, *IEEE Trans. Microwave Theory Techniques*, 44, 1540–1547, 1996.
71. P. Petre and M. Swaminathan, Spectral domain technique using surface wave excitation for the analysis of interconnects, *IEEE Trans. Microwave Theory Techniques*, 42, 1744–1749, 1994.
72. B.L. Ooi, M.S. Leong, P.S. Kooi, and T.S. Yeo, Enhancements of the spectral-domain approach for analysis of microstrip Y-junction, *IEEE Trans. Microwave Theory Techniques*, 45, 1800–1805, 1997.
73. K. Sabetfakhri and L.P.B. Katehi, Analysis of general class of open dielectric waveguides by spectral-domain technique, *Proc. IEEE-MTT Symposium*, 3, 1523–1526, 1993.
74. T. Itoh, Analysis of microstrip resonators, *IEEE Trans. Microwave Theory Techniques*, MTT-22, 946–952, 1974.
75. S.-H. Song, H.-B. Lee, H.-K. Jung, S.-Y. Hahn, K.-S. Lee, C. Cheon, and H.-S. Kim, Spectral domain analysis of the spiral inductor on multilayer substrates, *IEEE Trans. Magnetics*, 33(2), 1488–1491, 1997.
76. T.M. Weller, K.J. Herrick, and L.P.B. Katehi, Quasi-static design technique for mm-wave micromachined filters with lumped elements and series stubs, *IEEE Trans. Microwave Theory Techniques*, 45, 931–938, 1997.
77. Y.L. Chow, N. Hojjat, S. Safavi-Naeini, and R. Faraji-Dana, Spectral Green's functions for multilayer media in a convenient computational form, *IEE Proc. Microwave, Antennas and Propagation*, 145(1), 85–91, 1998.
78. M. Farina and T. Rozzi, Spectral domain approach to 2D modeling of open planar structures with thick lossy conductors, *IEE Proc. Microwave, Antennas and Propagation*, 147(5), 321–324, 2000.
79. M. Lecce and I. Montrosset, Analysis of integrated mode-locked semiconductor lasers: spectral domain approach, *IEE Proc. Optoelectronics*, 148(56), 266–272, 2001.

80. H. Yamamoto et al., Full-wave analysis for propagation characteristics of cylindrical coplanar waveguides with finite thickness of conductor, *IEEE Trans. Microwave Theory and Techniques*, 53(6-II), 2187–2195, 2005.
81. P. Baccarelli et al. Novel modal properties and relevant scanning behaviors of phased-arrays of microstrip leaky-wave antennas, *IEEE Trans. Antennas and Propagation*, 51(12), 3228–3238, 2003.
82. M. Vrancken and G.A.E. Vandenbosch, Hybrid dyadic-mixed-potential and combined spectral-space domain integral-equation analysis of quasi-3-D structures in stratified media, *IEEE Trans. Microwave Theory and Techniques*, 51(1-I), 216–225, 2003.
83. G. Plaza, F. Mesa, and F. Medina, On the use of SDA for the analysis of boxed planar lines with complex media, *IEEE Trans. Microwave Theory and Techniques*, 49(7), 1365–1368, 2001.
84. K.-Y. Su and J.-T. Kuo, An efficient analysis of shielded single and multiple coupled microstrip lines with the nonuniform Fast Fourier transform (NUFFT) technique, *IEEE Trans. Microwave Theory and Techniques*, 52(1-I), 90–96, 2004.
85. Y.C. Shih and K.G. Gray, Convergence of numerical solutions of step-type waveguide discontinuity problems by modal analysis, *Proc. IEEE-MTT Symposium*, 1, 233–235, 1983.
86. M. Leroy, On the convergence of numerical results in modal analysis, *IEEE Trans. Antennas and Propagation*, 31, 655–659, 1983.
87. A. Wexler, Solution of waveguide discontinuities by modal analysis, *IEEE Trans. Microwave Theory Techniques*, 15, 508–517, 1967.
88. W. Wessel, T. Sieverding, and F. Arndt, Mode-matching analysis of general waveguide multiport junctions, *Proc. IEEE-MTT Symposium*, 3, 1273–1276, 1999.
89. U. Balaji and R. Vahldieck, Mode-matching analysis of circular-ridged waveguide discontinuities, *IEEE Trans. Microwave Theory Techniques*, 46(2), 191–195, 1998.
90. K.L. Chan and S.R. Judah, Mode-matching analysis of a waveguide junction formed by a circular and a larger elliptic waveguide, *IEE Proc. Microwaves, Antennas and Propagation*, 145(1), 123–127, 1998.
91. A.S. Omar and K. Schunemann, Transmission matrix representation of finite discontinuity, *IEEE Trans. Microwave Theory Techniques*, 33, 830–835, 1985.
92. R. Vahldieck and W.J.R. Hoefer, Finline and metal insert filters with improved passband separation and increased stopband attenuation, *IEEE Trans. Microwave Theory Techniques*, 33, 1333–1339, 1985.
93. W. Menzel and I. Wolff, A method for calculating the frequency-dependent properties of microstrip discontinuities, *IEEE Trans. Microwave Theory Techniques*, 25, 107–112, 1977.
94. T.S. Chu, T. Itoh, and Y.-C. Shih, Comparative study of mode-matching formulations for microstrip discontinuity problems, *IEEE Trans. Microwave Theory Techniques*, 33, 1018–1023, 1985.
95. F. Alessandri, G. Baini, M. Mongiardo, and R. Sorrentino, A 3D mode matching technique for the efficient analysis of coplanar MMIC discontinuities with finite metallization thickness, *IEEE Trans. Microwave Theory Techniques*, 41, 1625–1629, 1993.
96. R. Schmidt and P. Russer, Modeling of cascaded coplanar waveguide discontinuities by the mode-matching approach, *IEEE Trans. Microwave Theory Techniques*, 43, 2910–2917, 1995.
97. Y.C. Shih, Design of waveguide E-plane filters with all metal insert, *IEEE Trans. Microwave Theory Techniques*, 32, 695–704, 1984.
98. F. Arndt et al. E-plane integrated filters with improved stopband attenuation, *IEEE Trans. Microwave Theory Techniques*, 32, 1391–1394, 1984.
99. F. Arndt et al. Computer-optimized multisection transforms between rectangular waveguides of adjacent frequency bands, *IEEE Trans. Microwave Theory Techniques*, 32, 1479–1484, 1984.
100. F. Arndt et al. Optimized E-plane T-junction series power divider, *IEEE Trans. Microwave Theory Techniques*, 35, 1052–1059, 1987.
101. R. Mehran, Computer-aided design of microstrip filters considering dispersion, loss and discontinuity effects, *IEEE Trans. Microwave Theory Techniques*, 27, 239–245, 1979.
102. L. Accatino, G. Bertin, and M. Mongiardo, Elliptical cavity resonators for dual-mode narrow-band filters, *IEEE Trans. Microwave Theory Techniques*, 45, 2393–2401, 1997.
103. J. Bornemann and F. Arndt, Rigorous design of evanescent-mode E-plane finned waveguide bandpass filters, *Proc. IEEE-MTT Symposium*, 2, 603–606, 1989.
104. O.A. Peverini, R. Tascone, M. Baralis, G. Virone, D. Trinchero, and R. Orta, Reduced-order optimized mode-matching CAD of microwave waveguide components, *IEEE Trans. Microwave Theory Tech.*, 52(1), Part 2, 311–318, 2004.

105. Z. Jiang and Z. Shen, Mode-matching analysis of large aperture coupling and its applications to the design of waveguide directional couplers, *IEE Proc. Microwaves, Antennas and Propagation*, 150(6), 422–428, 2003.
106. A.K. Bhattacharyya, On the convergence of MoM and mode matching solutions for infinite array and waveguide problems, *IEEE Trans. Antennas and Propagation*, 51(7), 1599–1606, 2003.
107. L. Prkna, M. Hubalek, and J. Ctyroky, Field modeling of circular microresonators by film mode matching, *IEEE Journal of Selected Topics in Quantum Electronics*, 11(1), 217–223, 2005.
108. D.C. Love and E.J. Rothwell, A mode-matching approach to determine the shielding properties of a doubly periodic array of rectangular apertures in a thick conducting screen, *IEEE Trans. Electromagnetic Compatibility*, 48(1), 121–133, 2006.
109. L. Prkna, M. Hubalek, and J. Ctyroski, Vectorial eigenmode solver for bent waveguides based on mode matching, *IEEE Photonics Technology Letters*, 16(9), 2057–2059, 2004.
110. S.P. Skobelev and P.S. Kildal, Mode-matching modeling of a hard conical quasi-TEM horn realized by an EBG structure with strips and vias, *IEEE Trans. Antennas and Propagation*, 53(1), Part 1, 139–143, 2005.
111. A.S. Jugessur, P. Pottier, and R.M. De La Rue, One-dimensional periodic photonic crystal microcavity filters with transition mode-matching features, embedded in ridge waveguides, *Electronics Letters*, 39(4), 367–369, 2003.
112. F.R. Cooray, Analysis of radiation from a horn with a superquadric aperture, *IEEE Trans. Antennas and Propagation*, 53(10), 3255–3261, 2005.
113. K.C. Hwang and H.J. Eom, Tunable notch filter of ferrite-filled grooves in parallel plates, *IEEE Microwave and Wireless Components Letters*, 15(5), 363–365, 2005.
114. S. Blair, and J. Goeckeritz, Effect of vertical mode matching on defect resonances in one-dimensional photonic crystal slabs, *Journal of Lightwave Technology*, 24(3), 1456–1461, 2006.
115. P. Bolli, G.G. Gentili, L. Lucci, R. Nesti, G. Pelosi, and G. Toso, A hybrid perturbative technique to characterize the coupling between a corrugated horn and a reflector dish, *IEEE Trans. Antennas and Propagation*, 54(2), Part 2, 595–603, 2006.
116. C.L. Holloway, M.A. Mohamed, E.F. Kuester, and A. Dienstfrey, Reflection and transmission properties of a metafilm: with an application to a controllable surface composed of resonant particles, *IEEE Trans. Electromagnetic Compatibility*, 47(4), 853–865, 2005.
117. T. Ueda and M. Tsutsumi, Left-handed transmission characteristics of rectangular waveguides periodically loaded with ferrite, *IEEE Trans. Magnetics*, 41(10), 3532–3537, 2005.
118. A. Ioffe, M. Thiel, and A. Dreher, Analysis of microstrip patch antennas on arbitrarily shaped multilayers, *IEEE Trans. Antennas and Propagation*, 51(8), 1929–1935,2003.
119. S. Tao, H. Heng-Tung, K.A. Zaki, A.E. Atia, and T.G. Dolan, Full-wave design of canonical waveguide filters by optimization, *IEEE Trans. Microwave Theory Tech.*, 51(2), Part. 1, 504–511, 2003.
120. T.X. Wu and D.L. Jaggard, Scattering of chiral periodic structure, *IEEE Trans. Antennas and Propagation*, 52(7), 1859–1870, 2004.
121. M. Ozkar and A. Mortazawi, Electromagnetic modeling and optimization of spatial power combiners/dividers with hard horns, *IEEE Trans. Antennas and Propagation*, 53(1), Part 1, 144–150, 2005.
122. J.A. Deibel, M.D. Escarra, and D.M. Mittleman, Photoconductive terahertz antenna with radial symmetry, *Electronics Letters*, 41(5), 226–228, 2005.
123. C. Acar and A.M. Shkel, Nonresonant micromachined gyroscopes with structural mode-decoupling, *IEEE Sensors Journal*, 3(4), 497–506, 2003.
124. A.P. Moneda, M.P. Ioannidou, and D.P. Chrissoulidis, Radio-wave exposure of the human head: analytical study based on a versatile eccentric spheres model including a brain core and a pair of eyeballs, *IEEE Trans. Biomedical Engineering*, 50(6), 667–676, 2003.
125. A. Alu and N. Engheta, Guided modes in a waveguide filled with a pair of single-negative (SNG), double-negative (DNG), and/or double-positive (DPS) layers, *IEEE Trans. Microwave Theory Tech.*, 52(1), Part 1, 199–210, 2004.
126. S. Maci, M. Caiazzo, A. Cucini, and M. Casaletti, A pole-zero matching method for EBG surfaces composed of a dipole FSS printed on a grounded dielectric slab, *IEEE Trans. Antennas and Propagation*, 53(1), Part 1, 70–81, 2005.
127. K.S. Yee, Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media, *IEEE Trans. Antennas and Propagation*, 14, 302–307, 1966.

128. R. Holland, THREDE: A free-field EMP coupling and scattering code, *IEEE Trans. Nuclear Science*, 24, 2416–2421, 1977.
129. A. Taflove and M.E. Brodwin, Numerical solution of steady-state electromagnetic scattering problems using the time-dependent Maxwell's equations, *IEEE Trans. Microwave Theory Techniques*, 23, 623–630, 1975.
130. J.-P. Berenger, A perfectly matched layer for the absorption of electromagnetic waves, *Computational Physics*, 114, 185–200, 1994.
131. D.S. Katz, E.T. Thiele, and A. Taflove, Validation and extension to three dimensions of the Berenger PML absorbing boundary conditions for FDTD meshes, *IEEE Microwave and Guided Wave Letters*, 4, 344–346, 1994.
132. E. Tentzeris, M. Krumpholz, N. Dib, J.-G. Yook, and L.P.B. Katehi, FDTD characterization of waveguide probe structures, *IEEE Trans. Microwave Theory Techniques*, 46(10), 1452–1460, 1998.
133. M. Werthen, M. Rittweger, and I. Wolff, FDTD simulation of waveguide junctions using a new boundary condition for rectangular waveguides, *Proc. 24th European Microwave Conf.*, Cannes, France, 1715–1719, 1994.
134. A. Taflove, Computation of the electromagnetic fields and induced temperatures within a model of the microwave-irradiated human eye, Ph.D. Dissertation, Department of Electrical Engineering, Northwestern University, Evanston, IL, June 1975.
135. J. Chen, C. Wu, T.K. Lo, K.-L. Wu, and J. Litva, Using linear and nonlinear predictors to improve the computational efficiency of the FDTD algorithm, *IEEE Trans. Microwave Theory Techniques*, 42, 1992–1997, 1994.
136. V. Jandhyala, E. Michielssen, and R. Mittra, On the performance of different AR methods in the spectral estimation of FDTD waveforms, *Microwave and Optical Technology Letters*, 7, 690–692, 1994.
137. M. DePourcq, Field and power density calculations in closed microwave systems by 3D finite differences, *IEE Proc. H: Microwaves, Antennas and Propagation*, 132, 360–368, 1985.
138. E.A. Navarro, V. Such, B. Gimeno, and J.L. Cruz, T-junctions in square coaxial waveguide: an FDTD approach, *IEEE Trans. Microwave Theory Techniques*, 42, 347–350, 1994.
139. C. Wang, B.-Q. Gao, and C.-P. Ding, Q factor of a resonator by the FDTD method incorporating perturbation techniques, *Electronics Letters*, 29, 1866–1867, 1993.
140. G.-C. Liang, Y.-W. Liu, and K.K. Mei, Full-wave analysis of coplanar waveguide and slotline using FDTD, *IEEE Trans. Microwave Theory Techniques*, 37, 1949–1957, 1989.
141. D.M. Sheen, S.M. Ali, M.D. Abouzahra, and J.A. Kong, Application of the 3D FDTD method to the analysis of planar microstrip circuits, *IEEE Trans. Microwave Theory Techniques*, 38, 849–857, 1990.
142. A.C. Cangellaris and D.B. Wright, Analysis of the numerical error caused by the stair-stepped approximation of a conducting boundary in FDTD simulations of electromagnetic phenomena, *IEEE Trans. Antennas and Propagation*, 39, 1518–1525, 1991.
143. C.-W. Lam, S.M. Ali, and P. Nuytkens, Three-dimensional modeling of multichip module interconnects, *IEEE Trans. Components, Hybrids and Manufacturing Technology*, 16, 699–704, 1993.
144. M. Piket-May, A. Taflove, and J. Baron, FDTD modeling of digital signal propagation in 3D circuits with passive and active loads, *IEEE Trans. Microwave Theory Techniques*, 42, 1514–1523, 1994.
145. R.J. Luebbers, K.S. Kunz, M. Schneider, and F. Hunsberger, An FDTD near zone to far zone transformation, *IEEE Trans. Antennas and Propagation*, 39, 429–433, 1991.
146. K.L. Shlager and G.S. Smith, Comparison of two FDTD near-field to near-field transformations applied to pulsed antenna problems, *Electronics Letters*, 31, 936–938, 1995.
147. J.G. Maloney, G.S. Smith, and W.R. Scott, Jr., Accurate computation of the radiation from simple antennas using the FDTD method, *IEEE Trans. Antennas and Propagation*, 38, 1059–1069, 1990.
148. R. Luebbers and K. Kunz, FDTD calculations of antenna mutual coupling, *IEEE Trans. Electromagnetic Compatibility*, 34, 357–359, 1992.
149. R.J. Luebbers and J. Beggs, FDTD calculation of wideband antenna gain and efficiency, *IEEE Trans. Antennas and Propagation*, 40, 1403–1407, 1992.
150. P.A. Tirkas and C.A. Balanis, FDTD method for antenna radiation, *IEEE Trans. Antenna and Propagation*, 40, 334–340, 1992.
151. K. Uehara and K. Kagoshima, Rigorous analysis of microstrip phased array antennas using a new FDTD method, *Electronics Letters*, 30, 100–101, 1994.

152. M.A. Jensen and Y. Rahmat-Samii, Performance analysis of antennas for hand-held transceivers using FDTD, *IEEE Trans. Antennas and Propagation*, 42, 1106–1113, 1994.
153. A Taflove and K.R. Umashankar, Radar cross section of general three-dimensional scatterers, *IEEE Trans. Electromagnetic Compatibility*, 25, 433–440, 1983.
154. C.L. Britt, Solution of electromagnetic scattering problems using time-domain techniques, *IEEE Trans. Antennas and Propagation*, 37, 1181–1192, 1989.
155. N. Bushyager, B. McGarvey, and E.M. Tentzeris, Introduction of an adaptive modeling technique for the simulation of RF structures requiring the coupling of Maxwell's, mechanical and solid-state equations, *Applied Computational Electromagnetics (ACES) Society Journal*, 17(1), 104–111, 2002.
156. N. Bushyager, K. Lange, M. Tentzeris, and J. Papapolymerou, Modeling and optimization of RF-MEMS reconfigurable tuners with computationally efficient time-domain techniques, *Proc. 2002 IEEE-IMS Symposium*, Seattle, WA, II.883–886, 2002.
157. W. Sui, D.A. Christensen, and C.H. Durney, Extending the 2D FDTD method to hybrid electromagnetic systems with active and passive lumped elements, *IEEE Trans. Microwave Theory Techniques*, 40, 724–730, 1992.
158. P. Ciampolini, P. Mezzanotte, L. Roselli, and R. Sorrentino, Accurate and efficient circuit simulation with lumped-element FDTD, *IEEE Trans. Microwave Theory Techniques*, 44, 2207–2215, 1996.
159. C.-N. Kuo, B. Housmand, and T. Itoh, Full-wave analysis of packaged microwave circuits with active and nonlinear devices: an FDTD approach, *IEEE Trans. Microwave Theory Techniques*, 45, 819–826, 1997.
160. M.A. Alsunaidi, S.M. Sohel Imtiaz, and S.M. El-Ghazaly, Electromagnetic wave effects on microwave transistors using a full-wave time-domain model, *IEEE Trans. Microwave Theory Techniques*, 44, 799–808, 1996.
161. V.A. Thomas, M.E. Jones, M. Piket-May, A. Taflove, and E. Harrigan, The use of SPICE lumped circuits as sub-grid models for FDTD analysis, *IEEE Microwave and Guided Wave Letters*, 4, 141–143, 1994.
162. S. Dey and R. Mittra, A locally conformal finite-difference time-domain algorithm for modeling three-dimensional perfectly conducting objects, *IEEE Microwave and Guided Wave Letters*, 7, 273–275, 1997.
163. S. Dey and R. Mittra, A modified locally conformal finite-difference time-domain algorithm for modeling three-dimensional perfectly conducting objects, *Microwave and Optical Technology Letters*, 17, 349–352, 1998.
164. J.G. Maloney and M.P. Kesler, Analysis of Periodic Structures, Chapter 6 in "Advances in Computational Electrodynamics: The Finite-Difference Time-Domain Method", A. Taflove, Ed., Norwood, MA, Artech House, 1998.
165. O.R. Painter, K. Lee, A. Scherer, A. Yariv, J.D. O'Brien, P.D. Dapkus, and I. Kim, Two-dimensional photonic band-gap defect mode laser, *Science*, 284, 1819–1821, 1999.
166. F. Zheng, Z. Chen, and J. Zhang, Towards the development of a three-dimensional stable finite-difference time-domain method, *IEEE Trans. Microwave Theory Tech.*, 48(9), 1550–1558, 2000.
167. H. Rao, R. Scarmozzino, and R.M. Osgood Jr., An improved ADI-FDTD method and its application to photonic simulations, *IEEE Photonics Technology Letters*, 14(4), 477–479, 2002.
168. S.W. Staker, C.L. Holloway, A.U. Bhobe, and M. Piket-May, Alternating-direction implicit (ADI) formulation of the finite-difference time-domain (FDTD) method: algorithm and material dispersion implementation, *IEEE Trans. Electromagnetic Compatibility*, 45(2), 156–166, 2003.
169. C.T.M. Choi and S.-H. Sun, Numerical performance and applications of the envelope ADI–FDTD method, *IEEE Trans. Microwave Theory and Techniques*, 54(1), 256–264, 2006.
170. M. Changning and Z. Chen, Dispersion analysis of the three-dimensional complex envelope ADI-FDTD method, *IEEE Trans. Antennas and Propagation*, 53(3), 971–976, 2005.
171. P. Kosmas and C.M. Rappaport, FDTD-based time reversal for microwave breast cancer detection—localization in three dimensions, *IEEE Trans. Microwave Theory and Techniques*, 54(4), Part. 2, 1921–1927, 2006.
172. K. Chamberlin and D. Vidacic, Analysis of finite-differencing errors to determine cell size when modeling ferrites and other lossy electric and magnetic materials using FDTD, *IEEE Trans. Electromagnetic Compatibility*, 46, 617–623, 2004.

173. A. Akyurtlu and D.H. Werner, Modeling of transverse propagation through a uniaxial bianisotropic medium using the finite-difference time-domain technique, *IEEE Trans. Antennas and Propagation*, 52(12), 3273–3279, 2004.
174. Kwan-Ho Lee, Chi-Chih Chen, F.L. Teixeira, and R. Lee, Modeling and investigation of a geometrically complex UWB GPR antenna using FDTD, *IEEE Trans. Antennas and Propagation*, 52(8), 1983–1991, 2004.
175. Gui-Rong Zhou and Xun Li, Wave equation-based semivectorial compact 2-D-FDTD method for optical waveguide modal analysis, *Journal of Lightwave Technology*, 22(2), 677–683, 2004.
176. J. Shibayama, M. Muraki, J. Yamauchi, and H. Nakano, Comparative study of several time-domain methods for optical waveguide analyses, *Journal of Lightwave Technology*, 23(7), 2285–2293, 2005.
177. En-Xiao Liu, Er-Ping Li, and Le-Wei Li, Hybrid FDTD-MPIE method for the simulation of locally inhomogeneous multilayer LTCC structure, *IEEE Microwave and Wireless Components Letters*, 15(1), 42–44, Jan. 2005.
178. S.V. Georgakopoulos, C.R. Birtcher, C.A. Balanis, and R.A. Renaut, HIRF penetration and PED coupling analysis for scaled fuselage models using a hybrid subgrid FDTD(2,2)/FDTD(2,4) method, *IEEE Trans. Electromagnetic Compatibility*, 45(2), 293–305, 2003.
179. S. Schmidt and G. Lazzi, Use of the FDTD thin-strut formalism for biomedical telemetry coil designs, *IEEE Trans. Microwave Theory and Techniques*, 52(8), Part 2, 1952–1956, 2004.
180. A. Karwowski, Improving accuracy of FDTD simulations in layered biological tissues, *IEEE Microwave and Wireless Components Letters*, 14(4), 151–152, 2004.
181. J.P. Berenger, Long range propagation of lightning pulses using the FDTD method, *IEEE Trans. Electromagnetic Compatibility*, 47(4), 1008–1011, 2005.
182. M.K. Karkkainen and P.M.T. Ikonen, Finite-difference time-domain modeling of frequency selective surfaces using impedance sheet conditions, *IEEE Trans. Antennas and Propagation*, 53(9), 2928–2937, 2005.
183. P.F. Hsieh, T.-T. Wu, and J.-H. Sun, Three-dimensional phononic band gap calculations using the FDTD method and a PC cluster system, *IEEE Trans. Ultrasonics, Ferroelectrics and Frequency Control*, 53(1), 148–158, 2006.
184. H.-R. Chuang and L.-C. Kuo, 3-D FDTD design analysis of a 2.4-GHz polarization-diversity printed dipole antenna with integrated balun and polarization-switching circuit for WLAN and wireless communication applications, *IEEE Trans. Microwave Theory Techniques*, 51(2), Part 1, 374–381, 2003.
185. G. Emili, A. Schiavoni, M. Francavilla, L. Roselli, and R. Sorrentino, Computation of electromagnetic field inside a tissue at mobile communications frequencies, *IEEE Trans. Microwave Theory Techniques*, 51(1), Part. 1, 178–186, 2003.
186. B.-Z. Wang, Z.-J. Yu, W. Shao, and P.-P. Ding, Analyses of bounces on power distribution networks using an unconditionally stable finite-difference time-domain method, *IEEE Trans. Mobile Computing*, 5(1), 20–27, 2006.
187. J.J. Simpson and A. Taflove, Three-dimensional FDTD modeling of impulsive ELF propagation about the earth-sphere, *IEEE Trans. Antennas and Propagation*, 52(2), 443–451, 2004.
188. W. Hu and S.A. Cummer, An FDTD model for low and high altitude lightning-generated EM fields, *IEEE Trans. Antennas and Propagation*, 54(5), 1513–1522, 2006.
189. A. Akyurtlu and D.H. Werner, A novel dispersive FDTD formulation for modeling transient propagation in chiral metamaterials, *IEEE Trans. Antennas and Propagation*, 52(9), 2267–2276, 2004.
190. T. Kokkinos, C.D. Sarris, and G.E. Eleftheriades, Periodic FDTD analysis of leaky-wave structures and applications to the analysis of negative-refractive-index leaky-wave antennas, *IEEE Trans. Microwave Theory Techniques*, 54(4), Part 1, 1619–1630, 2006.
191. B. Donderici and F.L. Teixeira, Improved FDTD subgridding algorithms via digital filtering and domain overriding, *IEEE Trans. Antennas and Propagation*, 53(9), 2938–2951, 2005.
192. C.T.M. Choi and S.-H. Sun, A new multilevel subgridding scheme for two-dimensional FDTD method, *IEEE Trans. Magnetics*, 40(2), Part 2, 1025–1028, 2004.
193. S.-H. Sun and C.T.M. Choi, A new subgridding scheme for two-dimensional FDTD and FDTD (2,4) methods, *IEEE Trans. Magnetics*, 40(2), Part 2, 1041–1044, 2004.

194. J.P. Durbano, F.E. Ortiz, J.R. Humphrey, M.S. Mirotznik, and D.W. Prather, Hardware implementation of a three-dimensional finite-difference time-domain algorithm, *IEEE Antennas and Wireless Propagation Letters*, 2(1), 54–57, 2003.
195. Y.S. Rickard and N.K. Nikolova, Off-grid perfect boundary conditions for the FDTD method, *IEEE Trans. Microwave Theory Techniques*, 53(7), 2274–2283, 2005.
196. M.K. Karkkainen, Subcell FDTD modeling of electrically thin dispersive layers, *IEEE Trans. Microwave Theory Techniques*, 51(6), 1774–1780, 2003.
197. N.K. Nikolova, H.W. Tam, and M.H. Bakr, Sensitivity analysis with the FDTD method on structured grids, *IEEE Trans. Microwave Theory Techniques*, 52(4), 1207–1216, 2004.
198. S. Akhtarzad and P.B. Johns, Three-dimensional transmission-line matrix computer analysis of microstrip resonators, *IEEE Trans. Microwave Theory Techniques*, MTT-23, 990–997, 1975.
199. W.J.R. Hoefer, The transmission-line matrix method—theory and applications, *IEEE Trans. Microwave Theory Techniques*, MTT-33, 882–893, 1985.
200. C. Christopoulos, *The Transmission-Line Modeling Method*, IEEE/OUP, London, UK, 1995.
201. P.B. Johns, A symmetrical condensed node for the TLM-method, *IEEE Trans. Microwave Theory Techniques*, 35, 370–377, 1987.
202. L.R.A.X. de Menezes and W.J.R. Hoefer, Modeling of general (nonlinear) constitutive relationships in SCN TLM, *IEEE Trans. Microwave Theory Techniques*, MTT-44, 854–861, 1996.
203. M. Krumpholz and P. Russer, Discrete time-domain Green's functions for three-dimensional TLM modeling of the radiating boundary conditions, *9th Annual Review of Progress in ACES Digest*, 2, 458–466, 1993.
204. M. Righi and W.J.R. Hoefer, Efficient 3D-SCN-TLM diakoptics for waveguide components, *IEEE Trans. Microwave Theory Techniques*, 42, 2381–2384, 1994.
205. M. Krumpholz, C. Huber, and P. Russer, A field theoretical comparison of FDTD and TLM, *IEEE Trans. Microwave Theory Techniques*, 43, 1935–1950, 1995.
206. C. Eswarappa and W.J.R. Hoefer, One-way equation absorbing boundary condition for 3D TLM analysis of planar and quasi-planar structures, *IEEE Trans. Microwave Theory Techniques*, 42, 1669–1677, 1994.
207. J.L. Dubard and D. Pompei, A modified 3D-TLM variable node for the Berenger's perfectly matched layer implementation, *13th Annual Review of Progress in ACES Digest*, 3, 661–665, 1997.
208. L. Pierantoni, C. Tomassoni, and T. Rozzi, A new termination condition for the application of the TLM method to discontinuity problems in closed homogeneous waveguide, *IEEE Trans. Microwave Theory Tech.*, 50(11), 2513–2518, 2002.
209. M.I. Sobhy, M.W.R. Ng, R.J. Langley, and J.C. Batchelor, TLM simulation of patch antenna on magnetized ferrite substrate, *16th Annual Review of Progress in ACES Digest*, 562–569, 2000.
210. J.-L. Dubard and D. Pompei, Optimization of the PML efficiency in 3D TLM method, *IEEE Transactions on Microwave Theory Techniques*, 48(7), 1081–1088, 2000.
211. A.P. Duffy, J.L. Herring, T.M. Benson, and C. Christopoulos, Improved wire modeling in TLM, *IEEE Trans. Microwave Theory Techniques*, 42, 1978–1983, 1994.
212. F. Coccetti, L. Vietzorreck, V. Chtchekatourov, and P. Russer, A numerical study of MEMS capacitive switches using TLM, *16th Annual Review of Progress in ACES Digest*, 3, 580–587, 2000.
213. P.P.M. So and W.J.R. Hoefer, A TLM-SPICE interconnection framework for coupled field and circuit analysis in time-domain, *IEEE Trans. Microwave Theory Techniques*, 50(12), 2728–2733, 2002.
214. J. Paul, C. Christopoulos, and D.W.P. Thomas, Generalized material models in TLM—part I: materials with frequency-dependent properties, *IEEE Trans. Antennas and Propagation*, 47(10), 1528–1534, 1999.
215. J. Paul, C. Christopoulos, and D.W.P. Thomas, Generalized material models in TLM—part II: materials with anisotropic properties, *IEEE Trans. Antennas and Propagation*, 47(10), 1535–1542, 1999.
216. J. Paul, C. Christopoulos, and D.W.P. Thomas, Generalized material models in TLM—part III: materials with nonlinear properties, *IEEE Trans. Antennas and Propagation*, 50(7), 997–1004, 2002.
217. H. Dominguez, A. Raizer, and W.P. Carpes Jr., Electromagnetic fields radiated by a cellular phone in close proximity to metallic walls, *IEEE Trans. Magnetics*, 38(2), 793–796, 2002.
218. M. Walter, O. Pertz, and A. Beyer, A contribution to the modeling of longitudinally periodic waveguides by the help of the TLM method, *IEEE Trans. Microwave Theory Tech.*, 48(9), 1574–1576, 2000.

219. L. Pierantoni, A. DiDonato, and T. Rozzi, Full-wave analysis of photonic bandgap integrated optical components by the TLM-IE method, *Journal of Lightwave Technology*, 22(10), 2348–2358, 2004.
220. P.P.M. So, D. Huilian, and W.J.R. Hoefer, Modeling of metamaterials with negative refractive index using 2-D shunt and 3-D SCN TLM networks, *IEEE Trans. Microwave Theory Techniques*, 53(4), Part 2, 1496–1505, 2005.
221. P. Lorenz, J.V. Vital, B. Biscontini, and P. Russer, TLM-G-a grid-enabled time-domain transmission-line-matrix system for the analysis of complex electromagnetic structures, *IEEE Trans. Microwave Theory Techniques*, 53(11), 3631–3637, 2005.
222. P. Abolghasem, M.H. Bakr, W.J.R. Hoefer, and P.P.M. So, A transformed TLM formulation for adjoint variable-based sensitivity analysis, *IEEE Trans. Microwave and Wireless Components Letters*, 15(11), 817–819, 2005.
223. S.M. Ali, N.K. Nikolova, and M.H. Bakr, Sensitivity analysis with full-wave electromagnetic solvers based on structured grids, *IEEE Trans. Magnetics*, 40(3), 1521–1529, 2004.
224. P. Sewell, T.M. Benson, C. Christopoulos, D.W.P. Thomas, A. Vukovic, and J.G. Wykes, Transmission-line modeling (TLM) based upon unstructured tetrahedral meshes, *IEEE Trans. Microwave Theory Techniques*, 53(6), Part 1, 1919–1928, 2005.
225. P. Sewell, J.G. Wykes, T.M. Benson, C. Christopoulos, D.W.P. Thomas, and A. Vukovic, Transmission-line modeling using unstructured triangular meshes, *IEEE Trans. Microwave Theory Techniques*, 52(5), 1490–1497, 2004.
226. H. Tosi, J.L. Dubard, and J.-M. Laheurte, TLM modelling of reconfigurable antennas for multi-band operation, *IEE Proc. Microwaves, Antennas and Propagation*, 153(2), 147–151, 2006.
227. M.I. Yaich and M. Khalladi, An SCN-TLM model for the analysis of ferrite media, *IEEE Microwave and Wireless Components Letters*, 13(6), 217–219, 2003.
228. P. Sewell, Y.K. Choong, and C. Christopoulos, An accurate thin-wire model for 3-D TLM simulations, *IEEE Trans. Electromagnetic Compatibility*, 45(2), 207–217, May 2003.
229. K. Michishita, M. Ishii, A. Asakawa, S. Yokoyama, and K. Kami, Voltage induced on a test distribution line by negative winter lightning strokes to a tall structure, *IEEE Trans. Electromagnetic Compatibility*, 45(1), 135–140, 2003.
230. V. Janvani, J.D. Paul, A. Vukovic, T.M. Benson, and P. Sewell, TLM modelling of nonlinear optical effects in fibre Bragg gratings, *IEE Procs. Optoelectronics*, 151(4), 185–192, 2004.
231. D. Pasalic and R.Vahldieck, A hybrid drift-diffusion-TLM analysis of traveling-wave photodetectors, *IEEE Trans. Microwave Theory Techniques*, 53(9), 2700–2706, 2005.
232. D.H. Schaubert, D.R. Wilton, and A.W. Glisson, A tetrahedral method for electromagnetic scattering by arbitrarily shaped inhomogeneous dielectric bodies, *IEEE Trans. Antennas and Propagation*, 32, 77–85, 1984.
233. Z.J. Cendes, Vector finite elements for electromagnetic field computation, *IEEE Trans. Magnetics*, 27, 3958–3966, 1991.
234. A. Chatterjee, J.M. Jin, and J.L. Volakis, Edge-based finite elements and vector ABC's applied to 3D scattering, *IEEE Trans. Antennas and Propagation*, 41, 221–226, 1993.
235. A.F. Peterson and D.R. Wilton, Curl-conforming mixed-order edge elements for discretizing the 2D and 3D vector Helmholtz equation, in *Finite Element Software for Microwave Engineering*, T. Itoh, G. Pelosi and P.P. Silvester, Eds., Wiley, New York, 1996.
236. Z. Pantic-Tanner, J.S. Savage, D.R. Tanner, and A.F. Peterson, Two-dimensional singular vector elements for finite-element analysis, *IEEE Trans. Microwave Theory Techniques*, 46, 178–184, 1998.
237. R. Lee and A.C. Cangellaris, A study of discretization error in the finite element approximation of wave solutions, *IEEE Trans. Antennas and Propagation*, 40, 542–549, 1992.
238. G.S. Warren and W.R. Scott, An investigation of numerical dispersion in the vector finite element method, *IEEE Trans. Antennas and Propagation*, 42, 1502–1508, 1994.
239. L.W. Pearson, R.A. Whitaker, and L.J. Bahrmasel, An exact radiation boundary condition for the finite-element solution of electromagnetic scattering on an open domain, *IEEE Trans. Magnetics*, 25, 3046–3048, 1989.
240. J.P. Webb and V.N. Kanellopoulos, A numerical study of vector absorbing boundary conditions for the finite element solution of Maxwell's equations, *IEEE Microwave Guided Wave Letters*, 1, 325–327, 1991.
241. W. Sun and C.A. Balanis, Vector one-way absorbing boundary conditions for FEM applications, *IEEE Trans. Antennas and Propagation*, 42, 872–878, 1994.

242. J.M. Jin and V.V. Liepa, A note on hybrid finite element for solving scattering problems, *IEEE Trans. Antennas and Propagation*, 36, 1486–1490, 1988.
243. J.S. Wang, Analysis of microstrip discontinuities based on edge-element 3D FEM formulation and absorbing boundary conditions, *Proc. IEEE-MTT Symposium*, 2, 745–748, 1993.
244. J.-S. Wang and R. Mittra, Finite element analysis of MMIC structures and electronic packages using absorbing boundary conditions, *IEEE Trans. Microwave Theory Techniques*, 42, 441–449, 1994.
245. J.-G. Yook, N.I. Dib, and L.P.B. Katehi, Characterization of high frequency interconnects using finite difference time-domain and finite element methods, *IEEE Trans. Microwave Theory Techniques*, 42, 1727–1736, 1994.
246. A.C. Polycarpou, M.R. Lyons, and C.A. Balanis, Finite element analysis of MMIC waveguide structures with anisotropic substrates, *IEEE Trans. Microwave Theory Techniques*, 44, 1650–1663, 1996.
247. Z. Li, J.L. Volakis, and P.Y. Papalambros, Optimization of patch antennas on ferrite substrate using the finite element methods, *Proc. 1999 IEEE AP-Symposium,* II, 1026–1029, 1999.
248. Y. Zhu and A.C. Cangellaris, Robust finite-element solution of lossy and unbounded electromagnetic eigenvalue problems, *IEEE Trans. Microwave Theory Techniques*, 50(10), 2331–2338, 2002.
249. E.S. Hung and S.D. Senturia, Generating efficient dynamical models for MEMS from a few finite-element simulation runs, *Journal of Microelectromechanical Systems*, 8(3), 280–289, 1999.
250. A. Ammous, S. Ghedira, B. Allard, H. Morel, and D. Renault, Choosing a thermal model for electro-thermal simulation of power semiconductor devices, *IEEE Trans. Power Electronics*, 14(2), 300–307, 1999.
251. L.S. Andersen and J.L. Volakis, Adaptive multiresolution antenna modeling using hierarchical mixed-order tangential vector finite elements, *IEEE Trans. Antennas and Propagation*, 49(2), 211–222, 2001.
252. Y. Zhu and A.C. Cangellaris, Hierarchical multilevel potential preconditioner for fast finite-element analysis of microwave devices, *IEEE Trans. Microwave Theory Techniques*, 50(8), 1984–1989, 2002.
253. P. Liu, Y.-Q. Jin, An FEM approach with FFT accelerated iterative robin boundary condition for electromagnetic scattering of a target with strong or weak coupled underlying randomly rough surface, *IEEE Trans. Antennas and Propagation*, 53(12), 4137–4144, 2005.
254. G. Aiello, S. Alfonzetti, and E. Dilettoso, Finite-element solution of eddy-current problems in unbounded domains by means of the hybrid FEM-DBCI method, *IEEE Trans. Magnetics*, 39(3), Part 1, 1409–1412, 2003.
255. V.A. Chiriac and T.-Y.T. Lee, Thermal assessment of RF-integrated LTCC front end modules, *IEEE Trans. Advanced Packaging*, 27(3), 545–557, 2004.
256. M.M. Ilic, A.Z. Ilic, and B.M. Notaros, Higher order large-domain FEM modeling of 3-D multiport waveguide structures with arbitrary discontinuities, *IEEE Trans. Microwave Theory Techniques*, 52(6), 1608–1614, 2004.
257. G. Xu, H. Wu, S. Yang, S. Liu, L. Ying, Q. Yang, W. Yan, and M. Wang, 3-D electrical impedance tomography forward problem with finite element method, *IEEE Trans. Magnetics*, 41(5), 1832–1835, 2005.
258. P. Sergeant, U. Adriano, L. Dupre, O. Bottauscio, M. De Wulf, M. Zucca, and J.A.A. Melkebeek, Passive and active electromagnetic shielding of induction heaters, *IEEE Trans. Magnetics*, 40(2), Part 2, 675–678, 2004.
259. L. Pierantoni, A. Massaro, and T. Rozzi, Efficient modeling of 3-D photonic crystals for integrated optical devices, *IEEE Photonics Technology Letters*, 18(2), 319–321, 2006.
260. M.L. Palmeri, A.C. Sharma, R.R. Bouchard, R.W. Nightingale, and K.R. Nightingale, A finite-element method model of soft tissue response to impulsive acoustic radiation force, *IEEE Trans. Ultrasonics, Ferroelectrics and Frequency Control*, 52(10), 1699–1712, 2005.
261. J.-H. Lee and Q.H. Liu, An efficient 3-D spectral-element method for Schrodinger equation in nanodevice simulation, *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, 24(12), 1848–1858, 2005.
262. L.A. Rocha, E. Cretu, and R.F. Wolffenbuttel, Analysis and analytical modeling of static pull-in with application to MEMS-based voltage reference and process monitoring, *Journal of Microelectromechanical Systems*, 13(2), 342–354, 2004.
263. V. Schulz, Adjoint high-order vectorial finite elements for nonsymmetric transversally anisotropic waveguides, *IEEE Trans. Microwave Theory Techniques*, 51(4), Part 1, 1086–1095, 2003.

264. F. Bertazzi, F. Cappelluti, S. Donati Guerrieri, S.F. Bonani, and G. Ghione, Self-consistent coupled carrier transport full-wave EM analysis of semiconductor traveling-wave devices, *IEEE Trans. Microwave Theory Tech.*, 54(4), Part 1, 1611–1618, 2006.
265. L.E.R. Peterson and J.-M. Jin, Analysis of periodic structures via a time-domain finite-element formulation with a Floquet ABC, *IEEE Trans. Antennas and Propagation*, 54(3), 933–944, 2006.
266. S.N. Makarov, S.D. Kulkarni, A.G. Marut, and L.C. Kempel, Method of moments solution for a printed patch/slot antenna on a thin finite dielectric substrate using the volume integral equation [volume integral equation versus FEM], *IEEE Trans. Antennas and Propagation*, 54(4), 1174–1184, 2006.
267. X. Yang, Y. Wang, F. Liu, Q. Yang, and W. Yan, The use of neural networks combined with FEM to optimize the coil geometry and structure of transverse flux induction equipment, *IEEE Trans. Applied Superconductivity*, 14(2), 1854–1857, 2004.
268. W. Wang, Y. Lu, J.S. Fu, and Y.Z. Xiong, Particle swarm optimization and finite-element based approach for microwave filter design, *IEEE Trans. Magnetics*, 41(5), 1800–1803, 2005.
269. M.M. Ilic, A.Z. Ilic, and B.M. Notaros, Efficient large-domain 2-D FEM solution of arbitrary waveguides using p-refinement on generalized quadrilaterals, *IEEE Trans. Microwave Theory Techniques*, 53(4), Part. 1, 1377–1383, 2005.
270. S.-C. Lee, J.-F. Lee, and R. Lee, Hierarchical vector finite elements for analyzing waveguiding structures, *IEEE Trans. Microwave Theory Techniques*, 51(8), 1897–1905, 2003.
271. N. Marais and D.B. Davidson, Numerical evaluation of hierarchical vector finite elements on curvilinear domains in 2-D, *IEEE Trans. Antennas and Propagation*, 54(2), Part 2, 734–738, 2006.
272. K. Watanabe, H. Igarashi, and T. Honma, Comparison of geometric and algebraic multigrid methods in edge-based finite-element analysis, *IEEE Trans. Magnetics*, 41(5), 1672–1675, 2005.
273. C.A. Brebbia, *The Boundary Element Method for Engineers*, Pentech Press, London, 1978.
274. S. Kagami and I. Fukai, Application of boundary element method to EM field problems, *IEEE Trans. Microwave Theory Techniques*, 32, 455–461, 1984.
275. H. Cam, S. Toutain, P. Gelin, and G. Landrac, Study of Fabry–Perot cavity in the microwave frequency range by the boundary element method, *IEEE Trans. Microwave Theory Techniques*, 40, 298–304, 1992.
276. T.F. Eibert and V. Hansen, 3-D FEM/BEM-hybrid approach based on a general formulation of Huygen's principle for planar layered media, *IEEE Trans. Microwave Theory Techniques*, 45, 1105–1112, 1997.
277. K. Guillouard, M.F. Wong, V. Fouad Hanna, and J. Citerne, A new global time domain electromagnetic simulator of microwave circuits including lumped elements based on finite element method, *Proc. IEEE-MTT Symposium*, 3, 1239–1242, 1997.
278. K. Zhao, M.N. Vouvakis, and J.-F. Lee, Solving electromagnetic problems using a novel symmetric FEM-BEM approach, *IEEE Trans. Magnetics*, 42(4), 583–586, 2006.
279. H.T. Yu, S.L. Ho, M.Q. Hu, and H.C. Wong, Edge-based FEM-BEM for wide-band electromagnetic computation, *IEEE Trans. Magnetics*, 42(4), 771–774, 2006.
280. Y.-S. Liao, S.-W. Chyuan, and J.-T. Chen, FEM versus BEM, *IEEE Circuits and Devices Magazine*, 20(5), 25–34, 2004.
281. J.M. Jin and J.L. Volakis, A FE-boundary integral formulation for scattering by 3D cavity-blocked apertures, *IEEE Trans. Antennas and Propagation*, 39, 97–104, 1991.
282. S.D. Gedney, J.F. Lee, and R. Mittra, A combined FEM/MoM approach to analyze the plane wave diffraction by arbitrary gratings, *IEEE Trans. Antenna and Propagation*, 40, 363–370, 1992.
283. A. Musolino and M. Raugi, Dual formulations of a hybrid FEM/MoM method for nonlinear analysis of EM fields, *IEEE Trans. Magnetics*, 35(3), 1380–1383, 1999.
284. R.-B. Wu and T. Itoh, Hybridizing FDTD analysis with unconditionally stable FEM for objects of curved boundary, *Proc. 1995 IEEE-MTT Symposium*, 2, 833–836, 1995.
285. A. Monorchio and R. Mittra, A hybrid FE/FDTD technique for solving complex electromagnetic problems, *IEEE Microwave Guided Wire Letters*, 8, 93–95, 1998.
286. A.G. Engel and L.P.B. Katehi, Frequency and time domain characterization of microstrip-ridge structures, *IEEE Trans. Microwave Theory Techniques*, 41(8), 1251–1262, 1993.
287. H. Esteban et al. A new hybrid mode-matching method for the analysis of inductive obstacles and discontinuities, *Proc. 1999 IEEE-AP Symposium*, 2, 966–969, 1999.
288. D.C. Ross, J.L. Volakis, and H. Anastasiu, Hybrid FE-modal analysis of jet engine inlet scattering, *IEEE Trans. Antennas and Propagation*, 43(3), 277–285, 1995.

289. J. Rubio, J. Arroyo, and J. Zapata, Analysis of passive microwave circuits by using a hybrid 2D and 3D finite-element mode-matching method, *IEEE Trans. Microwave Theory Techniques*, 47(9), 1746–1749, 1999.
290. S. Lindenmeier, C. Christopoulos, and P. Russer, Thin wire modeling with TLM/MIE method, *16th Annual Review of Progress in ACES Digest*, 3, 587–594, 2000.
291. S. Lindenmeier, L. Pierantoni, and P. Russer, Hybrid space discretizing-integral equation methods for numerical modeling of transient interference, *IEEE Trans. Electromagnetic Compatibility*, 41(4), 425–430, 1999.
292. L. Pierantoni, G. Cerri, S. Lindemeier, and P. Russer, Theoretical and numerical aspects of the hybrid MoM-FDTD, TLM-IE and ARB methods for the efficient modeling of EMC problems, *Proc. 29 EuMC Conference*, 2, 313–316, 1999.
293. V. Catina, F. Arndt, and J. Brandt, Hybrid surface integral-equation/mode-matching method for the analysis of dielectric loaded waveguide filters of arbitrary shape, *IEEE Trans. Microwave Theory Techniques*, 53(11), 3562–3567, 2005.
294. K.-L. Wu, Y. Ding, and D.-G. Fang, A parameter extraction technique for MM/FEM analysis of multilayered RF passives using TRL calibration, *IEEE Microwave and Wireless Components Letters*, 14(9), 452–454, 2004.
295. F. Arndt, J. Brandt, V. Catina, J. Ritter, I. Rullhusen, J. Daulberg, U. Hilgefort, and W. Wessel, Fast CAD and optimization of waveguide components and aperture antennas by hybrid MM/FE/MoM/FD methods-state-of-the-art and recent advances, *IEEE Trans. Microwave Theory Techniques*, 52(1), Part 2, 292–305, 2004.
296. H.-H. Chen, Finite-element method coupled with method of lines for the analysis of planar or quasi-planar transmission lines, *IEEE Trans. Microwave Theory Techniques*, 51(3), 848–855, 2003.
297. F.J.C. Meyer, D.B. Davidson, U. Jakobus, and M.A. Stuchly, Human exposure assessment in the near field of GSM base-station antennas using a hybrid finite element/method of moments technique, *IEEE Trans. Biomedical Engineering*, 50(2), 224–233, 2003.
298. J. Rubio, M.A. Gonzalez, and J. Zapata, Generalized-scattering-matrix analysis of a class of finite arrays of coupled antennas by using 3-D FEM and spherical mode expansion, *IEEE Trans. Antennas and Propagation*, 53(3), 1133–1144, 2005.
299. W. Shao, B.-Z. Wang, and Z.-J. Yu, Space-domain finite-difference and time-domain moment method for electromagnetic simulation, *IEEE Trans. Electromagnetic Compatibility*, 48(1), 10–18, 2006.
300. A. Monorchio, A.R. Bretones, R. Mittra, G. Manara, and R.G. Martin, A hybrid time-domain technique that combines the finite element, finite difference and method of moment techniques to solve complex electromagnetic problems, *IEEE Trans. Antennas and Propagation*, 52(10), 2666–2674, 2004.
301. D. Staiculescu, C. You, L. Martin, W. Hwang, and M.M. Tentzeris, *Hybrid Electrical/Mechanical Optimization Technique Using Time-Domain Modeling, Finite Element Method and Statistical Tools*, accepted for presentation to the IEEE IMS 2006, San Fransisco, June 2006.
302. I. Daubechies, *Ten Lectures on Wavelets*, SIAM, USA, 1992.
303. G. Kaiser, *A Friendly Guide to Wavelets*, Springer Verlag, Berlin, 1994.
304. C. Burrus, R.A. Gopinath, and H. Guo, *Introduction to Wavelets and Wavelet Transforms: A Primer*, Prentice Hall, Englewood Cliffs, NJ, 1997.
305. J. Goswami and A. Chan, *Fundamentals on Wavelets: Theory, Algorithms and Applications*, John Wiley & Sons, New York, 1999.
306. W. Dahmen, A.J. Kurdila, and P. Oswald, *Multiscale Wavelet Methods for Partial Differential Equations*, Academic Press, New York, 1997.
307. S. Mallat, *Multiresolution Representation and Wavelets*, Ph.D. Thesis, University of Pennsylvania, Philadelphia, 1988.
308. Y. Meyer, *Ondelettes et Fonctions Splines*, Seminaire EDP, Ecole Polytechnique, Paris, December 1986.
309. L.L. Schumaker, *Spline Functions: Basic Theory*, Wiley-Interscience, New York, 1981.
310. H.C. Schweinler and E.P. Wigner, Orthogonalization methods, *Journal of Mathematical Physics*, 11, 1693–1694, 1970.
311. G. Battle, A block spin construction of ondelettes, Part I: Lemarie functions, *Comm. Math. Phys.*, 110, 601–615, 1987.

312. P.G. Lemarie, Ondelettes a localization exponentielles, *Journal of Mathématiques Pures et Appliquées*, 67, 227–236, 1988.
313. A. Haar, Zur theorie der orthogonalen funktionsysteme, *Mathematische Annalen*, 69, 331–371, 1910.
314. I. Daubechies, Orthonormal bases of compactly supported wavelets, *Communications on Pure and Applied Mathematics*, 41, 909–996, 1988.
315. G. Beylkin, R. Coifman, and V. Rokhlin, Fast wavelet transforms and numerical algorithms I, *Communications on Pure and Applied Mathematics*, 44, 141–183, 1991.
316. B.Z. Steinberg and Y. Leviatan, On the use of wavelet expansions in the method of moments, *IEEE Trans. Antennas and Propagation*, 41, 610–619, 1993.
317. G. Wang, On the utilization of periodic wavelet expansions in the moment methods, *IEEE Trans. Microwave Theory Techniques*, 43, 2495–2498, 1995.
318. J.C. Goswami, A.K. Chan, and C.K. Chui, On solving first-kind integral equations using wavelets on a bounded interval, *IEEE Trans. Antennas and Propagation*, 43, 614–622, 1995.
319. K. Sabetfakhri and L.P.B. Katehi, Analysis of integrated millimeter-wave and submillimeter-wave waveguides using orthonormal wavelet expansions, *IEEE Trans. Microwave Theory Techniques*, 42, 2412–2422, 1994.
320. G. Pan, M.V. Toupikov, and B.K. Gilbert, On the use of Coifman Intervallic Wavelets in the Method of Moments for the Construction of Wavelet Sparsified Matrices, *IEEE Trans. Antennas and Propagation*, 47(7), 1189–1200, 1999.
321. D. Zahn, K. Sarabandi, K.F. Sabet, and J.F. Harvey, Numerical Simulation of scattering from rough surfaces: a wavelet-based approach, *IEEE Trans. Antennas and Propagation*, 48(2), 246–253, 2000.
322. J. Zheng, Z.-F. Li, and X.-N. Qian, An efficient solver for the 3D capacitance of the interconnects in high speed digital circuit by the multiresolution method of moments, *IEEE Trans. Advanced Packaging*, 22(1), 9–15, 1999.
323. M. Krumpholz and L.P.B. Katehi, MRTD: New time domain schemes based on multiresolution analysis, *IEEE Trans. Microwave Theory Techniques*, 44(4), 555–561, 1996.
324. M.M. Tentzeris, J. Harvey, and L.P.B. Katehi, Time adaptive time-domain techniques for the design of microwave circuits, *IEEE Microwave Guided Wave Letters*, 9, 96–98, 1999.
325. E.M. Tentzeris, A. Cangellaris, L.P.B. Katehi, and J. Harvey, Multiresolution time-domain (MRTD) adaptive schemes using arbitrary resolutions of wavelets, *IEEE Trans. Microwave Theory Techniques*, 50(2), 501–516, 2002.
326. M. Krumpholz, H.G. Winful, and L.P.B. Katehi, Nonlinear time-domain modeling by MRTD, *IEEE Trans. Microwave Theory Techniques*, 45(3), 385–393, 1997.
327. R.L. Robertson, M. Tentzeris, and L.P.B. Katehi, Modelling of dielectric-loaded cavities using MRTD, *Int. Journal of Numerical Modeling, Special Issue on Wavelets in Electromagnetics*, 11, 55–68, 1998.
328. M. Tentzeris and L.P.B. Katehi, Space adaptive analysis of evanescent waveguide filters, *Proc. IEEE-MTT (IMS) Symposium*, vol. 1, Baltimore, MD, 481–484, 1998.
329. L. Roselli, M. Tentzeris, and L.P.B. Katehi, Nonlinear circuit characterization using a multiresolution time domain technique, *Proc. IEEE-MTT (IMS) Symposium*, vol. 2, Baltimore, MD, 1387–1390, 1998.
330. B. McGarvey, D. Staiculescu, E.M. Tentzeris, and J. Laskar, Adaptive modeling of complex packaging geometries using Haar-based MRTD algorithms, *Proc. 2000 IEEE European Microwave Conference*, pp. 413–416 (Vol. I), Paris, France, October 2000.
331. M. Fujii and W.J.R. Hoefer, A wavelet formulation of the finite-difference method: full-vector analysis of optical waveguide junctions, *IEEE Journal of Quantum Electronics*, 37(8), 1015–1029, 2001.
332. T. Dogaru and L. Carin, Time-domain sensing of targets buried under a gaussian exponential, or fractal rough surface, *IEEE Trans. Geoscience and Remote Sensing*, 39(8), 1807–1819, 2001.
333. Q. Cao and Y. Chen, Scaling-function based multiresolution time domain analysis for planar printed millimeter-wave integrated circuits, *IEE Proc. Microwave, Antennas and Propagation*, 148(3), 179–187, 2001.
334. M. Tentzeris, R. Robertson, J. Harvey, and L.P.B. Katehi, Stability and dispersion analysis of Battle-Lemarie based MRTD schemes, *IEEE Trans. Microwave Theory Techniques*, 47(7), 1004–1013, 1999.
335. K. Goverdhanam, M. Tentzeris, M. Krumpholz, and L.P.B. Katehi, An FDTD multigrid based on multiresolution analysis, *Proc. IEEE-AP Symposium*, vol. 1, Baltimore, MD, 352–355, 1996.
336. M. Fujii and W.J.R. Hoefer, Multiresolution analysis similar to the FDTD method-derivation and application, *IEEE Trans. Microwave Theory Techniques*, 46(12), 2463–2475, 1998.

337. C. Sarris and L.P.B. Katehi, Multiresolution time-domain (MRTD) schemes with space-time Haar wavelets, *Proc. IEEE-MTT (IMS) Symposium*, vol. 2, Anaheim, CA, 1459–1462, 1999.
338. C.D. Sarris and L.P.B. Katehi, Fundamental gridding-related dispersion effects in multiresolution time-doamin schemes, *IEEE Trans. Microwave Theory Techniques*, 49(12), 2248–2257, 2001.
339. Q. Cao, Y. Chen, and R. Mittra, Multiple image technique (MIT) and anisotropic perfectly matched layer (APML) in implementation of MRTD scheme for boundary truncations of microwave structures, *IEEE Trans. Microwave Theory Techniques*, 50(6), 1578–1589, 2002.
340. C.D. Sarris and L.P.B. Katehi, An efficient numerical interface between FDTD and Haar MRTD-formulation and applications, *IEEE Trans. Microwave Theory Techniques*, 51(4), Part 1, 1146–1156, 2003.
341. N. Bushyager, J. Papapolymerou, and M.M. Tentzeris, A composite cell-multiresolution time-domain technique for the design of antenna systems including electromagnetic band gap and via-array structures, *IEEE Trans. Antennas and Propagation,* 53(8), Part. 2, 2700–2710, 2005.
342. S. Goasguen, M.M. Tomeh, and S.M. El-Ghazaly, Electromagnetic and semiconductor device simulation using interpolating wavelets, *IEEE Trans. Microwave Theory Techniques*, 49(12), 2258–2265, 2001.
343. Y.A. Hussein and S.M. El-Ghazaly, Extending multiresolution time-domain (MRTD) technique to the simulation of high-frequency active devices, *IEEE Trans. Microwave Theory Techniques*, 51(7), 1842–1851, 2003.
344. X. Zhu, T. Dogaru, and L. Carin, Three-dimensional biorthogonal multiresolution time-domain method and its application to electromagnetic scattering problems, *IEEE Trans.Antennas and Propagation*, 51(5), 1085–1092, 2003.
345. X. Zhu and L. Carin, Application of the biorthogonal multiresolution time-domain method to the analysis of elastic-wave interactions with buried targets, *IEEE Trans. Geoscience and Remote Sensing*, 42(7), 1502–1511, 2004.
346. X. Zhu, T. Dogaru, and L. Carin, Analysis of the CDF biorthogonal MRTD method with application to PEC targets, *IEEE Trans. Microwave Theory Techniques*, 51(9), 2015–2022, 2003.
347. Z. Zhaoxian and J.S. Tyo, An adaptive time-domain integral equation method for transient analysis of wire scatterer, *IEEE Antennas and Wireless Propagation Letters*, 4, 147–150, 2005.
348. Hecht-Nielsen, R. *Neurocomputing*, Addison-Wesley, New York, 1990.
349. N. Chaiyaratana and A.M.S. Zalzala: Recent developments in evolutionary and genetic algorithms: theory and applications, *Second International Conference on Genetic Algorithms in Engineering Systems, Innovations and Applications*, 270–277, 1997.
350. D.C. Montgomery, *Design and Analysis of Experiments*, John Wiley & Sons, New York, 1997.
351. C. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd Ed., Springer Science + Business Media Inc., New York, 2004.
352. D. Staiculescu, N. Bushyager, L. Martin, and M.M. Tentzeris, Design of millimeter-wave metamaterials using deterministic full-wave electromagnetic simulators and path of ascent statistical optimizer, *2005 IEEE-ECTC Proc.*, vol. 1, Orlando, FL, 1565–1568, 2005.
353. D. Staiculescu, N. Bushyager, A. Obatoyinbo, L.J. Martin, and M.M. Tentzeris, Design and optimization of 3-D compact stripline and microstrip Bluetooth/WLAN balun architectures using the design of experiments technique, *IEEE Trans. Antenna and Propagation*, 53(5), 1805–1812, 2005.

29

Computer Aided Design of Passive Components

Daniel G. Swanson, Jr.
Tyco Electronics

Computer-aided design (CAD) of passive RF and microwave components has advanced slowly but steadily over the past four decades. The 1960s and 1970s were the decades of the mainframe computer. In the early years, CAD tools were proprietary, in-house efforts running on text-only terminals. The few graphics terminals available were large, expensive, and required a short, direct connection to the mainframe. Later in this period, commercial tools became available for use on in-house machines or through time sharing services. A simulation of a RF or microwave network was based on a combination of lumped and distributed elements. The elements were connected in cascade using ABCD parameters or in a nodal network using admittance- or Y-parameters. The connection between elements and the control parameters for the simulation were stored in a text file called a netlist. The netlist syntax was similar but unique for each software tool. The mathematical foundations for a more sophisticated analysis based on Maxwell's equations were being laid down in this same time period [1–4]. However, the computer technology of the day could not support effective commercial implementation of these more advanced codes.

The 1980s brought the development of the microprocessor and UNIX workstations. The UNIX workstation played a large role in the development of more sophisticated CAD tools. For the first time there was a common operating system and computer language (the C language) to support the development of cross-platform applications. UNIX workstations also featured large, bit mapped graphics displays for interaction with the user. The same microprocessor technology that launched the workstation also made the personal computer possible. Although the workstation architecture was initially more sophisticated, personal computer hardware and software has grown steadily more elaborate. Today, the choice between a workstation and a personal computer is largely a personal one. CAD tools in this time period were still based on lumped and distributed concepts. The innovations brought about by the cheaper, graphics-based hardware had largely to do with schematic capture and layout. Schematic capture replaced the netlist on the input side of the analysis and automatic or semi-automatic layout provided a quicker path to the finished circuit after analysis and optimization.

The greatest innovation in the 1990s was the emergence of CAD tools based on the direct solution of Maxwell's equations. Finally, there was enough computer horsepower to support commercial versions of

the codes that had been in development since the late 1960s and early 1970s. These codes are in general labeled electromagnetic field-solvers although any one code may be based on one of several different numerical methods. Sonnet ***em*** [5], based on the method of moments (MoM), was the first commercially viable tool designed for RF and microwave engineers. Only a few months later, Hewlett-Packard HFSS [6], a finite element method (FEM) code co-developed with Ansoft Corp., was released to the design community. All of these tools approximate the true fields or currents in the problem space by subdividing the problem into basic "cells" or "elements" that are roughly one tenth to one twentieth of a guide wavelength in size. For any guided electromagnetic wave, the guide wavelength is the distance spanned by one full cycle of the electric or magnetic field. The problem is to find the magnitude of the assumed current on each cell or the field at the junction of elements. The final solution is then just the sum of each small contribution from each basic unit. Most of these codes first appeared on UNIX workstations and then migrated to the personal computer, as that hardware became more powerful. In the later years of the 1990s, field-solver codes appeared that were developed on and for the personal computer. In the early years, the typical field-solver problem was a single discontinuity or some other structure that was small in terms of wavelengths. Today, groups of discontinuities, complete matching networks, or small parts of a multilayer printed circuit (PC) board are all suitable problems for a field-solver. Field-solver data in S-parameter form is typically imported into a circuit simulator and combined with lumped and distributed models to complete the analysis of the structure.

29.1 Circuit Theory Based CAD

CAD of low frequency circuits is at least 30 years old and microwave circuits have been analyzed by computer for at least 20 years. At very low frequencies, we can connect inductors, capacitors, resistors, and active devices in a very arbitrary way. The lumped lowpass filter shown in Figure 29.1 is a simple example. This very simple circuit has only three nodes. Most network analysis programs will form an admittance matrix (*Y*-matrix) internally and invert the matrix to find a solution. The Y-matrix is filled using some fairly simple rules. A shunt element connected to node two generates an entry at Y_{22}. A series element connected between nodes two and three generates entries at Y_{22}, Y_{23}, Y_{32}, and Y_{33}. A large ladder network with sequential node numbering results in a large tri-diagonal matrix with many zeros off axis.

$$\mathbf{Y} = \begin{bmatrix} j\omega C_1 - j\dfrac{1}{\omega L_2} & j\dfrac{1}{\omega L_2} & 0 \\ j\dfrac{1}{\omega L_2} & j\omega C_3 - j\dfrac{1}{\omega L_2} - j\dfrac{1}{\omega L_4} & j\dfrac{1}{\omega L_4} \\ 0 & j\dfrac{1}{\omega L_4} & j\omega C_5 - j\dfrac{1}{\omega L_4} \end{bmatrix}$$

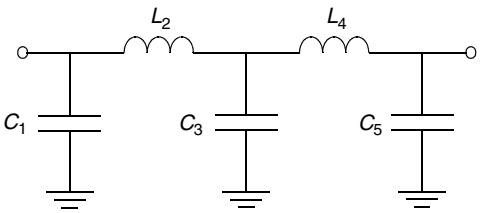


FIGURE 29.1 Lumped element lowpass filter or matching network.

The Y-matrix links the known source currents to the unknown node voltages. **I** is a vector of source currents. Typically the input node is excited with a one amp source and the rest of the nodes are set to zero. **V** is the vector of unknown node voltages. To find **V**, we invert the matrix **Y** and multiply by the known source currents.

$$\mathbf{I} = \mathbf{YV}$$

$$\mathbf{V} = \mathbf{Y}^{-1}\mathbf{I}$$

The time needed to invert an $N \times N$ matrix is roughly proportional to N^3. Filling and inverting the Y-matrix for each frequency of interest will be very fast, in this case, so fast it will be difficult to measure the computation time unless we specify a very large number of frequencies. This very simple approach might be good up to 1 MHz or so.

In our low-frequency model there is no concept of wavelength or even physical size. Any phase shift we compute is strictly due to the reactance of the component, not its physical size. There is also no concept of radiation; power can only be dissipated in resistive components. As we move into the HF frequency range (1–30 MHz) the real components we buy will have significant parasitics. Lead lengths and proximity to the ground plane become very important and our physical construction techniques will have a big impact on the results achieved.

By the time we reach VHF frequencies (50–150 MHz) we are forced to adopt distributed concepts in the physical construction and analysis of our circuits. The connections between components become transmission lines and many components themselves are based on transmission line models. Our simple lowpass circuit might become a cascade of low and high impedance transmission lines, as seen in Figure 29.2.

If this was a microstrip circuit, we would typically specify the substrate parameters and the width and length of each transmission line. We have ignored the step discontinuities due to changes in line width in this simplified example. Internally, the software would use analytical equations to convert our physical dimensions to impedances and electrical lengths. The software might use a *Y*-matrix, a cascade of ABCD parameter blocks, or a cascade of scattering-parameter (S-parameter) blocks for the actual analysis. At the ports, we typically ask for S-parameters referenced to the system impedance.

Notice that we still have a small number of nodes to consider. Our circuit is clearly distributed but the solution time does not depend on its size in terms of wavelengths. Any phase shift we compute is directly related to the physical size of the network. Although we can include conductor and substrate losses, there is still no radiation loss mechanism. It is also difficult to include enclosure effects; there may be box resonances or waveguide modes in our physical implementation. There is also no mechanism for parasitic coupling between our various circuit models.

The boundary between a lumped circuit point of view and a distributed point of view can be somewhat fuzzy. A quick review of some rules of thumb and terminology might be helpful. One common rule of

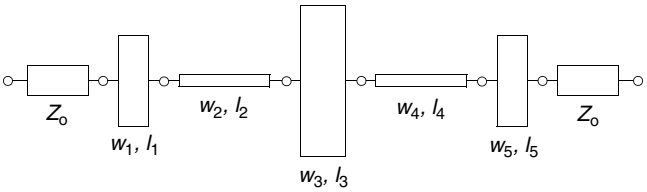
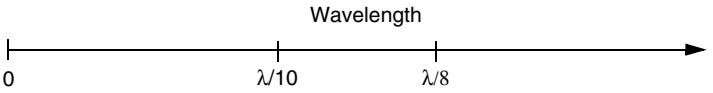


FIGURE 29.2 Distributed lowpass filter circuit. Step discontinuities are ignored.



Less than λ/10	Grey area	λ/8 or Greater
Lumped		Distributed
L, C, R, G		Transmission lines
Voltage, current		[S], [Z], [Y]
No radiation		Radiation possible
Only reactance can shift phase of V or I		Physical distance can shift phase of V or I
Fields rise and fall at same time all through the structure		There is phase shift in the fields across the structure

FIGURE 29.3 The transition between lumped and distributed behavior and some common terminology.

thumb says that the boundary between lumped and distributed behavior is somewhere between a tenth and an eighth of a guide wavelength. Remember that wavelength in inches is defined by

$$\lambda = \frac{11.803}{\sqrt{\varepsilon_{eff} \cdot f}}$$

where ε_{eff} is the effective dielectric constant of the medium and f is in GHz. At 1 GHz, $\lambda = 11.803$ inches in air and $\lambda = 6.465$ inches for a 50 Ω line on 0.014-inch thick FR4. FR4 is a common, low cost printed circuit board material for digital and RF circuits. In Figure 29.3 we can relate the physical size of our structure to the concept of wavelength and to some common terminology. Again, the boundary between purely lumped and purely distributed behavior is not always distinct.

29.2 Field Theory Based CAD

A field-solver based solution is an alternative to the previous distributed, circuit theory based approach. The field-solver takes a more microscopic view of any distributed geometry. Any field-solver we might employ must subdivide the geometry based on guide wavelength. Typically we need 10 to 30 elements or cells per guide wavelength to capture the fields or currents in our structure. Figure 29.4 shows a typical mesh generated by Agilent Momentum [7] for our microstrip lowpass filter example. Narrow cells are used on the edges of the strip to capture the spatial wavelength, or highly nonuniform current distribution across the width of the strips. This Method of Moments code has subdivided the microstrip metal and will solve for the current on each small rectangular or triangular patch. The default settings for mesh generation were used.

For this type of field-solver there is a strong analogy between the *Y*-matrix description we discussed for our lumped element circuit and what the field-solver must do internally. Imagine a lumped capacitor to ground at the center of each "cell" in our field-solver description. Series inductors connect these capacitors to each other. Coupling between non-adjacent cells can be represented by mutual inductances. So we have to fill and invert a matrix, but this matrix is now large and dense compared to our simple, lumped element circuit Y-matrix. For the mesh in Figure 29.4, $N = 474$ and we must fill and invert an $N \times N$ matrix.

One reason we turn to the field-solver is because it can potentially include all electromagnetic effects from first principles. We can include all loss mechanisms including surface waves and radiation.

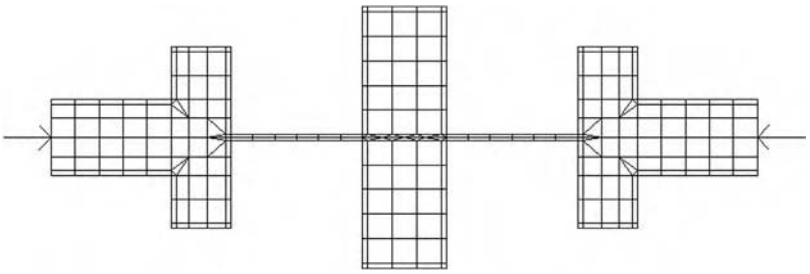


FIGURE 29.4 A typical MoM mesh for the distributed lowpass filter circuit. The number of unknowns, N is 474. (Momentum, Agilent EEsof EDA, Santa Rosa, CA.)

We can also include parasitic coupling between elements and the effects of compacting a circuit into a small space. The effects of the package or housing on our circuit performance can also be included in the field-solver analysis. However, the size of the numerical problem is now proportional to the structure size in wavelengths. The details of how enclosures are included in our analysis will vary from solver to solver. In some tools an enclosure is part of the basic formulation. In other tools, the analysis environment is "laterally open"; there are no sidewalls, although there may be a cover. One of the exciting aspects of field-solvers is the ability to observe fields and currents in the circuit, which sometimes leads to a deeper understanding of how the circuit actually operates. However, the size of the numerical problem will also be greater using a field-solver versus circuit theory, so we must carefully choose which pieces of global problem we will attack with the field-solver.

Although our discussion so far has focused on planar, distributed circuits there are actually three broad classes of field-solver codes. The 2D cross-section codes solve for the modal impedance and phase velocity of 1 to N strips with a uniform cross-section. This class of problem includes coupled microstrips, coupled slots, and conductors of arbitrary cross-section buried in a multilayer PC board. These tools use a variety of numerical methods including MoM, FEM and the spectral domain method (SDM). Field-solver engines that solve for multiple strips in a layered environment are built into several linear and nonlinear simulators. A multistrip model of this type is a building block for more complicated geometries like Lange couplers, spiral inductors, baluns, and many distributed filters. The advantage of this approach is speed; only the 2D cross-section must be discretized and solved.

The second general class of codes mesh or subdivide the surfaces of planar metals. The assumed environment for these surface meshing codes is a set of homogeneous dielectric layers with patterned metal conductors at the layer interfaces. Vertical vias are available to form connections between metal layers. There are two fundamental formulations for these codes, closed box and laterally open. In the closed box formulation the boundaries of the problem space are perfectly conducting walls. In the laterally open formulation, the dielectric layers extend to infinity. The numerical method for this class of tool is generally MoM. Surface meshing codes can solve a broad range of strip- and slot-based planar circuits and antennas. Compared to the 2D cross-section solvers, the numerical effort is considerably higher.

The third general class of codes meshes or subdivides a 3D volume. These volume meshing codes can handle virtually any three-dimensional object, with some restrictions on where ports can be located. Typical problems are waveguide discontinuities, various coaxial junctions, and transitions between different guiding systems, such as transitions from coax to waveguide. These codes can also be quite efficient for computing transitions between layers in multilayer PC boards and connector transitions between boards or off the board. The more popular volume meshing codes employ the FEM, the finite difference time domain (FDTD) method, and the transmission line matrix (TLM) method. Although the volume meshing codes can solve a very broad range of problems, the penalty for this generality is total solution

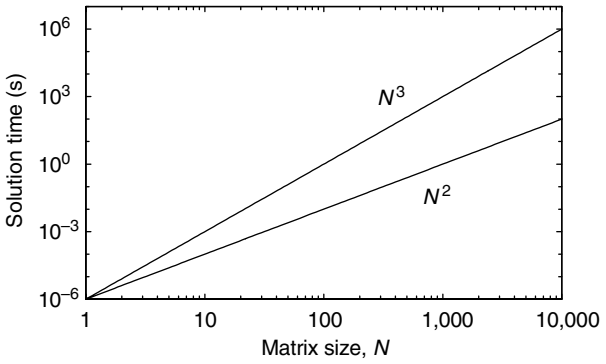


FIGURE 29.5 Solution time as a function of matrix size, N. Solution time for circuit simulators, MoM field-solvers, and FEM field-solvers is roughly proportional to N^3.

time. It typically takes longer to set up and run a 3D problem compared to a surface meshing or cross-section problem. Sadiku [8] has compiled a very thorough introduction to many of these numerical methods.

29.3 Solution Time for Circuit Theory and Field Theory

When we use circuit theory to analyze a RF or microwave network, we are building a Y-matrix of dimension N, where N is the number of nodes. A typical amplifier or oscillator design may have only a couple of dozen nodes. Depending on the solution method, the solution time is proportional to a factor between N^2 and N^3. When we talk about a "solution" we really mean matrix inversion. In Figure 29.5 we have plotted solution time as a function of matrix size N. The vertical time scale is somewhat arbitrary but should be typical of workstations and personal computers today.

When we use a MoM field-solver, a "small" problem has a matrix dimension of $N = 300$–600. Medium size problems may be around $N = 1500$ and large problems quickly get into the $N = 2000$–3000 range. Because of the N^2/N^3 effect, the solution time is impacted dramatically as the problem size grows. In this case we can identify two processes, filling the matrix with all the couplings between cells and inverting or solving that matrix. So we are motivated to keep our problem size as small as possible. The FEM codes also must fill and invert a matrix. Compared to MoM, the matrix tends to be larger but more sparse.

The time domain solvers using FDTD or TLM are exceptions to the N^2/N^3 rule. The solution process for these codes is iterative; there is no matrix to fill or invert with these solvers. Thus the memory required and the solution time grow more linearly with problem size in terms of wavelengths. This is one reason these tools have been very popular for radar cross-section (RCS) analysis of ships and airplanes. However, because these are time stepping codes, we must perform a fast Fourier transform (FFT) on the time domain solution to get the frequency domain solution. Closely spaced resonances in the frequency domain require a large number of time samples in the time domain. Therefore, time stepping codes may not be the most efficient choice for structures like filters, although there are techniques available to speed up convergence. Veidt [9] presents a good summary of how solution time scales for various numerical methods.

29.4 A Hybrid Approach to Circuit Analysis

If long solution times prevent us from analyzing complete circuits with a field-solver, what is the best strategy for integrating these tools into the design process? The best approach is to identify the key pieces of the problem that need the field-solver, and to do the rest with circuit theory. Thus the final result is a "hybrid solution" using different techniques, and even different tools from different vendors.

As computer power grows and software techniques improve, we can do larger and larger pieces of the problem with a field-solver. A simple example will help to demonstrate this approach. The circuit in Figure 29.6 is part of a larger RF printed circuit board. In one corner of the board we have a branchline coupler, a resistive termination, and several mitered bends.

Using the library of elements in our favorite linear simulator, there are several possible ways to subdivide this network for analysis (see Figure 29.7). In this case we get about 21 nodes in our circuit. Solution time is roughly proportional to N^3, so if we ignore the overhead of computing any of the individual models, we would expect the solution to come back very quickly. But we have clearly neglected several things in our analysis. Parasitic coupling between the arms of the coupler, interaction between the discontinuities, and any potential interaction with the package have all been ignored. Some of our analytical models may not be as accurate as we would like, and in some cases a combination of models may not accurately describe our actual circuit. If this circuit were compacted into a much denser layout, all of the effects mentioned above would become more pronounced.

Each one of the circuit elements in our schematic has some kind of analytic model inside the software. For a transmission line, the model would relate physical width and length to impedance and electrical length

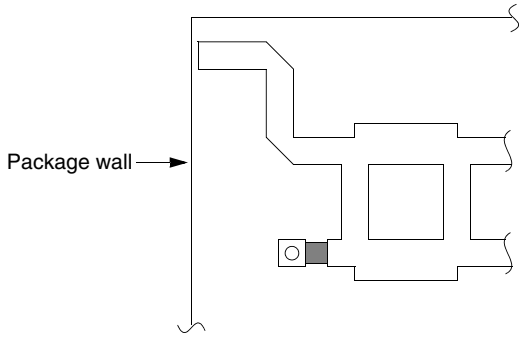


FIGURE 29.6 Part of an RF printed circuit board which includes a branchline coupler, a resistive termination to ground, and several mitered bends.

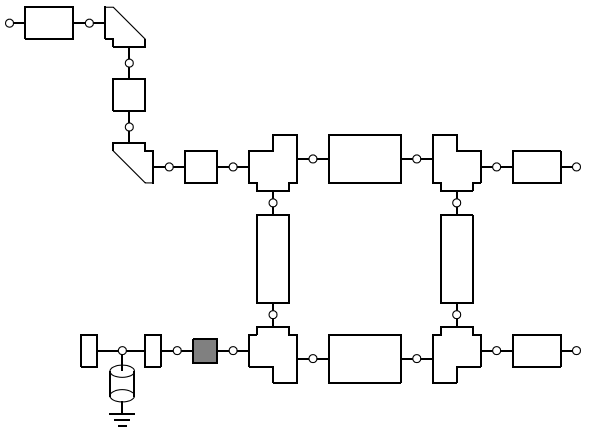


FIGURE 29.7 The layout in Figure 29.6 has been subdivided for analysis using the standard library elements found in many circuit-theory-based simulators.

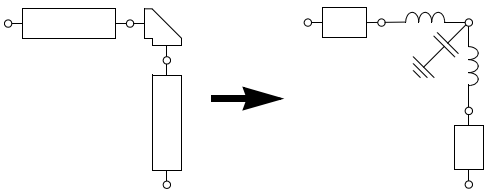


FIGURE 29.8 The equivalent circuit of a microstrip mitered bend. The physical dimensions are mapped to an equivalent lumped element circuit.

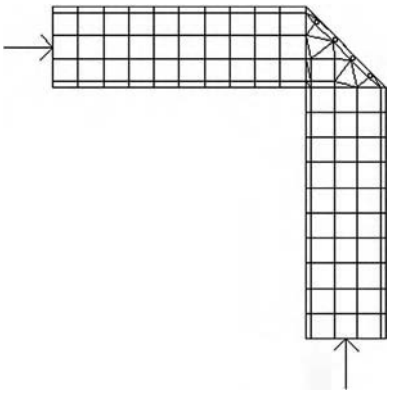


FIGURE 29.9 A typical MoM mesh for the microstrip mitered bend. The number of unknowns, N, is 221. (Momentum, Agilent EEsof EDA, Santa Rosa, CA.)

through a set of closed form equations. For a discontinuity like the mitered bend, the physical parameters might be mapped to an equivalent lumped element circuit (Figure 29.8), again through a set of closed form equations. The field-solver will take a more microscopic view of the same mitered bend discontinuity. Any tool we use will subdivide the metal pattern using 10 to 30 elements per guide wavelength. The sharp inside corner where current changes direction rapidly will force an even finer subdivision. If we want to solve the bend discontinuity individually, we must also connect a short length of series line to each port. Agilent Momentum generated the mesh in Figure 29.9. The number of unknowns is 221. If the line widths are not variable in our design, we could compute this bend once, and use it over and over again in our circuit design.

Another potential field-solver problem is in the corner of the package near the input trace. You might be able to include the box wall effect on the series line, but wall effects are generally not included in discontinuity models. However, it is quite easy to set up a field-solver problem that would include the microstrip line, the mitered bend, and the influence of the walls. The project in Figure 29.10 was drawn using Sonnet ***em***. The box walls to the left and top in the electromagnetic simulation mimic the true location of the package walls in the real hardware. There are 360 unknowns in this simulation.

One of the more interesting ways to use a field-solver is to analyze groups of discontinuities rather than single discontinuities. A good example of this is the termination resistor and via [10,11] in our example circuit. A field-solver analysis of this group may be much more accurate than a combination of individual analytical models. We could also optimize the termination, then use the analysis data and the optimized geometry over and over again in this project or other projects. The mesh for the resistor via combination (Figure 29.11) was generated using Sonnet ***em*** and represents a problem with 452 unknowns.

Our original analysis scheme based on circuit theory models alone is shown in Figure 29.7. Although this will give us the fastest analysis, there may be room for improvement. We can substitute results in our field-solver for the elements near the package walls and for the resistor/via combination (Figure 29.12).

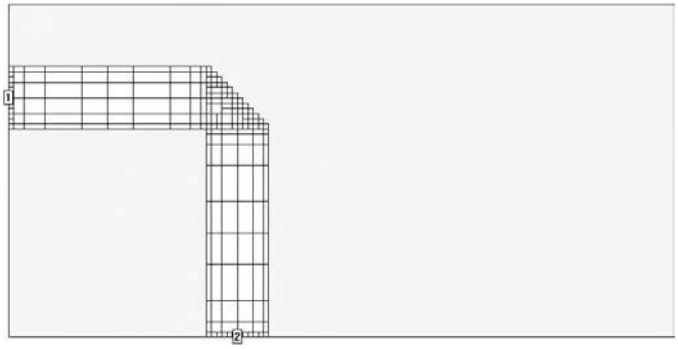


FIGURE 29.10 An analysis of the input line and the mitered bend in the presence of the package walls. The number of unknowns, *N*, is 360. (***em***™, Sonnet Software, Liverpool, NY.)

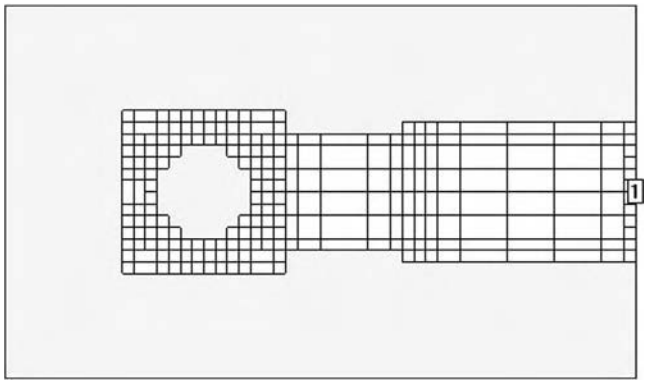


FIGURE 29.11 A MoM analysis of a group of discontinuities including a thin-film resistor, two steps in width, and a via hole to ground. The number of unknowns, *N*, is 452. (***em***™, Sonnet Software, Liverpool, NY.)

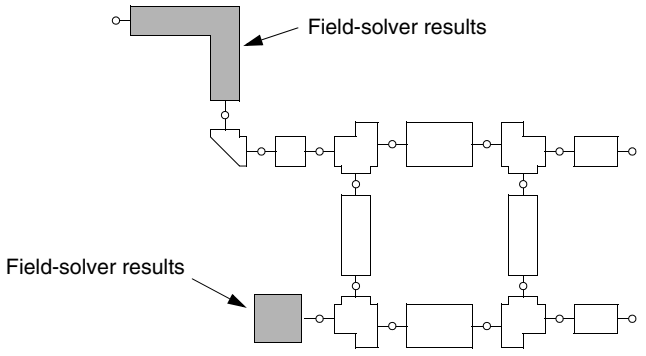


FIGURE 29.12 Substituting field-solver results into the original solution scheme mixes field-theory and circuit theory in a cost effective way.

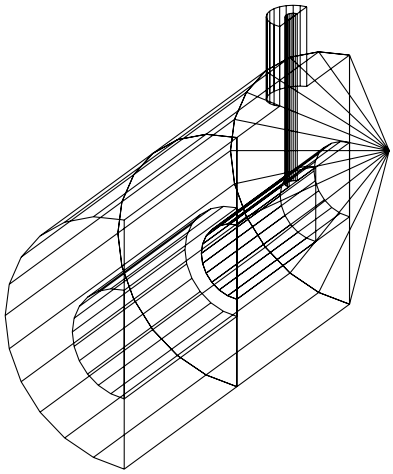


FIGURE 29.13 A right angle coax-to-coax transition that was optimized for return loss. The number of unknowns, N, is 8172. (HFSS, Ansoft Corp., Pittsburgh, PA.)

The data from the field-solver would typically be S-parameter files. This "hybrid" solution mixes field theory and circuit theory in a cost-effective way [12]. The challenge for the design engineer is to identify the critical components that should be addressed using the field-solver.

The hybrid solution philosophy is not limited to planar components; three-dimensional problems can be solved and cascaded as well. The right angle coax bend shown in Figure 29.13 is one example of a 3D component that was analyzed and optimized using Ansoft HFSS [13]. In this case we have taken advantage of a symmetry plane down the center of the problem in order to reduce solution time. This component includes a large step in inner conductor diameter and a Teflon sleeve to support the larger inner conductor. After optimizing two dimensions, the computed return loss is greater than –30 dB. The coax bend is only one of several problems taken from a larger assembly that included a lowpass filter, coupler, amplifier, and bandpass filter.

29.5 Optimization

Optimization is a key component of modern linear and nonlinear circuit design. Many optimization schemes require gradient information, which is often computed by taking simple forward or central differences. The extra computations required to find gradients become very costly if there is a field-solver inside the optimization loop. So it is important to minimize the number of field-solver analysis runs. It is also necessary to capture the desired changes in the geometry and pass this information to the field-solver. Bandler et al. [14,15] developed an elegant solution to both of these problems in 1993. The key concept was a "data pipe" program sitting between the simulator and the field-solver (see Figure 29.14). When the linear simulator calls for a field-solver analysis, the data pipe generates a new geometry file and passes it to the field-solver. In the reverse direction, the data pipe stores the analysis results and interpolates between data sets if possible. The final iterations of the optimization operate entirely on interpolated data without requiring any new field-solver runs. This concept was applied quite successfully to both surface meshing [16] and volume meshing solvers. The same basic rules that lead to successful circuit theory based optimization apply when a field-solver is in the loop as well. First, a good starting point leads to more rapid and consistent convergence. Second, it is important to limit the number of variables.

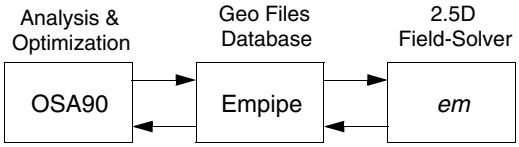


FIGURE 29.14 The first commercially successful optimization scheme which included a field-solver inside the optimization loop.

29.6 The Next Decade

The need for inexpensive wireless systems has forced the RF community to rapidly adopt low cost, multilayer PC board technology. In the simpler examples, most circuitry and components are mounted on the top layer while inner layers are used for routing of RF signals and DC bias. However, more complex examples can be found where printed passive components and discontinuities are located in one or more buried layers. Given the large number of variables in PC board construction it will be difficult for vendors of linear and nonlinear circuit simulators to support large libraries of passive models that cover all possible scenarios. However, a field-solver can be used to generate new models as needed for any novel layer stack up. Of course the user is also free to use the field-solver data to develop custom, proprietary models for his or her particular technology.

The traditional hierarchy of construction for RF systems has been a chip device, mounted to leaded package, mounted to printed circuit board located in system cabinet or housing. Today however, the "package" may be a multilayer low temperature co-fired ceramic (LTCC) substrate or a multilayer PC board using ball grid array (BGA) interconnects. Thus the boundary between package and PC board has blurred somewhat. No matter what the technology details, the problem remains to transfer a signal from the outside world into the system, onto the main system board, through the package, and into the chip. And of course there is an analogous connection from the chip back to the outside world. From this point of view, the problem becomes a complex, multilevel passive interconnect that must support not only the signal currents but also the ground currents in the return path. It is often the ground return path that limits package isolation or causes unexpected oscillations in active circuits [17]. The high-speed digital community is faced with very similar passive interconnect challenges at similar, if not higher frequencies and typically much higher signal densities. Again, there is ample opportunity to apply field-solver technology to these problems, although practical problem size is still somewhat limited. The challenge to the practitioner is to identify and correct problems at multiple points in the signal path.

29.7 Conclusion

At very low frequencies we can use lumped element models to describe our circuits. Connection lengths and device parasitics are not issues. At higher frequencies we use distributed models to capture the effects of guide wavelength, but spurious couplings between elements and other effects due to circuit compaction are typically not captured. A field-solver can potentially capture all the macro and micro aspects of our circuit. It should capture spatial wavelength effects, guide wavelength, spurious couplings among elements, and interference among elements due to dense packing. Although the size of a practical field-solver problem is still somewhat small, there are many useful and cost effective problems that can be identified and solved using a combination of circuit theory based and field theory based CAD.

References

1. K. S. Yee, Numerical solution of initial boundary-value problems involving Maxwell's equations in isotropic media, *IEEE Trans. Ant. Prop.*, AP-14, 302–207, May 1966.

2. R. F. Harrington, *Field Computation by Moment Methods*, Macmillan, New York, 1968.
3. P. B. Johns and R. L. Beurle, Numerical solution of 2-dimensional scattering problems using a transmission-line matrix, *Proc. Inst. Electr. Eng.*, 118, 1203–1208, Sept. 1971.
4. P. Silvester, Finite element analysis of planar microwave networks, *IEEE Trans. Microwave Theory Tech.*, MTT-21, 104–108, Feb. 1973.
5. ***em***™, Sonnet Software, Liverpool, NY.
6. HFSS, Hewlett-Packard, Santa Rosa, CA and Ansoft, Pittsburgh, PA.
7. Momentum, Agilent EEsof EDA, Santa Rosa, CA.
8. M. Sadiku, *Numerical Techniques in Electromagnetics*, CRC Press, Boca Raton, 1992.
9. B. Veidt, Selecting 3D electromagnetic software, *Microwave J.*, 126–137, Sept. 1998.
10. M. Goldfarb and R. Pucel, Modeling via hole grounds in microstrip, *IEEE Microwave and Guided Wave Lett.*, 1, 135–137, June 1991.
11. D. Swanson, Grounding microstrip lines with via holes, *IEEE Trans. Microwave Theory Tech.*, MTT-40, 1719–1721, Aug. 1992.
12. D. Swanson, Using a microstrip bandpass filter to compare different circuit analysis techniques, *Int. J. MIMICAE*, 5, 4–12, Jan. 1995.
13. HFSS, Ansoft Corp., Pittsburgh, PA.
14. J. W. Bandler, S. Ye, R. M. Biernacki, S. H. Chen, and D. G. Swanson, Jr., Minimax microstrip filter design using direct em field simulation, *IEEE MTT-S Int. Microwave Sympos. Dig.*, 889–892, 1993.
15. J. W. Bandler, R. M. Biernacki, S. H. Chen, D. G. Swanson, Jr., and S. Ye, Microstrip filter design using direct em field simulation, *IEEE Trans. Microwave Theory Tech.*, MTT-42, 1353–1359, July 1994.
16. D. Swanson, Optimizing a microstrip bandpass filter using electromagnetics, *Int. J. MIMICAE*, 5, 344–351, Sept. 1995.
17. D. Swanson, D. Baker, and M. O'Mahoney, Connecting MMIC chips to ground in a microstrip environment, *Microwave J.*, 58–64, Dec. 1993.

30

Nonlinear RF and Microwave Circuit Analysis

Michael B. Steer
North Carolina State University

John F. Sevic
Maury Microwave Corporation

The two most popular circuit-level simulation technologies are embodied in SPICE-like simulators, operating entirely in the time domain, and in harmonic balance (HB) simulators, which are hybrid time and frequency domain simulators. Neither is ideal for modeling RF and microwave circuits and in this chapter their concepts and bases of operation will be explored with the aim of illuminating the limitations and advantages of each. All of the technologies considered here have been implemented in commercial microwave simulators. An effort is made to provide sufficient background for these to be used to full advantage.

Simulation of digital and low frequency analog circuits at the component level is performed using SPICE, or commercial equivalents, and this has proved to be very robust. The operation of SPICE will be considered in detail later, but in essence SPICE solves for the state of the circuit at a time point and then uses this state to estimate the state of the circuit at the next time point (and so is referred to as a time-marching technique). The state of the circuit at the new time point is iterated to minimize error. This process captures the transient response of a circuit and the algorithm obtains the best waveform estimate. That is, the best estimate of the currents and voltages in the circuit at each time point are obtained. The accurate calculation of the waveform in a circuit is what we want in low pass circuits such as digital and low frequency analog circuits. However with RF and microwave circuits, especially in communications, it is more critical to accurately determine the spectrum of a signal (i.e., the frequency components and their amplitudes) than the precise waveform. In part this is because regulations require

strict control of spurious spectral emissions so as not to interfere with other wireless systems, and also because the generation of extraneous emissions compromises the demodulation and detection of communication signals by other radios in the same system. The primary distortion concern in radio is spectrum spreading or more specifically, adjacent channel interference. In-band distortion is also important especially with base station amplifiers where filtering can be used to eliminate spectral components outside the main channel. Distortion is largely the result of the nonlinear behavior of transmitters and so characterization of this phenomenon is important in RF design. In addition, provided that the designer has confidence in the stability and well-behaved transient response of a circuit, it is only necessary to determine its steady-state response. In order to determine the steady-state response using a time-marching approach, it is necessary to determine the RF waveform for perhaps millions of RF cycles, including the full transient interval, so as to extract the superimposed modulated signal. The essential feature of HB is that a solution form is assumed, in particular, a sum of sinusoids and the unknowns to be solved for are the amplitudes and phases of these sinusoids. The form of the solution then allows simplification of the equations and determination of the unknown coefficients. HB procedures work well when the signal can be described by a simple spectrum. However, it does not enable the transient response to be determined exactly.

In the following sections we will first look at the types of signals that must be characterized and identify the information that must be extracted from a circuit simulation. We will then look at transient SPICE-like simulation and HB simulation. Both types of analyses have restrictions and neither provides a complete characterization of an RF or microwave circuit. However, there are extensions to each that improve their basic capabilities and increase applicability. We will also review frequency domain analysis techniques as this is also an important technique and forms the basis of behavioral modeling approaches.

30.1 Modeling RF and Microwave Signals

The way nonlinear effects are modeled and characterized depends on the properties of the input signal. Signals having frequency components above a few hundred megahertz are generally regarded as RF or microwave signals. However, the distinguishing features that identify RF and microwave circuits are the design methodologies used with them. Communication systems generally have a small operating fractional bandwidth—rarely is it much higher than 10%. Generated or monitored signals in sensing systems (including radar and imaging systems) generally have small bandwidths. Even broadband systems including instrumentation circuits and octave (and more) bandwidth amplifiers have passband characteristics. Thus RF and microwave design and modeling technology has developed specifically for narrowband systems.

The signals to be characterized in RF and microwave circuits are either correlated, in the case of communication and radar systems, or uncorrelated noise in the case of many imaging systems. We are principally interested in handling correlated signals as uncorrelated noise is nearly always very small and can be handled using relatively straightforward linear circuit analysis techniques. There are two families of correlated signals, one being discrete tone and the other being digitally modulated. In the following, three types of signals will be examined and their response to nonlinearities described.

30.1.1 Discrete Tone Signals

Single tone signals, that is, a single sinewave, are found in frequency sources but such tones do not transmit information and must be modulated. Until recently, communication and radar systems used amplitude, phase, or frequency modulation (AM, PM, and FM, respectively) to put information on a carrier and transmission of the carrier was usually suppressed. These modulation formats are called analog modulation and the resulting frequency components can be considered as being sums of sinusoids. The signal and its response are then deterministic and a well-defined design methodology has been developed to characterize nonlinear effects. With multifrequency sinusoidal excitation consisting possibly of nonharmonically related (or noncommensurable) frequency components, the waveforms in the circuit

are not periodic yet the nonlinear circuit does have a steady-state response. Even considering a single-tone signal (a single sinewave) yields directly usable design information. However, being able to model the response of a circuit to a multitone stimulus increases the likelihood that the fabricated circuit will have the desired performance.

In an FM modulated scheme the transmitted signal can be represented as

$$x(t) = \cos\left\{\left[\omega_c + \omega_i(t)\right]t\right\} + \sin\left\{\left[\omega_c + \omega_q(t)\right]t\right\} \tag{30.1}$$

where the signal information is contained in $\omega_i(t)$ and can be adequately represented as a sum of sinewaves. The term $\omega_q(t)$ is the quadrature of $\omega_i(t)$, meaning that it is 90° out of phase. The net result is that $x(t)$ can also be represented as a sum of sinusoids. Other forms of analog modulation can be represented in a similar way. The consequence of this is that all signals in a circuit with analog modulation can be adequately represented as comprising discrete tones.

With discrete tones input to a nonlinear circuit, the output will also consist of discrete tones but will have components at frequencies that were not part of the input signal. Power series expansion analysis of a nonlinear subsystem illustrates the nonlinear process involved. When a single frequency sinusoidal signal excites a nonlinear circuit, the response "usually" includes the original signal and harmonics of the input sinewave. We say "usually," because if the circuit contains nonlinear reactive elements, subharmonics and autonomous oscillation could also be present. The process is more complicated when the excitation includes more than one sinusoid, as the circuit response may then include all sum and difference frequencies of the original signals. The term *intermodulation* is generally used to describe this process, in which power at one frequency, or group of frequencies, is transferred to power at other frequencies. The term intermodulation is also used to describe the production of sum and difference frequency components, or intermodulation frequencies, in the output of a system with multiple input sinewaves. This is a macroscopic definition of intermodulation as the generation of each intermodulation frequency component derives from many separate intermodulation processes. Here a treatment of intermodulation is developed at the microscopic level.

To begin with, consider a nonlinear system with output $y(t)$ described by the power series

$$y(t) = \sum_{l=1}^{\infty} a_l x(t)^l \tag{30.2}$$

where $x(t)$ is the input and is the sum of three sinusoids:

$$x(t) = c_1\cos(\omega_1 t) + c_2\cos(\omega_2 t) + c_3\cos(\omega_3 t). \tag{30.3}$$

Thus

$$x(t)^l = \left[c_1\cos(\omega_1 t) + c_2\cos(\omega_2 t) + c_3\cos(\omega_3 t)\right]^l. \tag{30.4}$$

This equation includes a large number of components the radian frequencies of which are the sum and differences of ω_1, ω_2, and ω_3. These result from multiplying out the term $[\cos(\omega_1 t)]^k\,[\cos(\omega_3 t)]^{l-p}$. For example

$$\begin{aligned}\cos(\omega_1 t)\cos(\omega_2 t)\cos(\omega_3 t) = \big[&\cos(\omega_1+\omega_2+\omega_3)t + \cos(\omega_1+\omega_2-\omega_3)t + \cos(\omega_1-\omega_2+\omega_3)t \\ &+ \cos(\omega_1-\omega_2-\omega_3)t\big]/4\end{aligned} \tag{30.5}$$

where the (radian) frequencies of the components are, in order $(\omega_1 + \omega_2 + \omega_3)$, $(\omega_1 + \omega_2 - \omega_3)$, $(\omega_1 - \omega_2 + \omega_3)$, and $(\omega_1 - \omega_2 - \omega_3)$. This mixing process is called intermodulation and the additional tones are called intermodulation frequencies with each separate component of the intermodulation process called an intermodulation product or IP. Thus when a sum of sinusoids is input to a nonlinear element additional frequency components are generated. In order to make the analysis tractable, the number of frequency components considered must be limited. With a two-tone input, the frequencies generated are integer combinations of the two inputs, for example, $f = mf_1 + nf_2$. One way of limiting the number of frequencies is to consider only the combinations of m and n such that

$$|m| + |n| \leq p_{\text{MAX}} \tag{30.6}$$

assuming that all products of order greater than p_{MAX} are negligible. This is called a triangular truncation scheme and is depicted as shown in Figure 30.1. The alternative rectangular truncation scheme is shown in Figure 30.2 and is defined by

$$|m| \leq m_{\text{MAX}} \quad \text{and} \quad |n| \leq n_{\text{MAX}}. \tag{30.7}$$

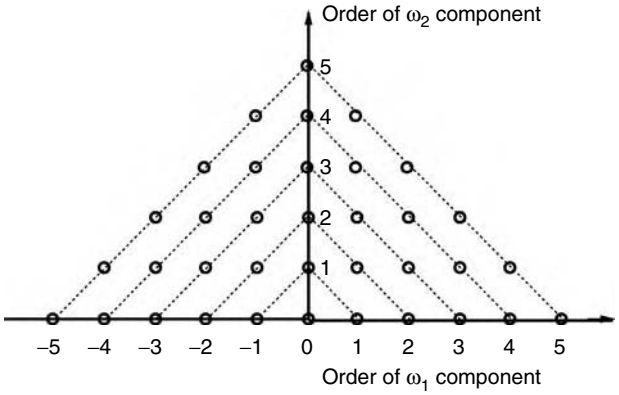


FIGURE 30.1 A triangular scheme for truncating higher order tones.

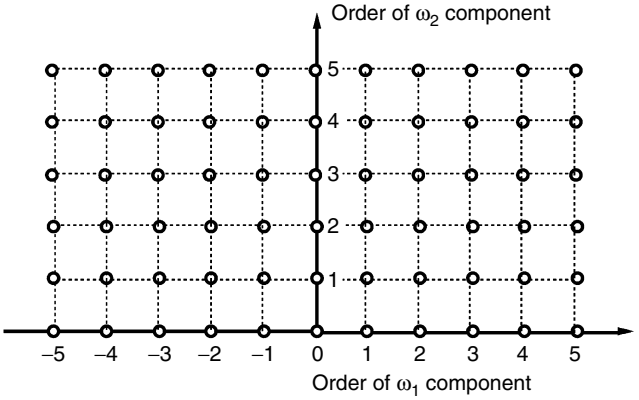


FIGURE 30.2 A rectangular scheme for truncating higher order tones. Here $m_{\text{max}} = 5 = n_{\text{max}}$.

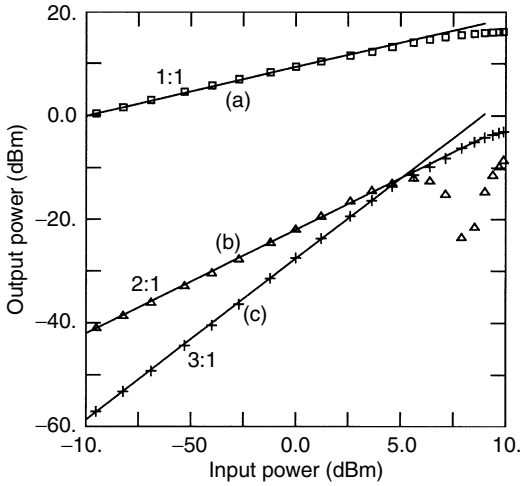


FIGURE 30.3 Measured (markers) and simulated (lines) response of a class A MESFET amplifier to a single tone input: (a) is the fundamental output; (b) is the second harmonic response; and (c) is the third harmonic response.

With one-tone excitation, the spectra of the input and output of a nonlinear circuit consists of a single tone at the input and the original, fundamental tone, and its harmonics. Here intermodulation converts power at f_1 to power at DC (this intermodulation is commonly referred to as rectification), and to power at the harmonics ($2f_1$, $3f_1$, …), as well as to power at f_1. Simply squaring a sinusoidal signal will give rise to a second harmonic component. The measured and simulated responses of a class A amplifier operating at 2 GHz are shown in Figure 30.3. This exhibits classic responses. At low signal levels the fundamental response has a slope of 1:1 with respect to the input signal level—corresponding to the linear response. Initially the second harmonic varies as the square of the input fundamental level and so has a 2:1 slope on the log-log plot. This is because the dominant IP contributing to the second harmonic level at low input powers is second order. Similarly the third harmonic response has a 3:1 slope because the dominant IP here is third order. As the input power increases, the second harmonic exhibits classic nonlinear behavior which is observed with many intermodulation tones and results from the production of a second, or more significant IP tone, which is due to higher order intermodulation than the dominant IP. In this situation, the dominant and additional IPs vectorially combine, with the result that the tone almost cancels out.

It is much more complicated to describe the nonlinear response to multifrequency sinusoidal excitation. If the excitation of an analog circuit is sinusoidal, then specifications of circuit performance are generally in terms of frequency domain phenomena, for example, intermodulation levels, gain, and the 1 dB gain compression point. However, with multi-frequency excitation by signals that are not harmonically related, the waveforms in the circuit are not periodic, although there is a steady-state response often called quasi-periodic.

Consider the nonlinear response of a system to the two-tone excitation shown in Figure 30.4. The frequencies f_1 and f_2 are, in general, nonharmonically related and components at all sum and difference frequencies $mf_1 + nf_2$, (m, $n = -\times,\ldots,-1,0,1,\ldots,\times$) of f_1 and f_2 will appear at the output of the system. If the nonlinear system has a quadratic nonlinearity, the spectrum of the output of the system is that of Figure 30.5. With a general nonlinearity, the spectrum of the output will contain a very large number of components. An approximate output spectrum is given in Figure 30.6. Also shown is a truncated spectrum that will be used in the following discussion. Most of the frequency components in the truncated spectrum of Figure 30.6 have names: DC (f_6) results from rectification; f_3, f_4, f_5, f_8, f_9, f_{10}, and f_{11} are called intermodulation frequency components; f_4, f_5 are commonly called image frequencies, or "third order" intermods, as well; f_1, f_2 are the input frequencies; and f_7, f_8 are harmonics.

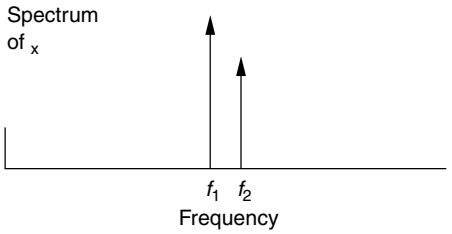


FIGURE 30.4 The spectrum of a two-tone signal.

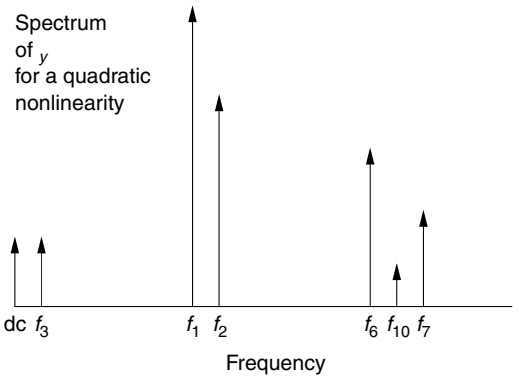


FIGURE 30.5 The spectrum at the output of a quadratic nonlinear system with a two-tone input.

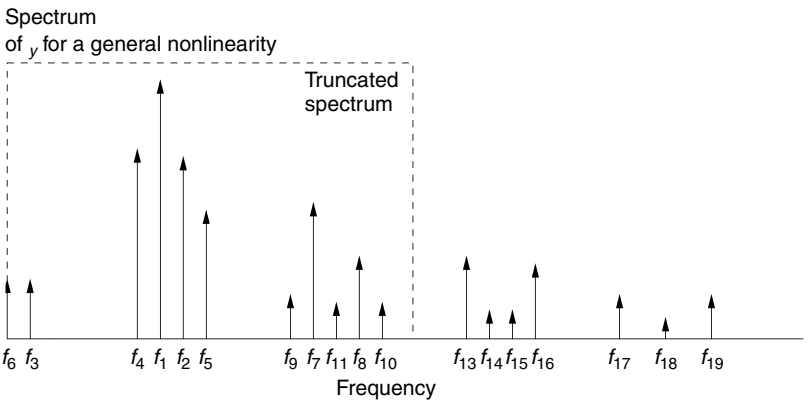


FIGURE 30.6 The approximate spectrum for a general nonlinearity with a two-tone input.

All of the frequencies in the steady-state output of the nonlinear system result from intermodulation—the process of frequency mixing. A classification of nonlinear behavior that closely parallels the way in which nonlinear responses are observed and specified is given below.

Gain Compression/Enhancement: Gain compression can be conveniently described in the time domain or in the frequency domain. Time domain descriptions refer to limited power availability or to limitations on voltage or current swings. At low signal levels, moderately nonlinear devices such as class A amplifiers behave linearly so that there is one dominant IP with a zero saturation term. As signal levels

increase, other IPs become important as harmonic levels increase. Depending on the harmonic loading condition, these IPs could be in phase with the original IP contributing to gain enhancement or out of phase contributing to gain compression.

Desensitization: Desensitization is the variation of the amplitude of one of the desired components due to the presence of another noncommensurable signal. This is an over-riding saturation effect affecting all output tones and comes out of the power series expansion.

Harmonic Generation: Harmonic generation is the most obvious result of nonlinear distortion and is identical to the process with a single-tone input.

Intermodulation: Intermodulation is the generation of spurious frequency components at the sum and difference frequencies of the input frequencies. In the truncated spectrum f_3, f_4, f_5, f_9, f_{10}, and f_{11} are intermodulation frequencies. Numerically $f_4 = 2f_1 - f_2$ and so this intermodulation tone is commonly called the lower third order intermod. There are other IPs that can contribute to the "third order" intermod that are not due to third order intermodulation. A particularly important intermodulation process begins with the generation of the difference frequency component $f_3 = f_2 - f_1$ as a second order IP. This is also referred to as the baseband component, envelope frequency, intermediate frequency, or difference frequency. This component then mixes with one of the original tones to contribute to the level of the "third order" intermod, for example, $f_4 = f_1 - f_3$, again a second order process. The corresponding contribution to the upper third order intermod f_5, that is, $f_5 = f_2 + f_3$, can (depending on the baseband impedance) have a phase that differs from the phase of the f_4 contribution and, in general, the result is that there can be asymmetry in the lower and upper third order intermod levels as the various IPs, at their respective frequencies, add vectorially.

Cross-modulation: Cross-modulation is modulation of one component by another noncommensurable component. Here it would be modulation of f_1 by f_2 or modulation of f_2 by f_1. However, with cross-modulation, information contained in the sidebands of one non-commensurable tone can be transferred to the other non-commensurable tone.

Detuning: Detuning is the generation of DC charge or DC current resulting in change of an active device's operating point. The generation of DC current with a large signal is commonly referred to as rectification. The effect of rectification can often be reduced by biasing using voltage and current sources. However, DC charge generation in nonlinear reactances is more troublesome as it can neither be detected nor effectively reduced.

AM–PM Conversion: The conversion of amplitude modulation to phase modulation (AM–PM conversion) is a troublesome nonlinear phenomenon in high frequency analog circuits and results from the amplitude of a signal affecting the delay through a system. Alternatively, the process can be understood by considering that at higher input levels, additional IPs are generated at the fundamental frequency and when these vectorially contribute to the fundamental response, phase rotation occurs.

Subharmonic Generation and Chaos: In systems with memory effects, that is, with reactive elements, subharmonic generation is possible. The intermodulation products for subharmonics cannot be expressed in terms of the input noncommensurable components. (Components are noncommensurable if they cannot be expressed as integer multiples of each other.) Subharmonics are initiated by noise, possibly a turn-on transient, and so in a steady-state simulation must be explicitly incorporated into the assumed set of steady-state frequency components. The lowest common denominator of the subharmonic frequencies then becomes the basis noncommensurable component. Chaotic behavior can only be simulated in the time domain. The nonlinear frequency domain methods as well as the conventional harmonic balance methods simplify a nonlinear problem by imposing an assumed steady-state on the nonlinear circuit solution problem. Chaotic behavior is not periodic and so the simplification is not valid in this case. Together with the ability to simulate transient behavior, the capability to simulate chaotic behavior is the unrivaled realm of time domain methods.

Except for chaotic behavior, all nonlinear behavior with discrete tones can be viewed as an intermodulation process with IPs (the number of significant ones increasing with increasing signal level) adding vectorially. Understanding this process provides valuable design insight and is also the basis of frequency domain nonlinear analysis.

30.1.2 Digitally Modulated Signals

A digitally modulated signal cannot be represented by discrete tones and so nonlinear behavior cannot be adequately characterized by considering the response to a sum of sinusoids. Nonlinear effects with digital signals are difficult to describe as the signals themselves appear to be random, but there is an underlying correlation. It is more appropriate to characterize a digitally modulated signal by its statistics, such as power spectral density, than by its component tones. Most current (and future) wireless communication systems use digital modulation, in contrast to first-generation radio systems, which were based on analog modulation. Digital modulation offers increased channel capacity, improved transmission quality, secure channels, and the ability to provide other value-added services. These systems present significant challenges to the RF and microwave engineer with respect to representation and characterization of digitally modulated signals, and also with respect to nonlinear analysis of digital wireless communication systems.

Amplifier linearity in the context of digital modulation is therefore most suitably characterized by measuring the degree of spectrum regeneration. This is done by comparing the power in the upper and lower adjacent channels to the power in the main channel: the adjacent-channel power ratio (ACPR). The spectrum of a digitally modulated signal is shown in Figure 30.7. This is the spectrum of a finite bit length digitally modulated signal and not the smooth spectrum of an infinitely long sequence often depicted.

30.2 Basics of Circuit Modeling

The solution, or simulation, of a circuit is obtained by solving a number of network equations developed by applying Kirchoff's current law (KCL) and Kirchoff's voltage law (KVL). There are two basic methods for developing the network equations for DC analysis, or steady-state analysis of linear circuits with sinusoidal excitation, based on Kirchoff's laws. These are the nodal formulation and mesh formulation of the network equations. The nodal formulation is best for electronic circuits as there are many fewer nodes than there are elements connecting the nodes. The nodal formulation, specifically node-voltage analysis, requires that the current in an element be expressed as a function of voltage. Some elements cannot be so described and so there is not a node-voltage description for them. Then the modified nodal approach is most commonly used wherein every element that can be described by an equation for current in terms of voltages is described in this way, and only for the exceptional elements are other constitutive relations considered. However, the general formulation approach can be illustrated by considering node-voltage analysis.

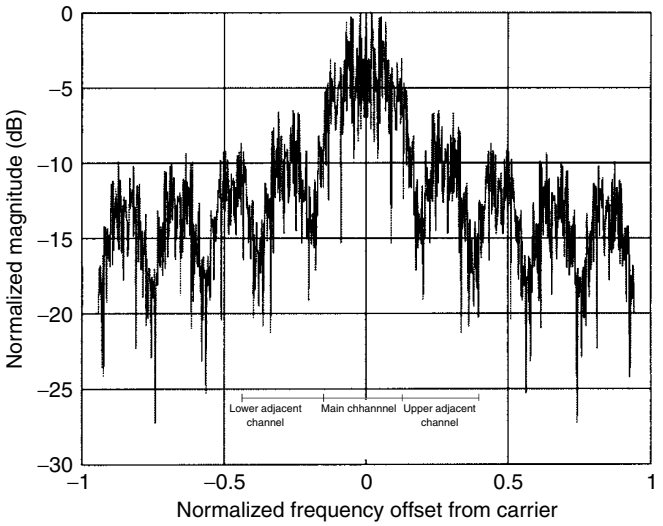


FIGURE 30.7 Spectrum of a digitally modulated signal.

The nodal formulation of the network equations is based on the application of KCL, which in its general form states that if a circuit is partitioned, then the total instantaneous current flowing into a partition is zero. This is an instantaneous requirement—physically it is only necessary that the net current flow be zero on average to ensure charge conservation. So this is an artificial constraint imposed by circuit analysis technology. The approach used in overcoming this restriction is to cast this issue as a modeling problem: it is the responsibility of the device modeler to ensure that a model satisfies KCL instantaneously. This results in many of the modeling limitations that are encountered. A general network is shown in Figure 30.8. The concept here is that every node of the circuit is pulled to the outside of the main body of the network. The main body contains only the constitutive relations and the required external nodes have the connectivity information to implement Kirchoff's laws. The result is that the constitutive relations are contained in the main body, but the variables, the node-voltages, and the external currents are clearly separated. This representation of a network enables as uniform a treatment as possible. It makes it very easy to add one element at a time to the network as variables are already defined. Indeed this is how all general purpose simulators work and the network equations are built up by inspection. Initially the network is defined with nothing in the main body and only the variables defined. Then each element is considered in turn and the describing relations added to the evolving network equation matrix. This representation serves us well when it comes to harmonic balance. Applying KCL to each of the nodes of the network the following matrix network equation is obtained:

$$\mathbf{YV} = \mathbf{J}. \tag{30.8}$$

Here **Y** is the nodal admittance matrix of the network, **V** is the vector of node voltages (i.e., voltages at the nodes each referred to the reference node), and **J** is the vector of external current sources at each node. Expanding the matrix equation:

$$\begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1N} \\ y_{21} & y_{22} & \cdots & y_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ y_{N1} & y_{N2} & \cdots & y_{NN} \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_N \end{bmatrix} = \begin{bmatrix} J_1 \\ J_2 \\ \vdots \\ J_N \end{bmatrix}. \tag{30.9}$$

We will see this utilized in the formulation of SPICE and HB analyses.

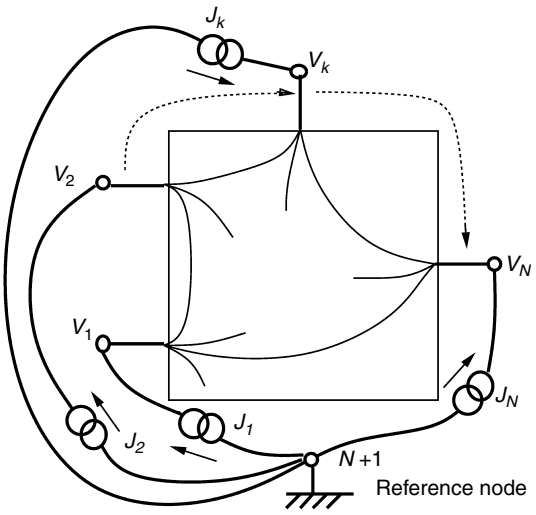


FIGURE 30.8 General network.

30.3 Time-Domain Circuit Simulation

The principal advantage of simulating circuits in the time domain is that it most closely resembles the real world. Phenomena such as chaos, instability, subharmonic generation, and parametric effects can be accurately simulated without the a priori knowledge of the spectral components of the signals in a circuit.

30.3.1 Direct Integration of the State Equations

The most direct method for analyzing nonlinear circuits is numerical integration of the differential equations describing the network. By applying Kirchoff's voltage and current laws and using the characteristic equations for the circuit elements (generally using the modified nodal formulation), the state equations can be written as a set of coupled first-order differential equations:

$$\dot{\mathbf{X}} = f(\mathbf{X}, t) \tag{30.10}$$

where, for example, the time derivative of a quantity such as voltage or current is a function of time and of the voltages and currents in the circuit. More generally the state equations are rearranged and written in the implicit form

$$g(\dot{\mathbf{X}}, \mathbf{X}, t) = 0 \tag{30.11}$$

where $\mathbf{X} = [X_1, X_2, \ldots, X_N]^T$ is a set of voltages and currents, typically at different nodes and different time instants. The general formulation of Equation 30.11 is discretized in time and solved using a numerical integration procedure. This modeling approach can be used with many systems as well as circuits and was the only approach considered in the early days of circuit simulation (in the 1960s). Unfortunately, it was not robust except for the simplest of circuits. SPICE-like analysis, considered next, solves the same problem but in a much more robust way.

30.3.2 SPICE: Associated Discrete Circuit Modeling

SPICE is the most common of the time domain methods used for nonlinear circuit analysis. This method is fundamentally the same as that just described in that the state equations are integrated numerically, however the order of operations is changed. The time discretization step is applied directly to the equations describing the circuit element characteristics. The nonlinear differential equations are thereby converted to nonlinear algebraic equations. Kirchoff's voltage and current laws are then applied to form a set of algebraic equations solved iteratively at each time point.

Converting the differential equations describing the element characteristics into algebraic equations changes the network from a nonlinear dynamic circuit to a nonlinear resistive circuit. In effect, the differential equations describing the capacitors and inductors, for example, are approximated by resistive circuits associated with the numerical integration algorithm. This modeling approach is called associated discrete modeling or just companion modeling. The term "associated" refers to the model's dependence upon the integration method while "discrete" refers to the model's dependence on the discrete time value.

The numerical integration algorithm is the means by which the element characteristics are turned into difference equations. Three low order numerical integration formulas are commonly used: the Forward Euler formula, the Backward Euler formula, and the Trapezoidal Rule. A generalization of these to higher order is called the weighted integration formula from which the Gear Two method, available in some SPICE simulators, is derived. In all methods the aim is to estimate the state of a circuit at the next time instant from the current state of the circuit and derivative information. In one dimension and denoting the current state by x_0 and the next state by x_1, the basic integration step is

$$x_1 = x_0 + hx'. \tag{30.12}$$

The formulas differ by the method used to estimate x'.

In the Forward Euler formula, $x' = x'_\phi$ is used and the basic numerical integration step (Equation 30.12) becomes

$$x_1 = x_0 + hx'_0. \tag{30.13}$$

Numerical integration using the forward Euler formula is called a predictor method as information about the behavior of the waveform at time t_0, x'_0, is used to predict the waveform at t_1.

In the Backward Euler Formula, $x' = x'_1$ is used and the discretized numerical integration equation becomes

$$x_1 = x_0 + hx'_1. \tag{30.14}$$

The obvious problem here is how to determine x'_1 when x_1 is not known. The solution is to iterate as follows: (1) assume some initial value for x_1 (e.g., using the Forward Euler formula); and (2) iterate to satisfy the requirement $x'_1 = f(x_1, t)$. Discretization using the Backward Euler formula is therefore called a predictor-corrector method.

In the Trapezoidal Rule, $x' = (x'_0 + x'_1)/2$ is used and the discretized numerical integration equation becomes

$$x_1 = x_0 + h\left(x'_0 + x'_1\right)\Big/2. \tag{30.15}$$

So the essence of the trapezoidal rule is that the slope of the waveform is taken as the average of the slope at the beginning of the time step and the slope at the end of the time step determined using the Backward Euler formula.

There is a significant difference in the numerical stability, accuracy, and run times of these methods, although all will be stable with a small enough step size. Note that stability is a different issue than whether or not the correct answer is obtained. The Backward Euler and Trapezoidal Rules will always be stable and these are the preferred integration methods. The Forward Euler method of discretization does not always result in a numerically stable method. This can be understood by considering that the Forward Euler method always predicts the response into the future and does not improve on the guess using other information that can be obtained. The Backward Euler and Trapezoidal Rule approaches use a prediction of the future state of a waveform, but then require iteration to correct any error and use derivative information as well as instantaneous information to achieve this. Generally, when any simulation strategy is first developed, predictor methods are used. However, in the long run, predictor-corrector methods are always adopted as these have much better overall performance in terms of stability and accuracy but do require much more development effort. Except for the Forward Euler method, none of the other methods are clearly the best choice in all circumstances, and experimentation should occur. Generally, we can say that for RF and microwave circuits that have resonant bandpass-pass characteristics, the Trapezoidal Rule tends to result in an over-damped response and the Backward Euler method results in an under-damped response. The effect of this on accuracy, the prime requirement, is not consistent and must be investigated for a specific circuit.

30.3.3 Associated Discrete Model of a Linear Element

The development of the associated discrete model (ADM) of an element begins with a time discretization of the constitutive relation of the element. The development for a linear capacitor is presented here as an example. The simplest algorithm to use in developing this discretization is the Backward Euler integration formula. The Backward Euler algorithm for solving the differential equation

$$\dot{x} = f\left(x\right) \tag{30.16}$$

with step size $h = t_{n+1} - t_n$ is

$$x_{n+1} = x_n + hf\left(x_{n+1}\right) = x_n + h\dot{x}_{n+1} \tag{30.17}$$

where the subscript n refers to the nth time sample. The discretization is performed for each and every element independently by replacing the differential equation of Equation 30.16 by the constitutive relation of the particular elements.

For a linear capacitor, the charge on the capacitor is linearly proportional to the voltage across it so that $q = Cv$. Thus

$$i\left(t\right) = \frac{dq}{dt} = C\frac{dv}{dt} = C\dot{v}$$

or

$$v_{n+1} = \frac{1}{C} i_{n+1} \tag{30.18}$$

where the reference convention for the circuit quantities are defined in Figure 30.9.

Substituting Equations 30.18 into 30.19 and rearranging leads to the discretized Backward Euler model of the linear capacitor:

$$i_{n+1} = \frac{C}{h} v_{n+1} - \frac{C}{h} v_n. \tag{30.19}$$

This equation has the form

$$i_{n+1} = g_{eq} v_{n+1} - i_{eq} \tag{30.20}$$

and so is modeled by a constant conductance $g_{eq} = C/h$ in parallel with a current source $i_{eq} = -C/hv_n$ that depends on the previous time step, as shown in Figure 30.10. The associated discrete circuit models for all other elements are developed in the same way, but of course the development is usually much more complicated, especially for nonlinear and multiterminal elements, but the approach is the same. The final circuit combining the ADM of all of the elements is linear with resistors and current sources, as well as a few special elements such as voltage sources. This circuit is especially compatible with the nodal-formulation described by Equation 30.9. The linear circuit is then solved repeatedly with the circuit elements updated at each step and, if the circuit voltage and current quantities change by less than a specified tolerance, the time step advanced.

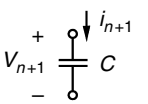


FIGURE 30.9 Reference direction for the circuit quantities of a capacitor.

The feature that distinguishes the associated discrete modeling approach from integration of the state equations for the system is that the discretization and particularly the Newton iteration is performed at the individual element level rather than at the top system level. The most important aspect of this is that special convergence treatments can be applied locally. For example, a diode has an exponential relationship between current and voltage and is the most difficult characteristic to handle. With the top-level systems-of-equations approach, any convergence

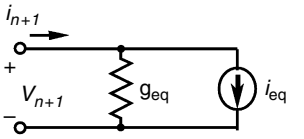


FIGURE 30.10 The associated discrete model of a two-terminal element.

scheme developed would need to be applied to all elements in a circuit, not just to the problem elements. In the associated discrete modeling, many local steps can be taken to improve convergence properties. This can include limiting the voltage and current changes from one iteration step to another. The scheme adopted depends on the characteristics of a particular element and heuristics developed in using it. It is this focus on local convergence control and embedding specific element knowledge in the element code that makes the SPICE approach so successful.

30.3.4 The Shooting Method

As has been mentioned, time-marching simulation has problems in determining the steady-state response because of the long simulation times that are involved. There is one elegant solution when the excitation is a sinusoid so that the response is known to be periodic. For strictly periodic excitation, shooting methods are often used to bypass the transient response altogether. This is advantageous in situations that would require many iterations for the transient components to die out. It is assumed that the nonlinear circuit has a periodic solution and that the solution can be determined by finding an initial state such that transients are not excited. If $\mathbf{x}(t)$ is the set of state variables obtained by a time-domain analysis, the boundary value constraint for periodicity is that $\mathbf{x}(t) = \mathbf{x}(t + T)$, where T is the known period. A series of iterations at time points between t and $t + T$ can be performed for a given set of initial conditions, and the condition for periodicity checked. Thus, in the shooting method, the problem of solving the state equations is converted into the two-point boundary value problem

$$\begin{aligned} \mathbf{x}(0) &= \mathbf{x}(T) \\ \mathbf{x}(T) &= \int_0^T \mathbf{f}(\mathbf{x}, \tau)\, d\tau + \mathbf{x}(0). \end{aligned} \tag{30.21}$$

If $\mathbf{x}(t) \neq \mathbf{x}(t + T)$ then a new set of initial conditions can then be determined using a gradient method based upon the error in achieving a periodic solution. Once the sensitivity of the circuit to the choice of initial conditions is established in this way, a set of initial conditions that establishes steady-state operation can be determined; this set is, of course, the desired solution. This iterative procedure can be implemented using the Newton's method iteration

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \left[\mathbf{I} - \frac{\partial \mathbf{x}^k(T)}{\partial \mathbf{x}^k(0)}\right]^{-1} \left[\mathbf{x}^k(0) - \mathbf{x}^k(T)\right] \tag{30.22}$$

where the superscripts refer to iteration numbers and $\mathbf{x}^k(T)$ is found by integrating the circuit equations over one period from the initial state $\mathbf{x}^k(0)$.

To begin the analysis, the period (T) is determined and the initial state ($\mathbf{x}^k(0)$) is estimated. Using these values, the circuit equations are numerically integrated from $t = 0$ to $t = T$ and the necessary derivatives calculated. Then, the estimate of the initial state is updated using the Newton iteration (Equation 30.22). This process is repeated until $\mathbf{x}(0) = \mathbf{x}(T)$ is satisfied within a reasonable tolerance.

Shooting methods are attractive for problems that have small periods. Unlike the direct integration methods, the circuit equations are only integrated over one period (per iteration). They are therefore more efficient, provided that the initial state can be found in a number of iterations that is smaller than the number of periods that must be simulated before steady-state is reached in the direct methods. Unfortunately, shooting methods can only be applied to find periodic solutions. Also, shooting methods become less attractive for cases where the circuit has a large approximate period, for example, when

several nonharmonic signals are present. The computation becomes further complicated when transmission lines are present, because functional initial conditions are then required to establish the initial conditions at every point along the line (corresponding to the delayed instants in time seen at the ports of the line).

In multitone situations when only one signal is large and when operating frequencies are not so high that distributed effects are important, the large tone response can be captured using the shooting method and then the frequency conversion method described in the next section can be used to determine the response with the additional small signals present.

30.3.5 Frequency Conversion Matrix Methods

In many multitone situations, one of two or more impressed non-commensurate tones is large while the others are much smaller. In a mixer, a large local oscillator, LO (which is generally 20 dB or more larger than the other signals) pumps a nonlinearity, while the effect of the other signals on the waveforms at the nonlinearities is negligible. The pumped time-invariant nonlinearity can be replaced by a linear time-varying circuit without an LO signal. The electrical properties of the time-varying circuit are described by a frequency domain conversion matrix. This conversion matrix relates the current and voltage phasors of the first order sidebands with each other. In other words, by performing a fast, single-tone shooting method or harmonic balance analysis with only the LO impressed upon it, the AC operating point of the mixer may be determined and linearized with respect to small-signal perturbations about this point. This information is already available in the Jacobian, which is essentially a gradient matrix relating the sensitivity of one dependent variable to another independent variable. A two-tone signal can be rewritten to group the LO waveform, $x_{LO}(t)$ terms and the first order sidebands as

$$x(t,j) = x_{LO}(t) + \mathrm{Re}\left\{\sum_{p=0}^{N_A} \mathbf{X}_{p,1} e^{j\left(p\omega_{LO}+\omega_{RF}\right)} + \sum_{p=0}^{N_A} \mathbf{X}_{p,-1} e^{j\left(p\omega_{LO}-\omega_{RF}\right)}\right\} \tag{30.23}$$

where $\mathbf{X}_{p,1}$ and $\mathbf{X}_{p,-1}$ are vectors of the spectral components at the first order sidebands of the pth harmonic of the LO. For voltage controlled nonlinearities, the output quantities (the X's) are current phasors so that the expression relating the IF current to the RF voltage is

$$\left[\mathbf{I}_{p,1}, \mathbf{I}_{p,-1}\right]^{\mathrm{T}} = \mathbf{Y}_C \left[\mathbf{V}_{0,1}, \mathbf{V}_{0,-1}\right]^{\mathrm{T}}. \tag{30.24}$$

Here $\mathbf{Y}_C$ is the admittance conversion matrix and can be used in much the same manner as a nodal admittance matrix. Alternatively, for current-controlled nonlinearities the following could be used:

$$\left[\mathbf{V}_{0,1}, \mathbf{V}_{0,-1}\right]^{\mathrm{T}} = \mathbf{Z}_C \left[\mathbf{I}_{0,1}, \mathbf{I}_{0,-1}\right]^{\mathrm{T}} \tag{30.25}$$

where $\mathbf{Z}_C$ is the impedance conversion matrix. Nonlinearities with state variable descriptions or mixed voltage- and current-controlled descriptions require a combination of Equations 30.24 and 30.25 to derive a modified nodal admittance formulation.

30.3.6 Convolution Techniques

The fundamental difficulty encountered in integrating RF and microwave circuits in a transient circuit simulator arises because circuits containing nonlinear devices or time dependent characteristics must be

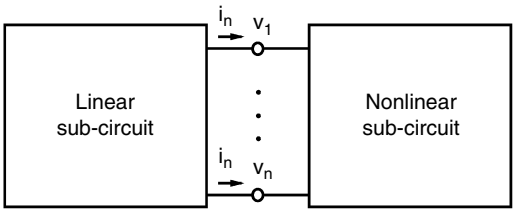


FIGURE 30.11 Circuit partitioned into linear and nonlinear sub-circuits.

characterized in the time domain while distributed elements such as transmission lines with loss, dispersion, and interconnect discontinuities are best simulated in the frequency domain. Convolution techniques are directed at the simulation of these circuits.

The procedure begins by partitioning the network into linear and nonlinear subcircuits as shown in Figure 30.11. In a typical approach the frequency domain admittance (y) parameter description of the distributed network is converted to a time domain description using a Fourier transform. This time domain description is then the Dirac delta impulse response of the distributed system. Using the method of Green's function, the system response is found by convolving the impulse response with the transient response of the terminating nonlinear load. Normally this requires that the impulse response be extended in time to include many reflections. While this technique can handle arbitrary distributed networks, a difficulty arises as the y parameters of a typical multiconductor array have a wide dynamic range. For a low loss, closely matched, strongly coupled system, the y parameters describing the coupling mechanism approach zero at low frequencies and become very large at high frequencies. Conversely, the transmission and self-admittance y parameters approach infinity at DC and zero at resonance frequencies. Both numerical extremes are important in describing the physical process of reflections and crosstalk. The dynamic range of the time domain solution is similarly large and values close to zero are significant in determining reflections and crosstalk. Consequently, aliasing in the frequency domain to time domain transformation can cause appreciable errors in the simulated transient response. The problem is considerably reduced by using resistive padding at the linear-nonlinear circuit interface to reduce the dynamic range of the variables being transformed. The effect of the padding can then be removed in subsequent iteration.

30.4 Harmonic Balance: Mixed Frequency and Time Domain Simulation

The HB procedure has emerged as a practical and efficient tool for the design and analysis of nonlinear circuits in steady-state with sinusoidal excitation. The harmonic balance method is a technique that allows efficient solution of the steady-state response of a nonlinear circuit. For example, the steady-state response of a circuit driven by one or more sinusoidal signals is also a sum of sinusoids and includes tones at frequencies other than those of the input sinusoids (e.g., harmonics and difference frequencies). The response does not need to be periodic to be steady-state and with narrowband systems it is common to call the response to a complicated narrowband input as being quasi-periodic. Usually we are not interested in the transient response of the circuit such as when the power supply is turned on or when a signal is first applied. Thus much of the behavior of the circuit is not of interest. The harmonic balance procedure is a technique to extract just the information that is required to describe the steady-state response. The method may also be compared to the solution of a homogeneous, ordinary differential equation. A solution that is the sum of sinusoids of unknown amplitudes is substituted into the differential equation. Using the orthogonality of the sinusoids, the resulting problem simplifies to solving a set of nonlinear algebraic equations for the amplitudes of the sinusoids. There are several methods of solving for (complex) amplitudes, which will be discussed later in this chapter.

The HB method formulates the system of nonlinear equations in either domain (although more typically the time domain), with the linear contributions calculated in the frequency domain and the nonlinear contributions in the time domain. This is a distinct advantage for microwave circuits, in that distributed and dispersive elements are then much more readily modeled analytically or using alternative electromagnetic techniques based in the frequency domain.

While it is common to refer to the nonlinear calculations as being in the time domain, the most usual HB implementations require that the nonlinear elements be described algebraically, that is without derivatives or other memory effects. Thus a nonlinear resistor is described, for example, as a current as a nonlinear function of instantaneous voltage. So given the voltage across the nonlinear resistor at a particular time, the current that flows at the same instant can be calculated. A nonlinear capacitive element must be expressed as a charge which is a nonlinear function of instantaneous voltage. Then a sequence of charge values in time is Fourier transformed so that phasors of charge are obtained. Each phasor of charge is then multiplied by the appropriate $j\omega$ to yield current phasors.

30.4.1 Problem Formulation

The HB method seeks to match the frequency components (harmonics) of current at the interface of two sub-circuits—one linear and one nonlinear. The sub-circuits are chosen in such a way that nonlinear elements are partitioned into one sub-circuit, linear elements into another, and (in some approaches) sources into a third (see Figure 30.12). The edges at the linear/nonlinear interface connect the two circuits and define corresponding nodes; current flowing out of one circuit must equal that flowing into the other. Every node in the nonlinear circuit is "pulled out" of the nonlinear sub-circuit so that it is at the interface and becomes part of the error function formulation. Matching the frequency components in each edge satisfies the continuity equation for current. The current at each edge is obtained by a process of iteration so that dependencies are satisfied for both the linear and nonlinear sides of the circuit.

The unknowns are found by forming an error function—typically the KCL error at the linear/nonlinear interface. This error function is minimized by adjusting the voltages at the interface. Every node in the nonlinear sub-circuit is therefore considered to be connected to the linear sub-circuit. If the total circuit has N nodes, and if $\mathbf{v}$ is the vector of node voltage waveforms, then applying KCL to each node yields a system of equations

$$f(\mathbf{v},t)=\mathbf{i}\left[\mathbf{v}(t)\right]+\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{q}\left[\mathbf{v}(t)\right]+\int_{-\infty}^{t} y(t-\tau)\mathbf{v}(\tau)\mathrm{d}\tau+\mathbf{i}_S(t)=0 \tag{30.26}$$

where the nonlinear circuit is chosen to contain only voltage-controlled resistors and capacitors for representational ease. The quantities $\mathbf{i}$ and $\mathbf{q}$ are the sum of the currents and charges entering the nodes from the nonlinearities, $\mathbf{y}$ is the matrix impulse response of the linear circuit with all the nonlinear devices removed, and $\mathbf{i}_S$ are the external source currents.

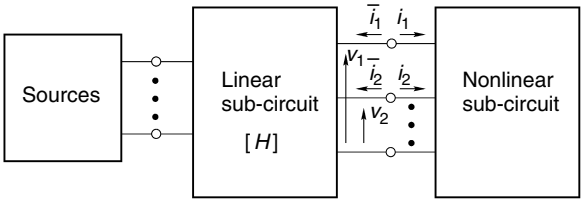


FIGURE 30.12 Circuit partitioned into linear, nonlinear, and source sub-circuits.

In the frequency domain, the convolution integral maps into **YV**, where **V** contains the Fourier coefficients of the voltages at each node and at each harmonic, and **Y** is a block node admittance matrix for the linear portion of the circuit. The system of Equation 30.26 then becomes, on transforming into the frequency domain

$$\mathbf{F}(\mathbf{V}) = \mathbf{IV} + \Omega\mathbf{QV} + \mathbf{YV} + \mathbf{I}_S = 0 \tag{30.27}$$

where Ω is a matrix with frequency coefficients (terms such as $j\Omega_k$) representing the differentiation step. The notation here uses small letters to represent the time domain waveforms and capital letters for the frequency domain spectra. This equation is, then, just KCL in the frequency domain for a nonlinear circuit. HB seeks a solution to Equation 30.27 by matching harmonic quantities at the linear-nonlinear interface. The first two terms are spectra of waveforms calculated in the time domain via the nonlinear model, that is,

$$\mathbf{F}(\mathbf{V}) = \Im i\left(\Im^{-1}\mathbf{V}\right) + \Omega\Im q\left(\Im^{-1}\mathbf{V}\right) + \mathbf{YV} + \mathbf{I}_S = 0 \tag{30.28}$$

where $\Im$ is the Fourier transform and $\Im^{-1}$ is the inverse Fourier transform.

The solution of Equation 30.28 can be obtained by several methods. One method, known as relaxation, uses no derivative information and is relatively simple and fast, but is not robust. In a relaxation method the error function is taken to zero by adjusting current phasors or voltage phasors on successive iterations using what is in effect very limited derivative information. Alternatively, gradient methods can be used to solve either a system of equations (e.g., using a quasi-Newton method) or to minimize an objective function using a quasi-Newton or search method.

The Newton and quasi-Newton methods require derivative information to guide the error minimization process. Calculation of these derivatives is computationally intensive and generally the equations for these require considerable development. As with all the harmonic balance methods, the number of nodes used can be reduced by "burying" internal nodes within the linear network, which then becomes a single multiterminal sub-circuit as far as harmonic balance is concerned. The system of equations is then reduced accordingly. Once the "interfacial" node voltages are known, any internal node voltage can be found by using simple linear analysis and the full *y* matrix for the linear circuit.

30.4.2 Multitone Analysis

The problem with multitone analysis reduces to implementing a method to perform the multifrequency Fourier transform operations required in solving Equation 30.28. This also requires developing the multifrequency Jacobian required in a Newton-like procedure. Time–frequency conversion for multitone signals can be achieved using nested Fourier transform operations. This is implemented using the multidimensional fast Fourier transform (MFFT). Application of the MFFT, in which the Fourier coefficients are themselves periodic in the other dimensions, requires that the multiple tones (in each dimension) be truly orthogonal, that is, not integer multiples of each other. If the two tones are frequency degenerate, then the method fails because orthogonality of the bases is a requirement for determining the Fourier coefficients in that basis. In such a case, one of the tones is slightly shifted to ensure that the technique can be applied.

The most general and easily programmed of the Fourier transform techniques applied to the HB method is the almost-periodic discrete Fourier transform (APDFT) algorithm. After truncation, consider the K arbitrarily spaced frequencies 0, ω_1, ω_2, …, ω_{K-1} generated by the nonlinearity. Then $\sum_{k=0}^{K-1} X_k^C \cos \omega_1 t_1 + X_k^2 \sin \omega_1 t_1 = x(t)$ may be sampled at S time points, resulting in a set of S equations and $2K - 1$ unknowns:

$$\begin{bmatrix} 1 & \cos\omega_1 t_1 & \sin\omega_1 t_1 & \cdots & \cos\omega_{K-1} t_1 & \sin\omega_{K-1} t_1 \\ 1 & \cos\omega_2 t_2 & \sin\omega_2 t_2 & \cdots & \cos\omega_{K-1} t_2 & \sin\omega_{K-1} t_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & \cos\omega_1 t_S & \sin\omega_2 t_S & \cdots & \cos\omega_{K-1} t_S & \sin\omega_{K-1} t_S \end{bmatrix} \begin{bmatrix} X_0 \\ X_1^C \\ X_1^S \\ \vdots \\ X_{K-1}^C \\ X_{K-1}^S \end{bmatrix} = \begin{bmatrix} x(t_1) \\ x(t_1) \\ \vdots \\ x(t_S) \end{bmatrix}. \quad (30.29)$$

The number of samples S must be at least $2K-1$ to uniquely determine the coefficients. This equation may compactly be written as

$$\Gamma^{-1}\mathbf{X} = x \text{ or } \Gamma x = \mathbf{X} \quad (30.30)$$

where Γ and Γ^{-1} are known as an APDFT. Thus the multifrequency transform can be performed as a matrix operation but spectrum mapping and fast Fourier transformation is much faster.

Combining the above procedures yields the time invariant form of HB. This is referred to as just HB. This technique is very efficient for simulating circuits with just a few active devices and a few tones, as then there are only a few unknowns. Problems arise as the number of active devices increases or the number of tones becomes large as the size of the problem increases significantly. Still, digitally modulated signals can be reasonably modeled by considering a very large number of tones.

30.4.3 Method of Time-Varying Phasors

Harmonic balance using time-variant phasors (also known as transient envelope) is ideally suited to the representation and characterization of circuits with digitally modulated signals. In contrast to time-variant HB, where the assumed phasor solution was time invariant, we instead assume a solution of the form

$$\mathbf{V}_k(j\omega) = \text{real}\left\{\sum_{m=0}^{n} V_m(t)\exp\left[jm\omega(t) + \phi_m(t)\right]\right\} \quad (30.31)$$

where in general the amplitude, frequency, and phase of each term are allowed to vary with respect to time. If $V_m(t)$ varies slowly with respect to the carrier frequency, we are in essence solving for the envelope of the signal at each node without the requisite memory requirements of time-invariant HB, or the frequency domain dynamic range and resolution problems of time domain methods. Taking the Fourier transform of each summation term in Equation 30.31 results in a highly resolved power spectral density distribution approximation of the digitally modulated signal, not an ill-conditioned approximation, as with time-invariant HB.

30.5 Frequency Domain Analysis of Nonlinear Circuits

Frequency domain characterization of RF and microwave circuits directly provides the types of performance parameters required in communication systems as well as many other applications of RF and microwave circuits. The HB method uses Fourier transformation to relate sequences of instantaneous current, voltage, and charge to their (frequency domain) phasor forms. In frequency domain nonlinear analysis techniques alternative mappings are used. There are many types of mappings for arriving at a set of (say, current) phasors as a nonlinear function of another set of (say, voltage) phasors.

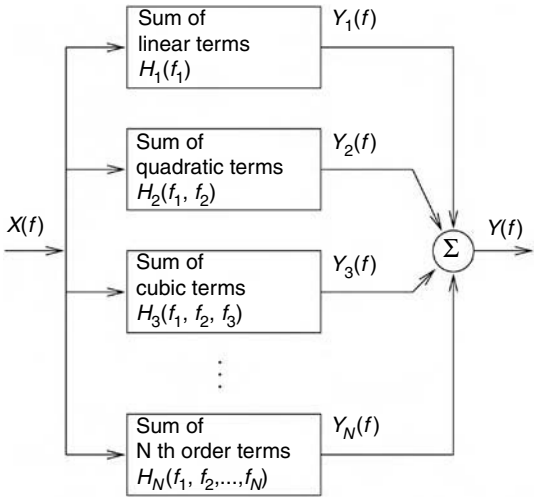


FIGURE 30.13 Mapping concept of frequency domain analysis.

The common underlying principle of frequency domain nonlinear analysis techniques is that the spectrum of the output of a broad class of nonlinear circuits and systems can be calculated directly given the input spectrum input to the nonlinear system. The mapping operation is depicted in Figure 30.13 and is the concept behind most RF and microwave behavioral modeling approaches. Some techniques determine an output frequency component by summing calculations of individual intermodulation products. For example, the product of two tones is, in the time domain, the product of two sinusoids. The trigonometric expansion of this yields two intermodulation products with frequencies that are the sum and difference, respectively, of the frequencies of the tones. Power series techniques use trigonometric identities to expand the power series and calculate each intermodulation product individually. Algorithms sum these by frequency to yield the output spectrum. At the coarse end of the scale are Volterra series-based techniques that evaluate groups of intermodulation products at a single frequency. Some frequency-domain nonlinear analysis techniques are noniterative, although these are restricted to unilateral systems. Others, known as frequency domain spectral balance techniques, are iterative, being the frequency domain equivalent of the harmonic balance techniques. The term spectral balance is used to distinguish the frequency domain techniques from the harmonic balance techniques as the latter term has come to be solely applied to mixed time and frequency domain methods in which Fourier transformation is used. Intermediate between these extremes are techniques that operate by converting a nonlinear element into a linear element shunted by a number of controlled current sources. This process is iterative and at each iteration a residual nonlinear element is left that reduces from one iteration to another. This is the basis of one of the Volterra series analysis techniques called the method of nonlinear currents, which is also discussed in the next section.

30.5.1 Volterra Analysis

Expanding on Volterra analysis illustrates the concepts behind functional analysis of circuits in the frequency domain. Volterra series have the form

$$G(x) = \sum_{n=0}^{\infty} F_n(x) \tag{30.32}$$

where $F_n(x)$ is a regular homogeneous functional such that

$$F_n(x) = \int_a^b \cdots \int_a^b h_n(\chi_1, \chi_2, \ldots, \chi_n) x(\chi_1) x(\chi_2) \cdots x(\chi_n) d\chi_1 d\chi_2 \cdots d\chi_n \tag{30.33}$$

and the functions $h_n(\chi_1, \chi_2, \ldots, \chi_n)$ are known as the nth order Volterra kernels. It can be used as a time domain description (with the χ's replaced by t) of many nonlinear systems including nonlinear microwave circuits that do not exhibit hysteresis. In this case, the nth order kernel, h_n, is called the nonlinear impulse response of the circuit of order n. Equation 30.33 is then interpreted as an n dimensional convolution of an nth order impulse response (h_n) and the input signal (x). The total response $G(x)$ is the summation of the different order responses $F_n(x)$. Note that for a linear system there is only a first order response so that the total response of the system is the conventional convolution integral

$$G(x) = F_0 + \int_a^b h_1(\tau) d\tau \tag{30.34}$$

where F_0 is just a DC offset.

The important concept here is that the total response of a signal is the summation of a number of responses of different order. This scheme only works if, as the order n increases, the contribution to the response gets smaller and eventually insignificant. The reason this works for many RF and microwave circuits is that the response is close to linear and nonlinear behavior is a departure from linearity. A weak nonlinearity could be represented with just the first few terms of such a series.

In analyzing nonlinear circuits it is not necessary to deal with the Volterra series which, here, is in the time domain. Mostly the frequency domain form is used, which is expressed in terms of Volterra nonlinear transfer functions. Mathematically these are obtained by taking the n-fold Fourier transform of h_n:

$$H_n(f_1, f_2, \ldots, f_n) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_n(\tau_1, \tau_2, \ldots, \tau_n) e^{-j2(f_1\tau_1 + \ldots \tau_n)} d\tau_1 d\tau_2 \cdots d\tau_n \tag{30.35}$$

where H_n is called the nonlinear transfer function of order n. The time-domain input–output relation $y(t) = f[x(t)]$ can be put in the form

$$y(t) = \sum_{n=1}^{\infty} y_n(t) \tag{30.36}$$

where

$$y_n(t) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_n(\tau_1, \ldots, \tau_n) x(t - \tau_1) \cdots x(t - \tau_n) d\tau_1 \cdots d\tau_n \tag{30.37}$$

and $x(t)$ is the input. Taking the n-fold Fourier transform of both sides we have an expression for the spectrum of the nth order component of the output

$$\mathbf{Y}_n = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} H_n(f_1, \ldots, f_n) \delta(f - f_1 - \cdots - f_n) \prod_{i=1}^{n} X(f_i) df_i \tag{30.38}$$

where $\mathbf{X}_n(f)$ is the Fourier transform of $x(t)$, $\mathbf{Y}_n(f)$ is the Fourier transform of $y_n(t)$, and $\delta(\cdot)$ is the delta function. This expresses the nth order terms of the output as a function of the input spectrum. The order

of the terms refers to the fact that multiplication of the input by a constant A results in multiplication of the nth order terms by A^n. Then a frequency domain series for the output can be written as

$$Y(f) = \sum_{n=1}^{\infty} Y_n(f) \tag{30.39}$$

in terms of the input spectrum and the nonlinear transfer functions. $\mathbf{Y}_n(f)$ is the nth order response and corresponds to the response of the nth order term in the power series description of the nonlinearity.

The method of nonlinear currents enables the direct calculation of the response of a circuit with nonlinear elements that are described by a power series. Here a circuit is first solved for its linearized response described by zero and first order Volterra nonlinear transfer functions. Considering only the linearized response allows standard linear circuit nodal admittance matrix techniques to be used. The second order response, described by the second order Volterra nonlinear transfer functions can then be represented by controlled current sources. Thus the second order sources are used as excitations again enabling linear nodal admittance techniques to be used. The process is repeated for the third- and higher-order node voltages and is easily automated in a general purpose microwave simulator. The process is terminated at some specified order of the Volterra nonlinear transfer functions. This is a noniterative technique, but relies on rapid convergence of the Volterra series restricting its use to moderately strong nonlinear circuits.

30.6 Summary

SPICE is at its best when simulating large circuits as memory and computation time increase a little more than linearly after a circuit reaches a certain size. However, to determine the response to sinusoidal excitation requires simulation over a great many cycles until the transient response has died down. A major problem in itself is determining when the steady state has been achieved. A similar problem occurs with narrowband modulated signals, which can have many millions of RF cycles before the response appears to be steady state. For example, a typical sequence length for the (digitally modulated) GSM format is 0.577 ms, although the time step would be on the order of 50 ps to capture the fundamental and harmonics of a 1 GHz carrier. This results in 10^7 time-points and hence the same order discrete Fourier transforms. Fourier transformation, for example, using a fast Fourier transform, of the simulated waveform is required to determine its spectral content. This is not too complex a task if the exciting signal is a single frequency, but if the signal driving the nonlinear circuit has non-commensurable frequency components or is digitally modulated, then the procedure is more difficult and the effect of numerical noise is exaggerated. Even low-level numerical noise may make it impossible to extract a low-level tone in the presence of a large tone. The ability to detect a small tone defines the dynamic range of a simulator in RF and microwave applications and SPICE-analysis has poor performance in this case.

There is also a fundamental approximation error present in the SPICE algorithm due to what amounts to a z-domain approximation to the frequency domain characteristics of the circuit. The consequence is that time steps must be short for reasonable dynamic range. This also makes it particularly difficult to represent circuits with strongly varying narrowband frequency response. Recent extensions to SPICE—the shooting method with the frequency conversion method and convolution techniques—increase the applicability of SPICE to RF and microwave circuits. In spite of the difficulties, SPICE remains the only method of determining the transient response of a circuit.

HB analysis of circuits achieves significant computation savings by assuming that the signals in a circuit are steady state, described by a sum of sinusoids. The coefficients and phases of these sinusoids are solved for and not the transient response. HB has a significant computation time advantage over SPICE for small to medium RF and microwave circuits. However, the time increases rapidly as circuit size increases. HB lends itself well to optimization and to analysis of multifunction circuits including amplifiers, oscillators, mixers, frequency converters, and numerous types of control circuits such as limiters and

switches, if transient effects are not of concern. Another major advantage of the HB method is that linear circuits can be of practically any size, with no significant decrease in speed if additional internal nodes are added, or if elements of widely varying time constants are used (such is not the case with time domain simulators). Two extensions, separately implemented, also increase the usefulness of HB. The method of time-variant phasors enables digitally modulated signals to be handled. The second extension using matrix-free methods enables Harmonic Balance to handle very rich spectra and thus also approximately treat digitally modulated signals.

All of the techniques discussed here have been implemented in circuit simulators developed for RF and microwave circuit modeling. Many other simulator technologies exist, but these are within the overall framework of the discussion here. The reader is directed to the Further Information list for exploration of other technologies and for greater detail on those treated here.

Further Information

The bases of circuit simulation are described in J. Vlach and K. Singhal, *Computer Methods for Circuit Analysis and Design*, Van Nostrand Reinhold, 1983, ISBN 0442281080; R.M. Kielkowski, *Inside Spice*, McGraw-Hill, 1998, ISBN 0079137121; and L. T. Pillage, R. A. Rohrer, and C. Visweswariah, *Electronic Circuit and System Simulation Methods*, McGraw-Hill, 1995, ISBN 0070501696. These three books are oriented toward SPICE-like analysis. Details on the algorithms used in SPICE are given in A. Vladimirescu, *The SPICE Book*, J. Wiley, 1994, ISBN 0471609269, and the techniques used in developing the associated discrete models used in SPICE in P. Antognetti and G. Massobrio, *Semiconductor Device Modeling with SPICE*, McGraw-Hill, 1988, ISBN 0070021538. In addition to the above, a short discussion of SPICE errors relevant to modeling RF and microwave circuits is contained in A. Brambilla and D. D'Amore, The simulation errors introduced by the SPICE transient analysis, *IEEE Trans. on Circuits and Systems-I: Fundamental Theory and Application,* 40, 57–60, January 1993. Circuit simulations oriented toward microwave circuit simulation are described in J. Dobrowolski, *Computer-Aided Analysis, Modeling, and Design of Microwave Networks: the Wave Approach,* Artech House, 1996, ISBN 0890066698; P. J. C. Rodrigues, *Computer-Aided Analysis of Nonlinear Microwave Circuits,* Artech House, 1998, ISBN 0890066906; and G. D. Vendelin, A. M. Pavio, and U L. Rohde, *Microwave Circuit Design Using Linear and Nonlinear Techniques,* Wiley, 1990, ISBN 0471602760. As well as providing a treatment of microwave circuit simulation, the following book provides a good treatment of Volterra analysis: S. A. Maas, *Nonlinear Microwave Circuits*, IEEE Press, 1997, ISBN 0780334035. Simulation of microwave circuits with digitally modulated signals is given in J. F. Sevic, M. B. Steer and A. M. Pavio, Nonlinear analysis methods for the simulation of digital wireless communication systems, *Int. J. on Microwave and Millimeter Wave Computer Aided Engineering*, 197–216, May 1996. A review of frequency domain techniques for microwave circuit simulation is given in M. B. Steer, C. R. Chang and G. W. Rhyne, Computer aided analysis of nonlinear microwave circuits using frequency domain spectral balance techniques: the state of the art, *Int. J. on Microwave and Millimeter Wave Computer Aided Engineering*, 1, 181–200, April 1991.

31

Computer Aided Design (CAD) of Microwave Circuitry

Ron Kielmeyer
RF Micro Devices

31.1 Introduction

The growth of personal communication and the internet industries along with the need for portability has resulted in an ever increasing demand for low cost high volume microwave circuitry. The commercialization of GaAs and SiGe wafer processing and the simultaneous reduction in the physical size of the silicon devices has enabled the development of complex microwave circuitry which can, no longer, be designed without the aid of sophisticated computer aided design (CAD) circuit simulators. This chapter will discuss the typical steps involved in a design cycle, some basic requirements a CAD program should have, and a look at the theory behind the most popular CAD programs in use today.

31.2 Initial Design

A typical flow chart for any design cycle is shown in Figure 31.1. The design cycle obviously begins with the identification of the need to create a circuit and/or system function such as an amplifier, a mixer, or a whole receiver. These circuits or systems must satisfy certain characteristics or specifications. Once the specifications are known, passive and active devices from the best technology to achieve these specifications are chosen. Circuit topologies may be explored simultaneously with device selection. Device vendors will usually provide a computer representation for their devices. This representation might be as simple as optimum matching impedances or measured *s*-parameters but could be as complex as a full nonlinear

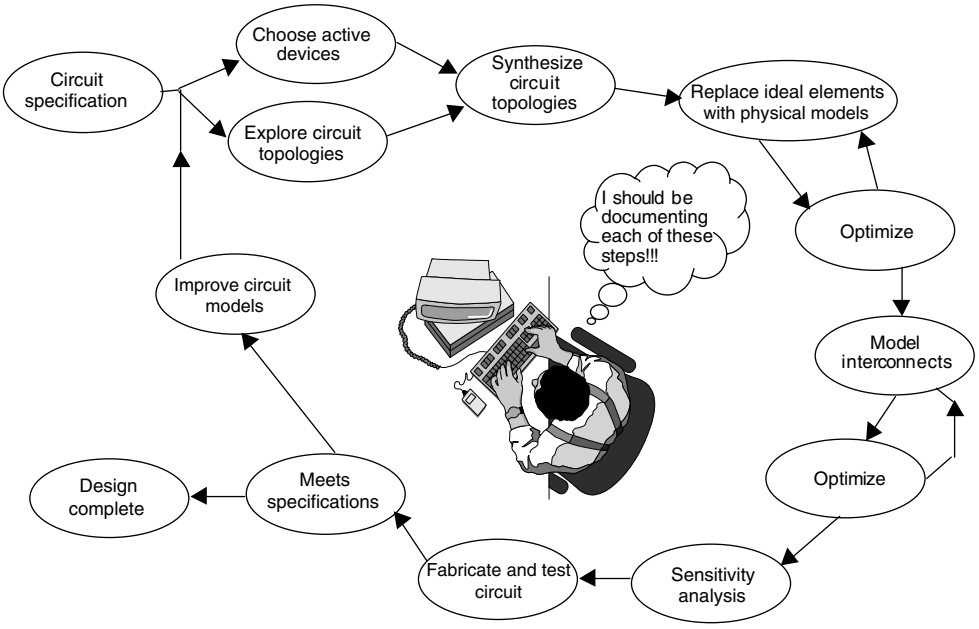


FIGURE 31.1 The design cycle.

mathematical model. With more complex and accurate device models, computer simulations are more useful because they can yield more information about how a circuit will behave and could possibly identify problems and specification deficiencies, early on, in the design cycle.

Synthesis programs, if available, are used to determine the best possible performance and topology. With these programs, many different topologies can be explored or identified as possible candidates to realize the function. Because nonideal elements will alter the ideal response of the topologies, the ideal topologies generated by the synthesis program must exceed the design specifications. Since performance will only deteriorate from the idealized case, the circuit designer spending an inordinate amount of time optimizing the ideal element values is wasteful because ultimately the ideal elements will need to be replaced with accurate physical models and doing this early in the design cycle saves time. The designer also might choose to explore more than one available topology because one topology may be less sensitive to the physical models than another topology.

31.3 Physical Element Models

As was stated in Section 31.2, ideal elements must be replaced by models of the physical devices. These models are, typically, subcircuits made up of ideal elements. For example, an ideal capacitor, physically realized by a chip capacitor, must have a model that can account for the finite inductance of the terminals and the finite resistive loss inherent in the physical device. Figure 31.2 shows a model for a physical capacitor mounted on a printed circuit board. In this model, C is the desired or ideal capacitance while C_{pad} represents the shunt capacitance of the metal pads on the circuit board that the capacitor is soldered to R_{s} and L_{s} take into account the inductance and conductivity of the metal plates which form the capacitance. R_{par} and C_{par} account for the parallel resonant characteristics of the capacitor. The extra elements R_{s}, R_{par}, C_{par}, C_{pad}, and L_{s} are commonly referred to as "parasitic" elements. By replacing the ideal elements with nonideal models, one at a time, the designer can accomplish two important practicalities. First, the sensitivity of the circuit to the nonideal element can be evaluated. Second, an optimization step can be performed on the remaining ideal elements in order to bring the circuit back to within the design specifications.

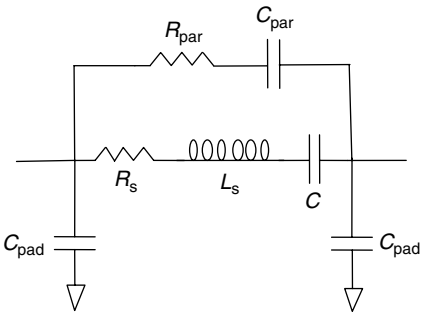


FIGURE 31.2 Physical model of a chip capacitor.

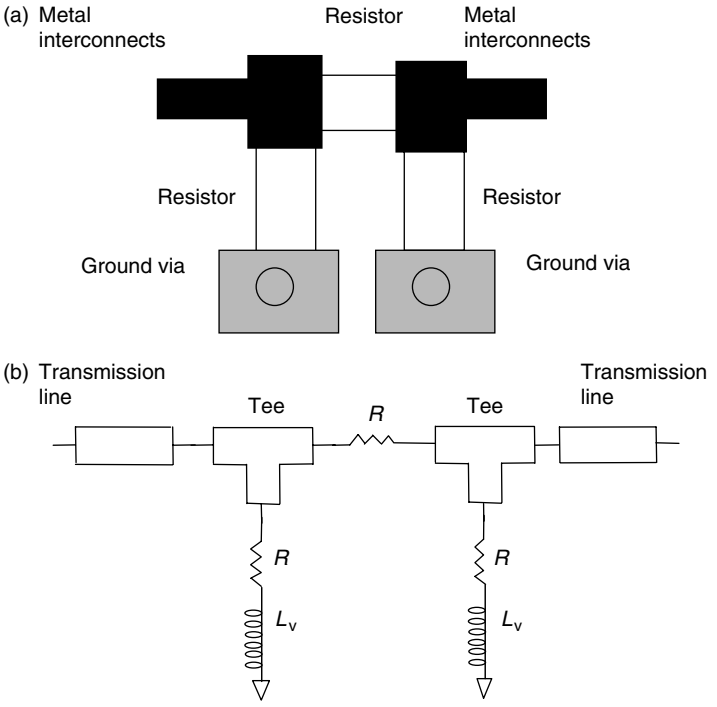


FIGURE 31.3 Simple PI pad resistive attenuator (a) physical layout and (b) CAD model.

31.3.1 Layout Effects

As the process of replacing the idealized elements continues, the physical layout of the circuit must be introduced into the analysis. Two physical elements cannot be connected with zero length metal patterns. How close the elements can be placed is usually determined by how close the manufacturing process can place them. The metal patterns used to interconnect the physical devices must be introduced in the form of transmission lines and/or transmission line junctions. Figure 31.3 illustrates a simple PI resistive attenuator as might be realized using MMIC or MIC technology. The metal interconnects are modeled as microstrip transmission lines, followed by a Tee junction model. The resistors in this model are modeled as ideal resistors at most practical frequencies but at very high microwave frequencies it can be modeled as very high loss transmission lines. The ground vias are represented as inductors. As each of these physical effects is introduced, an optimization step can be performed in an attempt to meet or exceed the design specifications.

31.4 Sensitivity to Process Variation

After transforming many possible design topologies, sensitivity of the circuit to the nonideal or parasitic elements of the circuits as well as to the actual element values must be determined. One measure of the sensitivity of a circuit is the percent yield. To determine the percent yield of a circuit, the element values, parasitic elements, and/or known physical tolerances, are treated as independent random variables. A range for each random variable is given and the computer analysis program iterates through random samples of the variables with their given ranges. This process is called a Monte Carlo analysis. The ideal outcome of the Monte Carlo analysis would be that the circuit passes all specifications under any combination of the physically possible random variable values. The percent yield can be improved by performing an optimization on the circuit. The optimization includes the ranges assigned to the random variables in the analysis. At each step in the optimization process, the mean values of the random variables are varied. The end result is an overall increase in the percent yield.

In applying Monte Carlo analysis, the identification of the true random variables is important to the accuracy of the analysis. For example, in the case of the physical capacitance, R_s, R_{par}, C_{par}, will vary with the actual value of the capacitance (C), while the solder pad capacitance C_{pad} is related to the width (P_w) and length (P_l) of the pad as well as the distance from the pad to the ground plane below (h) and the dielectric constant ε_r of the substrate it is attached to. However, P_w and P_l are further related to the etching tolerances δ used in manufacturing of the substate. Likewise the transmission line Z_o and electrical length (θ) used to connect multiple components are also related to δ, h, ε_r. For maximum accuracy, rather than treating the values of R_s, R_{par}, C_{par}, C_{pad}, Z_o, and θ as independent random variables, equations must be developed relating these values to δ, h, ε_r, and C.

Design of experiments (DOEs) is another way to measure the sensitivity of the circuit to the random variables. In performing the DOE, each of the random variable values are independently incremented from their mean values. The computer keeps track of the analyzed responses and does the tedious bookkeeping task of incrementing all of the variables. Once this process is complete, statistical techniques are used to determine the sensitivity of each of the random variables to each of the specifications. When important correlations between parameters are known to exist, it is important that DOE simulations take them into account. The use of explicit correlation descriptions will limit the scope of the required design space and provide more realistic results.

31.4.1 Improving the Computer Simulation

Once the circuit has been fabricated and the measurements are taken and all the specifications have been met then the design may be finished. However, a possible cost reduction design cycle may be desirable at this point. On the other hand, if the specifications have not been met then modifications will have to be implemented. The start of the new design iteration will be a refinement of the computer model. Presumably, the computer predicted that the design would meet or exceed all of the specifications before the commitment to fabricate was made. In this case, the model was not adequate to describe the actual circuit as implemented. The cause of the model inaccuracy must be determined. In some cases, the actual physical models are not adequate and must be refined. This is the case for most simple circuits. In other cases, usually involving multiple circuits or components, there are interactions between the elements which were not known before the circuit was fabricated.

One common problem is to assume that measured *s*-parameters accurately depict the actual physical environment that the device is being used in. Suppose, for example, that you are using measured *s*-parameters provided by the manufacturer of a three-stage amplifier circuit. Built into these measured *s*-parameters is the amount of grounding in the fixture used to obtain the *s*-parameters, the separation (or combination) of the power supply and the decoupling of the power supply lines feeding each of the individual stages. If the same physical characteristics of the test fixture that the module was measured in are implemented in the actual circuit being developed, then the simulations based on these *s*-parameters

will not be accurate. If the grounding or power supply decoupling, for example, is not as good as that in the test fixture then oscillations due to feedback could occur (and often do).

31.4.2 Design Tool Requirements

There are certain features that a microwave CAD tool must have in order to improve the productivity of the microwave designer (see Table 31.1). The most important is that it must be accurate. Once a circuit has been simulated, it must be fabricated and tested against the specifications for that circuit. If the computer simulation is inaccurate, the reasons why must be determined. The fabrication of the circuit and the debugging of the circuit are, by far, the most time consuming part of the design cycle. Therefore, the goal of any CAD design is to model the circuit so that only one fabrication cycle needs to be implemented.

Inaccurate simulations that are tool dependent can result from numerical difficulties, model inaccuracies, or a lack of acceptable models. Inaccurate simulations that are directly caused by the user usually result from failure to model the circuit correctly or the misuse of models by using them outside the range for which they are valid. The designer has little control over the numerical problems. In some cases, the designer can overcome model accuracy by knowing the limitations of the models being used and, if possible, compensating for them. However, since model accuracy and availability usually are the features that distinguish one program from the other, companies tend to treat the models as proprietary and, therefore, do not want to release information about the models and how they are implemented. This proprietary posturing can be a great disadvantage since the designer has no way of knowing how the model is implemented and when the range of validity for the model has been exceeded.

Programs usually provide a way to implement custom models. For example, many programs will allow the user to import *s*-parameter data files to represent a part of the circuit that cannot be accurately represented by standard models. These *s*-parameter files can be generated, for example, from a measurement of the actual discontinuity. Likewise, active device *s*-parameters can be measured at a specific bias point and imported for use in the design process. Measured data is usually limited to two port networks since multiport network analyzers are not, typically, available to the designer. For multiport passive networks, electromagnetic simulators are often used to create data file representations of the passive network. In some cases, programs will allow the designer access to the code or hooks into the code that can be used to implement custom models.

Ease of use is a program feature that can greatly increase the productivity of a designer. Most modern programs use a graphical interface to create a schematic representation of the circuit. For electromagnetic simulators, a drawing package is used to draw the geometric shapes. These drawing packages, at a minimum, must provide a means to implement complex, difficult to draw structures from basic shapes which are easy for a designer to draw. In addition to this, the ability to change dimensions without the need to redraw the shape can greatly decrease the time it takes to perform tolerance studies or to modify the structure in order to meet the desired performance criteria.

Because microwave circuit performance specifications can be time or frequency domain quantities, the circuit analysis program must provide an easy means to display the resulting data in both the time and frequency domain. The program must also be able to export the results into standard text or graphic formats for importation into word processors or view cell generating programs for documentation and presentation purposes.

TABLE 31.1 Summary of Desirable CAD Features

Analysis Features	Synthesis Features	User Interface Features
Accurate models	Filter/impedance matching	Schematic input
Optimization	constant noise circles	Text and graphical output
Import/export data	Constant gain circles	Easy documentation
Robust model library	Stability circles	Drawing complex shapes

Electromagnetic simulators need to be able to display field quantities such as current density on the conductors, E and H field intensity and direction, as well as outputting S, Z, or Y parameters for importation into circuit simulators. Data must be displayed in either graphical or text format.

Synthesis programs are available that provide impedance matching network topologies for amplifier/mixer design and filter response functions. These programs could be simple spreadsheet implementations of well-known design equations, or an electronic smith chart, or sophisticated implementations of impedance matching theory or filter design. The most sophisticated programs can provide networks based on both ideal and nonideal elements. The ability to display constant gain, noise circles, and stability circles can also be considered part of the synthesis capability since these are often used to determine the matching network impedances [1].

31.5 Time Domain versus Frequency Domain Simulation

A circuit can be analyzed in either the frequency domain or the time domain. SPICE is a program developed at UC Berkeley that analyzes a circuit in the time domain. Harmonic balance programs solve the circuit equations partially in the frequency domain and partially in the time domain. The choice of which program is used depends on the parameters that are specified. Since SPICE does the computations in the time domain, it is, quite naturally, used when the design parameters involve time-dependant quantities. SPICE is used for transient parameters such as the turn on time of an oscillator or amplifier, the switching time of a switch, or perhaps the impulse response of a circuit. SPICE can also solve for the DC bias point of the circuit and then perform a small signal, frequency domain analysis of the circuit about this bias point. By adding some special circuit elements to the file description, s-parameters can be computed from the results of this small signal analysis. However, many SPICE-based programs have added direct s-parameter output capability, in order, to accommodate the needs of microwave designers.

SPICE can be used to predict the effects of noise and distortion within the circuit. Using small signal, frequency domain analysis, the linear noise parameters of a circuit can be predicted. However, the noise due to mixing effects of nonlinearities within the circuit can not be predicted. Likewise the up or down conversion noise and phase noise of an oscillator are also not analyzed. Distortion analysis using SPICE is usually performed through the use of a transient analysis followed by conversion of a part of the time waveform into the frequency domain using a discrete Fourier transform (DFT). The microwave designer is typically interested in two types of distortions—the harmonic content of the time waveform and the intermodulation products caused by the excitation of the circuit by two signals typically close to each other in frequency.

Harmonic balance is used when the circuit is driven by periodic sources and when the input and output design parameters are specified in the frequency domain. The assumed periodicity of the circuit response avoids the need to compute the circuit response from time zero until the steady state response is obtained. Therefore, much less computer time is required to predict the circuit response. Since, the harmonic balance techniques were developed specifically to aid the microwave designer in the design of nonlinear circuits, the available programs are custom tailored to provide the results in a format familiar to the designer. For example, the input source for an amplifier can be swept in both frequency and power, nonlinear parameters such as the gain, 1 dB compression point, saturated power output, power added efficiency, and harmonic levels can all be displayed in a graphical or text form. Indeed these parameters are all natural artifacts of the computations. Other parameters which can easily be computed are intermodulation products, third order intercept point, noise side bands, and mixer conversion.

The harmonic balance method is a hybrid of the small signal, frequency domain analysis, and a nonlinear time domain analysis. The circuit is divided into two subcircuits. One subcircuit contains only those circuit elements which can be modeled in the small signal frequency domain. This subcircuit results in a Y-matrix representation that relates the frequency domain currents to the frequency domain voltages. At this point, the matrix can be reduced in size by eliminating voltages and currents for the nodes which are not connected to the nonlinear devices or the input and output nodes. These matrix computations need

only be performed once for each frequency and harmonic frequency of interest. The second subcircuit includes all of the active or nonlinear elements which are modeled in the time domain. These circuit models relate the instantaneous branch currents to the instantaneous voltages across the device nodes. These models are the same models used in SPICE programs.

The two subcircuits make up two systems of equations having equal node voltages whose branch currents must obey Kirchhoff's current law. The system of equations corresponding to the linear subcircuit is now solved by making an initial guess at the frequency spectrum of the node voltages. The node voltage frequency spectrum is, then, used to solve for the frequency spectrum of the branch currents of the linear subcircuit. In addition to this, the node voltage frequency spectrum is transformed into the time domain using a FFT algorithm. The result of this operation is a sampling of the periodic time voltage waveform. The sample voltages are applied to the time domain subcircuit resulting in time domain current waveforms which are, then, transformed into the frequency domain again using a FFT algorithm. The two frequency domain current spectrums are compared and, based on the error between them, the voltage spectrums are updated. This process is repeated until the error is sufficiently small.

Early implementations of harmonic balance programs used either an optimization routine to solve for the node voltage spectrums [2] or Newton's method [3]. The main advantage of Newton's method is that it uses the derivatives of the nonlinear device currents with respect to the node voltages to predict the next increment in the node voltages. By taking advantage of these derivatives, convergence can be achieved for a relatively large number of nonlinear devices. This method appears to work well as long as the nonlinearity of the system is not too severe.

31.6 Emerging Simulation Developments

Current research directed toward improving the implementations of harmonic balance programs is being directed toward techniques that can handle the large number of nonlinear devices typically found in integrated circuits. Currently, Krylov-subspace solutions have been implemented [4]. When Krylov-subspace techniques are used, the harmonic balance method can be used to solve circuit problems containing hundreds of transistors.

When the excitation of a circuit consists of multiple sinusoids, closed spaced in frequency, both SPICE and conventional harmonic balance methods tax computer hardware resources as they require large amounts of memory and computer time. A program must be able to efficiently handle this type of excitation in order to be able to predict the effects of spectral regrowth in digitally modulated circuits, as well as noise–power ratio simulations for these circuits. For these types of circuit analysis, the excitation consists of multiple sinusoids, closely spaced in frequency. Borich [5] has proposed a means to overcome these problems for harmonic balance programs by adjusting the sampling rate and the spacing between excitation carriers in order to reduce the computations of the multitone distorted spectra to an efficient one-dimensional FFT operation.

Envelope-following method [6] has been implemented to solve for circuits in which the excitation consists of a high frequency carrier modulated by a much slower information signal. The method performs a transient analysis consistent with the time scales of the information signal. At each time step, a harmonic balance analysis is performed at the harmonic frequencies of the carrier. This method can be used to study PPL phase noise, oscillator turn-on time, and mixer spectral regrowth due to digital modulation on the RF carrier [7].

References

1. D. George. Vendelin, *Design of Amplifiers and Oscillators by the S-Parameter Method*. John Wiley & Sons, New York, 1982
2. M. Nakhla, and J. Vlach. A piecewise harmonic balance technique for determinatoin of periodic response of nonlinear systems, *IEEE Transactions on Circuits and Systems*, Vol. CAS-23, No. 2, Feb. 1976.

3. K. Kundert, A. Sangiovanni-Vincentelli. Simulation of nonlinear circuits in the frequency domain, *IEEE Transactions on Computer-Aided Design*, Vol. CAD-5, No. 4, Oct. 1986, pp. 521–535.
4. R. Telichevesky, K. Kundert, and I. Elfadel, J. White. Fundamentals of Fast simulation algorithms for RF circuits, *Proceedings of the IEEE*, Vol. 95, No. 3, March 2007, pp. 600–621.
5. V. Borich, J. East, and G. Haddad. An efficient fourier transform algorithm for multitone harmonic balance, *IEEE Transactions on Microwave Theory and Techniques*, Vol. 47, No. 2, Feb. 1999, pp. 182–188.
6. P. Feldmann and J. Roychowdhury. Computation of circuit waveform envelopes using an efficient, matrix-decomposed harmonic balance algorithm. *ICCAD-1996 Digest of Technical Papers*, 10–14 Nov.1996, pp. 295–300.
7. K. Mayaram, D.C. Lee, S. Moinian, D. Rich, and J. Roychowdhury. Overview of computer-aided analysis tools for RFIC simulation: Algorithms, features, and limitations, *IEEE Custom Integrated Circuits Conference 1997*, 5–8 May 1997, pp. 505–512.

32

Nonlinear Transistor Modeling for Circuit Simulation

Walter R. Curtice
W. R. Curtice Consulting

32.1 Introduction

By definition, a transistor model is a simplified representation of the physical entity, constructed to enable analysis to be made in a relatively simple manner. It follows that models, although useful, may be wrong or inaccurate for some application. Designers must learn the useful range of application for each model.

It is interesting to note that all transistors are fundamentally nonlinear. That is, under any bias condition, one can always measure harmonic output power or intermodulation products at any input radio frequency (RF) power level, as long as the power is above the noise threshold of the measurement equipment. In that sense, the nonlinear model is more physical than the linear model.

TABLE 32.1 Types of Large-Signal Transistor Models

I.	Physical or "Physics-Based" device Models
II.	Measurement-Based Models
	1. Anaytical Models, such as SPICE Models
	2. Black Box Models
	Table-Based Models
	Artificial Neural Network (ANN) Models

The purpose of this work is to give a tutorial presentation of nonlinear transistor modeling. After reviewing the types of models, we will concentrate on equivalent-circuit models, of the type used in SPICE [1]. Recent improvements in models will be described and the modeling of temperature effects and the effects of traps will be discussed. Finally, parameter extraction and model verification is described.

32.1.1 Two-Dimensional Models

The models constructed for describing the nonlinear behavior of transistors fall into several distinctly different categories, as depicted in Table 32.1. The most complex is the "physics-based" model. Here electron and hole transport are described by fundamental transport and current continuity relationships and the physical geometry may be described in one-, two-, or even three-dimensional space. Electric field is found, by solution of Poisson's equation, consistent with the distribution of charge and boundary conditions. Such a model may use macro-physics, such as drift–diffusion equations [2–4], or more detailed descriptions, such as a particle-mesh model with scattering implemented using Monte-Carlo methods [5,6]. Two- and three-dimensional models must be used if geometrical effects are to be included. The matter of how much detail to put be into the model is often decided by the time it takes for the available computer to run a useful simulation using this model. In fact, as computers have increased their speed, modelers have increased the complexity of the model simulated.

The lengthy execution time required for Monte-Carlo analysis can be reduced by using electron temperature [7] as a measure of electron energy. Electron temperature is determined by the standard deviation of the energy distribution function and is well defined in the case of the displaced Maxwellian distribution function. Electron and hole transport coefficients are developed as a function of electron temperature, and nonequilibrium effects, such as velocity overshoot in GaAs, may be simulated in a more efficient manner. However, the solution of Poisson's equation still requires appreciable computational time.

BLAZE [8] is a good example of a commercial, physics-based device simulator that uses electron temperature models. BLAZE is efficient enough to model interaction of a device with simple circuits.

The physics-based model would be constructed with all known parameters and simulations of current control for DC and transient or RF operation then compared with measured data. Using the data, some transport coefficients or physical parameters would be fine-tuned for best agreement between the model and the data. This is the process of calibration of the model [9]. After calibration, simulations can be trusted to be of good accuracy as long as the model is not asked to produce effects that are not part of its construction. That is, if trapping effects [10] have not been incorporated into the model, the model will disagree with data when such effects are important. With the physics-based model, as with all others, a range of validity must be established.

Present physics-based models still require too much computational time to be used, to any extent, in circuit design work. Optimization of a circuit design will involve invoking the device model so frequently so as to be impractical with physics-based models. These models can be used if only one or two nonlinear transistors are used in a specific circuit, but typically, the circuit designer has a larger number of nonlinear devices.

Several quasi two-dimensional models [11,12] have been developed, which execute more efficiently. Initial results look good, but accuracy may depend upon the simplifications made in the development of the code and will vary with the application.

32.1.2 Measurement-Based Models

The next general category is that of "measurement-based" models. These are empirical models either constructed using analytical equations and called "analytical models" or else are based upon a look-up table developed from the measured data. The latter are called "table-based" models. Multi-dimensional spline functions are used to fit the data in some of these models [13] and only the coefficients need be stored.

In the case of analytical models, the coefficients of the equations serve as fitting parameters to permit the equations to approximate the measured data. Functions are usually chosen that have functional behavior similar to measured data so that the number of fitting parameters is reduced.

The advantages of analytical modes are: computational efficiency, automatic data smoothing, accommodation of device statistics, physical insight, and the ability to be modified in a systematic manner. Disadvantages are: restriction of behavior often due to use of over-simplified expressions, difficulty in parameter extraction and guaranteed nonphysical behavior in some operating condition. The nonphysical behavior is often associated with the use of a function, such as a polynomial, to fit data over a specific range of voltages and subsequent application of the model to voltages outside of this range. The function may not behave well outside of the fitting range. The best example of analytical models is the set of transistor models used in the various forms of the SPICE program. A major advantage of analytical models is that all the microwave nonlinear simulators provide some sort of user-defined model interface for analytical model insertion.

Table-based models have some properties of black-box models. The equations used result from fitting to the data, using splines or other such functions. These models can, therefore, "learn" the behavior of the nonlinear device and are ideal for applications where the functional form of the behavior is unknown. Table-based models are efficient but do not provide the user with any insight, since there is a minimal "circuit model." They have difficulty incorporating dispersive effects, such as "parasitic gating" due to traps and do not accommodate self-heating effects.

The model cannot be accurately extrapolated into regions where data was not taken, and the models are often limited in their application due to the particular coding used by their author. This means that users other than the author cannot tailor the model. Customization of models is important to improve the "performance" of a model. The first table-based model that has been widely used is the Root FET model [13].

32.1.3 Physical Parameter Models

One may argue that there is a class of models between physics-based and analytic models, namely, physical parameter models. A good example would be the Gummel–Poon model [14]. Here, analytical equations are used but the fitting parameters or equation coefficients have physical significance. For example, NF, the ideality factor of the emitter-base junction is one model parameter. This model is an analytical model but more useful device information may be gleamed from the value of the coefficients. This is often the case and the model is widely used for various forms of bipolar devices.

A useful physical parameter model for the AlGaAs/InGaAs/GaAs pHEMT has been published and verified by Daniel and Tayrani [15]. No information has been given on the range of validity of the analytical model, and it is expected that such a simple model will have inaccuracy when two-dimensional effects or non-equilibrium effects are important. It is interesting to note that the HEMT structure has less two-dimensional effects than the MESFET because of the sheet current layer produced in the HEMT.

There are many analytical models for which the coefficient has the name and dimensions of a physical parameter but the coefficient is only a fitting parameter and is not strongly related to the physical parameter.

Khatibzadeh and Trew [16] have presented one commonly called the Trew model and Ladbrooke [17] has presented a second model commonly called the Ladbrooke model. Ladbrooke's model is an extension of, the much earlier, Lehovec and Zuleeg [18] model and it is more empirical than physical, as Bandler [19] has shown. Such models must be tested to see how strong the relationship is between the model parameter and the physical parameter. It is a matter of the degree of correlation between the two quantities. It is dangerous to attach too much physical significance to these coefficients. One should verify the relationship before doing so.

Then, there is a unique case where physical parameters have been installed in a previously developed analytical model. The Statz–Pucel (analytical) GaAs MESFET model [20] has, then, been converted to a physical parameter model by D'Agostino et al. [21].

32.1.4 Neural Network Modeling

Rather recently, a new approach has been developed for the modeling of nonlinear devices and networks. It utilizes artificial neural networks, or ANNs. ANN models are similar to table-based or black-box models in that there is no assumption of particular analytical functions. As with table-based models, the ANN model "learns" the relationship between current and voltage from the data and model currents are efficiently calculated after application of voltages. ANN analysis can treat linear or nonlinear operation of devices or complex circuits. ANN models have many of the advantages and deficiencies of table-based models. An excellent special issue of *The International Journal of RF and Microwave Computer-Aided Engineering* [22] has been devoted to this modeling method. This approach is really a form a behavior modeling. Unfortunately, discussion of this approach is beyond the scope of this work.

32.2 Scope of this Work

The purpose of this work is to present a tutorial on the modeling of the nonlinear behavior of transistors. The scope is limited to nonlinear models useful for the development of circuit designs. It would not be possible to cover all the important material on other types of models, such as physical models, within this context. This chapter will deal primarily with nonlinear analytical models for MESFETs , (*p*)HEMTs, and HBTs. The emphasis is on GaAs device models, although many of these models are also used for transistors fabricated in GaN, InP, silicon, and other materials. The RF LDMOS power device is, also discussed, because it can be treated as a three-terminal device, much like a MESFET.

We address the concerns of analog and digital circuit designers who must choose between a wide variety of nonlinear models for transistors. Of particular concern here is the MMIC designer who must select the proper model to use for his GaAs microwave transistor. Even with very complex models, presently supplied in circuit simulators, some specific behaviors are not modeled and new model features are required. We will spend much time on SPICE models and SPICE-type models. We will inspect some of the recent models that incorporate important device effects, previously omitted, in models. We will discuss the modeling of gate charge as a function of local and remote voltages, the modeling of self-heating effects, the modeling of trapping effects, and model verification.

Unfortunately, the references presented will only be representative of the prior work because there is a wealth of papers in each area of transistor modeling. I apologize to any author not included, as there are, now and have been, many people working in this area. I do recommend some modeling tutorial articles, previously published. Trew [23] and Snowden [24], and Dortu et al. [25] have presented excellent reviews of SPICE-type transistor models and are recommended for reading.

32.3 Equivalent-Circuit or Compact Models

We concern ourselves here with equivalent-circuit models because they are formulated to be efficiently exercised in a circuit simulator and, thus, are efficient for circuit design and optimization. This is because

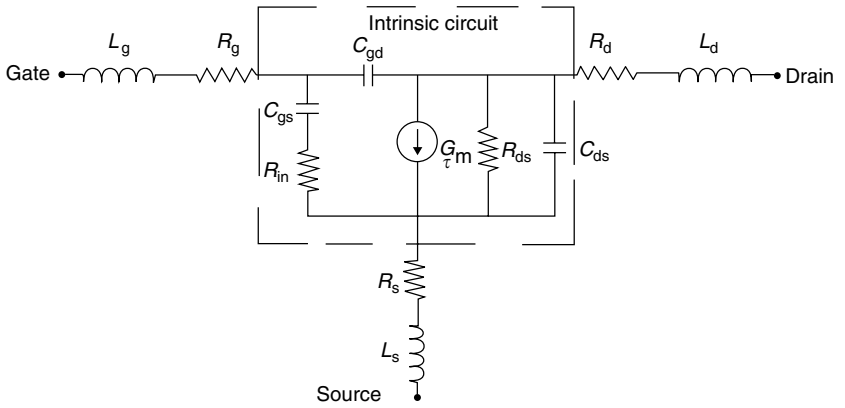


FIGURE 32.1 The conventional small-signal model for a MESFET, showing intrinsic and extrinsic elements.

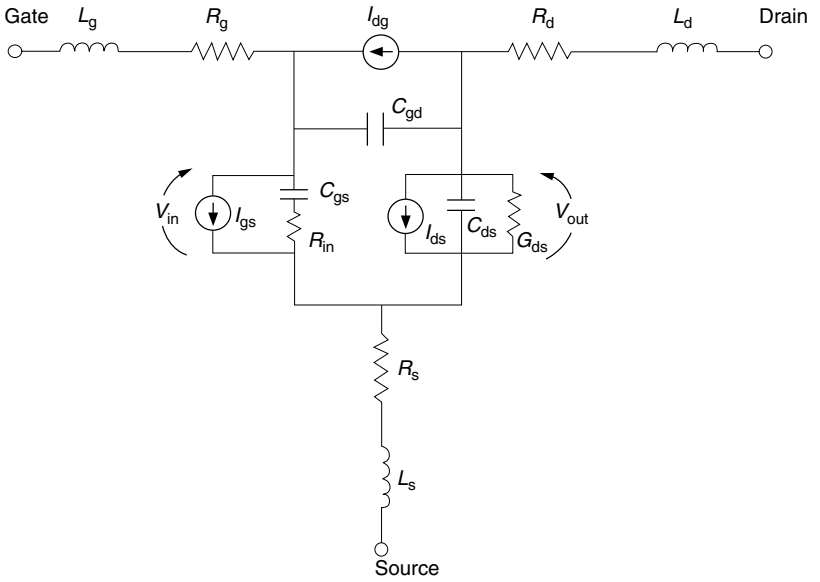


FIGURE 32.2 The conventional large-signal model for MESFET.

the simulator is accustomed to dealing with resistors, capacitors, inductors, and voltage or current-controlled sources. Such models are often referred to as "compact models."

One problem is that *de facto* standard models have evolved in the industry and, often, these models are inadequate to describe the device behavior. This is even true for small-signal equivalent-circuit models. Still, all circuit simulators utilize standard model topologies for small- and large-signal MESFET and *p*HEMT models. These models represent a minimum number of elements and are efficient for evaluation of transistor characteristics. Figures 32.1 and 32.2 show the conventional topologies for small- and large-signal simulation, respectively. It is conventional to separate the extrinsic parameters from the intrinsic device parameters, as shown in Figure 32.1. The intrinsic parameters are assumed to contain all the bias-dependent behavior and the extrinsic parameters are assumed to be of constant values.

Curtice and Camisa [26] and Vaitkus [27] have discussed the trade-offs that exist between simple models with a small number of parameters and complex models, with a large number of parameters. A primary concern is the significant increase in the uncertainty for each model parameter in a complex model. Vaitkus showed that a small increase in the number of elements in a small-signal model could

easily increase the uncertainty of critical elements beyond the standard deviation of the element value resulting from the fabrication process.

Byun et al. [28] and others assert that source resistance should be taken as bias-dependent. This decision is actually a choice up to the modeler. If source and drain resistance are taken as constant, then, the reference planes defining these resistances are taken as being close to the metal ohmic contacts and not too close to the Schottky contact. That is, no region that may become depleted of change is included. All bias-dependent behavior is then lumped into the intrinsic elements. This is the convention followed by most modelers. It results in a simpler model, with fewer parameters.

Unfortunately, many of the published nonlinear device models have inconsistencies with the conventional small-signal model. Some of the inconsistencies may go unnoticed but can cause design errors. A good example, that will be described later, is the modeling of capacitance as a function of two independent voltages. The small-signal model must contain "transcapacitance" elements to be consistent with the large-signal model.

The large-signal model should agree with the small-signal model but is not expected to be as efficient. Due to transconductance, some resistances and capacitances must be evaluated from functions of voltages, and therefore numerical evaluation will take more computational time. However, the behavior of the large-signal model will gracefully go from small- to large-signal in a good model.

32.4 SPICE Models and Application-Specific Models

SPICE was developed to help in the design of switching circuits. Thus, the SPICE transistor models were developed to model the time-domain behavior of devices in such circuits. However, transistors operating in RF analog circuits have a different locus of operation. For example, a simple class A amplifier will have locus of operation around its quiescent bias point. This should be compared with a logic circuit where the transistors go from biased-off condition (high voltage, low current) to strongly turned on condition (low voltage, high current). If the same transistor were used in these two applications, one would expect the SPICE model would approximate both behaviors, but not be optimum for either. In fact, if the model is fine-tuned to be more accurate for one of these applications, it will by default, be less accurate for the other.

For these reasons, accurate nonlinear transistor models will be application-specific. In order to make the model more general, the models can be made more complex and more model parameters will be added. This may result in poorer execution efficiency.

Clearly one goal of transistor modeling should be to keep the model simple and to keep the number of model parameters small so that the extraction of these parameters is more efficient. Designers prefer simple models for initial work and are willing to work with more complicated models for difficult design specifications. The SPICE transistor models serve the function of the initial, simple models. These models are also universally known by name and have history and familiarity associated with them. A designer attempting to use a new GaAs foundry would not be intimidated by obscure nonlinear models if SPICE transistor models were used in that foundry.

Many modelers have attempted to extend or enhance the SPICE models so that the accuracy is improved, particularly in microwave analog applications. Usually, the default model is the original SPICE model and the designer will feel comfortable with this approach. There are many examples. The SPICE Gummel–Poon model has been extended, by Samelis and Pavlidis [29] and others [30], for application to heterojunction bipolar devices. The JFET SPICE model was extended by Curtice [31] in 1980 for better application to GaAs MESFET logic circuits.

Because of the dominant use of harmonic-balance (HB) simulators for microwave applications and the increasing use of envelope-type simulations, many new equivalent-circuit models have been developed specifically for these simulators. Never the less, these models are "SPICE-type" models, and can also be executed in a time-domain simulation. The requirements of a model for SPICE are the same as for HB since the device is operated in the time domain in both simulators.

TABLE 32.2 SPICE Models

GaAs MESFET/HEMT	MOS Models	BJT Models	Diode Models
Curtice (Cubic & Quadratic)	BSIM 1, 2, 3, and 4	BJT (Gummel Poon)	P/N Diode
STATZ (Raytheon)	UC Berkeley 2 &3	MEXTRAM	PIN Diode
JFET (N & P)	PSP	VBIC	
TOM(TriQuint's Own Model)	MOSFET(various levels)	HICUM	
Materka			
Angelov			

The producers of commercial HB software recognized the need of users to customize their transistors models. All commercial packages now contain user-defined modeling interfaces that permit the installation of customized models into the transistor model library. The process of installing or customizing a model in SPICE is much more difficult and not available to the average user. However, the ease of installation of models into HB software has produced a wrath of new models for many transistor types.

Table 32.2 shows the typical array of SPICE equivalent-circuit models available as part of a commercial simulator software package. The models are categorized, in general, as diode models, GaAs MESFET or (*p*)HEMT models, MOS models, and bipolar device models. The list is not complete for any specific product but representative of the models available. The models listed in Table 32.2 are mostly, all commercial simulator products be it a version of SPICE or a HB simulator.

32.5 Improved Transistor Models for Circuit Simulation

Early SPICE models have shown a number of deficiencies. One problem is that the models developed before 1980 were developed for silicon devices, and they do not reflect the behavior of GaAs devices. Most SPICE models need to be customized to be accurate enough for present design requirements. With regard to GaAs MESFET and *p*HEMT modeling, the strong dependency of gate–source capacitance upon drain–source voltage as well as gate–source voltage, is not modeled in the early SPICE models. None of the standard SPICE models accommodate self-heating effects. These effects are more important in GaAs applications due to the poorer thermal conductivity of GaAs as compared to silicon. Some GaAs transistor exhibit important dispersion effects in transconductance as well as in drain admittance. All the GaAs models in Table 32.2 were added during the 1980s. These models represented major improvements; however, there remained deficiencies. These deficiencies are summarized below:

SPICE Large-Signal Model Deficiencies

- Insufficient accuracy for GaAs applications
- Poor modeling of nonlinear capacitance
- Poor modeling of self-heating effects
- No modeling of dispersion of transconductance
- No modeling of gds dispersion
- Model parameter extraction not defined
- Poor modeling of nonlinear effects dependent upon higher order derivatives
- No modeling of sub-threshold current
- No modeling of breakdown

The improved large-signal models of the 1990s exhibit some common features. It is quite popular to utilize analytical functions that have infinite number of derivatives. For example, in the modeling of *p*HEMTs , Angelov et al. [32] has relied heavily on the hyperbolic tangent function for current because its derivative with respect to gate–source voltage is a bell-shaped curve, much like transconductance in *p*HEMTs. All further derivatives also exit. The Parker [33] model also utilizes higher-order continuity in the drain current description and its derivatives.

Some SPICE models do have continuous derivatives but may not be accurate. The differences between the COBRA [34] model and the previous Materka [35] model are more evident when the derivatives of current (first through third) are compared. Since the COBRA model has derivatives closer to the data, Cojocaru and Brazil show that the model predicts intermodulation products more accurately.

Many SPICE models use a simple expression for junction capacitance in MESFETs and *p*HEMTs. However, capacitance values extracted from data show that the gate–source and gate–drain capacitance depends strongly on the remote voltage as well as the local voltage (the capacitance terminal voltage). The Statz (Raytheon), the TOM, and the EEFET3 SPICE models [36] all have detailed equations for the gate, drain, and source, charge as a function of local and remote voltages. Extraction of coefficients for these expressions is not simple and this will be discussed in a later section.

Often, a specific simulator's model gives reasonable agreement but lacks some capability required by the designer. For example, the SPICE Gummel–Poon model works well for many devices but lack self-heating features desirable for power bipolar devices. It is possible for a feature like this to be added to such models in a simple manner by using the model form in Verilog-A code [37].

All major commercial simulators now support transistor models in Verilog-A code. Since source Verilog-A code for most common transistor models is widely available, it is not a difficult procedure to add a small number of new features to an existing useful model. The code is, then, compiled and dynamically linked to the simulator. Execution time should be comparable to that of installed models.

A primary reason for the simplicity of the Verilog-A code is that the code compiler evaluates all derivative expressions needed. This is done with the same accuracy as for explicit derivative expressions. Thus, the source code for the model is much shorter and simpler.

32.6 Modeling Gate Charge as a Function of Local and Remote Voltages in MESFETs and HEMTs

Small-signal modeling of GaAs and InP MESFETs and HEMTs show that both C_{gd}, the gate–drain capacitance, and C_{gs}, the gate–source capacitance vary with change of V_{gs}, the gate–source voltage, and V_{ds}, the drain–source voltage. Thus, these capacitances are dependent upon the local, or terminal voltage, and a remote voltage. The dependency upon the local voltage is expected for capacitances, but the dependency upon the remote voltage leads to a term called "transcapacitance." The modeling of these capacitances can lead to nonphysical effects, if not handled properly, as Calvo et al. [38] have shown.

Simulators must work with charge functions whose derivatives are the capacitive terms. The conventional approach is to find some charge function for total gate charge, such as $Q_g\ (V_{gs},\ V_{ds})$. Then,

$$C_{11} = \text{Partial derivation of } Q_g \text{ with respect to } V_{gs}$$
$$C_{12} = \text{Partial derivation of } Q_g \text{ with respect to } V_{ds}$$

For consistency with small-signal *p*HEMT models

$$C_{11} = C_{gs} + C_{gd}$$
$$C_{12} = -C_{gd}$$

The problem remaining is to partition the total gate charge, Q_g, into charge associated with the gate–source region, Q_{gs}, and charge associated with the gate–drain region, Q_{gd}. Then, for large-signal modeling, the gate node has charge, Q_g, the source node has charge, $-Q_{gs}$, and the drain node has charge, $-Q_{gd}$, and

$$Q_g = Q_{gs} + Q_{gd}$$

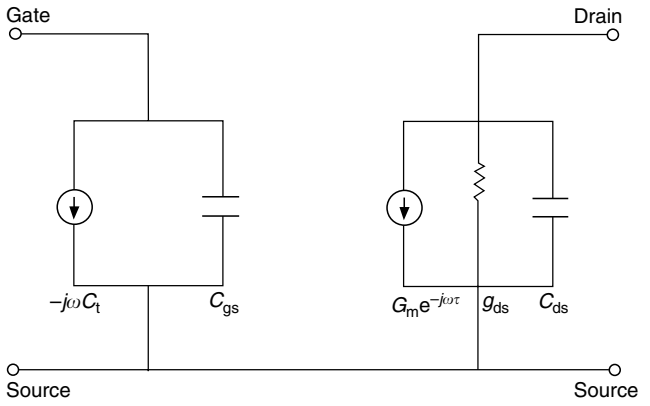


FIGURE 32.3 Jansen's topology for the intrinsic circuit.

One scheme for partitioning the charge is used in the EEFET model and described in the Advaced Design System Manual [36]. One advantage of this approach is that the model becomes symmetrical meaning that drain and source may be interchanged and the expressions are still valid.

A simpler approach is presented by Jansen et al. [39]. Jansen assumes that all of the gate charge is associated with the gate–source region and:

$$Q_{gd} = 0$$
$$Q_g = Q_{gs}$$

The topology for this approach is different than the conventional model. There is no drain–gate capacitance element. Instead, the transcapacitance term accounts for the conventional drain to gate capacitive effects. That is, a change in V_{ds} produces current in the gate–source region through the transcapacitiance term. Figure 32.3 shows the new topology for the intrinsic circuit.

The procedure for modeling the capacitive effects is the same for both approaches and is as follows:

1. Measure small-signal values of C_{gs} and C_{gd} as a function of V_{gs} and V_{ds}
2. Choose a Q_g function and optimize the coefficients of the Q_g expression for best fit of

$$C_{11} = C_{gs} + C_{gd}$$
$$C_{12} = -C_{gd}$$

A good example of the fitting functions and the type of fit obtained is given by Mallavarpu et al. [40] for a *p*HEMT.

In summary, large-signal capacitive effects are modeled by constructing a total gate charge function, $Q_g(V_{gs}, V_{ds})$, whose partial derivative approximates the measured capacitance functions. If device symmetry is important, the EEFET charge partitioning scheme may be used. For amplifier applications, the Jansen model is simplest to code and implement because there is no charge partitioning expression. However, the Jansen model uses a topology that is not conventional.

Jansen [39] has also discussed the conditions on the charge functions to maintain charge conservation. These conditions also assure that integration of capacitances by any path give the same charge value. In the simplest case of just a gate–source charge, $Q_g(V_{gs}, V_{ds})$, the charge conservation conditions are the

following:

$$C_{u1} \equiv \frac{\partial Q_g}{\partial V_{gs}}$$

$$C_{u2} \equiv \frac{\partial Qg}{\partial Vds}$$

$$\frac{\partial C_{u2}}{\partial V_{gs}} = \frac{\partial C_{u1}}{\partial V_{gs}}$$

If capacitance functions are formulated rather than charge functions, it is very difficult to be charge conservative. The best approach is to formulate charge functions with model parameters and to optimize their values for best fit to the capacitance data.

32.7 Modeling the Effects due to Traps

Electron and hole traps exist in GaAs materials and cause numerous effects during operation of a GaAs MESFET or HEMT transistor. The following is a brief listing of these effects:

Effects of Traps

- Dispersion in transconductance and output admittance
- Backgating
- Parasitic bipolar effects
- "Kinks" in the I–V relationship
- Surface gating
- Gate and drain lag effects during switching
- Light sensitivity
- Substrate current or lack of current pinch-off

These effects have been studied and circuit-level models developed to simulate the effects. In many cases, the details of the behavior have been made clear using two-dimensional simulation modeling. For example, Li and Dutton [10] used PISCES-IIB to show that the common EL2 trap causes dispersion in the output conductance of a GaAs MESFET up to several hundred Hz.

Golio [56] has a good discussion of the effects of frequency dispersion. The circuit-level modeling of dispersion is described by Cojocaru and Brazil [34]. They extend the previous conventional modeling of dispersion of the output conductance to include dispersion of the transconductance. The circuit is very simple. A second voltage-control current source in parallel with a resistance with capacitive coupling to the internal drain–source terminals. This enables the model's transconductance and drain–source conductance to be tailored for high frequencies using these new elements.

Figure 32.4 shows a "kink" in the family of I–V characteristics of a GaAs MESFET made by a commercial foundry. The kink occurs as a sudden change of slope at V_{ds}=5.5 V for gate–source voltage of –1 V. Notice that a large increase in gate current is present in the same region and that the gate is collecting holes.

This behavior is explained by Horio and Usarni [41] who use two-dimensional simulations to show that a small amount of avalanche breakdown in the presence of traps causes excess hole charge in the substrate that produces "kinks" in the low-frequency I–V data. Since the traps cannot be easily eliminated, the kinks may be removed by removing the conditions initiating avalanche breakdown.

Upon switching the gate voltage, the drain current of a MESFET will have lag effects in the microsecond and millisecond regions, which are produced by traps. Curtice et al. [42] have shown the circuit-level modeling of such gate lag effects as well as drain lag effects. Others, such as Kunihiro and Ohno [43], have also presented circuits for the modeling of drain lag effects.

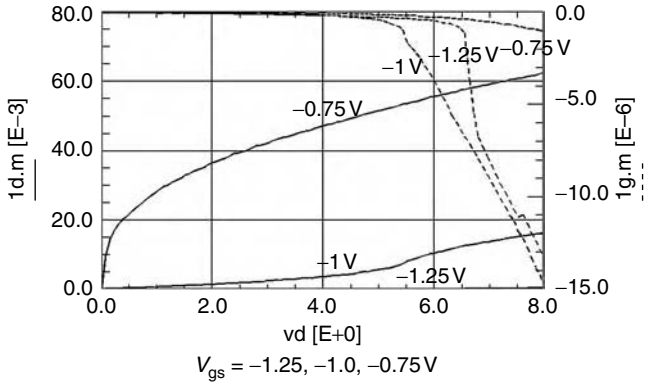


FIGURE 32.4 Current "kink" due to avalanche and traps in a GaAs MESFET.

The transistor model with such circuits may, then, be used to determine if the lag effects interfere with proper operation of the circuit. The work of Curtice et al. was directed toward GaAs digital circuits where the switching waveform must be of high quality. Using the new transistor model, one may determine not only if the circuit will perform, but also circuit changes that will permit operation in the presence of strong lag effects.

Light sensitivity has been described and modeled by Chakrubarti et al. [44] and by Madjar et al. [45]. Some circuit-level models are presented in their discussions.

In the microwave application arena, the principle difficulty with traps is that they cause difficulty in determining an accurate microwave model for the transistor. An excellent experimental study of surface gating effects is given by Teyssier et al. [46]. Surface gating means that the charge stored in surface states and traps influences the I–V behavior by acting as a second gate. The amount of charge stored in the traps will vary, depending upon the applied voltages, the ambient light, and the temperature and trapping time constants. Teyssier et al. show that measured trap capture time constants are quite different than trap-emission time constants. They show how they are able to accurately characterize the I–V behavior for RF operation by using 150 ns bias pulse width. They also describe how they characterize the thermal behavior using longer pulse width.

Many others have published data showing surface gating effects. See, for example, Platzker et al. [47]. The approach a modeler should use is to first determine how important such trapping effects are in the transistor operation. In many transistor designs, the active region is shielded sufficiently from surface charge so that negligible surface gating occurs. In that case, low-frequency I–V data and the resulting transconductance may be very predictive of microwave-frequency behavior. If the trapping effects are found to be of importance, then short-pulsed characterization is required and low frequency I–V data will not suffice.

In any case, device characterization must include the behavior changes due to self-heating and ambient temperature effects. The modeling of heating effects will be discussed next.

32.8 Modeling Temperature Effects and Self-Heating

Anholt and Swirhun [48] and others have documented the changes for GaAs MESFETs and HEMTs at elevated temperatures. However, the modeling of a device over a temperature range is often attempted using temperature coefficients. With regard to drain current and DC transconductance, the effects of elevated temperature are quite different at large channel current as compared to operation near pinch-off. At large channel current, the electron mobility decrease with temperature increase is most important, whereas, at low current, the decrease in the pinch-off voltage with temperature is most important. This produces the interesting effect that transconductance decreases with temperature at large current but

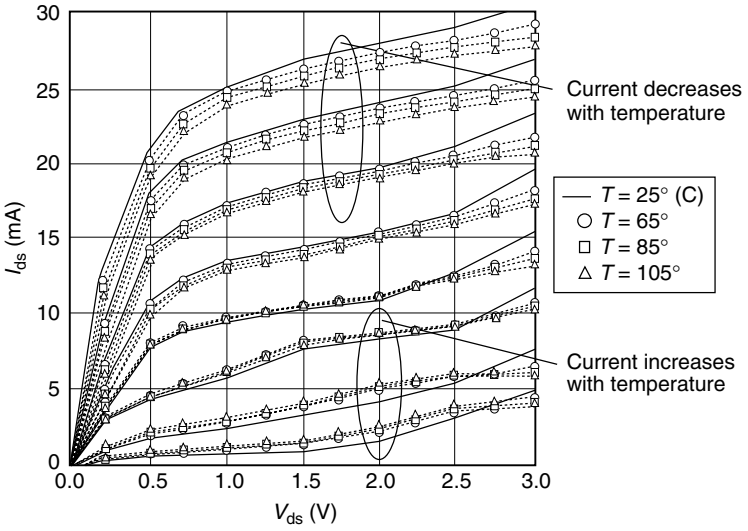


FIGURE 32.5 The behavior of DC drain current for a 0.25 μm PHEMT at four different temperatures.

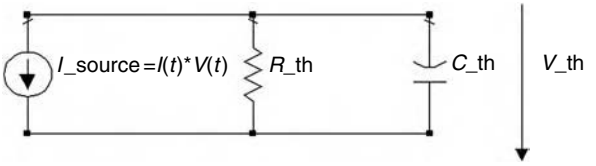


FIGURE 32.6 The thermal analog circuit.

increases with temperature at very low currents. This behavior is best modeled using a temperature analog circuit that will be described.

Figure 32.5 shows the behavior of DC drain current for a 0.25 μm *p*HEMT at four different ambient temperatures. Heating effects are obvious in Figure 32.5. Current (and transconductance) near positive V_{gs} is reduced and current (and transconductance) near pinch-off is increased.

The effects of temperature upon drain current in MESFETs and *p*HEMTs may be considered second-order but they are of first-order in bipolar devices. The reason is due to the exponential dependence of current upon temperature in a bipolar device. The gate current effects due to temperature in MESFETs and *p*HEMTs are of first-order for the same reason. Second-order and, even some, first-order changes with temperature may be modeled using linear temperature coefficients if the changes are reasonably linear.

Many devices are operated such that the self-heating effects are quite important. This can occur, for example, in a power amplifier design where the stand-by biasing current is low but the advent of input RF power leads to an increase in the drain current and output RF power. The ambient temperature may not change but the device will operate at a more elevated temperature with the application of the RF power. The use of temperature coefficients alone will not be accurate for such simulations.

More accurate modeling of self-heating effects in transistors circuit simulators has been done, for at least 10 years, using a thermal analog circuit, first used in SPICE applications. It has been found to work well for bipolar simulation as well as for MESFET and *p*HEMT simulations. The early studies were reported by Grossman and Oki [49], F. Q. Ye [50], and others [51–53]. The main differences in these early studies relate to the description of temperature effects in the transistor and not to the CAD model used for simulation in SPICE.

Figure 32.6 shows the thermal analog circuit used for analysis of thermal effects in a transistor. Each transistor must have its own thermal circuit. The thermal circuit is solved simultaneously with the transistor circuit. The transistor can be of any type but must have well defined descriptions of the model

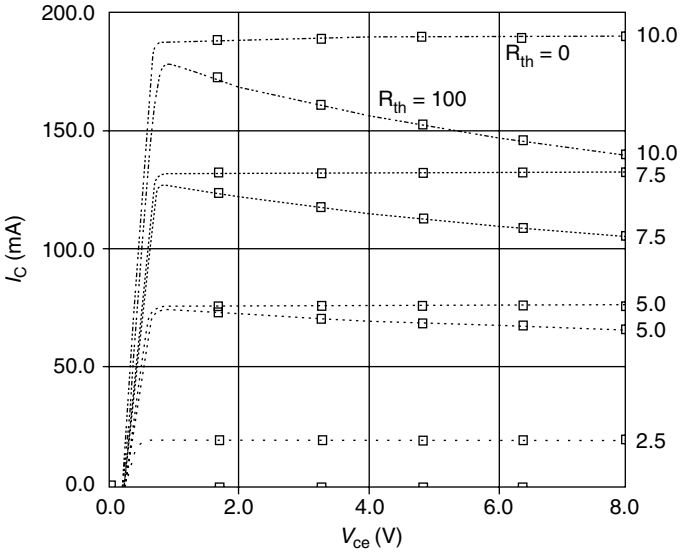


FIGURE 32.7 The collector *I–V* relationship for an HBT model with and without self-heating.

coefficients as a function of device temperature. The analog circuit consists of a current source, a resistor, *R*_th and a capacitor, *C*_th. *R*_th is the value of thermal resistance and the *R*_th * *C*_th time constant is the thermal time constant of the device. The current source in the thermal circuit, *I*_th, is equal in magnitude to the instantaneous internal dissipated power to the device. For DC biasing, *I*_th to the thermal circuit would be equal to the total biasing power to the device and the temperature rise, *V*_th, would be numerically equal to *I*_th * *R*_th. For RF or transient condition, the average temperature rise would be evaluated over some period of time including the effects of the thermal time constant. Thus, for an RF amplifier application, *I*_th would be equal to the DC biasing power plus the RF heating effects less the net RF power leaving the device. This "new" temperature is, then, used for the operating point of the transistor. The solution for a given operating condition is self-consistent, with the temperature rise producing the transistor current coefficients that result in that same temperature rise.

Convergence in a simulator is not assured because of the additional degree of freedom present. However, the thermal circuit is a low-pass filter and forces the temperature rise to change no faster than permitted by the thermal time constant. Since, the thermal time constant is usually long compared with the RF period or electrical tranisents changes, sequential solutions are, usually easily, obtained, such as for input RF power sweeps.

Analysis of a bipolar model with the thermal analog circuit has been used to explain the increased, low frequency gain observed in some GaAs HBTs that operate over a wide frequency range [54]. The cause is shown to be thermal modulation of the collector current at low frequencies.

In cases where the details of the thermal transient are important, the thermal circuit needs to be slightly more complex. In these cases the thermal circuit must have at least two time constants. The reason for this is that many authors [55] have shown that transient thermal behavior cannot be described with a single time constant, but one must have at least two time constants. This is a result, at least in part, of using the device in the real world where the influences of the die attach and packaging add their own *R*th/*C*th combinations.

Figure 32.7 shows the simulated collector *I–V* relationship for an HBT exhibiting self-heating effects. The current curves without heating effects are flat in the saturation region. Heating of the lattice reduces the electron mobility and, thus, reduces the collector current.

Figure 32.8 shows a thermal run-a-way condition for four-finger AlGaAs/GaAs HBT. Both data and simulated results are shown. The data shows that the device is in thermal run-a-way at VC = 8 V. The

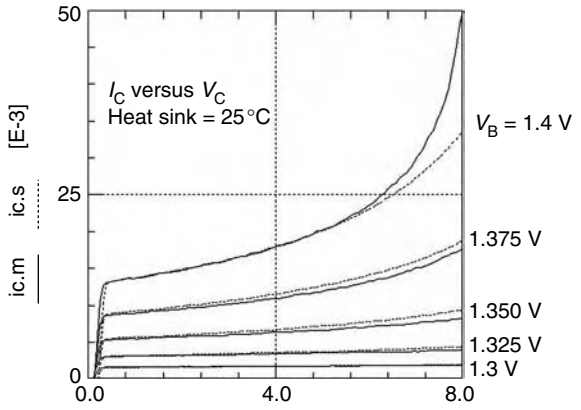


FIGURE 32.8 Model and data for thermal run-a-way condition of a four-finger GaAs HBT.

simulation is in good agreement with the data but does predict slightly larger run-a-way voltage. This difference is, most likely, due to the value of thermal resistance increasing with the device temperature.

32.9 Parameter Extraction for Compact Models

The extraction of parameters for a device model has become less laborious since the advent of new extraction and equipment control software, such as IC-CAP, by Agilent Technologies, Inc. and UTMOST by Sylvaco International. These programs provide data acquisition and parameter extraction to specific SPICE models. Such software, first, enables the engineer to collect I–V and RF data in a systematic fashion on each device tested. This provides consistency between I–V and RF data. The parameter extraction routines, then, permit the extraction to specific standard (SPICE) models, such as the Gunnel–Poon, or to customized models for which equations may be user-defined. Optimizers are used to provide the best fit between the data and the model.

Testing can be with pulsed biasing or DC. Heating effects can be separately studied using thermal chucks during testing. Teyssier et al. [46] have discussed the merits of long- and short-pulse testing. For the *p*HEMT device, a common method is to measure a pulsed drain current characteristic at 25°C and 125°C ambient temperature. The thermal resistance can be found by comparison with DC drain current data. The heating effect is accommodated in the model by using the temperature of operation from the thermal analog circuit to evaluate the drain current by interpolation of the two previously measured I–V characteristics, that is, the drain current at 25°C and 125°C.

It is, usually, necessary to use devices of small sizes for characterization and, then, scale the model to devices actually used in the circuit design. Most SPICE models provide scaling with device area. However, the scaling laws for devices should be verified. It is, usually, possible to scale MESFET and HEMT models accurately for a larger number of fingers of the same size. Golio [56] shows the scaling rules if the finger width is different. Because the biasing power may not be uniform on large devices and because the inter-electrode capacitance does not scale simply, more complicated scale rules may be found for relatively large devices at high frequencies, that is, above 4 GHz. Also, it is well known that thermal resistance does not scale with periphery.

32.10 Pulsed Current–Voltage Characterization of HEMTs

The use of the pulsed I–V characterization systems is quite useful for developing transistor models. The use of the dynamic I–V analyzer, or DiVA (Accent Optical Technologies Inc.), and the I–V and RF pulsed testing system (such as, Agilent Technologies now obsolete 85124A) have become more common in many

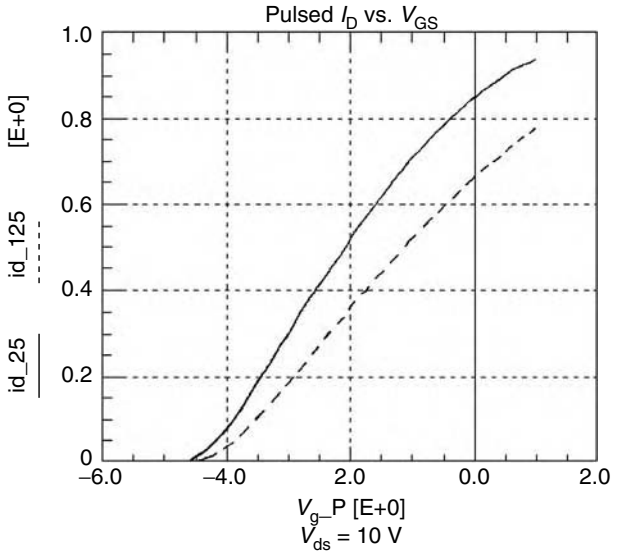


FIGURE 32.9 Drain current versus gate–source voltage for short pulsed conditions at $V_{ds} = 10$ V with quiescent $V_{ds} = 20$ V and $V_{gs} = -6$ V. Upper curve is for heat sink temperature of 25°C and lower curve is 125°C.

laboratories. Data may be taken using IC-CAP software control for the Agilent Technologies 85124A. However, for the DiVA system, the data is exported in IC-CAP data format to be used with IC-CAP for parameter extraction. The present DiVA system is limited to low power devices due to its current limit of 2A.

Pulsed biasing with pulse length less than 1 μs and for low duty cycle permits measurement of device currents without appreciable heating effects. Both gate– and drain–source voltages are pulsed.

These systems also permit separate setting of the quiescent state, that is, the voltage on the device's terminals between the biasing pulses. This permits control of the slow state trapping effects during the testing. Such testing is significantly different from standard gate-lag testing.

As Zhang et al. [57] and others have pointed out, trapping effects have the most influence at low drain–source voltages. The characterization of the transistor should be done at drain–source voltages that are large enough to avoid trapped-charge effects.

For example, Figure 32.9 shows the drain current control characteristic for a GaN *p*HEMT for a pulsed drain–source voltage of 10 V. The quiescent bias is drain–source voltage of 20 V and gate–source voltage of –6 V. The device is pinched-off between pulses. Two heat-sink temperatures are shown. They are 25°C and 125°C. This drain–source biasing voltage is significantly larger than the "knee" voltage for the device.

The currents are lower for 125°C mainly due to the lower electron mobility at the elevated temperature. However, it is also seen that the pinch-off voltage changes with temperature. A good *p*HEMT model for this device must include both effects and follow this data accurately. The next step in modeling would be to determine the thermal resistance of the device by comparison of the pulsed currents at different temperatures and the DC currents for the same condition, or by other testing means [57].

For a drain–source bias condition with low trapping effects, the final model's drain current (including self-heating effects) should agree well with the DC current data. In addition, the model's drain current characteristic should agree with the data of Figure 32.9 for the same bias conditions but with thermal resistance set to zero.

An interesting check of this modeling approach is to compare the external transconductance measured at RF frequency to that measured from the current control characteristic for the same bias conditions. The external RF transconductance can be found from the magnitude of Y21 at low-RF frequency, whereas

the current transductance is found by differentiation of the ID (V_{gs}) data for the same bias point. The agreement has been found to be within a few percent for GaN HEMTs with low trapping effects and low self-heating effects. Self-heating effects will reduce the transconductance calculated from DC data (either RF or I–V data). Trapping effects will reduce the transconductance for RF frequencies. In that case, the model may have to accommodate dispersion between the DC and RF transconductance. See, for example, the COBRA [34] model.

32.11 Model Verification

An excellent compact model will give good agreement with data for drain and gate current control in *p*HEMTs or collector and base currents for bipolar devices, good agreement for small-signal *s*-parameters at typical bias conditions and good agreement for large-signal measurements. The model must also be reasonably accurate for significant variation in ambient temperature. A limited range of scaling capability is useful for many device technologies.

Verification of the model's agreement at small signal levels is straightforward. However, verification of the large-signal performance is not.

The usual approach to verification of a large-signal model is to compare measured device performance with simulations under the same conditions. Initial verification should be comparison of a power sweep of the transistor at the application frequency and with no matching at the input and the output. For this test, we know that all harmonics at the input and output see 50 Ω impedance.

One should measure not only the output power at the fundamental, but also the power at second and third harmonics. If the model is not fully optimized, it will usually agree well with fundamental output power, gain, and efficiency, but not agree with the harmonic power production. The usual cause is due to poor modeling of the I–V relationship, however, in some cases, the nonlinear capacitive modeling may be the problem. After this problem is fixed, testing of third- and fifth-order intermodulation distortion (IMD) should be made, again in a 50-Ω system. If the third harmonic is in agreement, then the third-order IMD will agree since it is determined by the same device properties. Some further work may be required to get agreement for the fifth-order IMD.

In the previous test, it is important that the large-signal model be reasonably accurate at small-signal levels. It need not be as accurate as the best small-signal model for many power applications.

Further verification would involve power sweeps under tuned conditions. That is, the transistor may be tuned for best efficiency and the input and output tuner impedance measured. It is important to measure the tuner impedance for fundamental, second and third harmonics and to use these values in the simulation. The effects of the second harmonic voltage at either input [58] or output can be extremely important.

Testing of the load-pull characteristics should be made and compared with the model's behavior. Here, again, one has to be careful about the effects of harmonics. There are load-pull systems that operate separately on fundament and harmonics.

Further tests that may be important to the application are, for example, testing with various ambient temperatures, noise testing, switch-on (transient) testing, and others. The specific application of the transistor will dictate the importance of the agreement for each test as well as the RF frequencies and power and modulation to use for testing. The next section will give an example of verification of a large signal model for the LDMOS transistor.

32.12 Modeling the RF LDMOS Power Transistor

There are numerous silicon MOS models available for DC and RF modeling of silicon transistors. Because the silicon LDMOS transistor has become important for cost-effective consumer applications, many companies have developed nonlinear models specifically for this device. The device incorporates a *p*-type sinker diffusion used to ground the source to the substrate and, thus, the device can be treated as a

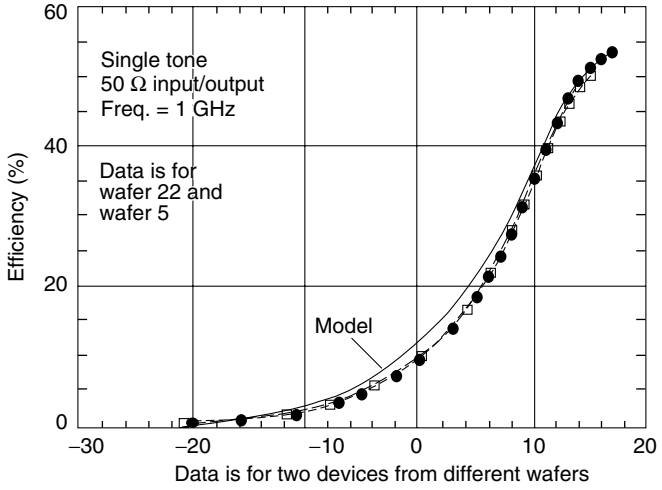


FIGURE 32.10 Efficiency prediction by the model and data for two RF LDMOS devices.

three-terminal device. This makes it possible to construct a much simpler model, one very similar to the SPICE models developed for the GaAs MESFET and *p*HEMT.

Perugupalli et al. [59] have used a SPICE circuit network incorporating the standard NMOS SPICE element. Motorola has utilized the Root [13] model developed for GaAs MESFETs for characterizing the device. However, self-heating effects, important for power applications, cannot be incorporated into this model. The Ho, Green, and Culbertson model [60], based upon the SPICE BSIM3v3 model sufferers from the same problem.

Miller et al. [61] developed analytical current equations for the LDMOS device which lead to the development of a new, simpler SPICE model by Curtice [62]. The model includes self-heating effects, accurate for both small- and large-signal simulations, and operates in transient or harmonic balance simulators. Figure 32.10 shows the model predicts power-added efficiency in excellent agreement with the data for an RF power sweep. A similar model based upon the same equations has been developed and verified by Heo et al. [63].

More recently, a new LDMOS model [64,65] was developed based upon the current control characteristics described by Fager et al. [66]. It is called the CMC model. The key advantage of the Fager–Pedro model is proper treatment of current in the four regions consisting of sub-threshold, quadratic, linear, and compressed. Figure 32.11 from Curtice et al. [64] shows how well the model tracks the data for optimum loading at 900 MHz.

Figure 32.12 from the same reference [64] shows IP3 data for the same device operated at V_{ds}= 27 V and 900 and 910 MHz. Here again, the agreement of the model and data is excellent.

A major feature of the CMC model is that it was found to scale very well to much larger devices. In fact, scaling from 1 to 30 W was accomplished to develop a model for the Cree UGF21030 device [65]. With a proper package model the simulations agree very well with data for this device.

32.13 Enhancing the Gummel–Poon Model for Use with GaAs and InP HBTs

The Gummel–Poon model, or the GP model [14] is a complex, physical parameter model with 55 parameters and is widely used. It was developed early for SPICE and all colleges and universities teach their electrical engineering students to use this model. Although the GP model has many parameters, the current expressions are relatively simple. In addition, the current parameters are more closely tied to material parameters rather than manufacturing tolerances, so that there is less variation in current control

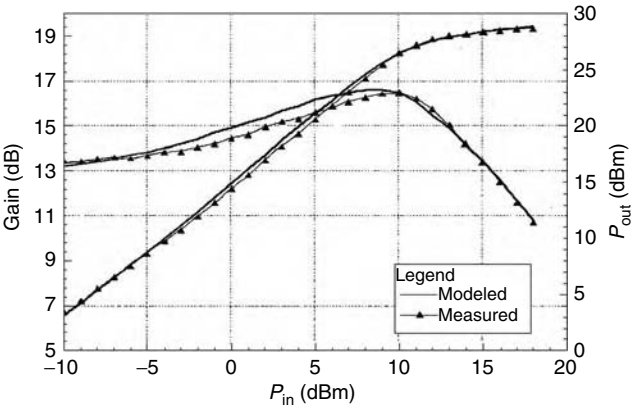


FIGURE 32.11 Output power and gain for model and data for optimum loading condition for a 1 W LDMOS device at 900 MHz.

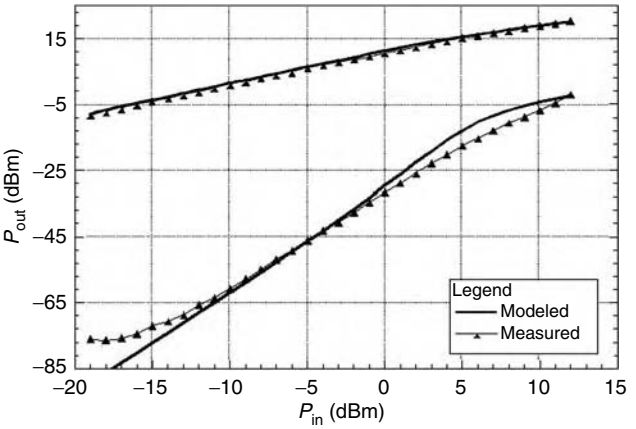


FIGURE 32.12 Fundamental and third-order IMD for model and data for operation at V_{ds}= 27 V with 900 and 910 MHz input signals.

characteristics than with MESFETs and *p*HEMTs. The standard bipolar device has less two-dimensional effects than do MESFETs and *p*HEMTs.

Much effort has been expended to improve the accuracy of compact BJT circuit models for silicon devices. Fossum [54] has reviewed the efforts to 1989 and it continues to this day.

Whether the bipolar device is silicon or a heterojunction device made with SiGe or GaAs based, etc., the bipolar action with current gain is the same physical process. So one expects some similarities in the analysis and modeling of the device. However, the heterojunction with the wide-band gap emitter causes the details of the analysis and model to have significant differences from a homojunction device.

The standard SPICE GP model has a number of major deficiencies that must be addressed before it can be used to accurately model an HBT in a large-signal microwave application. First of all, the SPICE code has silicon band gap parameters hard coded into it and this must be changed to produce the correct temperature effects upon the band gap. Next, collector to base avalanche breakdown must be added because it is important to most HBT applications. The GP model uses PTF, a phase function, to accommodate time delay associated with transconductance. It is more continental for microwave engineers to use the time delay term TAU, as used in MESFET models.

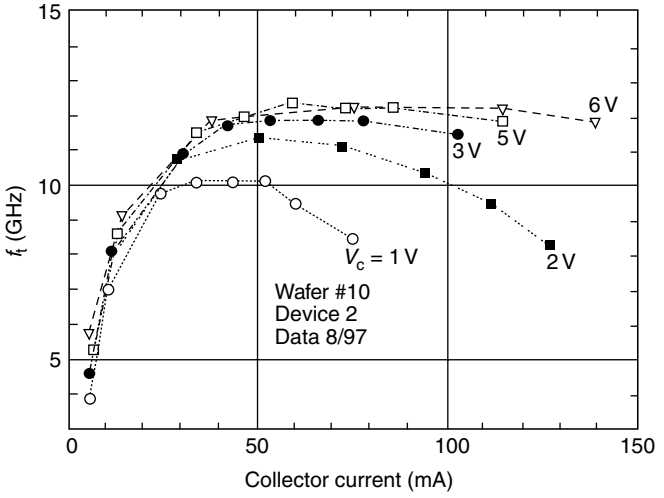


FIGURE 32.13 The behavior of Ft with biasing for a SiGe HBT.

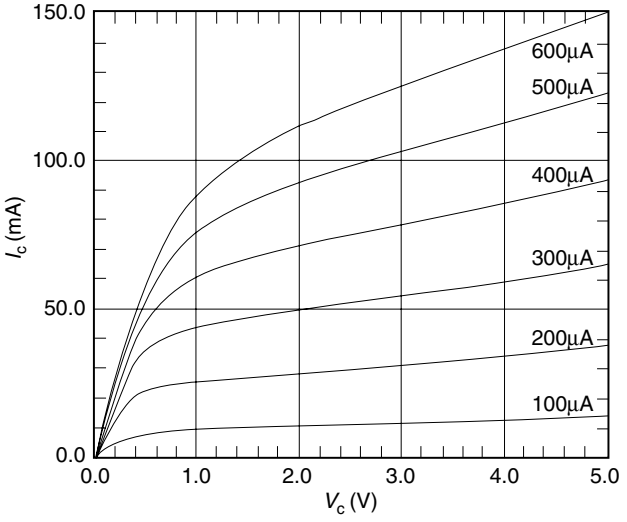


FIGURE 32.14 The *I–V* relationship for the device of Figure 32.7.

The parameter early voltage is not as important in GaAs modeling, as it is usually very large. This results because the base doping can be made an order of magnitude larger than for silicon devices, because of the wide band gap emitter. The large base doping reduces the importance of collector biasing upon the base region and, thus, upon the collector current.

There may be dispersion in the collector admittance in HBTs, so an RF conductive element may be needed between collector and emitter.

The manner in which Ft, the frequency for unity current gain, changes with voltage and current is quite different in GaAs device compared with silicon devices. Therefore, a new equation is needed for the total TF function of voltage and current. Most of the behavior of Ft with respect to collector voltage is related to the electron velocity–electric field (v–E) curve for the material. In the case of silicon devices, Ft increases with collector voltage and saturates until heating effects cause a decrease. Figure 32.13 show such behavior for a SiGe HBT. The *I–V* characteristic of the device is given in Figure 32.14.

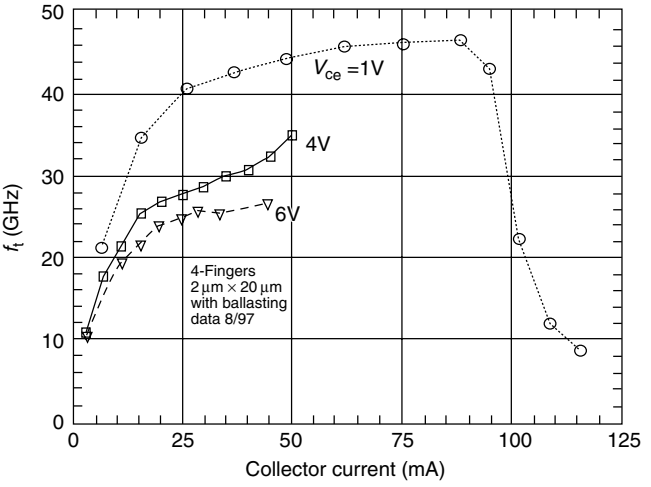


FIGURE 32.15 The behavior of Ft with biasing for GaAs HBT.

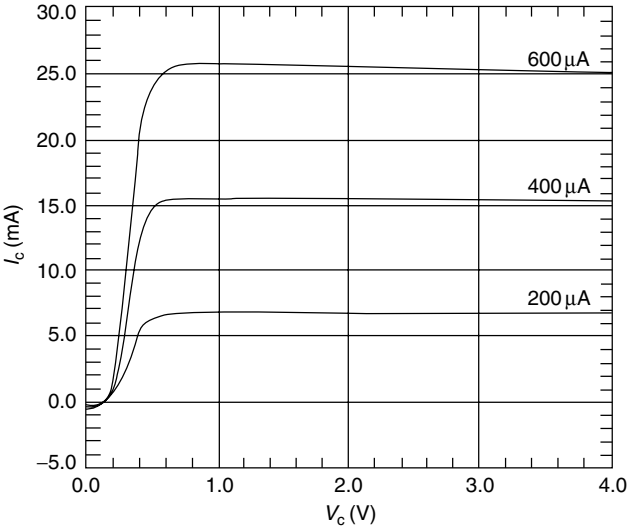


FIGURE 32.16 The *I–V* relationship for the device of Figure 32.9.

In the case of GaAs HBTs, Ft peaks at a low voltage and monotonically decrease with further increase of collector voltage. Figure 32.15 shows such data and the devices *I–V* characteristic is given in Figure 32.16.

This difference in behavior reflects the striking differences between silicon and GaAs electron velocity-field curves in the high-field region. Understandably, the same behavior occurs for comparison of silicon and GaAs MEFETs. That is, for silicon, Ft increases with drain–source voltage after the current knee. And for GaAs, Ft decreases with drain–source voltage after the current knee.

It is especially important that the emitter resistance be found accurately for the compact bipolar model. In addition, it should be consistent between DC and RF testing. An excellent procedure is the low-frequency impedance method of Maas and Tait [69]. The classic DC flyback method often produces poor values and is not reliable.

There are a multitude of other CAD models formulated to model the behavior of HBTs. Some examples are the VBIC model [67], the MEXTRAM model [68], and the HICUM [70] model and these are installed on most circuit simulators. All of the new models include self-heating effects because of their importance

to device operation and accurate modeling. Because of the significantly poorer thermal conductivity in GaAs compared to silicon and because of the higher power density for best operation in GaAs, self-heating effects are particularly important to the operation of the GaAs-based HBTs.

32.14 The Vector Nonlinear Network Analyzer

A large-signal, waveform measurement system has been used by many researchers to measure device characteristics dynamically. The equipment provides time-domain voltage and current waveforms during RF excitation of the transistor. The equipment is often called the vectorial nonlinear network analyzer, or VNNA.

Demmler and Tasker [71] have shown that it is possible to accurately determine the drain current relationship to gate voltage for RF excitation at 2 GHz of a MODFET. The characteristic time delay is found by adding delay until "looping" is minimized. Furthermore, the drain–source I–V relationship can also be evaluated at 2 GHz. There is some difference from that obtained from the dc data. Thus the VNNA provides large-signal transfer characteristics from which to do more accurate model extraction. Wei et al. [72] has also utilized this techniques to provide the data used for device model parameter extraction for a GaAs HBT.

Remley et al. [73] present an excellent discussion of the measurement consistency of such nonlinear measurement systems.

32.15 Foundry Models and Statistics

GaAs chip foundries provide design manuals that utilize small-signal as well as large-signal models. These are developed from measurements and statistical analysis of the data. However, these are guidelines for the designer and, often, the best procedure is to obtain foundry test devices and develop more accurate models based upon new data. Device uniformity has improved greatly and yield prediction is becoming more accurate.

In addition, software programs such as IC-CAP and others, provide statistical analysis of models extracted from test wafers. The model parameters used for statistical analysis must be significantly correlated to electrical behavior. For example, gate oxide thickness will strongly correlate to gate source capacitance and current cut-off frequency in an MOS transistor. But, polynomial coefficients used to describe current control will not have such correlation. Some software packages enable definition of principal components that are a new set of parameters constructed by linear combinations of the model parameters, but are independent. Also, some software packages can deal with non-Gaussian statistics, which is very common.

The net result from these statistical packages are a nominal model and boundary models. Such models then enable the designer to use a small number of models to improve the circuit design for yield and performance.

The fundamental problem is that most foundries continue to tune their device manufacturing processes. It is often the case that the process has been changed and previous statistics are no longer valid. However, the design engineer is better off using approximate guidelines as to statistical patterns and boundary models than none at all.

32.16 Future Nonlinear Transistor Models

One can expect that with the ever-increasing speed of computers, circuit simulators will be able to utilize more physics-based models. This will aid in the determining the effect of device design parameters upon chip yield and performance.

Improvements will be made in nonlinear model extraction software. The extraction parameters will be much less dependent upon the expertise of the user. There will be improved collection schemes for transistor model statistics.

Finally, one expects that the nonlinear models will be made to be more easily tailored for adaptation to specific device behaviors. One would like to start with a template for one of the standard nonlinear models and, then, tailor its behavior. Future simulators should make this procedure simpler than present procedures. The more frequent use of Verilog-A, for example, is a step in this direction.

References

1. Nagle, L. W., SPICE 2: A computer program to simulate semiconductor circuits, Electronics Research Laboratory, College of Engineering, University of California, Berkeley, Memo, ERL-M520, 1975.
2. Pinto, M. R., Conor, Rafferty, S. and Dutton, R. W., PIECES2—Poisson and continuity equation solver, Standord Electronics Laboratory, Technical Report, Stanford University, 1984.
3. Wada, T. and Frey, J., Physical basis of short-channel MESFET operation, *IEEE Transactions on Electronic Devices*, ED-26, 476, 1979.
4. Curtice, W. R., Analysis of the properites of three-terminal transferred electron logic gates, *IEEE Transactions on Electronic Devices*, ED-24, 1553, 1977.
5. Warriner, R. A., Computer simulation of gallium arsenide field-effect transistors using Monte-Carlo methods, *Solid-State Electronic Devices*, 1, 105, 1977.
6. Moglestue, C., A self-consistent Monte Carlo particle model to analyze semiconductor microcomponents of any geometry, *IEEE Transactions on CAD*, CAD-5, 326, 1986.
7. Curtice, W. R. and Yun, Y.-H., A temperature model for the GaAs MESFET, *IEEE Transactions on Electronic Devices*, ED-28, 954, 1981.
8. *ATLAS User's Manual*, Silvaco International, Santa Clara, CA, Version 4.0, 1995.
9. Curtice, W. R., Direct comparison of the electron-temperature model with the particle-mesh (Monte-Carlo) model for the GaAs MESFET, *IEEE Transactions on Electron Devices*, ED-29, 1942, Dec. 1982.
10. Li, Q. and Dutton, R. W. Numerical small-signal AC modeling of deep-level-trap related frequency-dependent output conductance and capacitance for GaAs MESFETs on semi-insulating substrates, *IEEE Transactions on Electronic Devices*, 38, 1285, 1991.
11. Snowden, D. M. and Pantoia, R. R., Quasi-two-dimensional MESFET simulation for CAD, *IEEE Transactions on Electronic Devices*, 36, 1989.
12. Morton, C. G., Atherton, J. S., Snowden, C. M., Pollard, R. D. and Howes. M. J., A large-signal physical HEMT model, *1996 International Microwave Symposium Digest*, 1759, 1996.
13. Root, D. E. et al., Technology independent large-signal non quasi-static FET models by direct construction from automatically characterized device data, *21st European Microwave Conference Proceedings*, 927, 1991.
14. Gummel and Poon, An integral charge-control relationship for bipolar transistors, *Bell System Technical Journal*, 49, 115, 1970.
15. Daniel, T. T. and Tayrani, R., Fast bias dependent device models for CAD of MMICs, *Microwave Journal*, 74, 1995.
16. Khatibzadeh , M. A. and R. J. Trew, R. J., A large-signal analytical model for the GaAs MESFET, *IEEE Transactions on Microwave Theory and Techniques,* 36, 231, 1988.
17. Ladbrooke, P. H., *MMIC Design: GaAs FETs and HEMTs*, Artech House, Inc., Boston, Chapter 6, 1989.
18. Lehovec, K. and Zuleeg, R., Voltage-current characteristics of GaAs JFETs in the hot electron range, *Solid State Electronics,* 13, 1415, 1970.
19. Bandler, J. W. et al., Statistical modeling of GaAs MESFETs, *1991 IEEE MTT-S International Microwave Symposium Digest,* 1 , 87, 1991.
20. Statz, H, Newman, P., Smith, I. W., Pucel, R. A. and Haus, H. A., GaAs FET device and circuit simulation in SPICE, *IEEE Transactions on Electronic Devices*, 34, 160, 1987.
21. D'Agostino, S. et al., Analytic physics-based expressions for the empirical earameters of the Statz-Pucel MESFET model, *IEEE Transactions on MTT*, MTT-40, 1576, 1992.
22. *International Journal of RF and Microwave Computer-Aided Engineering*, 9, No. 3, 1999.

23. Trew, R. J., MESFET models for microwave CAD applications, *International Journal of RF and Microwave Computer-Aided Engineering*, 1, 143, 1991.
24. Snowden, C. M., Nonlinear modeling of power FETs and HBTs, *International Journal of RF and Microwave Computer-Aided Engineering*, 6, 219, 1996.
25. Dortu, J.-M., Muller, J.-E., Pirola, M. and Ghione, G. Accurate large-signal GaAs MESFET and HEMT modeling for power MMIC amplifier design, *International Journal of RF and Microwave Computer-Aided Engineering*, 5, 195, 1995.
26. Curtice W. R. and Camisa, R. L. Self-consistent GaAs FET models for amplifier design and device diagnostics, *IEEE Transactions on Microwave Theory and Techniques*, MTT-32, 1573, 1984.
27. Vaitkus, R. L., Uncertainty in the Values of GaAS MESFET eqivalent circuit elements extracted from measured two-port scattering parameters, *Presented at 1983 IEEE Cornell Conference on High Speed Semiconductor Devices and Circuits*, Cornell University, 1983.
28. Byun, Y. H., Shur, M. S., Peczalski, A. and Schuermeyer, F. L., Gate voltage depenence of source and drain resistances, *IEEE Transactions on Electron Devices*, 35, 1241, 1998.
29. Samelis, A. and Pavlidis, D., Modeling HBT self-heating, *Applied Microwave and Wireless*, Summer Issue, 56, 1995.
30. Teeter, D. A. and Curtice, W. R., Comparison of hybrid pi and tee HBT circuit topologies and their relationship to large-signal modeling, *1997 IEEE MTT-S International Microwave Symposium Digest*, 2, 375, 1997.
31. Curtice, W. R., A MESFET model for use in the design of GaAs integrated circuits, *IEEE Transactions on Microwave Theory and Techniques*, 23, 448, 1980.
32. Angelov, I, Zirath, H. and Rorsman, N., New empirical nonlinear model for HEMT and MESFET and devices, *IEEE Transactions on Microwave Theory and Techniques*, 40, 2258, 1992.
33. Qu, G. and Parker, A. E., Continuous HEMT model for SPICE, *IEE Electronic Letters,* 32, 1321, 1996.
34. Cojocaru, V. I. and Brazil, T. J., A scalable general-purpose model for microwave FETs including the DC/AC dispersion effects, *IEEE Transactions on Microwave Theory and Techniques*, 12, 2248, 1997.
35. Materka, A. and Kacprzak, T., Computer calculation of large-signal GaAs FET amplifier characteristics, *IEEE Transactions on Microwave Theory and Techniques*, 33, 129, 1985.
36. Nonlinear Devices, Advanced Design System Documentation 2005A, August, 2005.
37. Kenneth S. Kundert and Olaf Zinke, *The Designer's Guide to Verilog-AMS*, Kluwer Academic Publishers, 2004.
38. Calvo, M. V., Snider, A. D. and Dunleavy, L. P., Resolving capacitor discrepancies between large and small signal FET models, *1995 IEEE MTT-S International Microwave Symposium*, 1251, 1995.
39. Ph. Jansen et al., Consistent small-signal and large-signal extraction techniques for heterojunction FET's, *IEEE Transaction on Microwave Theory and Techniques*, Vol. 43, No. 1, 87, 1995.
40. Mallavarpu, R. , Teeter. D. and Snow, M., The importance of gate charge formulation in large-signal pHEMT modeling, *GaAs IC Symposium Technical Digest*, 87, 1998.
41. Horio, K. and Usarni, K, Analysis of kink-related backgating effect in GaAs MESFETs, *IEEE Electron Devices Letters*, 16, 537, 1995.
42. Curtice, W. R., Bennett, J. H. , Suda, D. and Syrett, B. A., Modeling of current lag effects in GaAs IC's, *1998 IEEE MTT-S International Microwave Symposium Digest*, 2, 603, 1998.
43. K. Kunihiro and Y. Ohno, An equivalent circuit model for deep trap induced drain current transient behavior in HJFETs, *1994 GaAs IC Symposium Digest*, 267, 1994.
44. Chakrabarti, P., Shrestha, S. K., Srivastava, A. and Skxena, D., Switching characteristics of an optically controlled GaAs-MESFET, *Transactions on Microwave Theory and Techniques*, 42, 365, 1994.
45. Madjar, K., Paolella, A. and Herczfeld, P. R., Modeling the optical switching of MESFET's considering the external and internal photovoltaic effects, *Transactions on Microwave Theory and Techniques*, 42, 62, 1994.
46. Teyssier, J.-P., Bouysse, P., Ouarch, A., Barataud, D., Peyretaillade, T. and Quere, R.,40-GHz/150-ns versatile pulsed measurement system for microwave transistor isothermal characterization, *IEEE Transactions on Microwave Theory and Techniques*, 46, 2043, 1998.
47. A. Platzker et al., Characterization of GaAs devices by a versatile pulsed I–V measurement system, *1990 IEEE MTT Symposium Digest*, 1137.

48. Anlholt and Swirhun, Experimental characterization of the temperature dependence of GaAs FET equivalent circuits, *IEEE Transactions on Electron Devices*, 39, 2029, 1992.
49. Grossman, P. C. and Oki, A., A large signal dc model for GaAs/GaAlAs heterojunction bipolar transistors, *Proc. IEEE BCTM*, 258, 1989.
50. Ye, F. Q., A BJT model with self-heating for WATAND computer simulation, M.S. Thesis, Youngstown State University, 1990.
51. McAndrew, C. C., A complete and consistent electrical/thermal HBT model, *Proceedings of IEEE BCTM*, 200, 1992.
52. Corcoran, J., Poulton, K. and Knudsen, K., GaAs HBTs: an analog circuit design perspective, *Proceedings of IEEE BCTM*, 245, 1991.
53. Fox, R. M. and Lee, S.-G., Predictive modeling of thermal effects in BJTs, *Proceedings of IEEE BCTM*, 89, 1991.
54. Fossum, J. G., Modeling issues for advanced bipolar device/circuit simulation, *Proceedings of 1989 IEEE BCTM*, 234, 1989.
55. Mnif, M., Zimmer, T., Battaglia, J. L. and Ardouin, B., Modeling the self-heating effect in SiGe HBTs, *IEEE Proceedings of the 2002 BCTM*, IEEE Catalog No. 02CH37384, 96–99, 2002.
56. Golio, J. M., *Microwave MESFETs and HEMTs*, Artech House, 1991.
57. Zhang, A., Madjar, A., Osinsky, S., Hwang, C. M., Osinsky, A., Dabrian, A. M. and Hertog, B., Thermal resistance estimation of GaN high electron mobility transistors, *2003 Sarnoff Symposium Digest*, 2004.
58. Watanabe, S, Takatuka, S., Takagi, K., Kukoda, H. and Oda, Y., Simulation and experimental results of souce harmonic tuning on linearity of power GaAs FET, *1996 MTT-S International Microwave Symposium*, 1996.
59. Perugupalli, P., Trivedi, M., Shenai, K. and Leong, S. K., Modeling and characterization of 80v LDMOSFET for RF communications, *1997 IEEE BCTM*, 92, 1997.
60. Ho, M. C., Green, K., Culbertson, R., Yang, J. Y., Ladwig, D. and Ehnis, P., A physical large signal Si model for RF circuit design, *1997 MTT-S International Microwave Symposium Digest*, 1997.
61. Miller, M, Dinh, T. and Shumate, E., A new empirical large signal model for silicon RF LDMOS FET's, *1997 IEEE MTT Symposium on Technologies for Wireless Applications Digest*, Vancouver, Canada, 19, 1997.
62. Curtice, W. R., Pla, J. A., Bridges, D., Liang, T. and Shumate, E., A new dynamic electro-thermal model for silicon RF LDMOS FET's, *1999 IEEE MTT-S International Microwave Symposium Digest*, 1999.
63. Heo, D., Chen, E., Gebara, E., Yoo, S., Lasker, J. and Anderson, T., Temperature dependent MOSFET RF large signal model incorporating self-heating effects, *1999 MTT-S International Microwave Symposium Digest*, 1999.
64. Curtice, W. R., Dunleavy, L. and Pengelly, R., New LDMOS model delivers powerful transistor library, Part 1—The CMC Model, High Frequency Design, Oct., 2004.
65. Wood, S., Pengelly, R., Dunleavy, L., Clausen, W., Weller, T., Emmadi, L. New LDMOS model delivers powerful transistor library, Part 2—Library Applications, *High Frequency Design*, Nov., 2004.
66. Fager, Pedro, Carvalho and Zirath, Prediction of IMD in LDMOS Transistor Amplifiers Using aNew Large-Signal Model, *IEEE Transactions on Microwave Theory and Techniques*, 50, 12, 2834–2842, December 2002.
67. McAndrew, C. C., Seitchik, J., Bowers, D., Dunn, M., Foisy, M., Getreu, I., Moinian, S., Parker, J., van Wijnen, P. and Wagner, L., VBIC95: an improved vertical IC bipolar transistor model, *Proceedings of 1995 BCTM*, 170, 1995.
68. de Graaff, H. C. and Kloosterman, W. J., New formulation of the current and charge relations in bipolar transistor modeling for CACD purposes, *IEEE Transactions on Electronic Devices*, ED-32, 2415, 1986.
69. Maas, S. A. and Tait, D., Parameter-extraction method for heterojunction bipolar transistors, *IEEE Microwave and Guided Wave Letters*, 2.12, 502, 1992.
70. HICUM, A Scalable Physics-based Compact Bipolar Transistor Model, Schroter, M., see http://www.iee.et.tu-dresden.de/iee/eb/comp_mod.html
71. Demmler, M. and Tasker, P. J., A vector corrected on-wafer large-signal waveform system for novel characterization and onlinear modeling techniques for transistors, Presented at the workshop on New direction in Nonlinear RF and Microwave Characterization, *1996 International Microwave Symposium*, 1996.

72. Wei, C.-J., Lan, Y. E., Hwang, J. C. M., Ho, W.-J. and Higgins, J. A. Waveform-based modeling and characterization of microwave power heterojuction transistors, *IEEE Transactions on Microwave Theory and Techniques*, 43, 2898, 1995.
73. Remley, K. A., DeGroot, D. C., Jargon, J. A., and Gupta, K. C., A method to compare vector nonlinear analyzers, IEE MTT-S International Microwave Symposium Digest, pp. 1667–1670, 2001.

33

Behavioral Modeling of Microwave Power Amplifiers

Troels S. Nielsen
RF Micro Devices

33.1 Introduction

Today's ever-increasing demands for more bandwidth coupled with requirements for both high linearity and high efficiency create great challenges in the design of modern cellular handsets. The power amplifier (PA) in the transmitter chain is a particularly critical component. To prolong time between battery recharge, the PA must be very power efficient. Moreover, to comply with today's spectrally efficient modulation forms, the PA must also be very linear. To alleviate the fundamental tradeoff between efficiency and linearity in the PA, radio frequency (RF) designers may employ complex transmitter configurations [1–3]. In an early stage of development, such designs require accurate simulation models of the PA and other vital building blocks. For the modeling, the RF designer may choose from three different methodical approaches (Figure 33.1): models based on device physics, models based on equivalent circuits, and behavioral models. The three approaches are briefly summarized in the following. After that, we will focus on behavioral models only.

Physics-based models [4–6] describe the device in terms of its physical structure and predict performance from electromagnetic equations and charge transport definitions. In principle, such models may predict device performance, a priori without the need for actually fabricating the device itself. In practice, however, some parameters always need adjustment, a posteriori, from measurements. Physics-based

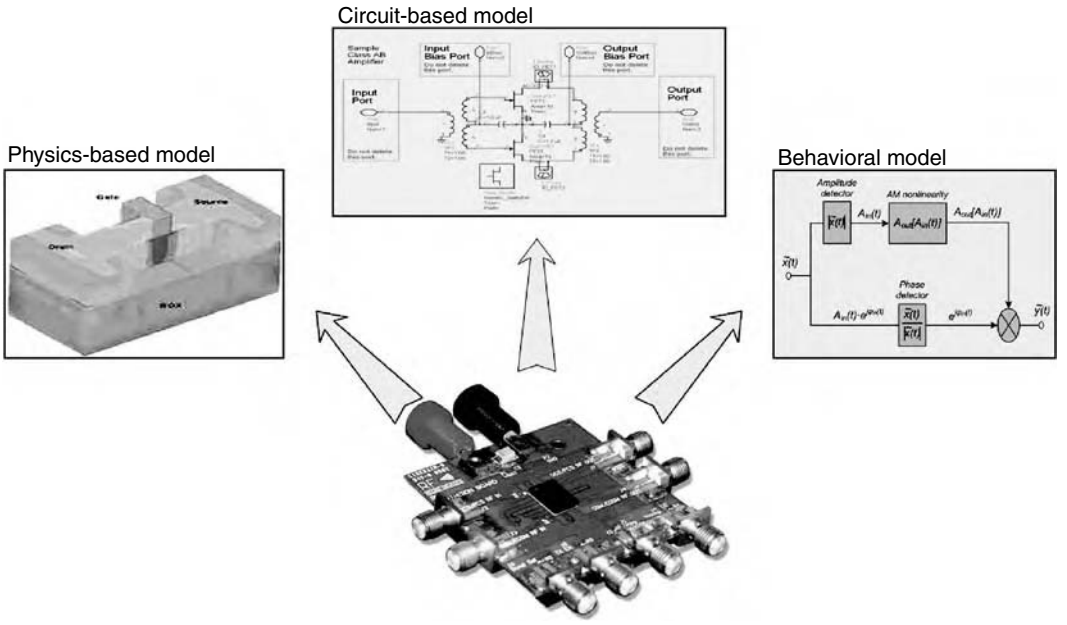


FIGURE 33.1 Three methodical approaches in microwave device modeling.

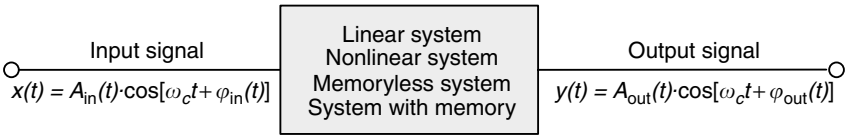


FIGURE 33.2 System block level representation with input and output signal definitions.

models tend to be computationally heavy and are, therefore, not very functional in large-scale system simulations.

Circuit-based models [7–10] rely on equivalent circuits. They consist of fundamental electrical circuit elements such as resistors, capacitors, and controlled sources. Because of their equivalent circuit base, the models require detailed knowledge of the internal device composition but apply well in circuit-level simulations. They may predict device performance very accurately. For large-scale system simulations, however, the circuit-based models are often too complex and computational intensive.

Behavioral models are extracted from data measured on fabricated devices. Typically, they assume no *a priori* knowledge of the internal device composition (black-box modeling) and rely exclusively on a set of wisely selected input–output measurements. The model base is a set of mathematical fitting functions, which approximate the input–output measurements. While behavioral models may vary greatly in complexity and in computational burden, they are often orders of magnitudes faster than physics-based models and circuit-based models. This makes them more useful in large-scale system simulations.

33.2 A Few System Classifications

Before initiating a discussion of the various approaches to PA behavioral modeling, it is convenient to recall some basic results from system identification theory. Figure 33.2 summarizes the four categories

of systems discussed in the following together with the input and output signal definitions used in the discussion.

33.2.1 Linear Systems vs. Nonlinear Systems

Linear systems are defined as those for which the superposition principle holds. More specifically, if the two inputs x_1 and x_2 are applied separately to a linear system and produce the two responses y_1 and y_2, respectively, then the response to the excitation $ax_1 + bx_2$ is $ay_1 + by_2$, where a and b are arbitrary constants, which may be real or complex, time-invariant, or time-varying. A simple example of such a system is a network consisting of linear resistors. In the frequency domain, the output of a linear system contains the same terms as the input signal.

Any system that does not obey the superposition principle is said to be nonlinear. In such systems, the output amplitude is a nonlinear function of the input amplitude (Figure 33.3). Simple nonlinearities may be described accurately by a polynomial input–output relation (here limited to the third order of the input amplitude).

$$y(t) = a_1 x(t) + a_2 x^2(t) + a_3 x^3(t), \tag{33.1}$$

where $a_1 - a_3$ are the real-valued nonlinearity coefficients, a_1 is the linear small-signal gain (defined in Figure 33.3 as the ratio of A_{out} to A_{in} at small-signal inputs), a_2 and a_3 are the nonlinear quadratic and cubic gain terms, respectively. The larger the quadratic and cubic coefficients, the more distorted the output waveform will be. In the frequency domain, the output of a nonlinear system will contain additional terms not present in the input signal. The quadratic term gives rise to a DC component and a second harmonic at twice the input frequency; the cubic term generates a third harmonic.

While harmonic distortion often is used to describe nonlinearities of analog circuits, certain devices (such as the PA) may call for additional linearity measures. Consider two tones of equal amplitude A at the angular frequencies ω_1 and ω_2

$$x(t) = A[\cos(\omega_1 t) + \cos(\omega_2 t)]. \tag{33.2}$$

Applying these two tones to Equation 33.1 results in the output

$$y(t) = a_1[A\cos(\omega_1 t) + A\cos(\omega_2 t)] + a_2[A\cos(\omega_1 t) + A\cos(\omega_2 t)]^2 + a_3[A\cos(\omega_1 t) + A\cos(\omega_2 t)]^3 \tag{33.3}$$

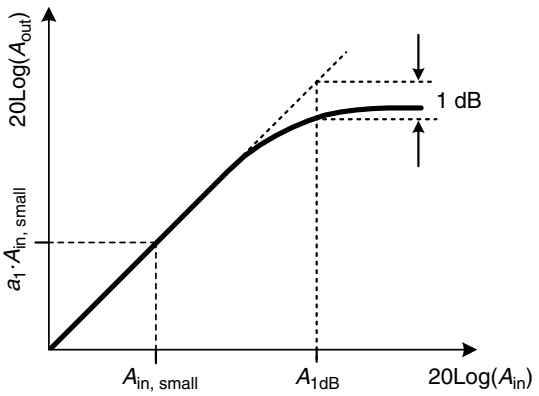


FIGURE 33.3 Input–output amplitude transfer function (AM/AM) of a simple third-order nonlinear system.

By expanding Equation 33.3 and discarding DC terms and harmonics of the input signal, we obtain the well-known third-order inter-modulation distortion (IMD) products

$$\begin{aligned} &\text{for } \omega = 2\omega_1 \pm \omega_2\text{:} \quad \frac{3a_3A^3}{4}\cos(2\omega_1 t + \omega_2 t) + \frac{3a_3A^3}{4}\cos(2\omega_1 t - \omega_2 t), \\ &\text{for } \omega = 2\omega_2 \pm \omega_1\text{:} \quad \frac{3a_3A^3}{4}\cos(2\omega_2 t + \omega_1 t) + \frac{3a_3A^3}{4}\cos(2\omega_2 t - \omega_1 t). \end{aligned} \tag{33.4}$$

The IMDs are located at frequencies above and below the two-tone input with frequency intervals equal to the separation of the two input carriers. Due to its simplicity and harshness (the envelope varies throughout the complete device dynamics), the two-tone test has become an almost universally accepted method for evaluating linearity [11–14].

33.2.2 Memoryless Systems versus Systems with Memory

The concept of memory in a system refers to the degree to which the current system output is influenced by the history of the system input. In a *memoryless system*, the output, $y(t_1)$, is solely dependent on the instantaneous input, $x(t_1)$. In a system with memory, the output is dependent not only on the input value at $t = t_1$ but also on past values of the input, $x\,(t<t_1)$. A system with memory is also sometimes referred to as a *dynamic system*.

Memory occurs because the system cannot dissipate its energy instantaneously. In electronic circuits, memory is associated with electric charge storage, magnetic flux storage, and delay elements [15–18]. Any energy-storing elements such as capacitors, inductors, or masses with thermal energy, therefore, add memory to the system. As an example, look at the simple current equation for a linear capacitor

$$i(t) = \lim_{\Delta t \to 0} \frac{q(t) - q(t-\Delta t)}{\Delta t} = C \lim_{\Delta t \to 0} \frac{v(t) - v(t-\Delta t)}{\Delta t} = C\frac{\mathrm{d}v(t)}{\mathrm{d}t}. \tag{33.5}$$

Here, the output current, $i(t)$, depends not only on the applied voltage $v(t)$ but also on past voltages (the accumulated capacitor charge).

The consequence of memory is that the system output response is no longer instantaneous. Instead, the output spreads over time and is given as the convolution product between system input and system impulse response. Figure 33.4 illustrates this. The step response of a memoryless system (here, a simple voltage division network consisting of two linear resistors) is attenuated, but in shape, an exact copy of the input signal. The output waveform of a system with memory (the same resistor network but now with a shunt capacitor added to the output) is clearly modified by the energy-storing element.

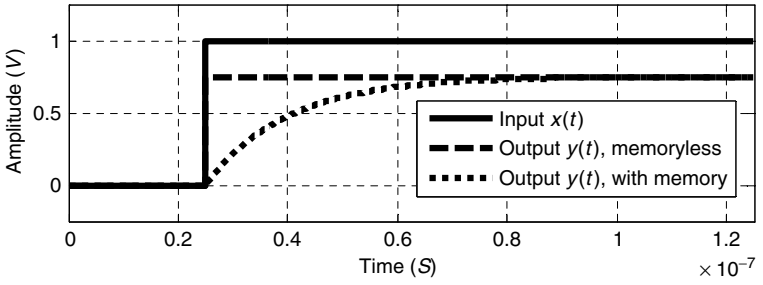


FIGURE 33.4 Simulated step responses of a system with and without memory. The memoryless system is a simple voltage division network made of linear resistors. The system with memory is the same resistor network but with the addition of a shunt capacitor at the output. The capacitor spreads the system response over time.

33.3 Memory Effects in Microwave PAs

In the literature, PA memory effects have traditionally been divided into two categories: Short-term memory effects and long-term memory effects [15–18]. The classification refers to the memory effect time constants. The first category has time constants in the range of the RF period whereas the second category has much longer time constants in the range of the period of the envelope modulation.

33.3.1 Short- and Long-Term Memory Effects

In general, a nonlinear memoryless system may cause only amplitude distortion and not phase distortion [19]. Such a system is therefore fully characterized by its AM/AM transfer function [20,21]. Any phase distortion is a sign of memory [22,23]. If a system has short-term memory effects, its time constants are much smaller than the reciprocal value of the maximum envelope frequency. At a certain time instant, the amount of amplitude and phase distortion may, therefore, be assumed dependent only on the input signal at that same time instant. Such a system may be characterized by its AM/AM and AM/PM transfer functions [20,21,24,25]. Long-term memory effects, often show as hysteresis in the AM/AM and AM/PM transfer functions (i.e., the system output is, no longer, a function of the instantaneous input only but also of the history of the input signal or, equivalently, of the modulation frequency of the input signal). A precise characterization of such a system requires more complex describing techniques [26,27].

Figure 33.5 shows typical locations of memory effects in bipolar and FET amplifiers [15]. As illustrated, LC time constants in the bias and matching networks, self heating in the active device, and feedback in the DC bias circuit are typical memory sources. Short-term memory effects are mainly caused by the LC networks [26,27–29], but also parasitics in the RF choke and the resonance frequency of the DC blocking capacitor [15] may contribute. Long-term memory effects are commonly attributed to the thermal time constants of the active device and to a long time constant in the DC bias circuit [16,23,15,29–32]. As an example, look at the simple BJT bias source in Figure 33.5, which forms a closed loop. If the collector current of the RF transistor Q_1 tends to increase, the voltage drop across R also increases and, consequently, the Q_2 base-to-emitter voltage decreases. This causes the, Q_2, collector current to decrease and, as a result, the base current of Q_1 decreases and holds the, Q_1, collector current at its original value. A similar self-adjustment occurs in the opposite direction [15]. Thermal memory effects occur because

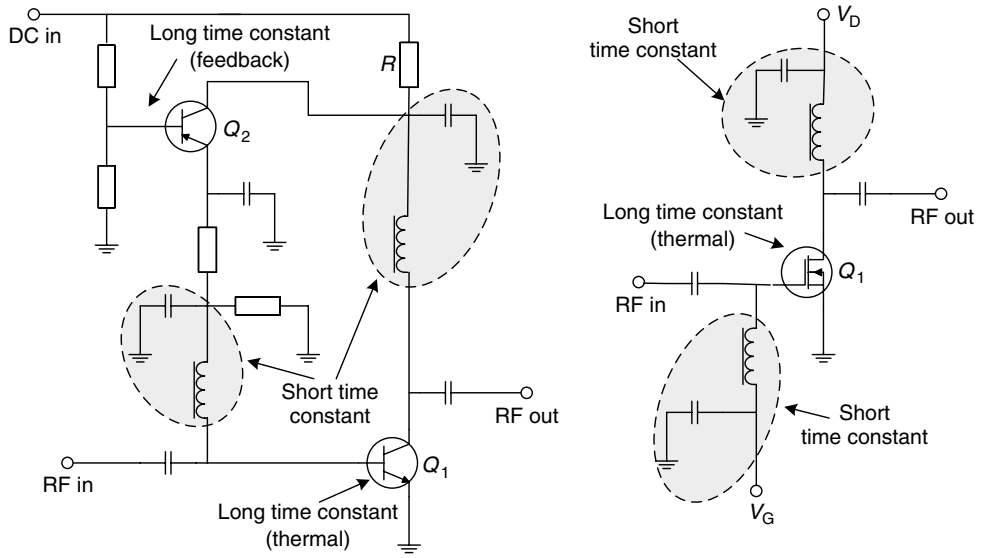


FIGURE 33.5 Typical memory locations in bias and matching networks of bipolar and FET amplifiers [15].

device characteristics are functions of temperature and, consequently, of power dissipated in the device [26–28]. Due to the finite mass of the active device and the surrounding package, the thermal impedance is not purely resistive but forms a distributed low-pass filter with a wide range of time constants [16,29].

Memory effects have been attempted reduced by design optimization or by adding compensation modules to an existing amplifier design. While memory effects due to the LC matching networks and memory effects related to the bias circuit may be reduced by careful circuit design [33], thermal memory effects are not so easily removed by design optimization [34–38]. In fact, the thermal memory effects may be reduced only by reducing the thermal impedance of the substrate, which requires unnecessarily large device geometries or the use of exotic materials [39–41]. Alternatively, certain designers prefer to measure the memory effects and, then, synthesize a dedicated compensation module. This was done by Kim and Konstantiou in Reference 42, where the compensation module was based on adding memory effects to a memoryless predistortion polynomial function. Also, Vuolevi et al. [43] has proposed a compensation scheme. This is based on envelope injection to reduce terminal impedance variations across the modulation frequency. In recent years, there has also been intensive research on memory effect compensation schemes using DSP techniques [44–46].

33.3.2 Memory Effects and the Two-Tone Test

A two-tone test signal with variable tone spacing has shown to be a valuable tool for identifying and understanding the underlying mechanisms responsible for generating memory effects [29,47–50]. From Equation 33.4, it can be seen that the IMD3 products of a memoryless system are independent of the tone spacing in Equation 33.2 and that the IMD3 amplitude increases exactly to the third power of the input amplitude. If the same two-tone signal is applied to a system with memory, amplitude and phase of the IMD3 products become functions of tone spacing (i.e., modulation frequency). As explained by Vuolevi and Rahkonen in References 29 and 47, this is the principal indication of memory effects.

In Reference 15, Bösch and Gatti proposed a two-tone measurement system, which uses two network analyzers to provide information about the intermodulation distortion products. However, this system does not provide information about how the memory effects affect the IMD products and, therefore, suffers from the practical ability to differentiate memory effects from (the much larger) fundamental signals. Youngoo et al. [51] proposed a system , which can measure the relative phase of the IMD3 products. Unfortunately, the accuracy of this system relies on a reference nonlinearity generator, which has shown to be difficult to realize in practice. Finally, Vuolevi and Rahkonen [29] proposed a setup capable of measuring both amplitude and phase of the IMD products. Although accurate, the main drawback of this approach is that it relies on a tedious and lengthy calibration procedure, which calls for many specialized instruments such as three signal generators, two vector network analyzers, two spectrum analyzers, and a number of down-converters.

33.4 Microwave PA Behavioral Models

In the majority of communication signals, the signal bandwidth is much smaller than the RF carrier frequency. In such (narrowband) signals, the carrier amplitude and carrier phase change slowly compared to the period of the carrier frequency. It, therefore, becomes more efficient to describe the amplitude and phase functions directly without inclusion of the (redundant) carrier information

$$x(t) = A(t)\cos[\omega_c t + \varphi(t)] = \mathrm{Re}\{\tilde{x}(t)\exp(j\omega_c t)\}, \tag{33.6}$$

where $\omega_c = 2\pi f_c$ is the angular carrier frequency and $\tilde{x}(t)$ is the complex envelope carrying the low-pass amplitude $A(t)$ and phase $\varphi(t)$ modulations

$$\tilde{x}(t) = A(t)\exp[j\varphi(t)]. \tag{33.7}$$

In the literature, we do find behavioral models, which operate on the complete RF signal and map $x(t)$ onto $y(t)$ [52]. However, the major part of published PA behavioral models are of the low-pass complex envelope equivalent type, in which $\tilde{x}(t)$ is mapped onto $\tilde{y}(t)$. All models discussed in the following are of the latter type. We divide the models into three categories, according to the presence of memory (Table 33.1): Memoryless models, models with short-term memory, and models with long-term memory.

TABLE 33.1 Classification of Microwave PA Behavioral Models

	PA Transfer Functions	Modeling Method
Memoryless system	AM/AM	Saleh model, power series
System with short-term memory effects	AM/AM, AM/PM	Saleh model, complex power series
System with long-term memory effects	Frequency dependent AM/AM, AM/PM	Volterra series, artificial neural network

33.4.1 Memoryless Models and Models with Short-Term Memory

Memoryless models and models with short-term memory are relatively simple. Typically, they rely on a narrowband representation of the PA where the AM/AM and AM/PM transfer functions are measured by sweeping the power of a single tone at the center frequency of the PA's pass-band [26,49]. Memoryless models describe the amplitude of the output envelope by a single algebraic function of instantaneous input envelope amplitude (Figure 33.6). Models with short-term memory describe the output complex envelope by two algebraic functions of instantaneous input envelope amplitude (Figure 33.7).

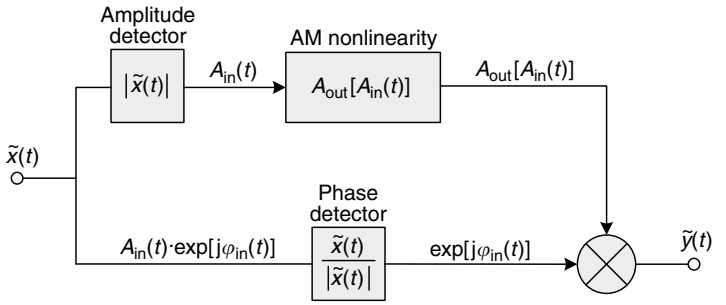


FIGURE 33.6 Block level representation of a memoryless behavioral model based on the AM/AM function $A_{out}[A_{in}(t)]$ [27].

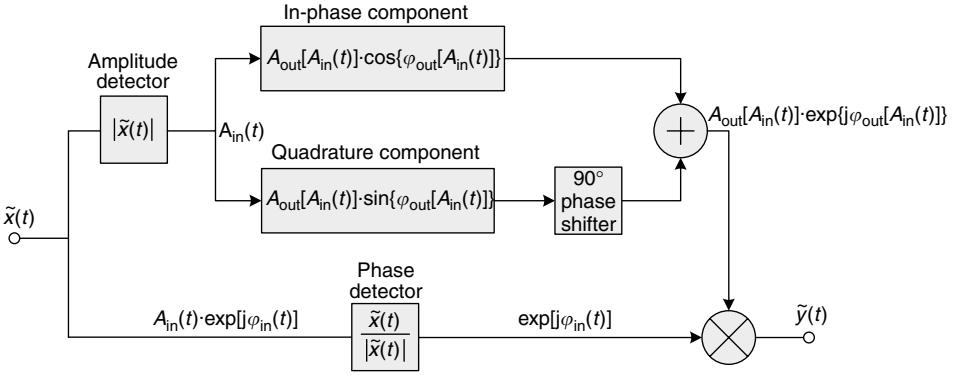


FIGURE 33.7 Block level representation of a behavioral model with short-term memory based on the AM/AM function $A_{out}[A_{in}(t)]$ and the AM/PM function $\varphi_{out}[A_{in}(t)]$ [27].

One commonly used model example is the Saleh model [53]

$$A_{\text{out}}[A_{\text{in}}(t)] = \frac{\alpha_A A_{\text{in}}(t)}{1 + \beta_A [A_{\text{in}}(t)]^2}$$
$$\varphi_{\text{out}}[A_{\text{in}}(t)] = \frac{\alpha_\varphi A_{\text{in}}(t)}{1 + \beta_\varphi [A_{\text{in}}(t)]^2} \tag{33.8}$$

For extraction of the model parameters, α_{A}, β_{A}, α_φ, and β_φ, Saleh provided equations derived from a mean square fitting criterion to measured AM/AM and AM/PM transfer functions. At the time of development, he demonstrated the model to provide a good mixture between accuracy and implementation complexity. For the modeling of modern microwave PAs, the simple model is less applicable.

Another widely used modeling approach is to approximate the PA AM/AM and AM/PM transfer functions by two power series [28,54]:

$$A_{\text{out}}[A_{\text{in}}(t)] = \sum_{i=1}^{I} c_i A_{\text{in}}^i(t),$$
$$\varphi_{\text{out}}[A_{\text{in}}(t)] = \sum_{i=1}^{I} d_i A_{\text{in}}^i(t). \tag{33.9}$$

where c_i and d_i are the model parameters. Significant value may be added to this type of model if the power series are driven by additional (independent) input variables [89–91]. As an example, we may include the effect of a PA load mismatch by adding load impedance magnitude M_{load} and load impedance phase angle P_{load} as inputs to $A_{\text{out}}(A_{\text{in}}, M_{\text{load}}, P_{\text{load}})$, $\varphi_{\text{out}}(A_{\text{in}}, M_{\text{load}}, P_{\text{load}})$. A straightforward approach is then to approximate the coefficients c_i and d_i as products of two new power series. c_i may, for example, be represented by

$$c_i = \sum_{k=0}^{l-1} \sum_{p=0}^{m-1} g_{i,k,p} M_{\text{load}}^k P_{\text{load}}^p \tag{33.10}$$

And the complete AM/AM distortion model then becomes

$$A_{\text{out}}(A_{\text{in}}, M_{\text{load}}, P_{\text{load}}) = \sum_{i=1}^{n} \sum_{k=0}^{l-1} \sum_{p=0}^{m-1} g_{i,k,p} M_{\text{load}}^k P_{\text{load}}^p A_{\text{in}}^i, \tag{33.11}$$

where A_{in} is assumed to vary as a function of time.

Figures 33.8 and 33.9 illustrate the significance of a load variation in a commercially available 3G-PP W-CDMA PA (RF3137 by RFMD [55,56]). Both 1 dB compression point and in-band transmit power are strong functions of the load impedance. A power series model much like the one described above, may capture this dependency very accurately. We extract the power series coefficients from a simple least mean squares (LMSs) fit to the measurements (pulsed large signal *s*-parameters [57]). The model order is chosen to 5 for the power series approximating the response to the input amplitude and to 4 for the two power series approximating the response to the load impedance. The choice of model orders reflects the fact that, for this particular PA, the effect of a change in load impedance is less pronounced than the effect of a change in input amplitude. Same orders were chosen for both the AM/AM and AM/PM models. This results in a total of 54 power series terms (after pruning out coefficients with low sensitivity) and a fitting error smaller than 5% at all measured data points. Fitted and measured transfer functions are plotted in Figure 33.10 for the nominal $50 + 0j\ \Omega$ load. Figures 33.11 and 33.12 verify the accuracy of the model in terms of measured and simulated 3G-PP W-CDMA ACPRs.

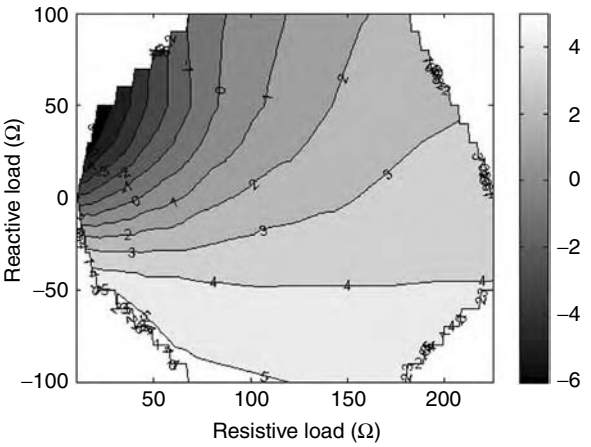


FIGURE 33.8 Measured PA 1 dB compression point (input referred, in dBm) as a function of complex load impedance.

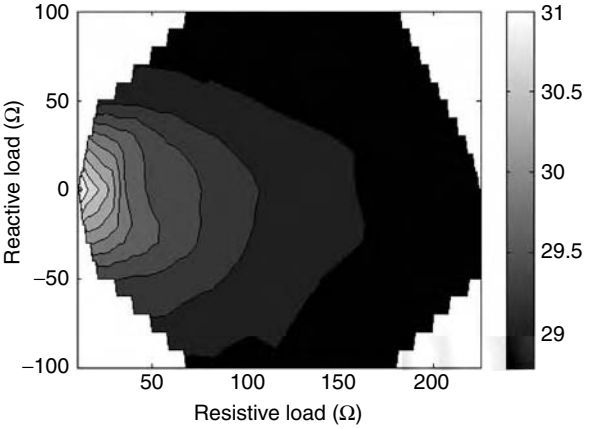


FIGURE 33.9 Measured PA in-band transmit power (in dBm) as a function of complex load impedance.

33.4.2 Models with Long-Term Memory

Models with long-term memory are generally more complex than memoryless models and models with short-term memory. Their output response is not only a function of input envelope amplitude but also of modulation frequency or, put equivalently, of input envelope amplitude history (Section 33.3). Models with long-term memory are, typically, based on either Volterra series [59] or artificial neural networks (ANNs) [58]. Volterra series are best suited for weakly nonlinear systems [47,60–65], whereas ANNs are more versatile and may be utilized also for highly nonlinear systems [66–69].

33.4.2.1 Two- and Three-Box Models

The Volterra series is a powerful modeling tool. Nevertheless, extraction of the Volterra kernels may be a difficult and very time consuming task. Reductions of the general Volterra series representation have, therefore, been proposed [70–72]. These rely on cascades of linear time-invariant (LTI) filters and memoryless nonlinearities (Figure 33.13).

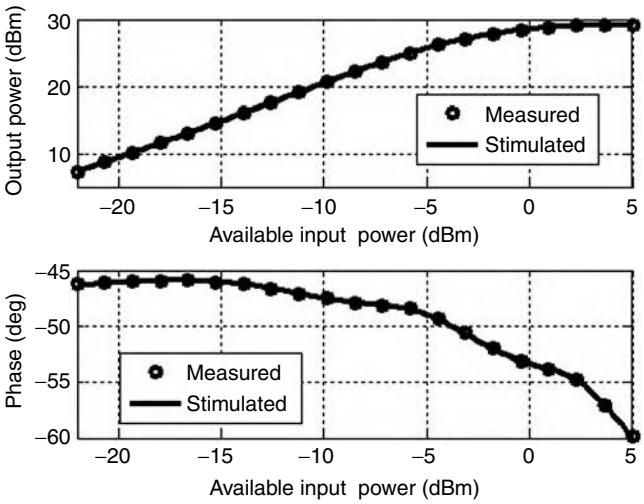


FIGURE 33.10 Measured and simulated PA AM/AM and AM/PM transfer functions with the nominal $50 + 0j\ \Omega$ load.

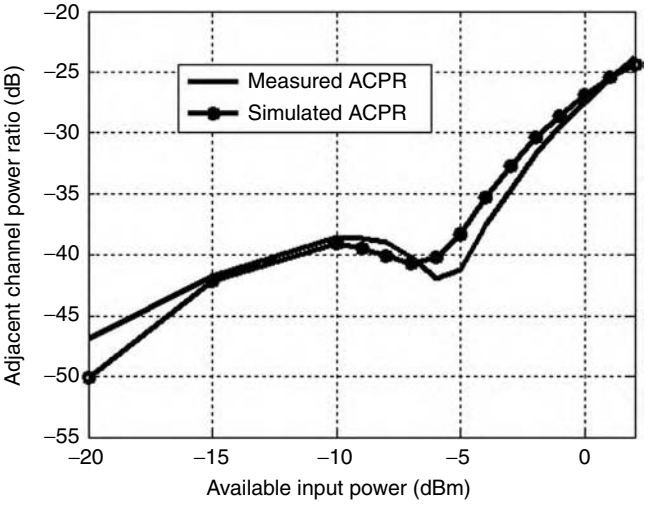


FIGURE 33.11 Measured and simulated PA ACPR as functions of available input power with the nominal $50 + 0j\ \Omega$ load.

If the LTI filter precedes the nonlinearity, the structure is called a Wiener model. Let us assume the input signal is the low-pass equivalent $\tilde{x}(t)$ containing Q sinusoidal tones with amplitudes A_q

$$\tilde{x}(t) = \sum_{q=1}^{Q} A_q \exp(j\omega_q t). \tag{33.12}$$

The output of the LTI filter is then given by

$$\tilde{s}(t) = \sum_{q=1}^{Q} A_q H(\omega_q) \exp(j\omega_q t), \tag{33.13}$$

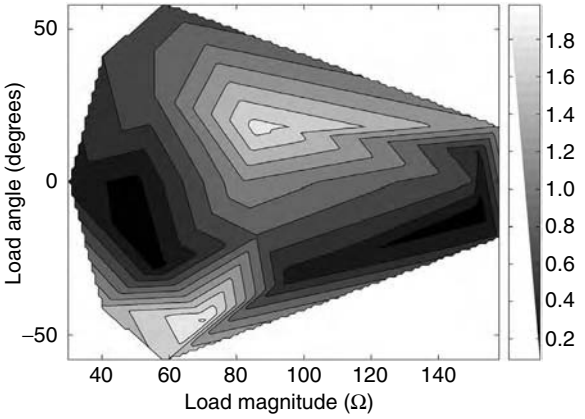


FIGURE 33.12 Deviation (in dB) between measured and simulated ACPR at the nominal 0 dBm PA drive level and across the measured load grid.

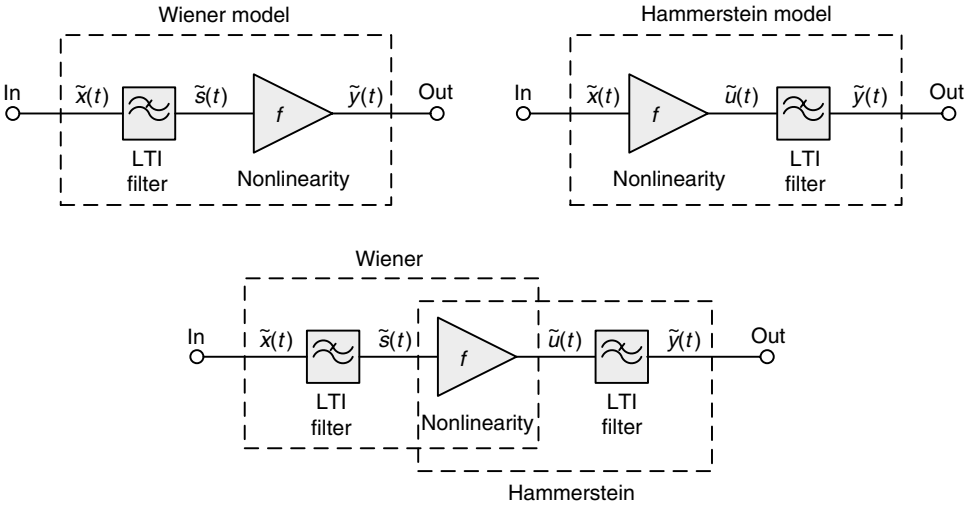


FIGURE 33.13 Block level representations of two- and three-box models. Depending on the cascade configuration of LTI filters and memoryless nonlinearities, these structures are also known as Wiener, Hammerstein, and Wiener–Hammerstein models [73].

where $H(\omega)$ is the transfer function of the LTI filter. Let us now approximate the memoryless nonlinearity by an Nth-order power series (Section 33.4.1)

$$f[\tilde{s}(t)] = \sum_{n=1}^{N} a_n \tilde{s}^n(t), \tag{33.14}$$

where a_n are the power series coefficients. The output of the Wiener models, then, becomes

$$\tilde{y}(t) = \sum_{n=1}^{N} a_n \left[\sum_{q=1}^{Q} A_q H(\omega_q) \exp(j\omega_q t) \right]^n. \tag{33.15}$$

Equation 33.15 is a reduction of the general Volterra series representation. This may be seen by expanding Equation 33.15

$$\tilde{y}(t) = \sum_{n=1}^{N} a_n \sum_{q1=1}^{Q} \sum_{q2=1}^{Q} \cdots \sum_{qn=1}^{Q} A_{q1} A_{q2} \cdots A_{qn} H(\omega_{q1}) \cdot \times H(\omega_{q2}) \cdots H(\omega_{qn}) \cdot \exp[j(\omega_{q1} + \omega_{q2} + \cdots + \omega_{qn})t], \tag{33.16}$$

where it becomes evident that nonlinearity and memory are described separately as the product of linear transfer functions $a_{\mathrm{n}} H(\omega_{\mathrm{q1}}) H(\omega_{\mathrm{q2}}) \ldots H(\omega_{\mathrm{qn}})$. The general Volterra series representation combines nonlinearity and memory in one describing function $H_{\mathrm{n}}(\omega_{\mathrm{q1}}, \omega_{\mathrm{q2}}, \ldots, \omega_{\mathrm{qn}})$ [13]. Thus, the power series representation is a special case of the Volterra series in which

$$H_n(\omega_{\mathrm{q1}}, \omega_{\mathrm{q2}}, \ldots, \omega_{\mathrm{qn}}) = a_n H(\omega_{\mathrm{q1}}) H(\omega_{\mathrm{q2}}) \cdots H(\omega_{\mathrm{qn}}). \tag{33.17}$$

To demonstrate the Wiener model functionality, let us place a pole in front of a simple, third-order power series nonlinearity (Figure 33.14). The pole is located at $f_{\mathrm{p}} = 1$ MHz, and simulated composite amplitude responses are shown in Figure 33.15. If the modulation frequency equals f_{p}, the pole attenuates the input envelope amplitude by exactly 3 dB. As the filter precedes the nonlinearity, the composite AM/AM transfer function is shifted the same 3 dB in input envelope amplitude (Figure 33.15, dashed curve trace). The AM/PM transfer function, which is not displayed here, shifts in a similar manner [73]. Thus, the Wiener model is applicable whenever variations in ω_{m} causes the memory in the PA to produce AM/AM and AM/PM transfer functions, which are similar in shape but shifted with respect to input envelope amplitude by an amount of $|H(\omega_{\mathrm{m}})|$.

Rearranging the structure and placing the nonlinearity in front of the pole gives us a Hammerstein model. As the filter is now placed after the nonlinearity, this structure can model a $|H(\omega_{\mathrm{m}})|$ vertical shift of the AM/AM transfer function (Figure 33.16) and a $\angle H(\omega_{\mathrm{m}})$ vertical shift of the AM/PM transfer function [73].

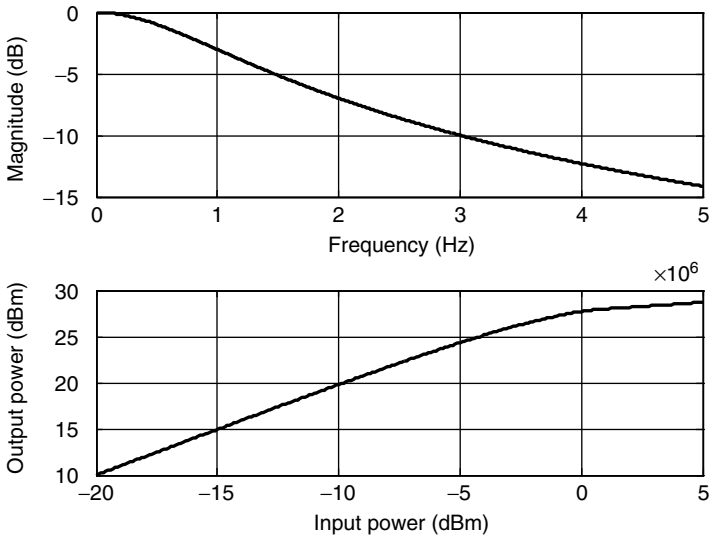


FIGURE 33.14 *Upper:* Amplitude response of a single pole placed at $f_{\mathrm{p}} = 1$ MHz. *Lower:* AM/AM transfer function of third-order power series nonlinearity.

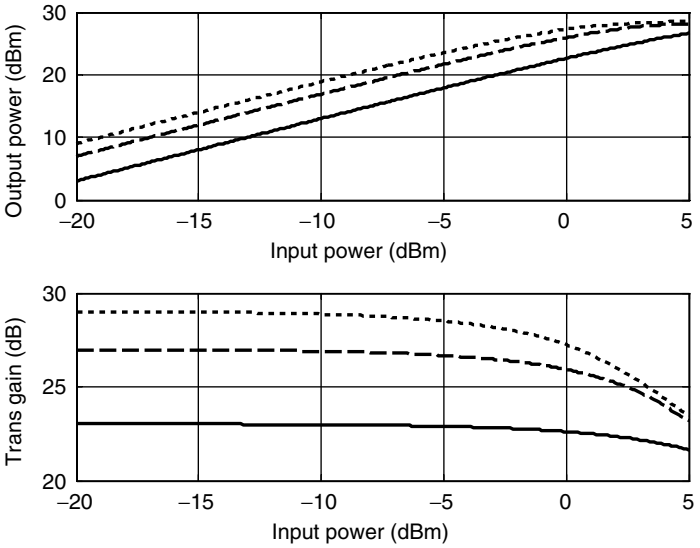


FIGURE 33.15 Simulated amplitude and gain responses of the Wiener model at modulation frequencies $\omega_m = \omega_p/2$ (dotted curve), $\omega_m = \omega_p$ (dashed curve), and $\omega_m = 2\omega_p$ (solid curve).

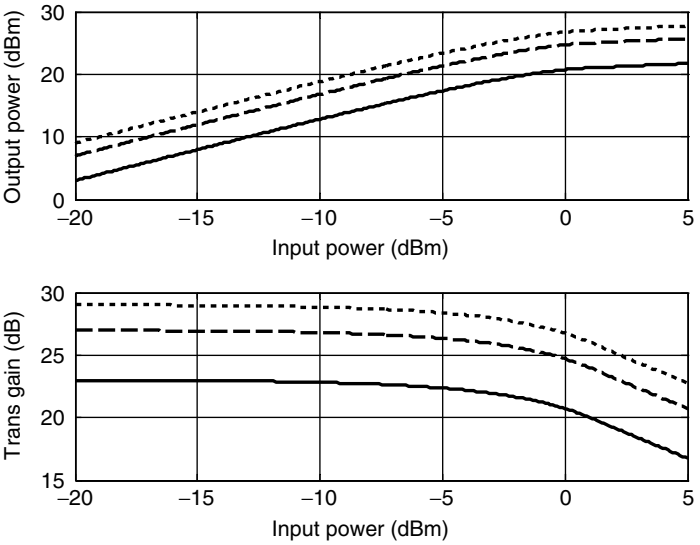


FIGURE 33.16 Simulated amplitude and gain responses of the Hammerstein model at $\omega_m = \omega_p/2$ (dotted curve), $\omega_m = \omega_p$ (dashed curve), and $\omega_m = 2\omega_p$ (solid curve).

If the nonlinear portion of a Wiener model is merged with the nonlinear portion of a Hammerstein model, the more versatile Wiener–Hammerstein model appears. This structure can model both horizontal and vertical shifts in the AM/AM and AM/PM transfer functions [73].

33.4.3 Artificial Neural Network Models

Artificial neural networks are information processing systems with designs inspired by studies of the human brain. They have, like the human brain, the ability to learn from observations and to generalize by

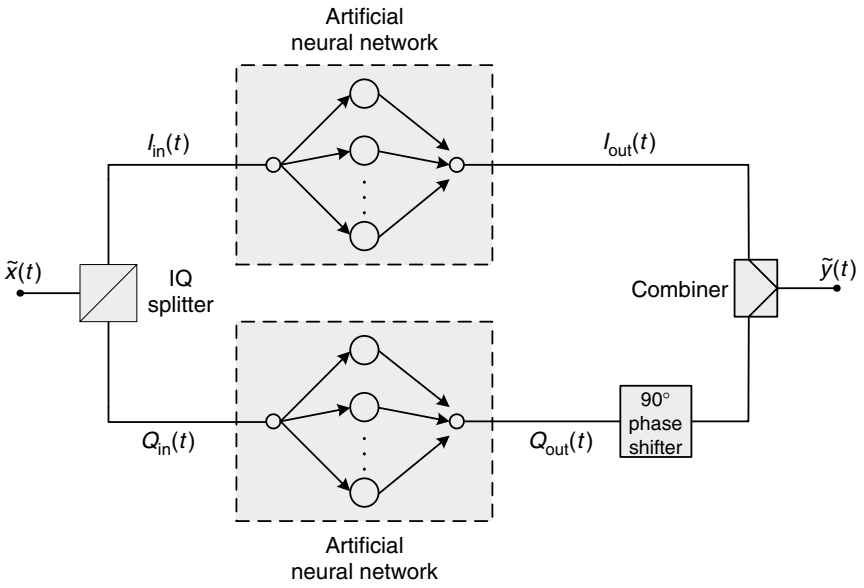


FIGURE 33.17 Block level representation of a Cartesian-format ANN-based PA behavioral model.

abstraction [74]. The *Universal Approximation Theorem* [75] proves that ANNs may be trained to approximate any degree of nonlinearity to any desired level of accuracy, provided a deterministic relationship between input and target exists. This makes the ANN a very attractive tool in the modeling of nonlinear device characteristics.

The conventional procedure in ANN modeling is to first convert the measured (or simulated) input–output signals into either Cartesian format (in-phase and quadrature components) or polar format (amplitude and phase components). Then, two real-valued ANNs are constructed to approximate the input–output relationship (demonstrated with Cartesian components in Figure 33.17).

In the literature, many different ANN topologies have been proposed for the modeling of microwave PAs [76–82]. Most of them rely on a multi-layer perceptron (MLP) structure (Figure 33.18). The MLP consists, as all neural networks do, of two basic components: The processing elements and the interconnections between them. The interconnections weigh inputs to the processing elements by weights W. The processing elements (typically referred to as neurons) consist of a summation node and a transfer function f. The summation node may add a bias b to the transfer function input. The transfer function may be linear or nonlinear. The neurons in a neural network are typically grouped into layers. The first and the last layers are called input and output layers, respectively. The remaining layers are called hidden layers. Key to the design of a neural network for function approximation is the number of neurons in the input and hidden layers. This determines the accuracy of the approximation and the total number of network weights and biases, which need to be optimized during the training procedure [83].

In its basic configuration of Figure 33.18, the MLP does not contain memory. The most straightforward approach to include memory is to add a tapped delay line (TDL) at the input layer (Figure 33.19). Such topologies have recently been utilized successfully in dynamic behavioral modeling of microwave PAs [84–88]. Note that we in Figure 33.19 have adopted the discrete time domain instead of the continuous one. We assume that time is a succession of uniform time samples of a convenient sampling period, which apply to all signals. Thus,

$$I_{in}(t) \Rightarrow I_{in}(n) \quad \text{and} \quad I_{out}(t) \Rightarrow I_{out}(n). \tag{33.18}$$

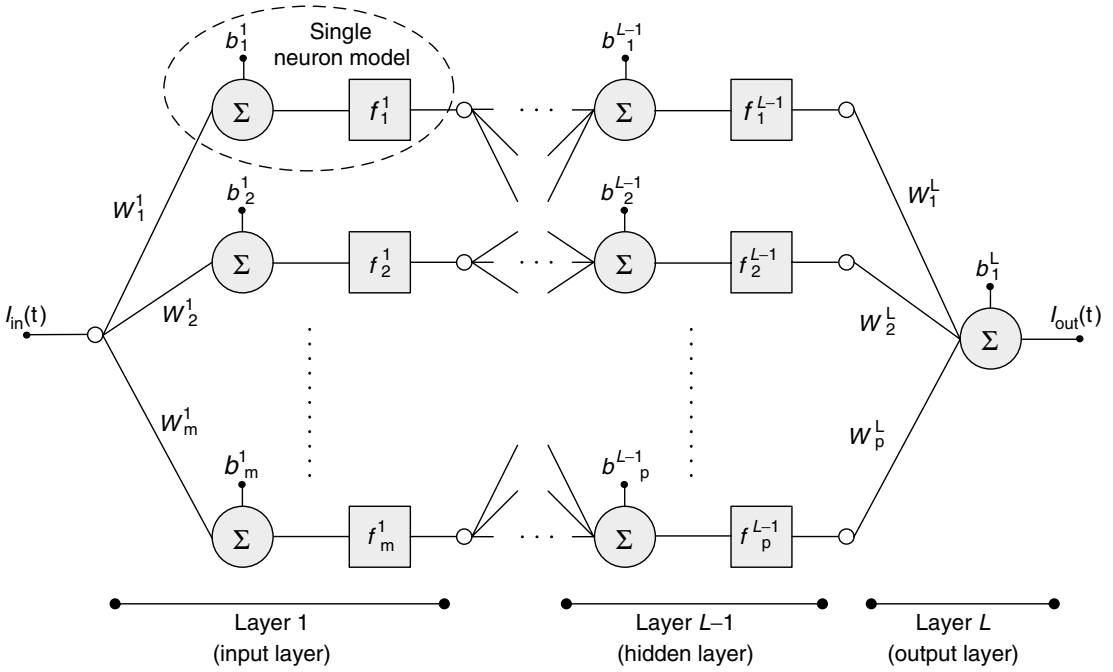


FIGURE 33.18 General MLP structure with input $I_{in}(t)$ and output $I_{out}(t)$. The MLP has a total of L layers. The first layer is the input layer, the L^{th} layer is the output layer, and layers 2 to $L-1$ are hidden layers. There are m neurons in the input layer, p neurons in layer $L-1$, and a single neuron in the output layer.

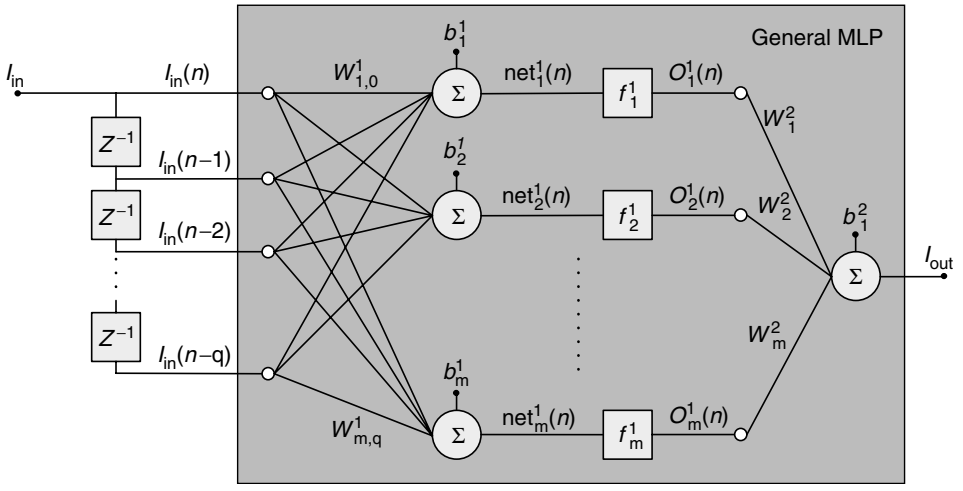


FIGURE 33.19 General time delayed neural network with input I_{in}, output I_{out}, and a tapped delay line (TDL) of length q at the input.

With addition of the input TDL, the network entries not only include the current value of the input signal, but also previous values. Assuming a TDL of length q, the baseband output I_{out} is in Figure 33.19 at time instant n a function of $I_{in}(n)$ and its q past values

$$I_{out}(n) = f_{MLP}[I_{in}(n), I_{in}(n-1), \ldots, I_{in}(n-q)] \tag{33.19}$$

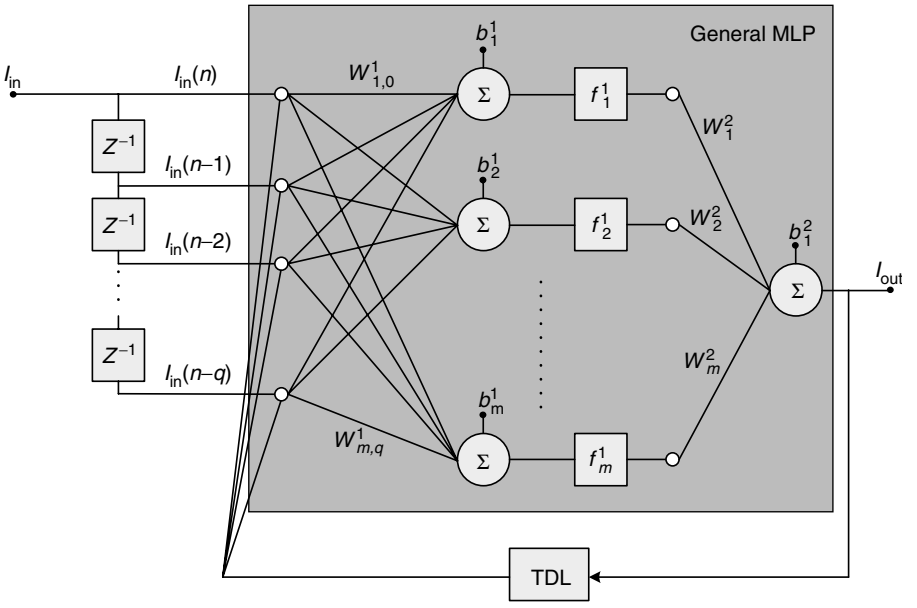


FIGURE 33.20 General recurrent neural network with input I_{in}, output I_{out}, a TDL of length q at the input layer, and a TDL in the feedback path from network output to network input.

where, after substituting f_{MLP} with its transfer function, we get

$$I_{out}(n) = \sum_{k=1}^{m} w_k^2 O_k^1(n) + b_1^2 \tag{33.20}$$

where

$$O_k^1(n) = f_k^1\left[\text{net}_k^1(n)\right] \tag{33.21}$$

and

$$\text{net}_k^1(n) = \sum_{i=0}^{q} w_{k,i}^1 I_{in}(n-i) + b_k^1 \tag{33.22}$$

The ability of the neural network to model nonlinear dynamics may be further enhanced by adding feedback connections from output to input (Figure 33.20) [87]. With a TDL of length p in the feedback path, the baseband output I_{out} becomes a function of the previous p output samples and q input samples

$$I_{out}(n) = f_{MLP}[I_{out}(n-1), I_{out}(n-2), \ldots, I_{out}(n-p), I_{in}(n), I_{in}(n-1), \ldots, I_{in}(n-q)] \tag{33.23}$$

Figure 33.21 shows dynamically measured AM/AM and AM/PM transfer functions of an RFMD proprietary GSM–EDGE PA [88]. A true GSM–EDGE signal was applied at the input of the PA and the distorted output signal was down-converted and demodulated. As illustrated, the measured transfer functions are no longer smooth curves as those we generally see for memoryless PAs. Especially the AM/PM transfer function show significant memory effects (Section 33.3). A neural network structure much like the one described above may accurately model the memory effects (Figures 33.22 and 33.23).

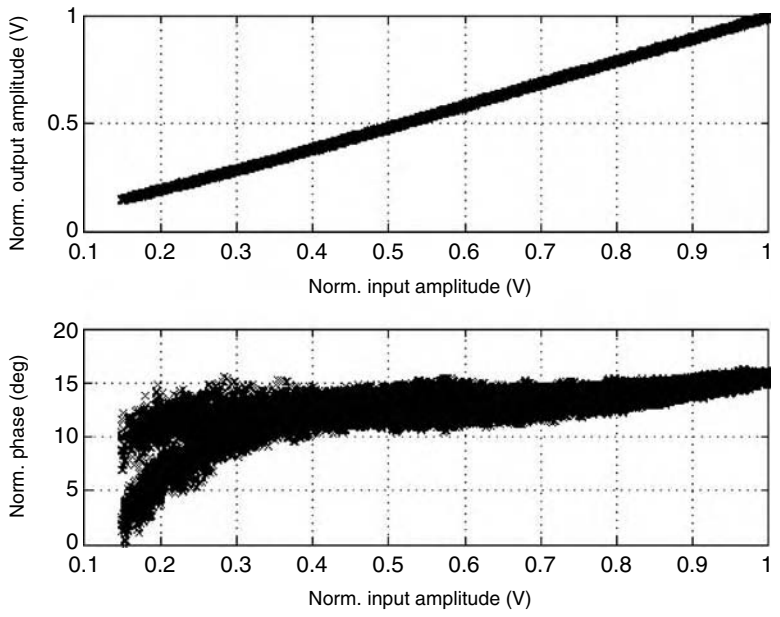


FIGURE 33.21 Dynamically measured AM/AM and AM/PM transfer functions of an RFMD proprietary GSM–EDGE PA [88]. Especially the measured AM/PM transfer function show significant memory effects.

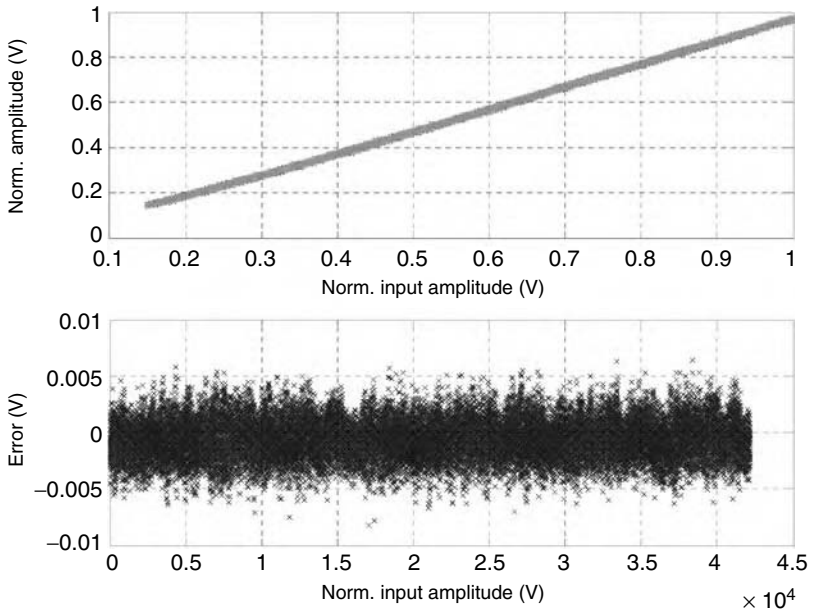


FIGURE 33.22 A time-delay neural network with feedback connections from output to input accurately models the measured AM/AM transfer function.

We use a two-layer model with 10 neurons in the input layer and a single neuron in the output layer. Two identical ANNs are used for in-phase and quadrature signal paths. The network weights and biases are found from a Levenberg–Marquardt training algorithm [58]. Viewed in time domain, the model output accurately tracks the in-phase and quadrature output trajectories (Figure 33.24).

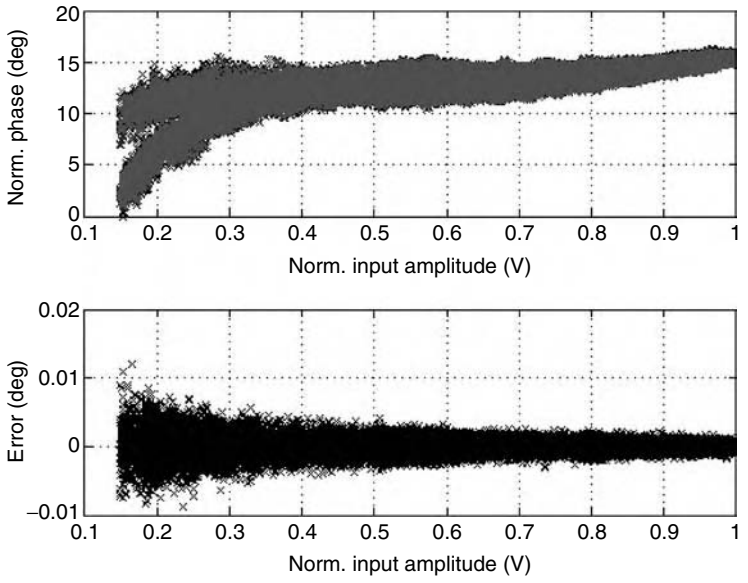


FIGURE 33.23 A time-delay neural network with feedback connections from output to input accurately models the measured AM/PM transfer function.

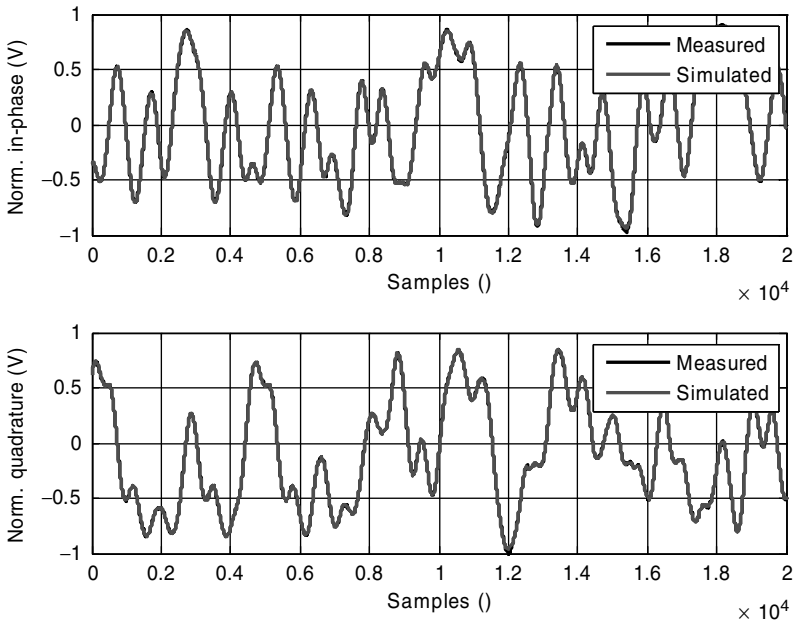


FIGURE 33.24 Time domain measured and simulated output waveforms. The neural network model tracks very accurately both in-phase and quadrature output trajectories.

33.5 Summary

This section has presented an overview of various behavioral modeling approaches for microwave PAs. From the review of a few basic results from system identification theory, PAs were categorized into three groups: Memoryless PAs, PAs with short-term memory, and PAs with long-term memory. The Saleh

model and the power series model were highlighted as good candidates for the two first PA categories. More complex models, based on Volterra series representations and ANNs, were in focus for the latter category. Two models, one based on a power series and another based on an artificial neural network, were successfully demonstrated on real PAs. Lastly, a comprehensive reference list has been reported to the reader.

References

1. S. Mann, M. Beach, P. Warr, and J. McGeehan, "Increasing the Talk-Time of Mobile Radios with Efficient Linear Transmitter Architectures," *Electronics and Communication Engineering Journal*, vol. 13, no. 2, 2001, pp. 65–76.
2. A. Katz, "Linearization: Reducing Distortion in Power Amplifiers," *IEEE Microwave Magazine*, vol. 2, no. 4, 2001, pp. 37–49.
3. F. H. Raab, P. Asbeck, S. Cripps, P. B. Kenington, Z. B. Popovic, N. Pothecary, J. F. Sevic, and O. N. Sokal, "Power Amplifier and Transmitters for RF and Microwave," *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, no. 3, 2002, pp. 814–826.
4. M. Schroter and L. Tzung-Yin, "Physics-Based Minority Charge and Transit Time Modeling for Bipolar Transistors," *IEEE Transactions on Electron Devices*, vol. 46, no. 2, 1999, pp. 288–300.
5. J. J. Liou and J. S. Yuan, "Physics-Based Large-Signal Hetrojunction Bipolar Transistor Model for Circuit Simulation," *IEEE Circuits, Devices and Systems*, vol. 138, no. 1, 1991, pp. 97–103.
6. C. M. Snowden, "Large-Signal Microwave Characterization of AlGaAs/GaAs HBT's Based on a Physical-Based Electrothermal Model," *IEEE Transactions on Microwave Theory and Techniques*, vol. 45, no. 1, 1997, pp. 58–71.
7. P. C. Grossman and J. Choma, "Large Signal Modeling of HBT's Including Self-Heating and Transient Time Effects," *IEEE Transactions on Microwave Theory and Techniques*, vol. 40, no. 3, 1992, pp. 449–464.
8. S. J. Spiegel, R. A. Hamm, A. Feygenson, and P. R. Smith," Extraction of the InP/GaInAs Hetrojunction Bipolar Transistor Small-Signal Equivalent Circuit," *IEEE Transactions on Electron Devices*, vol. 42, no. 6, 1995, pp. 1059–1064.
9. A. Samelis, "Modeling the Bias of the Dependence of the Base-Collector Capacitance of Power Hetrojunction Bipolar Transistors," *IEEE Transactions on Microwave Theory and Techniques*, vol. 47, no. 5, 1999, pp. 642–645.
10. S. Lee and A. Gopinath, "Parameter Extraction Technique for HBT Equivalent Circuit Using Cutoff Model Measurements," *IEEE Transactions on Microwave Theory and Techniques*, vol. 40, no. 3, 1992, pp. 574–577.
11. B. Razavi, *RF Microelectronics*. Prentice Hall, Upper Saddle River, NJ, 1998.
12. P. B. Kenington, *High-Linearity RF Amplifier Design*. Norwood, MA: Artech House, 2000.
13. S. A. Mass, *Nonlinear Microwave and RF Circuits*. Norwood, MA: Artech House, 2003.
14. C. J. Clark, C. P. Silva, A. A. Moulthrop, and M. S. Muha, "Power-Amplifier Characterization using a Two-Tone Measurement Technique," *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, no. 6, 2002, pp. 1590–1602.
15. W. Bösch and G. Gatti, "Measurement and Simulation of Memory Effects in Predistortion Linearizers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 37, no. 12, 1989, pp. 1885–1890.
16. J. Vuolevi and T. Rahkonen, *Distortion in RF Power Amplifiers*. Norwood, MA: Artech House, 2003.
17. J. S. Kenney, W. Woo, M. D. McKinley, H. Ku, and Y. C. Park, "Practical Limitations of Predistortion Linearization Systems for RF Power Amplifiers," *Proceedings of Asia-Pacific Microwave Conference*, Nov. 2003, FE1-1.
18. A. Rabany, L. Nguyen, and D. Rice, "Memory Effect Reduction for LDMOS Bias Circuits," *Microwave Journal*, vol. 46, no. 2, 2003, pp. 124–130.
19. J. S. Bendat, *Nonlinear System Techniques and Applications*. New York: Wiley, 1998.
20. P. Viszmuller, *RF Design Guide, Systems, Circuits, and Equations*. Norwood, MA: Artech House, 1995.
21. T. R. Turlington, *Behavioral Modeling of Nonlinear RF and Microwave Devices*. Boston, MA: Artech House, 2000.
22. J. Minkoff, "Intermodulation Noise in Solid-State Power Amplifiers for Wideband Signal Transmission," Nineth Communications Satellite Systems Conference, San Diego, CA, Mar. 7–11, 1982,

Collection of Technical Papers (A82-23476 09-32), New York, American Institute of Aeronautics and Astronautics, 1982, pp. 304–313.
23. J. C. Pedro and N. B. Carvalho, *Intermodulation Distortion in Microwave and Wireless Circuits*. Norwood, MA: Artech House, 2003.
24. O. Shimbo, "Effects of Intermodulation, AM-PM Conversion, and Additive Noise in Multicarrier TWT Systems," *Proceedings of IEEE*, vol. 59, Feb. 1971, pp. 230–238.
25. J. Minkoff, "The Role of AM-to-PM Conversion in Memoryless Nonlinear Systems," *IEEE Transactions on Communications*, vol. 33, no. 2, 1985, pp. 139–144.
26. M. Isaksson, D. Wisell, and D. Rönnow, "A Comparative Analysis of Behavioral Models for RF Power Amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 54, no. 1, 2006, pp. 348–359.
27. J. C. Pedro and S. A. Mass, "A Comparative Overview of Microwave and Wireless Power-Amplifier Behavioral Modeling Approaches," *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, no. 4, 2005, pp. 1150–1163.
28. H. Ku, M. D. McKinley, and J. S. Kenney, "Quantifying Memory Effects in RF Power Amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, no. 12, 2002, pp. 2843–2849.
29. J. H. K. Vuolevi and T. Rahkonen, "Measurement Technique for Characterizing Memory Effects in RF Power Amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 49, no. 8, 2001, pp. 1383–1389.
30. J. C. Pedro, N. B. Carvalho, and P. M. Lavrador, "Modeling Nonlinear Behavior of Band-Pass Memoryless and Dynamic Systems," *IEEE MTT-S International Microwave Symposium Digest*, vol. 3, June 2003, pp. 2133–2136.
31. P. Asbeck, H. Kobayashi, M. Iwamoto, G. Hanington, S. Nam, and L. E. Larson, "Argumented Behavioral Characterization for Modeling the Nonlinear Response of Power Amplifiers," *IEEE International Microwave Symposium Digest*, vol. 1, June 2002, pp. 135–138.
32. P. Draxler, I. Langmore, T. P. Hung, and P. M. Asbeck, "Time Domain Characterization of Power Amplifiers with Memory," *IEEE International Microwave Symposium Digest*, vol. 2, June 2003, pp. 803–806.
33. A. Rabany, L. Nguyen, and D. Rice, "Memory Effect Reduction for LDMOS Bias Circuits," *Microwave Journal*, vol. 46, no. 2, 2003, pp. 124–130.
34. E. Schurack, W. Rupp, T. Latzel, and A. Gottwald, "Analysis and Measurement of Nonlinear Effects in Power Amplifiers caused by Thermal Power Feedback," *IEEE International Symposium on Circuits and Systems*, vol. 2, 1992, pp. 758–761.
35. Y. Zhu, J. K. Twynam, M. Yagura, M. Hasegawa, T. Hasegawa, Y. Eguchi, A. Yamada, E. Suematsu, K. Sakuno, H. Sato, and N. Hashizume, "Analytical Model for Electrical and Thermal Transients of Self-Heating Semiconductor devices," *IEEE Transactions on Microwave Theory and Techniques*, vol. 46, 1996, pp. 1637–1641.
36. T. Hopkins and R. Tiziani, "Transient Thermal Impedance Considerations in Power Semiconductor Applications," *Automotive Power Electronics*, vol. 1, 1989, pp. 89–97.
37. D. T. Zweidinger, S. G. Lee, and R. M. Fox, "Compact Modeling of BJT Self-Heating in SPICE," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 12, Issue 9, 1983, pp. 1368–1375.
38. N. Le Gallou, J. M. Nebus, E. Ngoya, and H. Buret, "Analysis of Low Frequency Memory and Influence on Solid State HPA Intermodulation Characteristics," *IEEE International Microwave Symposium*, vol. 2, 2002, pp. 979–982.
39. S. Boumaiza and F. M. Ghannouchi, "Thermal Memory Effects Modeling and Compensation in RF Power Amplifiers and Predistortion Linearizers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, no. 12, 2003, pp. 2427–2433.
40. S. Boumaiza, J. Gauthier, and F. M. Ghannouchi, "Dynamic Electro-Thermal Behavioral Model for RF Power Amplifiers," *IEEE MTT-S International Microwave Symposium Digest*, vol. 1, June 2003, pp. 351–354.
41. S. Boumaiza and F. M. Ghannouchi, "Dynamic Nonlinear Distortion Characterization of Wireless Radio Transmitters," *Electron Devices for Microwave and Optoelectronic Applications 2004*, 2004, pp. 92–95.
42. J. Kim, K. Konstantiou, "Digital Predistortion of Wideband Signals Based on Power Amplifier Model with Memory," *Electron. Lett. Symp.*, vol. 37, no. 23, 2001, pp. 1417–1418.

43. J. Vuolevi, J. Manninen, and T. Rahkonen, "Cancelling the Memory Effects in RF Power Amplifiers," *IEEE International Symposium on Circuits and Systems*, vol. 1, May 2001, pp. 50–60.
44. W. Shan, L. Sundström, and B. Shi, "Spectral Sensitivity of Predistortion Linearizer Architectures to Filter Ripple," *Proceedings of IEEE Vehicular Technology Conference*, vol. 3, Oct. 2001, pp. 1570–1574.
45. C. Eun and E. J. Powers, "A New Volterra Predistorter Based on the Indirect Learning Architecture," *IEEE Transactions on Signal Processing*, vol. 45, no. 1, 1997, pp. 223–227.
46. L. Ding, G. T. Zhou, D. R. Morgan, Z. Ma, J. S. Kenney, J. Kim, and C. R. Giardina, "Memory Polynomial Predistortion Based on the Indirect Learning Architecture," *IEEE Global Telecommunication Conference*, vol. 1, Nov. 2002, pp. 967–971.
47. J. Vuolevi and T. Rahkonen, "Analysis of Third-Order Intermodulation Distortion in Common-Emitter BJT and HBT Amplifiers," *IEEE Transactions on Circuits and Systems II*, vol. 50, Dec. 2003, pp. 994–1001.
48. S. C. Cripps, *Advanced Techniques in RF Power Amplifier Design*, Norwood, MA: Artech House, 2002.
49. N. B. Carvalho and J. C. Pedro, "A Comprehensive Explanation of Distortion Sidebands Asymmetries," *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, no. 9, 2002, pp. 2090–2101.
50. D. J. Williams, J. Leckey, and P. J. Tasker, "A Study of the Effect of Envelope Impedance on Intermodulation Asymmetry Using a Two-Tone Time Domain Measurement System," *IEEE MTT-S International Microwave Symposium Digest*, June 2002, pp. 1841–1844.
51. Y. Youngoo, Y. Jeahyok, N. Joongjin, K. Bumman, and P. Myungkyu, "Measurement of Two-Tone Transfer Characteristics of High-Power Amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 49, Mar. 2001, pp. 568–571.
52. K. Yryssas and A. Samelis, "Behavioral Modeling of Nonlinear Radio-Frequency Power Amplifiers," *Asian-Pacific Conference Proceedings 2005, APMC 2005*, vol. 1, Dec. 2005, p. 4.
53. A. Saleh, "Frequency-Independent and Frequency-Dependent Nonlinear Models for TWT Amplifiers," *IEEE Transactions on Commununications*, vol. COM-29, no. 11, 1981, pp. 1715–1720.
54. J. L. Smith, "A Method to Predict the Level of Intermodulation Products in Broadband Amplifiers," *Microwave Journal*, vol. 46, no. 4, 2003, p. 64.
55. "Terminal Conformance Specification; Radio Transmission and Reception (FDD) (Release 1999)," in 3G-PP Technical Specification 34.121, v3.0.0 (2000–3). Available at: http://3gpp.org
56. RF Micro Devices, Greensboro, North Carolina, USA, Website: http://www.rfmd.com
57. T. S. Nielsen, S. Tawfik, T. Larsen, and S. Lindfors, "Behavioral Level Modeling of Power Amplifiers with Varying Antenna Load," *Vehicular Technology Conference, 2005, VTC 2005-Spring*, vol. 2, June 2005, pp. 963–967.
58. S. Haykin, *Neural Networks: A Comprehensive Foundation*. Upper Saddle River, NJ: Prentice-Hall, 1999.
59. M. Schetzen, *The Volterra & Wiener Theories of Nonlinear Systems*. New York: Wiley, 1998.
60. J. H. K. Vuolevi and T. Rahkonen, "Extraction of a Nonlinear AC FET Model using Small-Signal S-Parameters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, no. 5, 2002, pp. 1311–1315.
61. S. Narayanan, "Transistor Distortion Analysis Using Volterra Series Representation," *Bell System Technical Journal*, vol. 46, May 1967, pp. 991–1024.
62. R. A. Minasian, "Intermodulation Distortion Analysis of MESFET Amplifiers Using the Volterra Series Representation," *IEEE Transactions on Microwave Theory and Techniques*, vol. 28, Jan. 1980, pp. 1–8.
63. G. M. Lambrianou and C. S. Aitchison, "Optimization of Third Order Intermodulation Product and Output Power from X-Band MESFET Amplifier Using Volterra Series Analysis," *IEEE Transactions on Microwave Theory and Techniques*, vol. 33, Dec. 1985, pp. 1395–1403.
64. Y. Hu, J. C. Mollier, and J. Obregon, "A New Method of Third-Order Intermodulation Reduction in Nonlinear Microwave Systems," *IEEE Transactions on Microwave Theory and Techniques*, vol. 34, Feb. 1986, pp. 245–250.
65. S. A. Mass and A. Crosmun, "Minimization of Intermodulation Distortion in GaAs MESFET Small Signal Amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 37, Sept. 1989, pp. 1411–1417.
66. J. J. Xu, M. C. E. Yagoub, R. T. Ding, and Q. J. Zhang, "Neural-Based Dynamic Modeling of Nonlinear Microwave Circuits," *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, Dec. 2002, pp. 2769–2780.

67. Y. H. Fang, M. C. E. Yagoub, F. Wang, and Q. J. Zhang, "A New Macro-Modeling Approach for Nonlinear Microwave Circuits Based on Recurrent Neural Networks," *IEEE Transactions on Microwave Theory and Techniques*, vol. 48, Dec. 2000, pp. 2335–2344.
68. J. W. Nandler, M. A. Ismail, J. E. Rayas-Sanchez, and Q. J. Zhang, "New Directions in Model Development for RF/Microwave Components Utilizing Artificial Neural Networks and Space Mapping," *IEEE MTT-S International Microwave Symposium and Digest*, vol. 4, 1999, pp. 2572–2575.
69. Q. J. Zhang and K. C. Gupta, *Neural Networks for RF and Microwave Design.* Norwood, MA: Artech House, 2000.
70. L. Taijun, S. Boumaiza, and F. M. Ghannouchi, "Deembedding Static Nonlinearities and Accurately Identifying and Modeling Memory Effects in Wide-Band RF Transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, no. 11, 2005, pp. 3578–3587.
71. R. Raich and G. T. Zhou, "On the Modeling of Memory Nonlinear Effects of Power Amplifiers for Communication Applications," *IEEE Digital Signal Processing Workshop, 2002 and the 2nd Signal Processing Education Workshop*, Oct. 2002, pp. 7–10.
72. P. Gilabert, G. Montoro, and E. Bertran, "On the Wiener and Hammerstein Models for Power Amplifier Predistortion," *Microwave Conference Proceedings, 2005. APMC 2005. Asia-Pacific Conference Proceedings*, vol. 2, Dec. 2005, p. 4.
73. M. C. Jeruchim, P. Balaban, and K. S. Shanmugan, *Simulation of Communication Systems—Modeling, Methodology, and Techniques*, 2nd Edn., New York: Kluwer, 2000.
74. Q. J. Zhang and K. C. Gupta, *Neural Networks for RF and Microwave Design.* Norwood, MA: Artech House, 2000.
75. H. White, A. R. Gallant, K. Hornik, M. Stinchcombe, and J. Wooldrige, *Artificial Neural Networks: Approximation and Learning Theory.* Oxford, UK: Blackwell, 1992.
76. I. Santamaria, J. Ibanez, M. Lazaro, C. Pantaleon, and L. Vielvo, "Modeling Nonlinear Power Amplifiers from Subsampled Data: A Comparative Study Using Real Measurements," *Eurasip Journal Applied Signal Processing*, 2003, pp. 1219–1228.
77. M. Ibnkahla, N. J. Sombrin, and F. Castanie, "Neural Network Modeling and Identification of nonlinear Channels with Memory: Algorithms, Applications, and Analytical Models," *IEEE Transactions on Signal Processing*, vol. 46, no. 5, 1998, pp. 1208–1220.
78. J. Wood, D. E. Root, and N. B. Tuffilaro, "A Behavioral Modeling Approach to Nonlinear Model-Order Reduction for RF/Microwave ICs and Systems," *IEEE Transactions on Microwave Theory and Techniques*, vol. 52, no. 9, 2004, pp. 2274–2284.
79. J. Xu, M. C. E. Yagoub, R. Ding, and Q. J. Zhang, "Neural-Based Dynamic Modeling of Nonlinear Microwave Circuits," *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, no. 12, 2002, pp. 2769–2780.
80. J. de Villiers and E. Barnard, "Backpropagation Neural Nets with One and Two Hidden Layers," *IEEE Transactions on Neural Networks*, vol. 8, Jan. 1992, pp. 136–141.
81. Q. J. Zhang, K. C. Gupta, and V. K. Devabhaktuni, "Artificial Neural Networks for RF and Microwave Design—From Theory to Practice," *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, part 2, Apr. 2003, pp. 1339–1350.
82. V. Devabhaktuni, B. Chattaraj, M. C. E. Yagoub, and Q. J. Zhang, "Advanced Microwave Modeling Framework Exploiting Automatic Model Generation, Knowledge Neural Networks, and Space Mapping," *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, no. 7, 2003, pp. 1822–1833.
83. L. Prechelt, "Automatic Early-Stopping Using Cross Validation: Quantifying the Criteria," *Neural Networks*, vol. 11, 1998, pp. 761–767.
84. T. Liu, S. Boumaiza, and F. M. Ghannouchi, "Dynamic Behavioral Modeling of 3G Power Amplifiers using Real-Valued Time-Delay Neural Networks," *IEEE Transactions on Microwave Theory and Techniques*, vol. 52, no. 3, 2004, pp. 1025–1033.
85. G. Orengo, P. Colantonio, A. Serino, F. Giannini, G. Ghione, M. Pirola, and G. Stegmayer, "Time-Domain Neural Network Characterization for Dynamic Behavioral Models of Power Amplifiers," *European Gallium Arsenide and Other Semiconductor Application Symposium 2005*, *EGAAS 2005*, Oct. 2005, pp. 189–192.
86. T. Liu, S. Boumaiza, and F. M. Ghannouchi, "Application of Neural Networks to 3G Power Amplifier Modeling," *2005 IEEE International Joint Conference on Neural Networks IJCNN '05*, vol. 4, Aug. 2005, pp. 2378–2382.

87. D. Luongvinh and K. Youngwoo, "Behavioral Modeling of Power Amplifiers using Fully Recurrent Neural Networks," *IEEE MTT-S International Microwave Symposium Digest 2005*, June 2005, p. 4.
88. ETSI Tdoc SMG2 WPB 108/98, Ericsson, *EDGE Evaluation of 8-PSK*.
89. G. F. Pedersen and J. O. Nielsen, "Radiation Pattern Measurements of Mobile Phones Next to Different Head Phantoms," *IEEE Vehicular Technology Conference 2002 Proceedings*, VTC 2002-Fall, 2002 IEEE 56th, vol. 4, Sept. 2002, pp. 24–28.
90. A. Kumar, *"Fixed and Mobile Terminal Antennas,"* MA: Artech House, 1991.
91. M. B. Knudsen, G. F. Pedersen, B. G. H. Olsson, and K. Olesen, "Mobile Phone Measurements Including the Antenna," *IEEE Vehicular Technology Conference 2002 Proceedings*, VTC 2002-Spring, 2000 IEEE 51st, vol. 3, May 2000, pp. 15–18.

34

Technology CAD

Peter A. Blakey
Northern Arizona University

34.1 Introduction

Computer-aided design (CAD) helps engineers to respond to various issues associated with complexity. One is the manipulation and storage of large amounts of relatively simple information. Another is the impact of nonlinear physical phenomena on the behavior of components, circuits, and systems. Both issues are encountered in the field of microelectronics. They have triggered the development of two distinct but overlapping categories of CAD tools. Electronics CAD (ECAD) is used to design new products using an existing semiconductor technology. Technology CAD (TCAD) is used to accelerate the development of new semiconductor technologies. Among other functions, ECAD tools help manage large amounts of relatively simple information, such as the number, size, and locations of the polygons used to define lithographic masks. TCAD predicts the way in which nonlinear process and device physics impacts the performance of proposed technologies. ECAD and TCAD overlap in the area of SPICE-type circuit simulation. The subject matter of this chapter is TCAD and its application to the development of radio frequency (RF), microwave, and millimeter-wave semiconductor technologies.

34.2 An Overview of TCAD

The core tools of TCAD are process simulators and device simulators. Process simulators predict the structures that result from applying a specified sequence of processing steps to a specified initial structure. Device simulators predict the electrical behavior of specified structures. Process and device simulators are, frequently, used in combination to predict the impact of process variations on electrical behavior. The basic flow of information is indicated in Figure 34.1.

Process and device simulators encapsulate and integrate the knowledge of experts in a way that makes detailed knowledge and experience accessible to less expert process and device engineers. This function

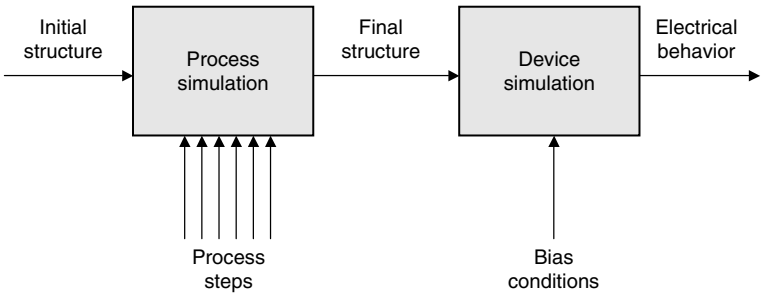


FIGURE 34.1 The basic flow of TCAD information.

of encapsulation and integration helps to ameliorate the shortage of experienced technical personnel and the comparatively narrow technical span of most experts.

Process and device simulators use numerical techniques to solve mathematical equations that describe the underlying physics of an experiment. They predict the results of experiments without requiring the experiments to be run. Predictive simulation is useful whenever it is time consuming, expensive, dangerous, or otherwise difficult to run real experiments. In the case of microelectronics development, it is both time-consuming and expensive to run real experiments in a semiconductor fab. As a result, it is often very cost-effective to substitute simulated experiments for some of the real experiments that would, otherwise, be required by purely empirical development procedures.

Process simulators incorporate physical models for phenomena such as lithography, ion implantation, diffusion, oxidation, deposition, and etching. Device simulators solve equations that express charge continuity and the dependency of the electrostatic potential on the distribution of charge. In order to solve the charge continuity equations, device simulators need models for charge transport and for the generation and recombination of charge carriers. Implementing process and device simulators is a challenging task that requires knowledge of physics, numerical methods, visualization techniques, user interfaces, and software engineering. Research groups in universities and in industry were the first to develop this type of software. However, such groups are not well suited to the task of maintaining the software and providing long-term support to users in industry. General-purpose process and device simulation software is now available from several commercial vendors who provide maintenance and support. Most microelectronics companies use commercial software as the foundation of their TCAD capabilities.

Commercial TCAD systems include tool integration utilities that make it convenient to use the core tools in combination, and task integration environments that automate large-scale simulation-based experimentation. The tool integration utilities include run time environments, visualization tools, structure editors, layout editors and interfaces, and optimization capabilities. Task integration environments usually support a virtual wafer fab (VWF) paradigm that makes it convenient to define and run simulated split lot experiments. After a simulated experiment is defined, a VWF environment automatically generates all the associated jobs, farms out the jobs across a computer network, and collects the results. It may also provide modeling capabilities that make it easy to analyze the simulated results.

Process and device simulators were implemented initially to assist the development of digital silicon VLSI technologies. In recent years, the capabilities of TCAD software have been extended to cover a range of different material systems, and to provide convenient RF-oriented analysis capabilities. The widespread use of epitaxially grown materials and low thermal budget processing means that process simulation for III–V heterostructure technologies can often be reduced to the definition of material layers and doping densities, with geometric specification of deposition and etch steps. This avoids difficult problems associated with process simulation of silicon technologies, which must account for moving boundaries during oxidation, complicated diffusion mechanisms, and so forth. Device simulation for advanced material systems often requires advanced models for charge transport, and it is usually necessary to account for the impact of deep level traps. The appropriate models have been incorporated into commercially available device simulators.

In principle, device simulators can predict any electrical behavior that can be measured. Simulators that are targeted towards the development of digital silicon VLSI technologies are primarily concerned with the calculation of DC I–V characteristics, but they also provide basic capabilities for calculating small-signal AC behavior and large-signal transient behavior. Mixed-mode device-and-circuit simulation capabilities are also available. Mixed-mode simulation is basically SPICE-like circuit simulation that uses numerical device simulation to model the behavior of one or more of the active devices in a circuit. Harmonic balance analysis of device operation implemented at the level of the underlying device physics, rather than at the circuit level, is available from some vendors. Another type of simulation is device-level-noise analysis that includes all relevant transport phenomena and generation and recombination processes.

In the context of RF and microwave technology development, the initial use of TCAD is to calculate DC I–V characteristics, in order to meet target specifications for threshold voltages, knee voltages, breakdown voltages, current drive capability, leakage currents, and low-frequency transconductance and output conductance. The next step is to calculate frequency-dependent small-signal terminal parameters in order to extract and assess the behavior of quasi-static device capacitances and the impact of non-quasi-static effects. Simulated small-signal data is, often, processed into conventional figures of merit such as f_T and f_{max}, and is also used to investigate small-signal stability behavior. Although large-signal simulation of RF devices is technically feasible, run times soon become excessive. It is usually more cost-effective to extract compact device models from simulated DC and small-signal data, using the same procedures that are used with experimental data, and then use compact device models for subsequent simulation-based investigation of circuit design issues.

In the same way that VWF environments mirror experimental technology development procedures, computational load-pull systems seek to mirror the functionality of a popular experimental procedure for the design of wireless power amplifiers. Computational load-pull systems predict RF output power, power added efficiency, gain, linearity, and tunability, and present the results as a function of load (or source) impedance plotted on a Smith chart. Although, it is possible to build such systems around TCAD tools, a more cost-effective architecture is built around compact device models. Device simulation can be linked to the compact model interface via automated procedures for data generation and model extraction.

In addition to predicting quantities that are measured routinely, simulation predicts the values of other quantities that are very difficult or impossible to measure. Examples of such quantities are the doping profiles, carrier profiles, and electrostatic potential distributions within a semiconductor device. By examining these quantities, engineers gain insight into the subtleties of device operation. This insight can guide the direction of future development, and it can lead to the development of simple analytic theories that capture the essence of device operation.

Predictive, physics-based simulation is different from data modeling. Simulation predicts the data that an experiment would yield if it were run. Modeling provides a compact description of data that comes from either real or simulated experiments. Data models are used to interpolate, but can not be used to extrapolate. Expectations regarding "accuracy" are quite different for data models and for physics-based simulators. The agreement between modeled and experimental data is normally very good. The agreement between simulated data and experimental data is subject to several sources of uncertainty that can lead to significant discrepancies between simulated data and experimental data. These sources of uncertainty are discussed in a later section.

34.3 Economic Impact of TCAD

The direct economic benefits of TCAD occur when simulated experiments either substitute for or eliminate the need for a significant fraction of the real experiments that are, otherwise, required by traditional, empirical, development procedures. Simulated experiments are almost always performed more rapidly and more cheaply than real experiments. The intelligent use of TCAD yields significant reductions in cycle

times and development costs. Reduced development times lead, in turn, to a faster time to market, which often translates into higher profits.

The magnitude of the direct financial benefits depends both on the skill with which TCAD is deployed and on the characteristics of the market for a given technology. TCAD is most useful when the market is high-volume, and both performance- and cost-driven. Until the mid-1990s, this meant that TCAD was applied primarily to the development of silicon VLSI technologies. As a result of the wireless revolution, TCAD is now being used to accelerate the development of analog, RF, microwave, and millimeter-wave technologies.

The high costs of semiconductor fabs, material and personnel, mean that reducing development costs by even a few percent is enough to justify significant investments in TCAD capabilities. When TCAD is used effectively it can yield direct savings in development times and development costs of around 50%, and can result in enhanced profits (due to faster time to market) that are much greater than the direct cost savings. Savings of this magnitude can, in principle, justify investments of tens of millions dollars in TCAD capabilities. The investments that companies make in TCAD have traditionally been much smaller than this. This is due to a chronic shortage of TCAD engineers who (a) possess an appropriately broad range of skills, spanning detailed device and process physics through industrial development methodologies to system applications and (b) are willing to remain indefinitely in a job function that is usually perceived as being peripheral to the core business of a company.

34.4 Limitations of TCAD

Although the use of TCAD is capable of providing major economic benefits, achieving the potential benefits is by no means trivial. The key is to use TCAD intelligently, in ways that acknowledge and circumvent the limitations and pitfalls that can trap unwary users. This section provides an overview of some of the limitations and pitfalls.

For certain widely used types of simulation, the underlying physics is well established and, for practical purposes, "exact." For example, electromagnetic simulation is based on Maxwell's equations, thermal simulation is based on the heat flow equation, and the simulation of lumped element circuits is based on Kirchoff's Current and Voltage Laws. The underlying physics of process and device simulation is less well defined. For each physical phenomena (e.g., ion implantation, or electron transport) there is a hierarchy of models of varying complexity. All of these models introduce approximations. In general, the more complicated models provide a more complete description, but are more difficult to implement and involve longer (often much longer) computation times. The need to select appropriate accuracy-efficiency tradeoffs can make it difficult for non-experts to utilize TCAD effectively.

After a physical model has been selected, it is necessary to define values for the parameters associated with the model. These parameters must usually be defined over a range of different conditions (e.g., for different temperatures, doping concentrations, or electric fields). The values of the parameters may be obtained from experiments or from theory. In either case, there is often considerable uncertainty in the accuracy of the model parameters.

Process and device simulators solve systems of coupled, nonlinear equations that do not have general analytic solutions. The equations are, therefore, solved using numerical techniques that are implemented in computer software. Numerical techniques provide an approximate solution that is defined on a discrete "mesh" of points. The choice of mesh and numerical techniques has a major impact on the accuracy and efficiency with which solutions are obtained. Using a finer mesh provides solutions that have higher accuracy but are calculated more slowly. Numerical techniques for solving nonlinear equations employ sequences of approximations that are expected to converge to a solution. In some cases, the sequence of approximations fails to converge. Because of these issues, users of TCAD need some expertise in defining meshes and numerical techniques that provide acceptable tradeoffs between accuracy, efficiency, and convergence.

Discrepancies between measured and simulated results arise because of difficulties in specifying equivalent experiments. For example, a real diffusion step may specify a thermal ramp in terms of the times and temperatures that are dialed in as input to a controller. Process simulation will, normally, use the same times and temperatures. However, in the real experiment the temperature of the wafer as a function of time is different from the nominally specified ramp. The temperature as a function of time may also vary from wafer to wafer within a lot, and from point to point on a given wafer. Another example is provided by device simulation. The lattice temperature used in device simulation is often assumed to be 300 K, but the real lattice temperature is a function of both the true ambient temperature and any self-heating that occurs. Other sources of uncertainty are associated with systematic measurement errors and repeatability.

In order to apply TCAD effectively, users need technical skills that encompass physics, numerical techniques, and the application domain. They need to understand and employ effective application methodologies; and they need good judgment to assess the validity and usefulness of TCAD predictions that are impacted by several sources of uncertainty. They also need to be able to "sell" their insights to process and device engineers who do not understand the role of simulation and are skeptical of its value; and they need to be comfortable with having a secondary support role within a technology development organization.

34.5 The Role of Calibration

By calibrating simulated results to experimental results, it is possible to ameliorate some of the issues that were outlined in the previous section. The existence of high-quality calibration is a prerequisite for using TCAD for certain applications. Achieving a good calibration involves a considerable amount of effort, and poor calibration can exacerbate the problems outlined previously. In view of the importance of calibration to the effective use of TCAD, a review of the associated issues is presented in this section.

Calibration attempts to reduce some of the errors associated with simulation. The measure of the errors is taken as the differences between simulated and measured results. Calibration seeks to reduce these errors by adjusting the values of the model parameters used by the simulators. In essence, it is assumed that the various sources of errors can be partially compensated for by using appropriate "effective" values of model parameters. A good calibration must use a sufficiently complete set of output variables, that is, quantities for which the predicted and measured results are to be brought into better agreement. It must also use an appropriate set of input variables, that is, model parameters whose values are adjusted to improve the agreement between predicted and measured values. A good calibration must not allow the adjusted values of the model parameters to assume implausible values.

Calibration is often performed using ad-hoc procedures that lead to unsatisfactory results. A very common error is to adjust either too few input variables, or the wrong input variables. A related error is to allow input variables to assume values that are implausible or unphysical. Another common problem is that some key output variables are neglected, and the associated errors diverge as a result of the calibration attempt. Calibration needs to be set up as a clearly defined optimization problem with multiple constrained input variables. Because small-change parameter sensitivity matrices can be calculated using TCAD, it is certainly possible to use advanced optimization techniques. However, it is usually more efficient to use a sequential, step-by-step physics-driven calibration procedure in which only a subset of the input and output variables are varied at each step.

Figure 34.2 indicates how a calibration of a GaAs MESFET simulation can be performed in a logical sequence. The first step matches the pinch-off voltage by adjusting the sheet charge. The second step adjusts the low-field mobility to fit the on-resistance in the linear region. The third step adjusts the high-field mobility parameters to fit the current in the saturation region; and the fourth step adjusts the impact ionization coefficients (and if necessary, the tunneling parameters) to match the breakdown characteristics. The principle of progressing from zero bias, through low bias, to high bias, to breakdown, applies to the calibration of other devices also. As a practical point, it is advisable to use pulsed I–V

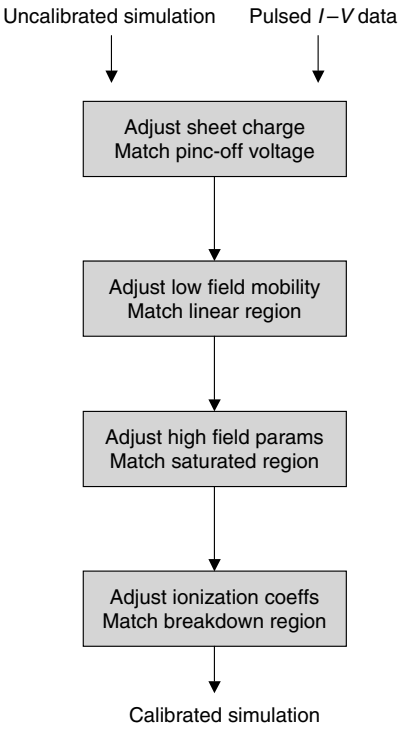


FIGURE 34.2 Sequential calibration of a MESFET simulation.

characteristics, rather than DC I–V characteristics, when performing a TCAD calibration, in order to bypass the effects of self-heating.

Performing a good calibration takes a lot of effort, but certain applications of TCAD only become viable after a good calibration has been achieved. Regardless of how much effort is expended on calibration, there will still be residual errors. This is because calibration does not remove approximations in the underlying physics; it simply allows a portion of their impact to be absorbed into the values of the model parameters. No set of model parameters can correct for an underlying physical model that is inadequate. The converse of this observation is that the degree to which a calibration is successful provides insight into the adequacy of the underlying physics.

34.6 Applications of TCAD

The block diagram shown in Figure 34.3 provides a high-level view of the activities associated with traditional, empirical development of semiconductor technologies. The development process involves a make-and-measure cycle that is repeated until acceptable results are obtained. The process recipe is then transferred to a fab; and layout rules and electrical models for active and passive components are transferred to the circuit design community. Circuits are designed, masks are made, and the circuits are fabricated.

The block diagram shown in Figure 34.4 provides a high-level view of the activities that are associated with more modern simulation-oriented development strategies. Depending on the ways in which TCAD is used, modern development methodologies may be viewed as being "TCAD-influenced," "TCAD-driven" or "system-driven via TCAD." TCAD-influenced development uses process and device simulation to accelerate the make-and-measure phase of technology development. TCAD-driven development makes more extensive use of process and device simulation in the areas of modeling, characterization, and

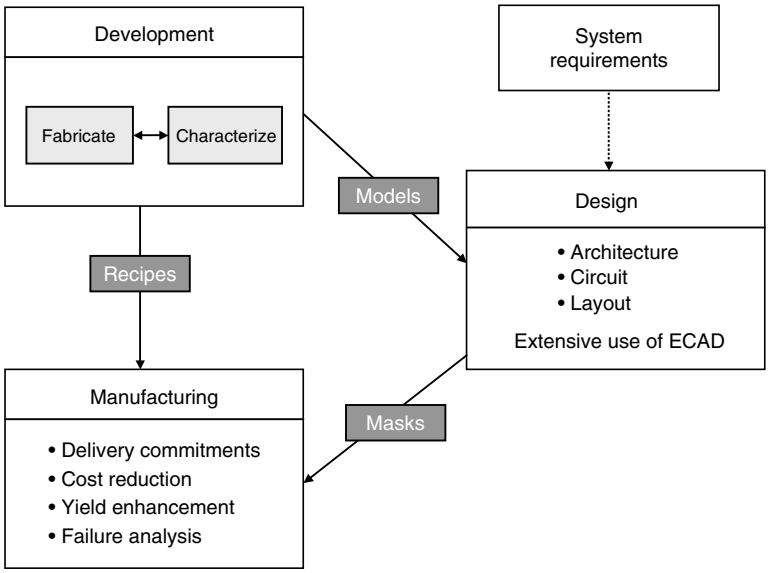


FIGURE 34.3 A high-level view of the traditional development methodology.

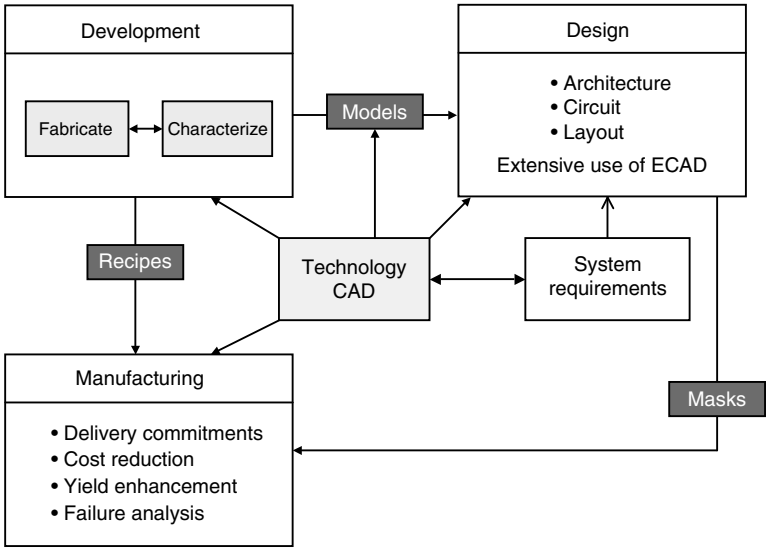


FIGURE 34.4 A high-level view of modern development methodologies.

manufacturing. System-driven development supplements TCAD-driven development by adding explicit bi-directional links to system requirements and basic architectural choices. Much of modern development is TCAD-influenced. Some companies use TCAD-driven methodologies, and a few companies have implemented aspects of system-driven methodologies.

It is useful to order individual TCAD activities by the amount of calibration required to achieve useful results. One such list, ordered by increasing calibration requirements, is as follows:

- Concept evaluation
- Technology discrimination
- Technology development

- Simulation-based model development
- Simulation-assisted characterization
- Statistical process design
- Technology transfer
- Process control
- Yield enhancement and failure analysis

Prior to the availability of TCAD, there was considerable reliance on simplified analytic models. Gross oversimplification of these models would lead to unrealistic performance projections, for example, to the prediction of useful transistor performance at THz frequencies. TCAD is now used to perform rapid evaluations of proposed new concepts. Accounting for Poisson's equation, one or more carrier continuity equations, the coupling between these equations, and the need for device contacts ensure more realistic estimates of device performance. Projections of wildly over-optimistic device performance have been substantially eradicated by the widespread availability of TCAD. Concept evaluation usually requires little or no calibration effort, since all that is required is a go/no-go assessment.

At the initial stages of technology development, it is sometimes necessary to select between two or more competing approaches. TCAD is well suited to this purpose. Calibration is usually not very critical because the errors that are introduced by simulation tend to affect both technologies approximately equally, and are unlikely to cause a simulation-based ranking to be different from an experimentally determined ranking.

Once an overall approach has been selected, TCAD is used to explore different sub-options, and to assist process optimization. This is, presently, the single most important application of TCAD. Simulation-based design and optimization can and should be done systematically, but in practice, it is often done using inefficient ad-hoc techniques. Application methodologies associated with simulation-based technology development are discussed in the next section. As the results of real experiments become available they should be fed back to refine the simulator calibrations, in order to improve the usefulness of simulation at each stage of design and optimization. A general goal is for the calibration achieved for one generation of a technology to provide a solid starting point for the simulation-based design and optimization of the next generation of the technology.

Technology CAD can also be used to assist modeling and characterization activities. In the area of device modeling, TCAD can be used as a reference standard that enables careful testing of assumptions made in the course of developing device models, and supports systematic assessments of the global accuracy of a device model. TCAD is especially useful for assessing the accuracy of large-signal models that are constructed from DC and small-signal AC data. A large-signal model is constructed from simulated DC and small-signal data. The large-signal behavior predicted by the model is then, compared to the large-signal behavior predicted directly using TCAD. The development and characterization of charge-based models is also facilitated by the use of TCAD, since the calculated charge distributions within a device can easily be integrated and partitioned, in accordance, with charge-based modeling approaches. TCAD can also be used to assist device characterization. For example, it is quite straightforward to predict the values of bias- and temperature-dependent source and drain access resistances. Simulation can also assist the interpretation of pulsed I–V measurements that are used to characterize thermal effects and low-frequency dispersive effects associated with deep-level traps.

When there is sufficiently high confidence in the quality of an underling calibration, the uses of TCAD can be expanded further. Calculation of process sensitivities provides a foundation for statistical process design, design centering, and yield modeling. Information concerning process sensitivity is also a useful input for process control. Simulation of the impact of large-change variations (e.g., the impact on electrical behavior of skipping a process step) can provide "signature" information that assists failure analysis. The ideal situation is when a comprehensive set of simulated data, together with an associated set of input data, forms an integral part of the technology transfer package that is supplied by a development organization to a manufacturing organization.

34.7 Application Protocols

Computer-aided design is often deployed using inefficient ad-hoc techniques. In order to use TCAD efficiently, it is helpful to develop, refine, and document sets of useful simulation-based procedures. The details of these procedures often vary considerably from one application (e.g., technology discrimination) to another (e.g., failure analysis). A description of a simulation-based procedure that is used for a particular application is referred to as an application protocol. From a high-level perspective, the development of simulation-based application protocols is often a two-stage procedure. In the first stage, the simulation-based procedures are set up to mirror the existing experimental procedures associated with traditional, empirical, technology development. In the second stage, there is explicit consideration of the additional possibilities that can be exploited due to the use of simulated, rather than real, experiments.

Since technology development is, presently, the most important application of TCAD, an example of an application protocol for technology development will be outlined. The first step is to mirror traditional, empirical development procedures that use split-lot experiments. This is the underlying idea of VWF methodology, variants of which have been implemented by several commercial vendors of TCAD software. VWF methodology makes it convenient to define and run large-scale simulated split-lot experiments. Users select from a broad range of built-in experimental designs, define split variables among input quantities, and specify the output quantities to be extracted. Once these variables have been specified, all of the input descriptions associated with the simulated experiment are generated automatically, and farmed out automatically within a networked computing environment. The results are collected automatically, stored in a database, and are conveniently viewed and analyzed using built-in tools.

The next step is to define ways in which a simulation-based approach can go beyond traditional development procedures. When using TCAD it is possible to use sophisticated optimization techniques that use the ability to calculate accurate derivatives in a multi-dimensional search space. Due to long cycle times, high costs, and noisy experimental data, this approach is not feasible in the real world. However, it is quite suitable for use in the simulation world, where results can be obtained quickly and reproducibly. The use of sophisticated optimization techniques focuses explicit attention on the need to define an objective function, and on the need to define all of the constraints on the input variables. (Split-lot experimentation also involves objective functions and constraints, but they are usually defined and implemented less formally.)

A focus on objective functions and constraints leads to a useful high-level view of technology development, in general, and the role of simulation, in particular. Technology development can be viewed as a search, within a multi-dimensional design space, for the point corresponding to an optimum design. Various techniques are available for narrowing down the regions of the design space that need to be searched. These techniques include analytic theory and scaling arguments, numerical simulation, and real experiments. Each of these techniques has its advantages and limitations. Analytic theory and scaling arguments are quick and easy to apply, and can eliminate large regions of the design space, but their absolute accuracy is poor. Simulation can provide much better accuracy, and is relatively fast and inexpensive, but is subject to model errors and numerical error. Experiments provide "real" data, but are expensive, time-consuming, and subject to experimental error. Analytic theory, simulation, and experiments can be viewed as a hierarchy of search techniques that can be applied sequentially and iteratively to maximize the efficiency of the search process.

A simplified view of the search for an optimum design in a two-dimensional space is presented in Figure 34.5. Analytic theories and scaling arguments are used to establish the boundaries of the region of the design space that needs to be searched. Simulation is then used to predict the optimum point in the design space. Due to modeling and numerical errors, this predicted optimum (indicated by the closed circle), is not exactly coincident with the true optimum point. The region around the predicted optimum is, therefore, searched experimentally, to determine the location of the true optimum (the open circle). According to this view, the goal of simulation is to predict an optimum design with sufficient accuracy that the true optimum can be found using only a small number of real experiments. Simulation, thus,

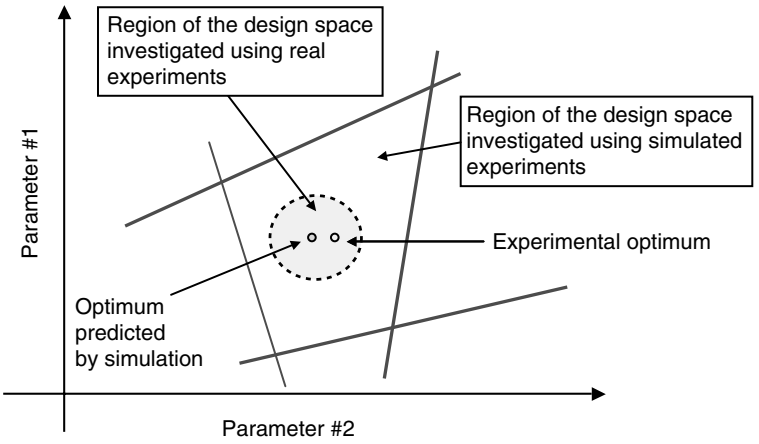


FIGURE 34.5 A simplified view of the search process.

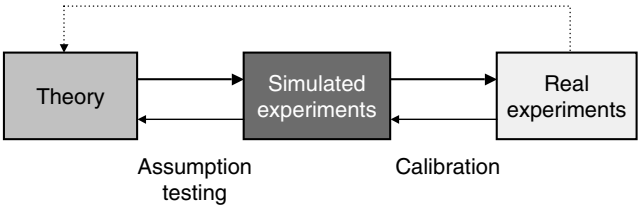


FIGURE 34.6 A three-level hierarchical approach to technology development.

does not eliminate the need for experiments; but it is realistic to expect a reduction of 50–80% in the number of real experiments that would be required if simulation were not used.

Figure 34.6 provides another view that indicates how theory, simulated experiments, and real experiments are linked in modern development procedures. The forward arrows indicate the basic sequence of development activities, as outlined in the previous paragraph. The reverse arrows indicate that information is fed back to allow simulator calibration and assumption testing. Subsets of this view correspond to other development methodologies. For example, eliminating simulation results in a "traditional" development methodology, in which analytic theory and scaling arguments are used to bracket a design space that is then searched experimentally. Eliminating theory results in a simulation-reliant development methodology that may be inefficient in its search of an unconstrained design space.

Similar approaches can be used to develop application protocols for the other uses of TCAD.

34.8 Conclusions

Technology CAD provides benefits associated with faster development times, lower development costs, enhanced physical insight, and the encapsulation and integration of expert knowledge. These benefits are especially important in the context of high-performance, high-volume, consumer-oriented applications. TCAD is very well established as an integral part of the procedures used to develop silicon VLSI technologies, and is assuming an important role in the development of semiconductor technologies for RF, microwave, and millimeter-wave applications.

Although the benefits of using TCAD can be very great, achieving the benefits is not easy. Some of the challenges, such as the definition of physical models, model parameters, calibration procedures,

and application protocols, are primarily technical in nature. Other challenges, such as a shortage of trained manpower that is able to apply TCAD effectively, are associated with economic and social factors. Companies that are able to overcome these challenges can obtain a competitive advantage. Conversely, companies that do not deploy TCAD effectively can expect to find themselves at a competitive disadvantage.

APPENDIX A

Mathematics, Symbols, and Physical Constants

Greek Alphabet

	Greek Letter	Greek Name	English Equivalent		Greek Letter	Greek Name	English Equivalent
A	α	Alpha	a	N	ν	Nu	n
B	β	Beta	b	Ξ	ξ	Xi	x
Γ	γ	Gamma	g	O	o	Omicron	ŏ
Δ	δ	Delta	d	Π	π	Pi	p
E	ε	Epsilon	ĕ	P	ρ	Rho	r
Z	ζ	Zeta	z	Σ	σ	Sigma	s
H	η	Eta	ē	T	τ	Tau	t
Θ	θ ϑ	Theta	th	Y	υ	Upsilon	u
I	ι	Iota	i	Φ	ϕ φ	Phi	ph
K	κ	Kappa	k	X	χ	Chi	ch
Λ	λ	Lambda	l	Ψ	ψ	Psi	ps
M	μ	Mu	m	Ω	ω	Omega	ō

International System of Units (SI)

The International System of units (SI) was adopted by the 11th General Conference on Weights and Measures (CGPM) in 1960. It is a coherent system of units built form seven *SI base units,* one for each of the seven dimensionally independent base quantities: they are the meter, kilogram, second, ampere, kelvin, mole, and candela, for the dimensions length, mass, time, electric current, thermodynamic temperature, amount of substance, and luminous intensity, respectively. The definitions of the SI base units are given below. The *SI derived units* are expressed as products of powers of the base units, analogous to the corresponding relations between physical quantities but with numerical factors equal to unity.

In the International System there is only one SI unit for each physical quantity. This is either the appropriate SI base unit itself or the appropriate SI derived unit. However, any of the approved decimal prefixes, called *SI prefixes,* may be used to construct decimal multiples or submultiples of SI units.

It is recommended that only SI units be used in science and technology (with SI prefixes where appropriate). Where there are special reasons for making an exception to this rule, it is recommended always to define the units used in terms of SI units. This section is based on information supplied by IUPAC.

Definitions of SI Base Units

Meter—The meter is the length of path traveled by light in vacuum during a time interval of 1/299 792 458 of a second (17th CGPM, 1983).

Kilogram—The kilogram is the unit of mass; it is equal to the mass of the international prototype of the kilogram (3rd CGPM, 1901).

Second—The second is the duration of 9 192 631 770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the cesium-133 atom (13th CGPM, 1967).

Ampere—The ampere is that constant current which, if maintained in two straight parallel conductors of infinite length, of negligible circular cross-section, and placed 1 meter apart in vacuum, would produce between these conductors a force equal to 2×10^{-7} newton per meter of length (9th CGPM, 1948).

Kelvin—The kelvin, unit of thermodynamic temperature, is the fraction 1/273.16 of the thermodynamic temperature of the triple point of water (13th CGPM, 1967).

Mole—The mole is the amount of substance of a system which contains as many elementary entities as there are atoms in 0.012 kilogram of carbon-12. When the mole is used, the elementary entities must be specified and may be atoms, molecules, ions, electrons, or other particles, or specified groups of such particles (14th CGPM, 1971).

Examples of the use of the mole:

1 mol of H_2 contains about 6.022×10^{23} H_2 molecules, or 12.044×10^{23} H atoms
1 mol of HgCl has a mass of 236.04 g
1 mol of Hg_2Cl_2 has a mass of 472.08 g
1 mol of Hg_2^{2+} has a mass of 401.18 g and a charge of 192.97 kC
1 mol of $Fe_{0.91}S$ has a mass of 82.88 g
1 mol of e^- has a mass of 548.60 μg and a charge of –96.49 kC
1 mol of photons whose frequency is 10^{14} Hz has energy of about 39.90 kJ

Candela—The candela is the luminous intensity, in a given direction, of a source that emits monochromatic radiation of frequency 540×10^{12} hertz and that has a radiant intensity in that direction of (1/683) watt per steradian (16th CGPM, 1979).

Names and Symbols for the SI Base Units

Physical Quantity	Name of SI Unit	Symbol for SI Unit
Length	Meter	m
Mass	Kilogram	kg
Time	Second	s
Electric Current	Ampere	A
Thermodynamic temperature	Kelvin	K
Amount of substance	Mole	mol
Luminous intensity	Candela	cd

SI Derived Units with Special Names and Symbols

Physical Quantity	Name of SI Unit	Symbol for SI Unit	Expression in Terms of SI Base Units	
Frequency[1]	Hertz	Hz	s^{-1}	
Force	Newton	N	m kg s^{-2}	
Pressure, stress	Pascal	Pa	N m^{-2}	= m^{-1} kg s^{-2}
Energy, work, heat	Joule	J	N m	= m^2 kg s^{-2}
Power, radiant flux	Watt	W	J s^{-1}	= m^2 kg s^{-3}
Electric charge	Coulomb	C	A s	
Electric potential, electromotive force	Volt	V	J C^{-1}	= m^2 kg s^{-3} A^{-1}

Physical Quantity	Name of SI Unit	Symbol for SI Unit	Expression in Terms of SI Base Units	
Electric resistance	Ohm	Ω	V A^{-1}	= m^2 kg s^{-3} A^{-2}
Electric conductance	Siemens	S	Ω^{-1}	= m^{-2} kg^{-1} s^3 A^2
Electric capacitance	Farad	F	C V^{-1}	= m^{-2} kg^{-1} s^4 A^2
Magnetic flux density	Tesla	T	V s m^{-2}	= kg s^{-2} A^{-1}
Magnetic flux	Weber	Wb	V s	= m^2 kg s^{-2} A^{-1}
Inductance	Henry	H	V A^{-1} s	= m^2 kg s^{-2} A^{-2}
Celsius temperature[2]	Degree Celsius	°C	K	
Luminous flux	Lumen	lm	cd sr	
Illuminance	Lux	lx	cd sr m^{-2}	
Activity (Radioactive)	Becquerel	Bq	s^{-1}	
Absorbed dose (of radiation)	Gray	Gy	J kg^{-1}	= m^2 s^{-2}
Dose equivalent (dose equivalent index)	Sievert	Sv	J kg^{-1}	= m^2 s^{-2}
Plane angle	Radian	rad	1	= m m^{-1}
Solid angle	Steradian	sr	1	= m^2 m^{-2}

[1]For radial (circular) frequency and for angular velocity the unit rad s^{-1}, or simply s^{-1}, should be used, and this may not be simplified to Hz. The unit Hz should be used only for frequency in the sense of cycles per second.

[2]The Celsius temperature θ is defined by the equation:

$$\theta/°C = T/K - 273.15$$

The SI unit of Celsius temperature interval is the degree Celsius, °C, which is equal to the kelvin, K. °C should be treated as a single symbol, with no space between the ° sign and the letter C. (The symbol °K, and the symbol °, should no longer be used.)

Units in Use Together with the SI

These units are not part of the SI, but it is recognized that they will continue to be used in appropriate contexts. SI prefixes may be attached to some of these units, such as milliliter, ml; millibar, mbar; megaelectronvolt, MeV; kilotonne, ktonne.

Physical Quantity	Name of Unit	Symbol for Unit	Value in SI Units
Time	Minute	min	60 s
Time	Hour	h	3600 s
Time	Day	d	86 400 s
Plane angle	Degree	°	(π/180) rad
Plane angle	Minute	′	(π/10 800) rad
Plane angle	Second	″	(π/648 000) rad
Length	Ångstrom[1]	Å	10^{-10} m
Area	Barn	b	10^{-28} m^2
Volume	Litre	l, L	$dm^3 = 10^{-3}$ m^3
Mass	Tonne	t	$Mg = 10^3$ kg
Pressure	Bar[1]	bar	10^5 Pa = 10^5 N m^{-2}
Energy	Electronvolt[2]	eV (= $e \times$ V)	$\approx 1.60218 \times 10^{-19}$ J
Mass	Unified atomic mass unit[2,3]	u (= $m_a(^{12}C)/12$)	$\approx 1.66054 \times 10^{-27}$ kg

[1]The ångstrom and the bar are approved by CIPM for "temporary use with SI units," until CIPM makes a further recommendation. However, they should not be introduced where they are not used at present.

[2]The values of these units in terms of the corresponding SI units are not exact, since they depend on the values of the physical constants e (for the electronvolt) and N_a (for the unified atomic mass unit), which are determined by experiment.

[3]The unified atomic mass unit is also sometimes called the dalton, with symbol Da, although the name and symbol have not been approved by CGPM.

Physical Constants

General

Equatorial radius of the earth = 6378.388 km = 3963.34 miles (statute).
Polar radius of the earth, 6356.912 km = 3949.99 miles (statute).
1 degree of latitude at 40° = 69 miles.
1 international nautical mile = 1.15078 miles (statute) = 1852 m = 6076.115 ft.
Mean density of the earth = 5.522 g/cm^3 = 344.7 lb/ft^3.
Constant of gravitation $(6.673 \pm 0.003) \times 10^{-8}$ cm^3 gm^{-1} s^{-2}.
Acceleration due to gravity at sea level, latitude 45° = 980.6194 cm/s^2 = 32.1726 ft/s^2.
Length of seconds pendulum at sea level, latitude 45° = 99.3575 cm = 39.1171 in.
1 knot (international) = 101.269 ft/min = 1.6878 ft/s = 1.1508 miles (statute)/h.
1 micron = 10^{-4} cm.
1 ångstrom = 10^{-8} cm.
Mass of hydrogen atom = $(1.67339 \pm 0.0031) \times 10^{-24}$ g.
Density of mercury at 0°C = 13.5955 g/ml.
Density of water at 3.98°C = 1.000000 g/ml.
Density, maximum, of water, at 3.98°C = 0.999973 g/cm^3.
Density of dry air at 0°C, 760 mm = 1.2929 g/l.
Velocity of sound in dry air at 0°C = 331.36 m/s – 1087.1 ft/s.
Velocity of light in vacuum = $(2.997925 \pm 0.000002) \times 10^{10}$ cm/s.
Heat of fusion of water 0°C = 79.71 cal/g.
Heat of vaporization of water 100°C = 539.55 cal/g.
Electrochemical equivalent of silver 0.001118 g/s international amp.
Absolute wavelength of red cadmium light in air at 15°C, 760 mm pressure = 6438.4696 Å.
Wavelength of orange-red line of krypton 86 = 6057.802 Å.

π Constants

π = 3.14159 26535 89793 23846 26433 83279 50288 41971 69399 37511
$1/\pi$ = 0.31830 98861 83790 67153 77675 26745 02872 40689 19291 48091
π^2 = 9.8690 44010 89358 61883 44909 99876 15113 53136 99407 24079
$\log_e \pi$ = 1.14472 98858 49400 17414 34273 51353 05871 16472 94812 91531
$\log_{10} \pi$ = 0.49714 98726 94133 85435 12682 88290 89887 36516 78324 38044
$\log_{10} \sqrt{2}\,\pi$ = 0.39908 99341 79057 52478 25035 91507 69595 02099 34102 92128

Constants Involving *e*

e = 2.71828 18284 59045 23536 02874 71352 66249 77572 47093 69996
$1/e$ = 0.36787 94411 71442 32159 55237 70161 46086 74458 11131 03177
e^2 = 7.38905 60989 30650 22723 04274 60575 00781 31803 15570 55185
$M = \log_{10} e$ = 0.43429 44819 03251 82765 11289 18916 60508 22943 97005 80367
$1/M = \log_e 10$ = 2.30258 50929 94045 68401 79914 54684 36420 67011 01488 62877
$\log_{10} M$ = 9.63778 43113 00536 78912 29674 98645 –10

Numerical Constants

$\sqrt{2}$ = 1.41421 35623 73095 04880 16887 24209 69807 85696 71875 37695
$\sqrt[3]{2}$ = 1.25992 10498 94873 16476 72106 07278 22835 05702 51464 70151
$\log_e 2$ = 0.69314 71805 59945 30941 72321 21458 17656 80755 00134 36026
$\log_{10} 2$ = 0.30102 99956 63981 19521 37388 94724 49302 67881 89881 46211
$\sqrt{3}$ = 1.73205 08075 68877 29352 74463 41505 87236 69428 05253 81039
$\sqrt[3]{3}$ = 1.44224 95703 07408 38232 16383 10780 10958 83918 69253 49935
$\log_e 3$ = 1.09861 22886 68109 69139 52452 36922 52570 46474 90557 82275
$\log_{10} 3$ = 0.47712 12547 19662 43729 50279 03255 11530 92001 28864 19070

Symbols and Terminology for Physical and Chemical Quantities

Name	Symbol	Definition	SI Unit
		Classical Mechanics	
Mass	m		kg
Reduced mass	μ	$\mu = m_1 m_2/(m_1 + m_2)$	kg
Density, mass density	ρ	$\rho = M/V$	kg m^{-3}
Relative density	d	$d = \rho/\rho^{\theta}$	l
Surface density	ρ_A, ρ_S	$\rho_A = m/A$	kg m^{-2}
Specific volume	ν	$\nu = V/m = 1/\rho$	m^3 kg^{-1}
Momentum	p	$\boldsymbol{p} = m\boldsymbol{v}$	kg m s^{-1}
Angular momentum, action	L	$\mathrm{l} = \mathrm{r} \times \mathrm{p}$	J s
Moment of inertia	I, J	$I = \Sigma m_i r_i^2$	kg m^2
Force	F	$\boldsymbol{F} = \mathrm{d}\boldsymbol{p}/\mathrm{d}t = m\boldsymbol{a}$	N
Torque, moment of a force	$\boldsymbol{T}$, $(\boldsymbol{M})$	$\mathrm{T} = \mathrm{r} \times \mathrm{F}$	N m
Energy	E		J
Potential energy	E_p, V, Φ	$E_\mathrm{p} = -\int \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{s}$	J
Kinetic energy	E_k, T, K	$e_\mathrm{k} = (1/2)mv^2$	J
Work	W, w	$w = \int \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{s}$	J
Hamilton function	H	$H(q, p) = T(q, p) + V(q)$	J
Lagrange function	L	$L(q, \dot{q})$ T$(q, \dot{q}) - V(q)$	J
Pressure	p, P	$p = F/A$	Pa, N m^{-2}
Surface tension	γ, σ	$\gamma = \mathrm{d}W/\mathrm{d}A$	N m^{-1}, J m^{-2}
Weight	G, (W, P)	$G = mg$	N
Gravitational constant	G	$F = Gm_1m_2/r^2$	N m^2 kg^{-2}
Normal stress	σ	$\sigma = F/A$	Pa
Shear stress	τ	$\tau = F/A$	Pa
Linear strain, relative elongation	ε, e	$\varepsilon = \Delta l/l$	l
Modulus of elasticity, Young's modulus	E	$E = \sigma/\varepsilon$	Pa
Shear strain	γ	$\gamma = \Delta x/d$	l
Shear modulus	G	$\mathrm{G} = \tau/\gamma$	Pa
Volume strain, bulk strain	θ	$\theta = \Delta V/V_0$	l
Bulk modulus, compression modulus	K	$K = -V_0(\mathrm{d}p/\mathrm{d}V)$	Pa
Viscosity, dynamic viscosity	η μ	$\tau_{x,z} = \eta(\mathrm{d}v_x/\mathrm{d}z)$	Pa s
Fluidity	ϕ	$\phi = 1/\eta$	m kg^{-1} s
Kinematic viscosity	ν	$\nu = \eta/\rho$	m^2 s^{-1}
Friction coefficient	μ, (f)	$F_\mathrm{frict} = \mu F_\mathrm{norm}$	l
Power	P	$P = \mathrm{d}W/\mathrm{d}t$	W
Sound energy flux	P, P_a	$P = \mathrm{d}E/\mathrm{d}t$	W
Acoustic factors			
Reflection factor	ρ	$\rho = P_\mathrm{t}/P_0$	l
Acoustic absorption factor	α_a, (α)	$\alpha_\mathrm{a} = 1 - \rho$	l
Transmission factor	τ	$\tau = P_\mathrm{tr}/P_0$	l
Dissipation factor	δ	$\delta = \alpha_\mathrm{a} - \tau$	l

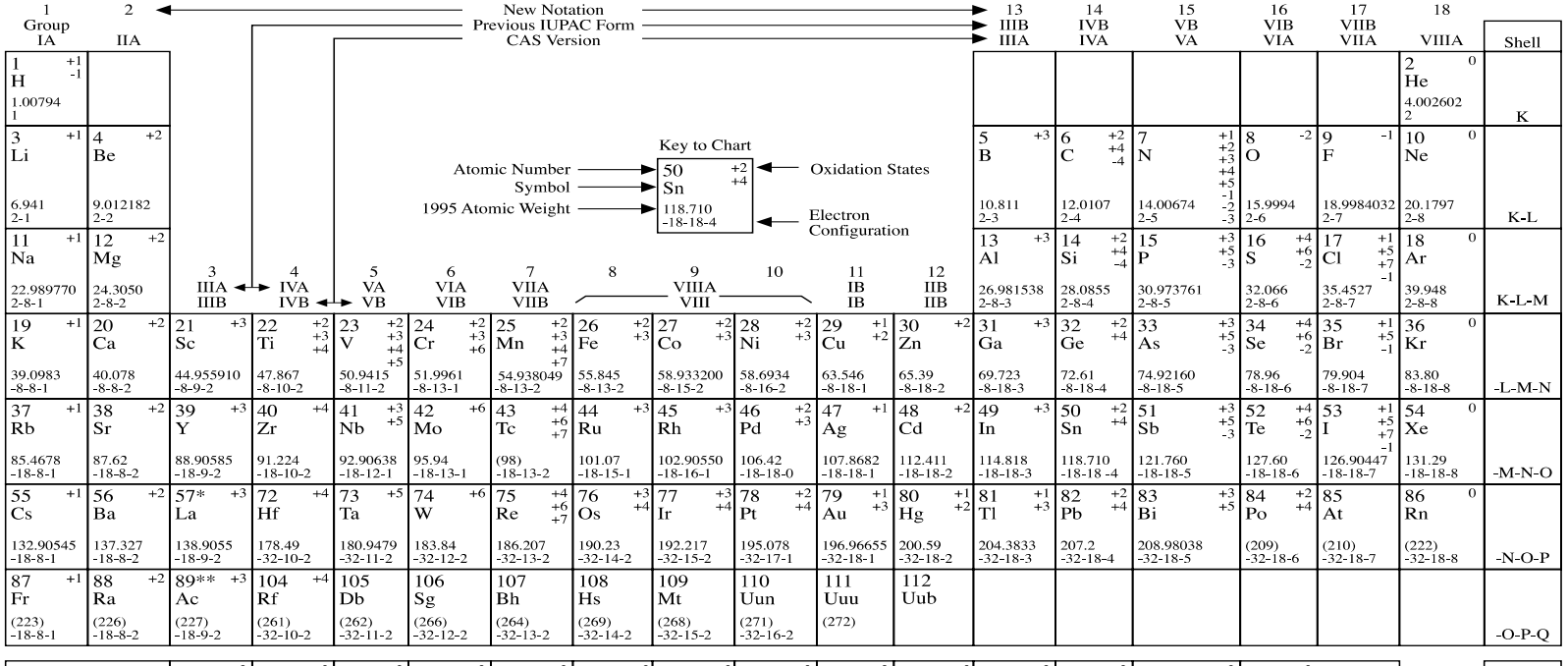
Fundamental Physical Constants

Summary of the 1986 Recommended Values of the Fundamental Physical Constants

Quantity	Symbol	Value	Units	Relative Uncertainty (ppm)
Speed of light in vacuum	c	299 792 458	ms^{-1}	(exact)
Permeability of vacuum	μ_o	$4\pi \times 10^{-7}$	$N\ A^{-2}$	
		= 12.566 370614 …	$10^{-7}\ N\ A^{-2}$	(exact)
Permittivity of vacuum	ε_o	$1/\mu_o c^2$		
		= 8.854 187 817 …	$10^{-12}\ F\ m^{-1}$	(exact)
Newtonian constant of gravitation	G	6.672 59(85)	$10^{-11}\ m^3\ kg^{-1}\ s^{-2}$	128
Planck constant	h	6.626 0755(40)	10^{-34} J s	0.60
$h/2\pi$	$\hbar$	1.054 572 66(63)	10^{-34} J s	0.60
Elementary charge	e	1.602 177 33(49)	10^{-19} C	0.30
Magnetic flux quantum, $h/2e$	Φ_o	2.067 834 61(61)	10^{-15} Wb	0.30
Electron mass	m_e	9.109 3897(54)	10^{-31} kg	0.59
Proton mass	m_p	1.672 6231(10)	10^{-27} kg	0.59
Proton–electron mass ratio	m_p/m_e	1836.152701(37)		0.020
Fine-structure constant, $\mu_o ce^2/2h$	α	7.297 353 08(33)	10^{-3}	0.045
Inverse fine-structure constant	α^{-1}	137 035 9895(61)		0.045
Rydberg constant, $m_e c\alpha^2/2h$	R_∞	10 973 731.534(13)	m^{-1}	0.0012
Avogadro constant	$N_{A'}\ L$	6.022 1367(36)	$10^{23}\ mol^{-1}$	0.59
Faraday constant, $N_A e$	F	96 485.309(29)	$C\ mol^{-1}$	0.30
Molar gas constant	R	8.314 510(70)	$J\ mol^{-1}\ K^{-1}$	8.4
Boltzmann constant, R/N_A	k	1.380 658(12)	$10^{-23}\ J\ K^{-1}$	8.5
Stafan–Boltzmann constant, $(\pi^2/60)k^4/\hbar^3 c^2$	σ	5.670 51(19)	$10^{-8}\ W\ m^{-2}\ K^{-4}$	34
Non-SI units used with SI				
Electronvolt, $(e/C)J = \{e\}J$	eV	1.602 17733(40)	10^{-19} J	0.30
(Unified) atomic mass unit, 1 u = $m_u = 1/12m(^{12}C)$	u	1.660 5402(10)	10^{-27} kg	0.59

Note: An abbreviated list of the fundamental constants of physics and chemistry based on a least-squares adjustment with 17 degrees of freedom. The digits in parentheses are the one-standard-deviation uncertainty in the last digits of the given value. Since the uncertainties of many entries are correlated, the full covariance matrix must be used in evaluating the uncertainties of quantities computed from them.

PERIODIC TABLE OF THE ELEMENTS



The new IUPAC format numbers the groups from 1 to 18. The previous IUPAC numbering system and the system used by Chemical Abstracts Service (CAS) are also shown. For radioactive elements that do not occur in nature, the mass number of the most stable isotope is given in parentheses.

References

1. G. J. Leigh, Editor, *Nomenclature of Inorganic Chemistry*, Blackwell Scientific Publications, Oxford, 1990.
2. *Chemical and Engineering News*, 63(5), 27, 1985.
3. Atomic Weights of the Elements, 1995, *Pure & Appl. Chem.*, 68, 2339, 1996.

Electrical Resistivity

Electrical Resistivity of Pure Metals

The first part of this table gives the electrical resistivity, in units of $10^{-8}\ \Omega\,m$, for 28 common metallic elements as a function of temperature. The data refer to polycrystalline samples. The number of significant figures indicates the accuracy of the values. However, at low temperatures (especially below 50 K) the electrical resistivity is extremely sensitive to sample purity. Thus the low-temperature values refer to samples of specified purity and treatment.

The second part of the table gives resistivity values in the neighborhood of room temperature for other metallic elements that have not been studied over an extended temperature range.

Electrical Resistivity in $10^{-8}\ \Omega\,m$

T/K	Aluminum	Barium	Beryllium	Calcium	Cesium	Chromium	Copper
1	0.000100	0.081	0.0332	0.045	0.0026		0.00200
10	0.000193	0.189	0.0332	0.047	0.243		0.00202
20	0.000755	0.94	0.0336	0.060	0.86	0.00280	
40	0.0181	2.91	0.0367	0.175	1.99		0.0239
60	0.0959	4.86	0.067	0.40	3.07		0.0971
80	0.245	6.83	0.075	0.65	4.16		0.215
100	0.442	8.85	0.133	0.91	5.28	1.6	0.348
150	1.006	14.3	0.510	1.56	8.43	4.5	0.699
200	1.587	20.2	1.29	2.19	12.2	7.7	1.046
273	2.417	30.2	3.02	3.11	18.7	11.8	1.543
293	2.650	33.2	3.56	3.36	20.5	12.5	1.678
298	2.709	34.0	3.70	3.42	20.8	12.6	1.712
300	2.733	34.3	3.76	3.45	21.0	12.7	1.725
400	3.87	51.4	6.76	4.7		15.8	2.402
500	4.99	72.4	9.9	6.0		20.1	3.090
600	6.13	98.2	13.2	7.3		24.7	3.792
700	7.35	130	16.5	8.7		29.5	4.514
800	8.70	168	20.0	10.0	34.6	5.262	
900	10.18	216	23.7	11.4		39.9	6.041

T/K	Gold	Hafnium	Iron	Lead	Lithium	Magnesium	Manganese
1	0.0220	1.00	0.0225		0.007	0.0062	7.02
10	0.0226	1.00	0.0238		0.008	0.0069	18.9
20	0.035	1.11	0.0287		0.012	0.0123	54
40	0.141	2.52	0.0758		0.074	0.074	116
60	0.308	4.53	0.271		0.345	0.261	131
80	0.481	6.75	0.693	4.9	1.00	0.557	132
100	0.650	9.12	1.28	6.4	1.73	0.91	132
150	1.061	15.0	3.15	9.9	3.72	1.84	136
200	1.462	21.0	5.20	13.6	5.71	2.75	139
273	2.051	30.4	8.57	19.2	8.53	4.05	143
293	2.214	33.1	9.61	20.8	9.28	4.39	144
298	2.255	33.7	9.87	21.1	9.47	4.48	144
300	2.271	34.0	9.98	21.3	9.55	4.51	144
400	3.107	48.1	16.1	29.6	13.4	6.19	147
500	3.97	63.1	23.7	38.3		7.86	149
600	4.87	78.5	32.9			9.52	151
700	5.82		44.0			11.2	152
800	6.81		57.1			12.8	
900	7.86					14.4	

(continued)

T/K	Molybdenum	Nickel	Palladium	Platinum	Potassium	Rubidium	Silver
1	0.00070	0.0032	0.0200	0.002	0.0008	0.0131	0.00100
10	0.00089	0.0057	0.0242	0.0154	0.0160	0.109	0.00115
20	0.00261	0.0140	0.0563	0.0484	0.117	0.444	0.0042
40	0.0457	0.068	0.334	0.409	0.480	1.21	0.0539
60	0.206	0.242	0.938	1.107	0.90	1.94	0.162
80	0.482	0.545	1.75	1.922	1.34	2.65	0.289
100	0.858	0.96	2.62	2.755	1.79	3.36	0.418
150	1.99	2.21	4.80	4.76	2.99	5.27	0.726
200	3.13	3.67	6.88	6.77	4.26	7.49	1.029
273	4.85	6.16	9.78	9.6	6.49	11.5	1.467
293	5.34	6.93	10.54	10.5	7.20	12.8	1.587
298	5.47	7.12	10.73	10.7	7.39	13.1	1.617
300	5.52	7.20	10.80	10.8	7.47	13.3	1.629
400	8.02	11.8	14.48	14.6			2.241
500	10.6	17.7	17.94	18.3			2.87
600	13.1	25.5	21.2	21.9			3.53
700	15.8	32.1	24.2	25.4			4.21
800	18.4	35.5	27.1	28.7			4.91
900	21.2	38.6	29.4	32.0			5.64

T/K	Sodium	Strontium	Tantalum	Tungsten	Vanadium	Zinc	Zirconium
1	0.0009	0.80	0.10	0.00016		0.0100	0.250
10	0.0015	0.80	0.102	0.000137	0.0145	0.0112	0.253
20	0.016	0.92	0.146	0.00196	0.039	0.0387	0.357
40	0.172	1.70	0.751	0.0544	0.304	0.306	1.44
60	0.447	2.68	1.65	0.266	1.11	0.715	3.75
80	0.80	3.64	2.62	0.606	2.41	1.15	6.64
100	1.16	4.58	3.64	1.02	4.01	1.60	9.79
150	2.03	6.84	6.19	2.09	8.2	2.71	17.8
200	2.89	9.04	8.66	3.18	12.4	3.83	26.3
273	4.33	12.3	12.2	4.82	18.1	5.46	38.8
293	4.77	13.2	13.1	5.28	19.7	5.90	42.1
298	4.88	13.4	13.4	5.39	20.1	6.01	42.9
300	4.93	13.5	13.5	5.44	20.2	6.06	43.3
400		17.8	18.2	7.83	28.0	8.37	60.3
500		22.2	22.9	10.3	34.8	10.82	76.5
600		26.7	27.4	13.0	41.1	13.49	91.5
700		31.2	31.8	15.7	47.2		104.2
800		35.6	35.9	18.6	53.1		114.9
900			40.1	21.5	58.7		123.1

Electrical Resistivity of Pure Metals (continued)

Element	*T*/K	Electrical Resistivity $10^{-8}\ \Omega\,m$
Antimony	273	39
Bismuth	273	107
Cadmium	273	6.8
Cerium	290–300	82.8
Cobalt	273	5.6
Dysprosium	290–300	92.6
Erbium	290–300	86.0
Europium	290–300	90.0
Gadolinium	290–300	131
Gallium	273	13.6
Holmium	290–300	81.4
Indium	273	8.0
Iridium	273	4.7
Lanthanum	290–300	61.5
Lutetium	290–300	58.2
Mercury	273	94.1
Neodymium	290–300	64.3
Niobium	273	15.2
Osmium	273	8.1
Polonium	273	40
Praseodymium	290–300	70.0
Promethium	290–300	75
Protactinium	273	17.7
Rhenium	273	17.2
Rhodium	273	4.3
Ruthenium	273	7.1
Samrium	290–300	94.0
Scandium	290–300	56.2
Terbium	290–300	115
Thallium	273	15
Thorium	273	14.7
Thulium	290–300	67.6
Tin	273	11.5
Titanium	273	39
Uranium	273	28
Ytterbium	290–300	25.0
Yttrium	290–300	59.6

Electrical Resistivity of Selected Alloys

Values of the resistivity are given in units of $10^{-8}\ \Omega\,m$. General comments in the preceding table for pure metals also apply here.

	273 K	293 K	300 K	350 K	400 K
Alloy—Aluminum-Copper					
Wt % Al					
99[a]	2.51	2.74	2.82	3.38	3.95
95[a]	2.88	3.10	3.18	3.75	4.33
90[b]	3.36	3.59	3.67	4.25	4.86
85[b]	3.87	4.10	4.19	4.79	5.42
80[b]	4.33	4.58	4.67	5.31	5.99
70[b]	5.03	5.31	5.41	6.16	6.94
60[b]	5.56	5.88	5.99	6.77	7.63
50[b]	6.22	6.55	6.67	7.55	8.52
40[c]	7.57	7.96	8.10	9.12	10.2
30[c]	11.2	11.8	12.0	13.5	15.2
25[f]	16.3[aa]	17.2	17.6	19.8	22.2
15[h]	—	12.3	—	—	—
19[g]	10.8[aa]	11.0	11.1	11.7	12.3
5[e]	9.43	9.61	9.68	10.2	10.7
1[b]	4.46	4.60	4.65	5.00	5.37
Alloy—Aluminum-Magnesium					
Wt % Al					
99[c]	2.96	3.18	3.26	3.82	4.39
95[c]	5.05	5.28	5.36	5.93	6.51
90[c]	7.52	7.76	7.85	8.43	9.02
85	—	—	—	—	—
80	—	—	—	—	—
70	—	—	—	—	—
60	—	—	—	—	—
50	—	—	—	—	—
40	—	—	—	—	—
30	—	—	—	—	—
25	—	—	—	—	—
15	—	—	—	—	—
10[b]	17.1	17.4	17.6	18.4	19.2
5[b]	13.1	13.4	13.5	14.3	15.2
1[a]	5.92	6.25	6.37	7.20	8.03
Alloy—Copper-Gold					
Wt % Cu					
99[c]	1.73	1.86[aa]	1.91[aa]	2.24[aa]	2.58[aa]
95[c]	2.41	2.54[aa]	2.59[aa]	2.92[aa]	3.26[aa]
90[c]	3.29	4.42[aa]	3.46[aa]	3.79[aa]	4.12[aa]
85[c]	4.20	4.33	4.38[aa]	4.71[aa]	5.05[aa]
80[c]	5.15	5.28	5.32	5.65	5.99
70[c]	7.12	7.25	7.30	7.64	7.99
60[c]	9.18	9.13	9.36	9.70	10.05
50[c]	11.07	11.20	11.25	11.60	11.94
40[c]	12.70	12.85	12.90[aa]	13.27[aa]	13.65[aa]
30[c]	13.77	13.93	13.99[aa]	14.38[aa]	14.78[aa]
25[c]	13.93	14.09	14.14	14.54	14.94
15[c]	12.75	12.91	12.96[aa]	13.36[aa]	13.77
10[c]	10.70	10.86	10.91	11.31	11.72
5[c]	7.25	7.41[aa]	7.46	7.87	8.28
1[c]	3.40	3.57	3.62	4.03	4.45

	273 K	293 K	300 K	350 K	400 K
Alloy—Copper-Nickel					
Wt % Cu					
99[c]	2.71	2.85	2.91	3.27	3.62
95[c]	7.60	7.71	7.82	8.22	8.62
90[c]	13.69	13.89	13.96	14.40	14.81
85[c]	19.63	19.83	19.90	2032	20.70
80[c]	25.46	25.66	25.72	26.12[aa]	26.44[aa]
70[i]	36.67	36.72	36.76	36.85	36.89
60[i]	45.43	45.38	45.35	45.20	45.01
50[i]	50.19	50.05	50.01	49.73	49.50
40[c]	47.42	47.73	47.82	48.28	48.49
30[i]	40.19	41.79	42.34	44.51	45.40
25[c]	33.46	35.11	35.69	39.67[aa]	42.81[aa]
15[c]	22.00	23.35	23.85	27.60	31.38
10[c]	16.65	17.82	18.26	21.51	25.19
5[c]	11.49	12.50	12.90	15.69	18.78
1[c]	7.23	8.08	8.37	10.63[aa]	13.18[aa]
Alloy—Copper-Palladium					
Wt % Cu					
99[c]	2.10	2.23	2.27	2.59	2.92
95[c]	4.21	4.35	4.40	4.74	5.08
90[c]	6.89	7.03	7.08	7.41	7.74
85[c]	9.48	9.61	9.66	10.01	10.36
80[c]	11.99	12.12	12.16	12.51[aa]	12.87
70[c]	16.87	17.01	17.06	17.41	17.78
60[c]	21.73	21.87	21.92	22.30	22.69
50[c]	27.62	27.79	27.86	28.25	28.64
40[c]	35.31	35.51	35.57	36.03	36.47
30[c]	46.50	46.66	46.71	47.11	47.47
25[c]	46.25	46.45	46.52	46.99[aa]	47.43[aa]
15[c]	36.52	36.99	37.16	38.28	39.35
10[c]	28.90	29.51	29.73	31.19[aa]	32.56[aa]
5[c]	20.00	20.75	21.02	22.84[aa]	24.54[aa]
1[c]	11.90	12.67	12.93[aa]	14.82[aa]	16.68[aa]
Alloy—Copper-Zinc					
Wt % Cu					
99[b]	1.84	1.97	2.02	2.36	2.71
95[b]	2.78	2.92	2.97	3.33	3.69
90[b]	3.66	3.81	3.86	4.25	4.63
85[b]	4.37	4.54	4.60	5.02	5.44
80[b]	5.01	5.19	5.26	5.71	6.17
70[b]	5.87	6.08	6.15	6.67	7.19
60	—	—	—	—	—
50	—	—	—	—	—
40	—	—	—	—	—
30	—	—	—	—	—
25	—	—	—	—	—
15	—	—	—	—	—
10	—	—	—	—	—
5	—	—	—	—	—
1	—	—	—	—	—

	273 K	293 K	300 K	350 K	400 K
	Alloy—Gold-Palladium				
Wt % Au					
99[c]	2.69	2.86	2.91	3.32	3.73
95[c]	5.21	5.35	5.41	5.79	6.17
90[i]	8.01	8.17	8.22	8.56	8.93
85[b]	10.50[aa]	10.66	10.72[aa]	11.100[aa]	11.48[aa]
80[b]	12.75	12.93	12.99	13.45	13.93
70[c]	18.23	18.46	18.54	19.10	19.67
60[b]	26.70	26.94	27.01	27.63[aa]	28.23[aa]
50[a]	27.23	27.63	27.76	28.64[aa]	29.42[aa]
40[a]	24.65	25.23	25.42	26.74	27.95
30[b]	20.82	21.49	21.72	23.35	24.92
25[b]	18.86	19.53	19.77	21.51	23.19
15[a]	15.08	15.77	16.01	17.80	19.61
10[a]	13.25	13.95	14.20[aa]	16.00[aa]	17.81[aa]
5[a]	11.49[aa]	12.21	12.46[aa]	14.26[aa]	16.07[aa]
1[a]	10.07	10.85[aa]	11.12[aa]	12.99[aa]	14.80[aa]

	273 K	293 K	300 K	350 K	400 K
	Alloy—Iron-Nickel				
Wt % Fe					
99[a]	10.9	12.0	12.4	—	18.7
95[c]	18.7	19.9	20.2	—	26.8
90[c]	24.2	25.5	25.9	—	33.2
85[c]	27.8	29.2	29.7	—	37.3
80[c]	30.1	31.6	32.2	—	40.0
70[b]	32.3	33.9	34.4	—	42.4
60[c]	53.8	57.1	58.2	—	73.9
50[d]	28.4	30.6	31.4	—	43.7
40[d]	19.6	21.6	22.5	—	34.0
30[c]	15.3	17.1	17.7	—	27.4
25[b]	14.3	15.9	16.4	—	25.1
15[c]	12.6	13.8	14.2	—	21.1
10[c]	11.4	12.5	12.9	—	18.9
5[c]	9.66	10.6	10.9	—	16.1[aa]
1[b]	7.17	7.94	8.12	—	12.8

	273 K	293 K	300 K	350 K	400 K
	Alloy—Gold-Silver				
Wt % Au					
99[b]	2.58	2.75	2.80[aa]	3.22[aa]	3.63[aa]
95[a]	4.58	4.74	4.79	5.19	5.59
90[j]	6.57	6.73	6.78	7.19	7.58
85[j]	8.14	8.30	8.36[aa]	8.75	9.15
80[j]	9.34	9.50	9.55	9.94	10.33
70[j]	10.70	10.86	10.91	11.29	11.68[aa]
60[j]	10.92	11.07	11.12	11.50	11.87
50[j]	10.23	10.37	10.42	10.78	11.14
40[j]	8.92	9.06	9.11	9.46[aa]	9.81
30[a]	7.34	7.47	7.52	7.85	8.19
25[a]	6.46	6.59	6.63	6.96	7.30[aa]
15[a]	4.55	4.67	4.72	5.03	5.34
10[a]	3.54	3.66	3.71	4.00	4.31
5[i]	2.52	2.64[aa]	2.68[aa]	2.96[aa]	3.25[aa]
1[b]	1.69	1.80	1.84[aa]	2.12[aa]	2.42[aa]

	273 K	293 K	300 K	350 K	400 K
	Alloy—Silver-Palladium				
Wt % Ag					
99[b]	1.891	2.007	2.049	2.35	2.66
95[b]	3.58	3.70	3.74	4.04	4.34
90[b]	5.82	5.94	5.98	6.28	6.59
85[k]	7.92[aa]	8.04[aa]	8.08	8.38[aa]	8.68[aa]
80[k]	10.01	10.13	10.17	10.47	10.78
70[k]	14.53	14.65	14.69	14.99	15.30
60[i]	20.9	21.1	21.2	21.6	22.0
50[k]	31.2	31.4	31.5	32.0	32.4
40[m]	42.2	42.2	42.2	42.3	42.3
30[b]	40.4	40.6	40.7	41.3	41.7
25[k]	36.67[aa]	37.06	37.19	38.1[aa]	38.8[aa]
15[i]	27.08[aa]	26.68[aa]	27.89[aa]	29.3[aa]	30.6[aa]
10[i]	21.69	22.39	22.63	24.3	25.9
5[b]	15.98	16.72	16.98	18.8[aa]	20.5[aa]
1[a]	11.06	11.82	12.08[aa]	13.92[aa]	15.70[aa]

[a] Uncertainty in resistivity is ±2%.
[b] Uncertainty in resistivity is ±3%.
[c] Uncertainty in resistivity is ±5%.
[d] Uncertainty in resistivity is ±7% below 300 K and ±5% at 300 and 400 K.
[e] Uncertainty in resistivity is ±7%.
[f] Uncertainty in resistivity is ±8%.
[g] Uncertainty in resistivity is ±10%.
[h] Uncertainty in resistivity is ±12%.
[i] Uncertainty in resistivity is ±4%.
[j] Uncertainty in resistivity is ±1%.
[k] Uncertainty in resistivity is ±3% up to 300 K and ±4% above 300 K.
[m] Uncertainty in resistivity is ±2% up to 300 K and ±4% above 300 K.
[a] Crystal usually a mixture of α-hep and fcc lattice.
[aa] In temperature range where no experimental data are available.

Resistivity of Selected Ceramics (Listed by Ceramic)

Ceramic	Resistivity (Ω-cm)
Borides	
Chromium diboride (CrB_2)	21×10^{-6}
Hafnium diboride (HfB_2)	$10–12 \times 10^{-6}$ at room temp.
Tantalum diboride (TaB_2)	68×10^{-6}
Titanium diboride (TiB_2) (polycrystalline)	
85% dense	$26.5–28.4 \times 10^{-6}$ at room temp.
85% dense	9.0×10^{-6} at room temp.
100% dense, extrapolated values	$8.7–14.1 \times 10^{-6}$ at room temp.
	3.7×10^{-6} at liquid air temp.
Titanium diboride (TiB_2) (monocrystalline)	
Crystal length 5 cm, 39 deg. and 59 deg. orientation with respect to growth axis	$6.6 \pm 0.2 \times 10^{-6}$ at room temp.
Crystal length 1.5 cm, 16.5 deg. and 90 deg. orientation with respect to growth axis	$6.7 \pm 0.2 \times 10^{-6}$ at room temp.
Zirconium diboride (ZrB_2)	9.2×10^{-6} at 20° C
	1.8×10^{-6} at liquid air temp.
Carbides: boron carbide (B_4C)	0.3–0.8

Dielectric Constants

Dielectric Constants of Solids

These data refer to temperatures in the range 17–22°C.

Material	Freq. (Hz)	Dielectric Constant	Material	Freq. (Hz)	Dielectric Constant
Acetamide	4×10^8	4.0	Diphenylmethane	4×10^8	2.7
Acetanilide	—	2.9	Dolomite ⊥optic axis	10^8	8.0
Acetic acid (2° C)	4×10^8	4.1	Dolomite ‖	10^8	6.8
Aluminum oleate	4×10^8	2.40	Ferrous oxide (15° C)	10^8	14.2
Ammonium bromide	10^8	7.1	Iodine	10^8	4
Ammonium chloride	10^8	7.0	Lead acetate	10^8	2.6
Antimony trichloride	10^8	5.34	Lead carbonate (15° C)	10^8	18.6
Apatite ⊥optic axis	3×10^8	9.50	Lead chloride	10^8	4.2
Apatite ‖ optic axis	3×10^8	7.41	Lead monoxide (15° C)	10^8	25.9
Asphalt	$<3 \times 10^4$	2.68	Lead nitrate	6×10^7	37.7
Barium chloride (anhyd.)	6×10^7	11.4	Lead oleate	4×10^8	3.27
Barium chloride ($2H_3O$)	6×10^7	9.4	Lead sulfate	10^4	14.3
Barium nitrate	6×10^7	5.9	Lead sulfide (15° C)	16^8	17.9
Barium sulfate (15° C)	10^8	11.40	Malachite (mean)	10^{12}	7.2
Beryl ⊥optic axis	10^4	7.02	Mercuric chloride	10^8	3.2
Beryl ‖ optic axis	10^4	6.08	Mercurous chloride	10^8	9.4
Calcite ⊥optic axis	10^4	8.5	Naphthalene	4×10^8	2.52
Calcite ‖ optic axis	10^4	8.0	Phenanthrene	4×10^8	2.80
Calcium carbonate	10^4	6.14	Phenol (10° C)	4×10^8	4.3
Calcium fluoride	10^4	7.36	Phosphorus, red	10^8	4.1
Calcium sulfate ($2H_2O$)	10^4	5.66	Phosphorus, yellow	10^8	3.6
Cassiterite ⊥optic axis	10^{12}	23.4	Potassium aluminum sulfate	10^8	3.8
Cassiterite ‖ optic axis	10^{12}	24	Potassium carbonate (15° C)	10^8	5.6
d-Cocaine	5×10^8	3.10	Potassium chlorate	6×10^7	5.1
Cupric oleate	4×10^8	2.80	Potassium chloride	10^4	5.03
Cupric oxide (15° C)	10^8	18.1	Potassium chromate	6×10^7	7.3
Cupric sulfate (anhyd.)	6×10^7	10.3	Potassium iodide	6×10^7	5.6
Cupric sulfate ($5H_2O$)	6×10^7	7.8	Potassium nitrate	6×10^7	5.0
Diamond	10^8	5.5	Potassium sulfate	6×10^7	5.9

Material	Freq. (Hz)	Dielectric Constant	Material	Freq. (Hz)	Dielectric Constant
Quartz ⊥optic axis	3×10^7	4.34	Sodium carbonate ($10H_2O$)	6×10^7	5.3
Quartz ∥ optic axis	3×10^7	4.27	Sodium chloride	10^4	6.12
Resorcinol	4×10^8	3.2	Sodium nitrate	—	5.2
Ruby ⊥optic axis	10^4	13.27	Sodium oleate	4×10^8	2.75
Ruby ∥ optic axis	10^4	11.28	Sodium perchlorate	6×10^7	5.4
Rutile ⊥optic axis	10^8	86	Sucrose (mean)	3×10^8	3.32
Rutile ∥ optic axis	10^8	170	Sulfur (mean)	—	4.0
Selenium	10^8	6.6	Thallium chloride	10^4	46.9
Silver bromide	10^4	12.2	*p*-Toluidine	4×10^8	3.0
Silver chloride	10^4	11.2	Tourmaline ⊥optic axis	10^4	7.10
Silver cyanide	10^4	5.6	Tourmaline ∥ optic axis	10^4	6.3
Smithsonite ⊥optic axis	10^{12}	9.3	Urea	4×10^8	3.5
Smithsonite ∥ optic axis	10^{10}	9.4	Zircon ⊥, ∥	10^8	12
Sodium carbonate (anhyd.)	6×10^7	8.4			

Dielectric Constants of Ceramics

Material	Dielectric Constant 10^4 Hz	Dielectric strength Volts/mil	Volume Resistivity Ohm-cm (23° C)	Loss Factor[a]
Alumina	4.5–8.4	40–160	10^{11}–10^{14}	0.0002–0.01
Corderite	4.5–5.4	40–250	10^{12}–10^{14}	0.004–0.012
Forsterite	6.2	240	10^{14}	0.0004
Porcelain (dry process)	6.0–8.0	40–240	10^{12}–10^{14}	0.0003–0.02
Porcelain (wet process)	6.0–7.0	90–400	10^{12}–10^{14}	0.006–0.01
Porcelain, zircon	7.1–10.5	250–400	10^{13}–10^{15}	0.0002–0.008
Steatite	5.5–7.5	200–400	10^{13}–10^{15}	0.0002–0.004
Titanates (Ba, Sr, Ca, Mg, and Pb)	15–12.000	50–300	10^8–10^{13}	0.0001–0.02
Titanium dioxide	14–110	100–210	10^{13}–10^{18}	0.0002–0.005

Dielectric Constants of Glasses

Type	Dielectric Constant At 100 MHz (20° C)	Volume Resistivity (350° C megohm-cm)	Loss Factor[a]
Corning 0010	6.32	10	0.015
Corning 0080	6.75	0.13	0.058
Corning 0120	6.65	100	0.012
Pyrex 1710	6.00	2,500	0.025
Pyrex 3320	4.71	—	0.019
Pyrex 7040	4.65	80	0.013
Pyrex 7050	4.77	16	0.017
Pyrex 7052	5.07	25	0.019
Pyrex 7060	4.70	13	0.018
Pyrex 7070	4.00	1,300	0.0048
Vycor 7230	3.83	—	0.0061
Pyrex 7720	4.50	16	0.014
Pyrex 7740	5.00	4	0.040
Pyrex 7750	4.28	50	0.011
Pyrex 7760	4.50	50	0.0081
Vycor 7900	3.9	130	0.0023
Vycor 7910	3.8	1,600	0.00091
Vycor 7911	3.8	4,000	0.00072
Corning 8870	9.5	5,000	0.0085
G. E. Clear (silica glass)	3.81	4,000–30,000	0.00038
Quartz (fused)	3.75 4.1 (1 MHz)	—	0.0002 (1 MHz)

[a] Power factor × dielectric constant equals loss factor.

Properties of Semiconductors

H. Mike Harris *Georgia Tech Research Institute*

Semiconducting Properties of Selected Materials

Substance	Minimum Energy Gap (eV) R.T.	0 K	dE_g/dT $\times 10^4$ eV/°C	dE_g/dP $\times 10^6$ eV·cm²/kg	Density of States eleCtron Effective Mass m_{d_n} (m_o)	Electron Mobility and Temperature Dependence μ_n (cm²/V·s)	$-x$	Density of States Hole Effec-Tive Mass m_{d_p} (m_o)	Hole Mobility and Temperature Dependence μ_p (cm²/V·s)	$-x$
Si	1.107	1.153	–2.3	–2.0	1.1	1,900	2.6	0.56	500	2.3
Ge	0.67	0.744	–3.7	±7.3	0.55	3,80	1.66	0.3	1,820	2.33
αSn	0.08	0.094	–0.5		0.02	2,500	1.65	0.3	2,400	2.0
Te	0.33				0.68	1,100		0.19	560	
III–V Compounds										
AlAs	2.2	2.3				1,200			420	
AlSb	1.6	1.7	–3.5	–1.6	0.09	2..	1.5	0.4	500	1.8
GaP	2.24	2.40	–5.4	–1.7	0.35	300	1.5	0.5	150	1.5
GaAs	1.35	1.53	–5.0	+9.4	0.068	9,000	1.0	0.5	500	2.1
GaSb	0.67	0.78	–3.5	+12	0.050	5,000	2.0	0.23	1,400	0.9
InP	1.27	1.41	–4.6	+4.6	0.067	5,000	2.0		200	2.4
InAs	0.36	0.43	–2.8	+8	0.022	33,000	1.2	0.41	460	2.3
InSb	0.165	0.23	–2.8	+15	0.014	78,000	1.6	0.4	750	2.1
II–VI Compounds										
ZnO	3.2		–9.5	+0.6	0.38	180	1.5			
ZnS	3.54		–5.3	+5.7		180			5 (400°C)	
ZnSe	2.58	2.80	–7.2	+6		540			28	
ZnTe	2.26			+6		340			100	
CdO	2.5 ± 0.1		–6		0.1	120				
CdS	2.42		–5	+3.3	0.165	400		0.8		
CdSe	1.74	1.85	–4.6		0.13	650	1.0	0.6		
CdTe	1.44	1.56	–4.1	+8	0.14	1,200		0.35	50	
HgSe	0.30				0.030	20,000	2.0			
HgTe	0.15		–1		0.017	25,000		0.5	350	
Halite Structure Compounds										
PbS	0.37	0.28	+4		0.16	800		0.1	1,000	2.2
PbSe	0.26	0.16	+4		0.3	1,500		0.34	1,500	2.2
PbTe	0.25	0.19	+4	–7	0.21	1,600		0.14	750	2.2
Others										
ZnSb	0.50	0.56			0.15	10				1.5
CdSb	0.45	0.57	–5.4		0.15	300			2,000	1.5
Bi_2S_3	1.3					200			1,100	
Bi_2Se_3	0.27					600			675	
Bi_2Te_3	0.13		–0.95		0.58	1,200	1.68	1.07	510	1.95
Mg_2Si		0.77	–6.4		0.46	400	2.5		70	
Mg_2Ge		0.74	–9			280	2		110	
Mg_2Sn	0.21	0.33	–3.5		0.37	320			260	
Mg_3Sb_2		0.32				20			82	
Zn_3As_2	0.93					10	1.1		10	
Cd_3As_2	0.55				0.046	100,000	0.88			
GaSe	2.05		3.8						20	
GaTe	1.66	1.80	–3.6			14	–5			
InSe	1.8					9000				
TlSe	0.57		–3.9		0.3	30		0.6	20	1.5
$CdSnAs_2$	0.23				0.05	25,000	1.7			
Ga_2Te_3	1.1	1.55	–4.8							
α-In_2Te_3	1.1	1.2			0.7				50	1.1
β-In_2Te_3	1.0								5	
$Hg_5In_2Te_8$	0.5								11,000	
SnO_2									78	

Band Properties of Semiconductors

Part A. Data on Valence Bands of Semiconductors (Room Temperature)

Substance	Band curvature effective mass (expressed as fraction of free electron mass): Heavy Holes	Light Holes	"Split-off" Band Holes	Energy Separation of "Split-off" Band (eV)	Measured Light Hole Mobility ($cm^2/V \cdot s$)
Semiconductors with Valence Bands Maximum at the Center of the Brillouin Zone ("F")					
Si	0.52	0.16	0.25	0.044	500
Ge	0.34	0.043	0.08	0.3	1,820
Sn	0.3				2,400
AlAs					
AlSb	0.4			0.7	550
GaP				0.13	100
GaAs	0.8	0.12	0.20	0.34	400
GaSb	0.23	0.06		0.7	1,400
InP				0.21	150
InAs	0.41	0.025	0.083	0.43	460
InSb	0.4	0.015		0.85	750
CdTe	0.35				50
HgTe	0.5				350

Semiconductors with Multiple Valence Band Maxima

Substance	Number of Equivalent Valleys and Directions	Band Curvature Effective Masses: Longitudinal m_L	Transverse m_T	Anisotropy $K = m_L/m_T$	Measured (light) Hole Mobility $cm^2/V \cdot s$
PbSe	4 "L" [111]	0.095	0.047	2.0	1,500
PbTe	4 "L" [111]	0.27	0.02	10	750
Bi_2Te_3	6	0.207	~0.045	4.5	515

Part B. Data on Conduction Bands of Semiconductors (Room Temperature Data)

Single Valley Semiconductors

Substance	Energy Gap (eV)	Effective Mass (m_o)	Mobility ($cm^2/V \cdot s$)
GaAs	1.35	0.067	8,500
InP	1.27	0.067	5,000
InAs	0.36	0.022	33,000
InSb	0.165	0.014	78,000
CdTe	1.44	0.11	1,000

Multivalley Semiconductors

Substance	Energy Gap	Number of equivalent valleys and direction	Band curvature effective mass: Longitudinal m_L	Transverse m_T	Anisotropy $K = m_L/m_T$
Si	1.107	6 in [100] "Δ"	0.90	0.192	4.7
Ge	0.67	4 in [111] at "L"	1.588	~0.0815	19.5
GaSb	0.67	as Ge	~1.0	~0.2	~5
PbSe	0.26	4 in [111] at "L"	0.085	0.05	1.7
PbTe	0.25	4 in [111] at "L"	0.21	0.029	5.5
Bi_2Te_3	0.13	6			~0.05

References

1. L. Patrick and W. J. Choyke, *Physics Review Journal*, vol. 2, p. 2255, 1977.
2. W. J. Choyke, NATO ASI Ser. E, Applied Sciences, vol. 185, and references therein, 1990.
3. A. Suzuki, A. Ogura K. Furukawa, Y. Fuji, M. Shigeta and S. Nakajima, *Journal of Applied Physics*, vol. 64, p. 2818, 1988.
4. B. W. Wessels and J. C. Gatos, *Journal of Physics Chemical Solids*, vol. 15, p. 83, 1977.
5. M. M. Anikin, A. S. Zubrilov, A. A. Lebedev, A. M. Sterlchuck, and A. E. Cherenkov, *Fitz. Tekh. Polurpovdn.*, vol. 24, p. 467, 1992.
6. *CRC Materials Science and Engineering Handbook*, p. 304.
7. D. R. Lide, Properties of Semiconductors, *CRC Handbook of Chemistry and Physics*, pp. 12–98, 1999–2000 ed. CRC Press 2000.
8. http://vshields.jpl.nasa.gov/windchime.html
9. P. T. B. Schaffer and R. G. Naum, *Journal of Optical Society of America*, vol. 59, p. 1498, 1969.
10. P. T. B. Schaffer, *Applied Optics*, vol. 10, p. 1034, 1977.
11. M. N. Yoder, *Wide Bandgap Semiconductor Materials and Devices*, vol. 43, no.10, p.1634, 1966.
12. E. L. Kern, H. W. Hamill, H. W. Deem and H. D. Sheets, *Mater. Res. Bull.*, vol. 4, p. 107, 1969.
13. D. Morelli, J. Hermans, C. Bettz, W. S. Woo, G. L. Harris and C. Taylor, *Inst. Physics Conf. Ser.*, vol. 137, p. 313–6, 1990.
14. I. I. Parafenova, Y. M. Tairov and V. F. Tsvetkov, Sov. *Physics-Semiconductors*, vol. 24, no. 2, pp. 158–61, 1990.
15. http://www.ioffe.rssi.ru/SVA/NSM/Nano/index.html
16. C.W. Garland and K. C. Parks, *Journal Applied Physics*, vol. 33, p. 759, 1962.
17. http://mems.isi.edu/mems/materials/measurements.cgi?MATTAG=galliumarsenidegaas-bulk&PAGE_SIZE=20
18. M. C. Benjamin, C. Wang, R. F. Davis and R. J. Nemanich, *Applied Phys. Letter*, vol. 64, p. 3288, 1994.
19. S. N. Mohammad and H. Morkoc, *Prog. Quant. Electron.* vol. 20, p. 361, 1996.
20. A. S. Barker and M. Ilegems *Phys. Rev. B*, vol. 7, p. 743, 1973.
21. H. P. Maruska and J. J. Tietjen, A*ppl. Phys. Lett.* vol. 15, p. 327, 1969.

Resistance of Wires

The following table gives the approximate resistance of various metallic conductors. The values have been computed from the resistivities at 20°C, except as otherwise stated, and for the dimensions of wire indicated. Owing to differences in purity in the case of elements and of composition in alloys, the values can be considered only as approximations.

	Diameter			Diameter	
B. & S. Gauge	mm	mills 1 mil = .001 in	B. & S. gauge	mm	mills 1 mil = .001 in
10	2.588	101.9	26	0.4049	15.94
12	2.053	80.81	27	0.3606	14.20
14	1.628	64.08	28	0.3211	12.64
16	1.291	50.82	30	0.2546	10.03
18	1.024	40.30	32	0.2019	7.950
20	0.8118	31.96	34	0.1601	6.305
22	0.6438	25.35	36	0.1270	5.000
24	0.5106	20.10	40	0.07987	3.145

B. & S. No.	Ohms per cm	Ohms per ft	B. & S. No.	Ohms per cm	Ohms per ft
Advance (0° C) $Q = 48. \times 10^{-6}$ ohm cm			Brass $Q = 7.00 \times 10^{-6}$ ohm cm		
10	.000912	.0278	10	.000133	.00406
12	.00145	.0442	12	.000212	.00645
14	.00231	.0703	14	.000336	.0103
16	.00367	.112	16	.000535	.0163
18	.00583	.178	18	.000850	.0259
20	.00927	.283	20	.00135	.0412
22	.0147	.449	22	.00215	.0655
24	.0234	.715	24	.00342	.104
26	.0373	1.14	26	.00543	.166
27	.0470	1.43	27	.00686	.209
28	.0593	1.81	28	.00864	.263
30	.0942	2.87	30	.0137	.419
32	.150	4.57	32	.0219	.666
34	.238	7.26	34	.0348	1.06
36	.379	11.5	36	.0552	1.68
40	.958	29.2	40	.140	4.26
Aluminum $Q = 2.828 \times 10^{-6}$ ohm cm			Climax $Q = 87. \times 10^{-6}$ ohm cm		
10	.0000538	.00164	10	.00165	.0504
12	.0000855	.00260	12	.00263	.0801
14	.000136	.00414	14	.00418	.127
16	.000216	.00658	16	.00665	.203
18	.000344	.0105	18	.0106	.322
20	.000546	.0167	20	.0168	.512
22	.000869	.0265	22	.0267	.815
24	.00138	.0421	24	.0425	1.30
26	.00220	.0669	26	.0675	2.06
27	.00277	.0844	27	.0852	2.60
28	.00349	.106	28	.107	3.27
30	.00555	.169	30	.171	5.21
32	.00883	.269	32	.272	8.28
34	.0140	.428	34	.432	13.2
36	.0223	.680	36	.687	20.9
40	.0564	1.72	40	1.74	52.9
Constantan (0° C) $Q = 44.1 \times 10^{-6}$ ohm cm			Excello $Q = 92. \times 10^{-6}$ ohm cm		
10	.000838	.0255	10	.00175	.0533
12	.00133	.0406	12	.00278	.0847
14	.00212	.0646	14	.00442	.135
16	.00337	.103	16	.00703	.214
18	.00536	.163	18	.0112	.341
20	.00852	.260	20	.0178	.542
22	.0135	.413	22	.0283	.861
24	.0215	.657	24	.0449	1.37
26	.0342	1.04	26	.0714	2.18
27	.0432	1.32	27	.0901	2.75
28	.0545	1.66	28	.114	3.46
30	.0866	2.64	30	.181	5.51
32	.138	4.20	32	.287	8.75
34	.219	6.67	34	.457	13.9
36	.348	10.6	36	.726	22.1
40	.880	26.8	40	1.84	56.0

B. & S. No.	Ohms per cm	Ohms per ft	B. & S. No.	Ohms per cm	Ohms per ft
Copper, annealed $Q = 1.724 \times 10^{-6}$ ohm cm			German silver $Q = 33. \times 10^{-6}$ ohm cm		
10	.0000328	.000999	10	.000627	.0191
12	.0000521	.00159	12	.000997	.0304
14	.0000828	.00253	14	.00159	.0483
16	.000132	.00401	16	.00252	.0768
18	.000209	.00638	18	.00401	.122
20	.000333	.0102	20	.00638	.194
22	.000530	.0161	22	.0101	.309
24	.000842	.0257	24	.0161	.491
26	.00134	.0408	26	.0256	.781
27	.00169	.0515	27	.0323	.985
28	.00213	.0649	28	.0408	1.24
30	.00339	.103	30	.0648	1.97
32	.00538	.164	32	.103	3.14
34	.00856	.261	34	.164	4.99
36	.0136	.415	36	.260	.794
40	.0344	1.05	40	.659	20.1
Eureka (0° C) $Q = 47. \times 10^{-6}$ ohm cm			Gold $Q = 2.44 \times 10^{-6}$ ohm cm		
10	.000893	.0272	10	.0000464	.00141
12	.00142	.0433	12	.0000737	.00225
14	.00226	0.688	14	.000117	.00357
16	.00359	.109	16	.000186	.00568
18	.00571	.174	18	.000296	.00904
20	.00908	.277	20	.000471	.0144
22	.0144	.440	22	.000750	.0228
24	.0230	.700	24	.00119	.0363
26	.0365	1.11	26	.00189	.0577
27	.0460	1.40	27	.00239	.0728
28	.0580	1.77	28	.00301	.0918
30	.0923	2.81	30	.00479	.146
32	.147	4.47	32	.00762	.232
34	.233	7.11	34	.0121	.369
36	.371	11.3	36	.0193	.587
40	.938	28.6	40	.0487	1.48
Iron $Q = 10. \times 10^{-6}$ ohm cm			Manganin $Q = 44. \times 10^{-6}$ ohm cm		
10	.000190	.00579	10	.000836	.0255
12	.000302	.00921	12	.00133	.0405
14	.000481	.0146	14	.00211	.0644
16	.000764	.0233	16	.00336	.102
18	.00121	.0370	18	.00535	.163
20	.00193	.0589	20	.00850	.259
22	.00307	.0936	22	.0135	.412
24	.00489	.149	24	.0215	.655
26	.00776	.237	26	.0342	1.04
27	.00979	.299	27	.0431	1.31
28	.0123	.376	28	.0543	1.66
30	.0196	.598	30	.0864	2.63
32	.0312	.952	32	.137	4.19
34	.0497	1.51	34	.218	6.66
36	0.789	2.41	36	.347	10.6
40	.200	6.08	40	.878	26.8

B. & S. No.	Ohms per cm	Ohms per ft	B. & S. No.	Ohms per cm	Ohms per ft
Lead $Q = 22. \times 10^{-6}$ ohm cm			Molybdenum $Q = 5.7 \times 10^{-6}$ ohm cm		
10	.000418	.0127	10	.000108	.00330
12	.000665	.0203	12	.000172	.00525
14	.00106	.0322	14	.000274	.00835
16	.00168	.0512	16	.000435	.0133
18	.00267	.0815	18	.000693	.0211
20	.00425	.130	20	.00110	.0336
22	.00676	.206	22	.00175	.0534
24	.0107	.328	24	.00278	.0849
26	.0171	.521	26	.00443	.135
27	.0215	.657	27	.00558	.170
28	.0272	.828	28	.00704	.215
30	.0432	1.32	30	.0112	.341
32	.0687	2.09	32	.0178	.542
34	.109	3.33	34	.0283	.863
36	.174	5.29	36	.0450	1.37
40	.439	13.4	40	.114	3.47
Magnesium $Q = 4.6 \times 10^{-6}$ ohm cm			Monel Metal $Q = 42. \times 10^{-6}$ ohm cm		
10	.0000874	.00267	10	.000798	.0243
12	.000139	.00424	12	.00127	.0387
14	.000221	.00674	14	.00202	.0615
16	.000351	.0107	16	.00321	.0978
18	.000559	.0170	18	.00510	.156
20	.000889	.0271	20	.00811	.247
22	.00141	.0431	22	.0129	.393
24	.00225	.0685	24	.0205	.625
26	.00357	.109	26	.0326	.994
27	.00451	.137	27	.0411	1.25
28	.00568	.173	28	.0519	1.58
30	.00903	.275	30	.0825	2.51
32	.0144	.438	32	.131	4.00
34	.0228	.696	34	.209	6.36
36	.0363	1.11	36	.331	10.1
40	.0918	2.80	40	.838	25.6
*Nichrome $Q = 150. \times 10^{-6}$ ohm cm			Silver (18° C) $Q = 1.629 \times 10^{-6}$ ohm cm		
10	.0021281	.06488	10	.0000310	.000944
12	.0033751	.1029	12	.0000492	.00150
14	.0054054	.1648	14	.0000783	.00239
16	.0085116	.2595	16	.000124	.00379
18	.0138383	.4219	18	.000198	.00603
20	.0216218	.6592	20	.000315	.00959
22	.0346040	1.055	22	.000500	.0153
24	.0548088	1.671	24	.000796	.0243
26	.0875760	2.670	26	.00126	.0386
28	.1394328	4.251	27	.00160	.0486
30	.2214000	6.750	28	.00201	.0613
32	.346040	10.55	30	.00320	.0975
34	.557600	17.00	32	.00509	.155
36	.885600	27.00	34	.00809	.247
38	1.383832	42.19	36	.0129	.392
40	2.303872	70.24	40	.0325	.991

B. & S. No.	Ohms per cm	Ohms per ft	B. & S. No.	Ohms per cm	Ohms per ft
Nickel $Q = 7.8 \times 10^{-6}$ ohm cm			Steel, piano wire (0° C) $Q = 11.8 \times 10^{-6}$ ohm cm		
10	.000148	.00452	10	.000224	.00684
12	.000236	.00718	12	.000357	.0109
14	.000375	.0114	14	.000567	.0173
16	.000596	.0182	16	.000901	.0275
18	.000948	.0289	18	.00143	.0437
20	.00151	.0459	20	.00228	.0695
22	.00240	.0730	22	.00363	.110
24	.00381	.116	24	.00576	.176
26	.00606	.185	26	.00916	.279
27	.00764	.233	27	.0116	.352
28	.00963	.294	28	.0146	.444
30	.0153	.467	30	.0232	.706
32	.0244	.742	32	.0368	1.12
34	.0387	1.18	34	.0586	1.79
36	.0616	1.88	36	.0931	2.84
40	.156	4.75	40	.236	7.18
Platinum $Q = 10. \times 10^{-6}$ ohm cm			Steel, invar (35% Ni) $Q = 81. \times 10^{-6}$ ohm cm		
10	.000190	.00579	10	.00154	.0469
12	.000302	.00921	12	.00245	.0746
14	.000481	.0146	14	.00389	.119
16	.000764	.0233	16	.00619	.189
18	.00121	.0370	18	.00984	.300
20	.00193	.0589	20	.0156	.477
22	.00307	.0936	22	.0249	.758
24	.00489	.149	24	.0396	1.21
26	.00776	.237	26	.0629	1.92
27	.00979	.299	27	.0793	2.42
28	.0123	.376	28	.100	3.05
30	.0196	.598	30	.159	4.85
32	.0312	.952	32	.253	7.71
34	.0497	1.51	34	.402	12.3
36	.0789	2.41	36	.639	19.5
40	.200	6.08	40	1.62	49.3
Tantalum $Q = 15.5 \times 10^{-6}$ ohm cm			Tungsten $Q = 5.51 \times 10^{-6}$ ohm cm		
10	.000295	.00898	10	.000105	.00319
12	.000468	.0143	12	.000167	.00508
14	.000745	.0227	14	.000265	.00807
16	.00118	.0361	16	.000421	.0128
18	.00188	.0574	18	.000669	.0204
20	.00299	.0913	20	.00106	.0324
22	.00476	.145	22	.00169	.0516
24	.00757	.231	24	.00269	.0820
26	.0120	.367	26	.00428	.130
27	.0152	.463	27	.00540	.164
28	.0191	.583	28	.00680	.207
30	.0304	.928	30	.0108	.330
32	.0484	1.47	32	.0172	.524
34	.0770	2.35	34	.0274	.834
36	.122	3.73	36	.0435	1.33
40	.309	9.43	40	.110	3.35

B. & S. No.	Ohms per cm	Ohms per ft	B. & S. No.	Ohms per cm	Ohms per ft
Tin $Q = 11.5 \times 10^{-6}$ ohm cm			Zinc (0° C) $Q = 5.75 \times 10^{-6}$ ohm cm		
10	.000219	.00666	10	.000109	.00333
12	.000348	.0106	12	.000174	.00530
14	.000553	.0168	14	.000276	.00842
16	.000879	.0268	16	.000439	.0134
18	.00140	.0426	18	.000699	.0213
20	.00222	.0677	20	.00111	.0339
22	.00353	.108	22	.00177	.0538
24	.00562	.171	24	.00281	.0856
26	.00893	.272	26	.00446	.136
27	.0113	.343	27	.00563	.172
28	.0142	.433	28	.00710	.216
30	.0226	.688	30	.0113	.344
32	.0359	1.09	32	.0180	.547
34	.0571	1.74	34	.0286	.870
36	.0908	2.77	36	.0454	1.38
40	.230	7.00	40	.115	3.50

Credits

Except for the Properties of Semiconductors section, material in Appendix A was reprinted from the following sources:

D. R. Lide, Ed., *CRC Handbook of Chemistry and Physics,* 76th ed., Boca Raton, Fla.: CRC Press, 1992: International System of Units (SI), conversion constants and multipliers (conversion of temperatures), symbols and terminology for physical and chemical quantities, fundamental physical constants, classification of electromagnetic radiation.

W. H. Beyer, Ed., *CRC Standard Mathematical Tables and Formulae,* 29th ed., Boca Raton, Fla.: CRC Press, 1991: Greek alphabet, conversion constants and multipliers (recommended decimal multiples and submultiples, metric to English, English to metric, general, temperature factors), physical constants, series expansion, integrals, the Fourier transforms, numerical methods, probability, positional notation.

R. J. Tallarida, *Pocket Book of Integrals and Mathematical Formulas,* 2nd ed., Boca Raton, Fla.: CRC Press, 1991: Elementary algebra and geometry; determinants, matrices, and linear systems of equations; trigonometry; analytic geometry; series; differential calculus; integral calculus; vector analysis; special functions; statistics; tables of probability and statistics; table of derivatives.

J. F. Pankow, *Aquatic Chemistry Concepts,* Chelsea, Mich.: Lewis Publishers, 1991: Periodic table of the elements.

J. Shackelford and W. Alexander, Eds., *CRC Materials Science and Engineering Handbook,* Boca Raton, Fla.: CRC Press, 1992: Electrical resistivity of selected alloy cast irons, resistivity of selected ceramics.

APPENDIX B

Microwave Engineering Appendix

John P. Wendler
Tyco Electronics
Wireless Network Solutions

Attenuator Design Values

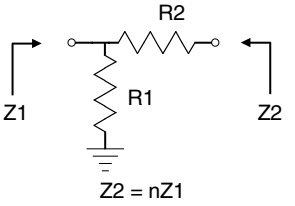


FIGURE B.1 Equivalent circuit for a minimum loss pad.

TABLE 1 Minimum Loss Matching Pad Resistance Values as a Function of Transformation Ratio For Z1 = 1 Ohm, 50 Ohms, and 75 Ohms

n Z2/Z1	r1 Z1 = 1	r2 Z1 = 1	R1 Z1 = 50	R2 Z1 = 50	R1 Z1 = 75	R2 Z1 = 75	Loss [dB]
1.1	3.3166	0.3317	165.8	16.6	248.7	24.9	2.7
1.2	2.4495	0.4899	122.5	24.5	183.7	36.7	3.8
1.3	2.0817	0.6245	104.1	31.2	156.1	46.8	4.5
1.4	1.8708	0.7483	93.5	37.4	140.3	56.1	5.2
1.5	1.7321	0.8660	86.6	43.3	129.9	65.0	5.7
1.6	1.6330	0.9798	81.6	49.0	122.5	73.5	6.2
1.7	1.5584	1.0909	77.9	54.5	116.9	81.8	6.6
1.8	1.5000	1.2000	75.0	60.0	112.5	90.0	7.0
1.9	1.4530	1.3077	72.6	65.4	109.0	98.1	7.3
2.0	1.4142	1.4142	70.7	70.7	106.1	106.1	7.7
2.1	1.3817	1.5199	69.1	76.0	103.6	114.0	8.0
2.2	1.3540	1.6248	67.7	81.2	101.6	121.9	8.2
2.3	1.3301	1.7292	66.5	86.5	99.8	129.7	8.5
2.4	1.3093	1.8330	65.5	91.7	98.2	137.5	8.7
2.5	1.2910	1.9365	64.5	96.8	96.8	145.2	9.0
2.6	1.2748	2.0396	63.7	102.0	95.6	153.0	9.2
2.7	1.2603	2.1424	63.0	107.1	94.5	160.7	9.4
2.8	1.2472	2.2450	62.4	112.2	93.5	168.4	9.6
2.9	1.2354	2.3473	61.8	117.4	92.7	176.1	9.8
3.0	1.2247	2.4495	61.2	122.5	91.9	183.7	10.0
3.1	1.2150	2.5515	60.7	127.6	91.1	191.4	10.1

TABLE 1 (continued)

n Z2/Z1	r1 Z1 = 1	r2 Z1 = 1	R1 Z1 = 50	R2 Z1 = 50	R1 Z1 = 75	R2 Z1 = 75	Loss [dB]
3.2	1.2060	2.6533	60.3	132.7	90.5	199.0	10.3
3.3	1.1978	2.7550	59.9	137.7	89.8	206.6	10.5
3.4	1.1902	2.8566	59.5	142.8	89.3	214.2	10.6
3.5	1.1832	2.9580	59.2	147.9	88.7	221.9	10.8
3.6	1.1767	3.0594	58.8	153.0	88.3	229.5	10.9
3.7	1.1706	3.1607	58.5	158.0	87.8	237.1	11.0
3.8	1.1650	3.2619	58.2	163.1	87.4	244.6	11.2
3.9	1.1597	3.3630	58.0	168.2	87.0	252.2	11.3
4.0	1.1547	3.4641	57.7	173.2	86.6	259.8	11.4
4.1	1.1500	3.5651	57.5	178.3	86.3	267.4	11.6
4.2	1.1456	3.6661	57.3	183.3	85.9	275.0	11.7
4.3	1.1415	3.7670	57.1	188.3	85.6	282.5	11.8
4.4	1.1376	3.8678	56.9	193.4	85.3	290.1	11.9
4.5	1.1339	3.9686	56.7	198.4	85.0	297.6	12.0
4.6	1.1304	4.0694	56.5	203.5	84.8	305.2	12.1
4.7	1.1271	4.1701	56.4	208.5	84.5	312.8	12.2
4.8	1.1239	4.2708	56.2	213.5	84.3	320.3	12.3
4.9	1.1209	4.3715	56.0	218.6	84.1	327.9	12.4
5.0	1.1180	4.4721	55.9	223.6	83.9	335.4	12.5
5.1	1.1153	4.5727	55.8	228.6	83.6	343.0	12.6
5.2	1.1127	4.6733	55.6	233.7	83.5	350.5	12.7
5.3	1.1102	4.7739	55.5	238.7	83.3	358.0	12.8
5.4	1.1078	4.8744	55.4	243.7	83.1	365.6	12.9
5.5	1.1055	4.9749	55.3	248.7	82.9	373.1	13.0
5.6	1.1034	5.0754	55.2	253.8	82.8	380.7	13.1
5.7	1.1013	5.1759	55.1	258.8	82.6	388.2	13.2
5.8	1.0992	5.2764	55.0	263.8	82.4	395.7	13.3
5.9	1.0973	5.3768	54.9	268.8	82.3	403.3	13.3
6.0	1.0954	5.4772	54.8	273.9	82.2	410.8	13.4
6.1	1.0937	5.5776	54.7	278.9	82.0	418.3	13.5
6.2	1.0919	5.6780	54.6	283.9	81.9	425.9	13.6
6.3	1.0903	5.7784	54.5	288.9	81.8	433.4	13.6
6.4	1.0887	5.8788	54.4	293.9	81.6	440.9	13.7
6.5	1.0871	5.9791	54.4	299.0	81.5	448.4	13.8
6.6	1.0856	6.0795	54.3	304.0	81.4	456.0	13.9
6.7	1.0842	6.1798	54.2	309.0	81.3	463.5	13.9
6.8	1.0828	6.2801	54.1	314.0	81.2	471.0	14.0
6.9	1.0814	6.3804	54.1	319.0	81.1	478.5	14.1
7.0	1.0801	6.4807	54.0	324.0	81.0	486.1	14.1
7.1	1.0789	6.5810	53.9	329.1	80.9	493.6	14.2
7.2	1.0776	6.6813	53.9	334.1	80.8	501.1	14.3
7.3	1.0764	6.7816	53.8	339.1	80.7	508.6	14.3
7.4	1.0753	6.8819	53.8	344.1	80.6	516.1	14.4
7.5	1.0742	6.9821	53.7	349.1	80.6	523.7	14.5
7.6	1.0731	7.0824	53.7	354.1	80.5	531.2	14.5
7.7	1.0720	7.1826	53.6	359.1	80.4	538.7	14.6
7.8	1.0710	7.2829	53.6	364.1	80.3	546.2	14.6
7.9	1.0700	7.3831	53.5	369.2	80.3	553.7	14.7
8.0	1.0690	7.4833	53.5	374.2	80.2	561.2	14.8
8.1	1.0681	7.5835	53.4	379.2	80.1	568.8	14.8
8.2	1.0672	7.6837	53.4	384.2	80.0	576.3	14.9
8.3	1.0663	7.7840	53.3	389.2	80.0	583.8	14.9
8.4	1.0654	7.8842	53.3	394.2	79.9	591.3	15.0
8.5	1.0646	7.9844	53.2	399.2	79.8	598.8	15.0
8.6	1.0638	8.0846	53.2	404.2	79.8	606.3	15.1

TABLE 1 (continued)

n Z2/Z1	r1 Z1 = 1	r2 Z1 = 1	R1 Z1 = 50	R2 Z1 = 50	R1 Z1 = 75	R2 Z1 = 75	Loss [dB]
8.7	1.0630	8.1847	53.1	409.2	79.7	613.9	15.2
8.8	1.0622	8.2849	53.1	414.2	79.7	621.4	15.2
8.9	1.0614	8.3851	53.1	419.3	79.6	628.9	15.3
9.0	1.0607	8.4853	53.0	424.3	79.5	636.4	15.3
9.1	1.0599	8.5855	53.0	429.3	79.5	643.9	15.4
9.2	1.0592	8.6856	53.0	434.3	79.4	651.4	15.4
9.3	1.0585	8.7858	52.9	439.3	79.4	658.9	15.5
9.4	1.0579	8.8859	52.9	444.3	79.3	666.4	15.5
9.5	1.0572	8.9861	52.9	449.3	79.3	674.0	15.6
9.6	1.0565	9.0863	52.8	454.3	79.2	681.5	15.6
9.7	1.0559	9.1864	52.8	459.3	79.2	689.0	15.7
9.8	1.0553	9.2865	52.8	464.3	79.1	696.5	15.7
9.9	1.0547	9.3867	52.7	469.3	79.1	704.0	15.7
10.0	1.0541	9.4868	52.7	474.3	79.1	711.5	15.8

$$R_1 = \frac{Z_1(n+\sqrt{n^2-n})}{n-1+\sqrt{n^2-n}} \qquad R_2 = Z_1\sqrt{n^2-n} \qquad \frac{P_O}{P_A} = \frac{1}{n}\left(\frac{1}{1+\sqrt{\frac{1}{n}}}\right)^2$$

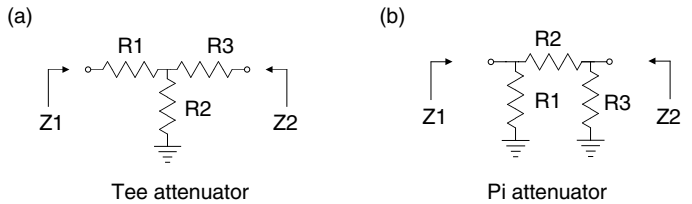


FIGURE B.2 (a) Equivalent circuit for a Tee attenuator; (b) Equivalent circuit for a Pi attenuator.

TABLE 2 Tee- and Pi-Pad Resistor Values for Zo = 1 Ohm and Zo = 50 Ohms

Loss [dB]	Voltage Atten	r1, r3, g1, g3 Z1 = Z2 = 1	r2, g2 Z1 = Z2 = 1	Tee R1, R3 Z1 = Z2 = 50	Tee R2 Z1 = Z2 = 50	Pi R1, R3 Z1 = Z2 = 50	Pi R2 Z1 = Z2 = 50
0.1	0.98855	0.0058	86.8570	0.3	4342.8	8686.0	0.6
0.2	0.97724	0.0115	43.4256	0.6	2171.3	4343.1	1.2
0.3	0.96605	0.0173	28.9472	0.9	1447.4	2895.6	1.7
0.4	0.95499	0.0230	21.7071	1.2	1085.4	2171.9	2.3
0.5	0.94406	0.0288	17.3622	1.4	868.1	1737.7	2.9
0.6	0.93325	0.0345	14.4650	1.7	723.2	1448.2	3.5
0.7	0.92257	0.0403	12.3950	2.0	619.7	1241.5	4.0
0.8	0.91201	0.0460	10.8420	2.3	542.1	1086.5	4.6
0.9	0.90157	0.0518	9.6337	2.6	481.7	966.0	5.2
1.0	0.89125	0.0575	8.6667	2.9	433.3	869.5	5.8
1.2	0.87096	0.0690	7.2153	3.4	360.8	725.0	6.9
1.4	0.85114	0.0804	6.1774	4.0	308.9	621.8	8.1
1.6	0.83176	0.0918	5.3981	4.6	269.9	544.4	9.3
1.8	0.81283	0.1032	4.7911	5.2	239.6	484.3	10.4
2.0	0.79433	0.1146	4.3048	5.7	215.2	436.2	11.6

TABLE 2 (continued)

Loss [dB]	Voltage Atten	r1, r3, g1, g3 Z1 = Z2 = 1	r2, g2 Z1 = Z2 = 1	Tee R1, R3 Z1 = Z2 = 50	Tee R2 Z1 = Z2 = 50	Pi R1, R3 Z1 = Z2 = 50	Pi R2 Z1 = Z2 = 50
2.2	0.77625	0.1260	3.9062	6.3	195.3	396.9	12.8
2.4	0.75858	0.1373	3.5735	6.9	178.7	364.2	14.0
2.6	0.74131	0.1486	3.2914	7.4	164.6	336.6	15.2
2.8	0.72444	0.1598	3.0490	8.0	152.5	312.9	16.4
3.0	0.70795	0.1710	2.8385	8.5	141.9	292.4	17.6
3.2	0.69183	0.1822	2.6539	9.1	132.7	274.5	18.8
3.4	0.67608	0.1933	2.4906	9.7	124.5	258.7	20.1
3.6	0.66069	0.2043	2.3450	10.2	117.3	244.7	21.3
3.8	0.64565	0.2153	2.2144	10.8	110.7	232.2	22.6
4.0	0.63096	0.2263	2.0966	11.3	104.8	221.0	23.8
4.2	0.61660	0.2372	1.9896	11.9	99.5	210.8	25.1
4.4	0.60256	0.2480	1.8921	12.4	94.6	201.6	26.4
4.6	0.58884	0.2588	1.8028	12.9	90.1	193.2	27.7
4.8	0.57544	0.2695	1.7206	13.5	86.0	185.5	29.1
5.0	0.56234	0.2801	1.6448	14.0	82.2	178.5	30.4
5.5	0.53088	0.3064	1.4785	15.3	73.9	163.2	33.8
6.0	0.50119	0.3323	1.3386	16.6	66.9	150.5	37.4
6.5	0.47315	0.3576	1.2193	17.9	61.0	139.8	41.0
7.0	0.44668	0.3825	1.1160	19.1	55.8	130.7	44.8
7.5	0.42170	0.4068	1.0258	20.3	51.3	122.9	48.7
8.0	0.39811	0.4305	0.9462	21.5	47.3	116.1	52.8
8.5	0.37584	0.4537	0.8753	22.7	43.8	110.2	57.1
9.0	0.35481	0.4762	0.8118	23.8	40.6	105.0	61.6
9.5	0.33497	0.4982	0.7546	24.9	37.7	100.4	66.3
10.0	0.31623	0.5195	0.7027	26.0	35.1	96.2	71.2
10.5	0.29854	0.5402	0.6555	27.0	32.8	92.6	76.3
11.0	0.28184	0.5603	0.6123	28.0	30.6	89.2	81.7
11.5	0.26607	0.5797	0.5727	29.0	28.6	86.3	87.3
12.0	0.25119	0.5985	0.5362	29.9	26.8	83.5	93.2
12.5	0.23714	0.6166	0.5025	30.8	25.1	81.1	99.5
13.0	0.22387	0.6342	0.4714	31.7	23.6	78.8	106.1
13.5	0.21135	0.6511	0.4425	32.6	22.1	76.8	113.0
14.0	0.19953	0.6673	0.4156	33.4	20.8	74.9	120.3
14.5	0.18836	0.6830	0.3906	34.1	19.5	73.2	128.0
15.0	0.17783	0.6980	0.3673	34.9	18.4	71.6	136.1
15.5	0.16788	0.7125	0.3455	35.6	17.3	70.2	144.7
16.0	0.15849	0.7264	0.3251	36.3	16.3	68.8	153.8
16.5	0.14962	0.7397	0.3061	37.0	15.3	67.6	163.3
17.0	0.14125	0.7525	0.2883	37.6	14.4	66.4	173.5
17.5	0.13335	0.7647	0.2715	38.2	13.6	65.4	184.1
18.0	0.12589	0.7764	0.2558	38.8	12.8	64.4	195.4
18.5	0.11885	0.7875	0.2411	39.4	12.1	63.5	207.4
19.0	0.11220	0.7982	0.2273	39.9	11.4	62.6	220.0
19.5	0.10593	0.8084	0.2143	40.4	10.7	61.8	233.4
20.0	0.10000	0.8182	0.2020	40.9	10.1	61.1	247.5
20.5	0.09441	0.8275	0.1905	41.4	9.5	60.4	262.5
21.0	0.08913	0.8363	0.1797	41.8	9.0	59.8	278.3
21.5	0.08414	0.8448	0.1695	42.2	8.5	59.2	295.0
22.0	0.07943	0.8528	0.1599	42.6	8.0	58.6	312.7
22.5	0.07499	0.8605	0.1508	43.0	7.5	58.1	331.5
23.0	0.07079	0.8678	0.1423	43.4	7.1	57.6	351.4
23.5	0.06683	0.8747	0.1343	43.7	6.7	57.2	372.4
24.0	0.06310	0.8813	0.1267	44.1	6.3	56.7	394.6
24.5	0.05957	0.8876	0.1196	44.4	6.0	56.3	418.2
25.0	0.05623	0.8935	0.1128	44.7	5.6	56.0	443.2

TABLE 2 (continued)

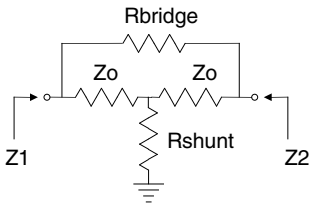
Loss [dB]	Voltage Atten	r1, r3, g1, g3 Z1 = Z2 = 1	r2, g2 Z1 = Z2 = 1	Tee R1, R3 Z1 = Z2 = 50	Tee R2 Z1 = Z2 = 50	Pi R1, R3 Z1 = Z2 = 50	Pi R2 Z1 = Z2 = 50
26.0	0.05012	0.9045	0.1005	45.2	5.0	55.3	497.6
27.0	0.04467	0.9145	0.0895	45.7	4.5	54.7	558.6
28.0	0.03981	0.9234	0.0797	46.2	4.0	54.1	627.0
29.0	0.03548	0.9315	0.0711	46.6	3.6	53.7	703.7
30.0	0.03162	0.9387	0.0633	46.9	3.2	53.3	789.8
31.0	0.02818	0.9452	0.0564	47.3	2.8	52.9	886.3
32.0	0.02512	0.9510	0.0503	47.5	2.5	52.6	994.6
33.0	0.02239	0.9562	0.0448	47.8	2.2	52.3	1116.1
34.0	0.01995	0.9609	0.0399	48.0	2.0	52.0	1252.5
35.0	0.01778	0.9651	0.0356	48.3	1.8	51.8	1405.4
36.0	0.01585	0.9688	0.0317	48.4	1.6	51.6	1577.0
37.0	0.01413	0.9721	0.0283	48.6	1.4	51.4	1769.5
38.0	0.01259	0.9751	0.0252	48.8	1.3	51.3	1985.5
39.0	0.01122	0.9778	0.0224	48.9	1.1	51.1	2227.8
40.0	0.01000	0.9802	0.0200	49.0	1.0	51.0	2499.8
41.0	0.00891	0.9823	0.0178	49.1	0.9	50.9	2804.8
42.0	0.00794	0.9842	0.0159	49.2	0.8	50.8	3147.1
43.0	0.00708	0.9859	0.0142	49.3	0.7	50.7	3531.2
44.0	0.00631	0.9875	0.0126	49.4	0.6	50.6	3962.1
45.0	0.00562	0.9888	0.0112	49.4	0.6	50.6	4445.6

Note: P_i values are duals of Tee values.

$$a = \sqrt{\frac{P_{z_2}}{P_{z_1}}} \qquad R_{1T} = \left(\frac{2}{(1-a^2)} - 1\right) Z_1 - \frac{2a}{(1-a^2)}\sqrt{Z_1 Z_2}$$

$$R_{2T} = 2\sqrt{Z_1 Z_2}\,\frac{a}{(1-a^2)}$$

$$R_{3T} = \left(\frac{2}{(1-a^2)} - 1\right) Z_1 - \frac{2a}{(1-a^2)}\sqrt{Z_1 Z_2}$$



Bridged-tee attenuator
Z1=Z2=Zo

FIGURE B.3 Equivalent circuit for a Bridged-T attenuator.

TABLE 3 Bridged-T Attenuator Resistance Values for Zo = 1 Ohm, 50 Ohms, 75 Ohms

Loss [dB]	Voltage Atten	Bridge Arm Z1 = Z2 = 1	Shunt Arm Z1 = Z2 = 1	Bridge Arm Z1 = Z2 = 50	Shunt Arm Z1 = Z2 = 50	Bridge Arm Z1 = Z2 = 75	Shunt Arm Z1 = Z2 = 75
0.1	0.98855	0.0116	86.3599	0.6	4318.0	6477.0	0.9
0.2	0.97724	0.0233	42.9314	1.2	2146.6	3219.9	1.7
0.3	0.96605	0.0351	28.4558	1.8	1422.8	2134.2	2.6
0.4	0.95499	0.0471	21.2186	2.4	1060.9	1591.4	3.5
0.5	0.94406	0.0593	16.8766	3.0	843.8	1265.7	4.4
0.6	0.93325	0.0715	13.9822	3.6	699.1	1048.7	5.4
0.7	0.92257	0.0839	11.9151	4.2	595.8	893.6	6.3

TABLE 3 (continued)

Loss [dB]	Voltage Atten	Bridge Arm Z1 = Z2 = 1	Shunt Arm Z1 = Z2 = 1	Bridge Arm Z1 = Z2 = 50	Shunt Arm Z1 = Z2 = 50	Bridge Arm Z1 = Z2 = 75	Shunt Arm Z1 = Z2 = 75
0.8	0.91201	0.0965	10.3650	4.8	518.3	777.4	7.2
0.9	0.90157	0.1092	9.1596	5.5	458.0	687.0	8.2
1.0	0.89125	0.1220	8.1955	6.1	409.8	614.7	9.2
1.2	0.87096	0.1482	6.7498	7.4	337.5	506.2	11.1
1.4	0.85114	0.1749	5.7176	8.7	285.9	428.8	13.1
1.6	0.83176	0.2023	4.9440	10.1	247.2	370.8	15.2
1.8	0.81283	0.2303	4.3428	11.5	217.1	325.7	17.3
2.0	0.79433	0.2589	3.8621	12.9	193.1	289.7	19.4
2.2	0.77625	0.2882	3.4692	14.4	173.5	260.2	21.6
2.4	0.75858	0.3183	3.1421	15.9	157.1	235.7	23.9
2.6	0.74131	0.3490	2.8656	17.4	143.3	214.9	26.2
2.8	0.72444	0.3804	2.6289	19.0	131.4	197.2	28.5
3.0	0.70795	0.4125	2.4240	20.6	121.2	181.8	30.9
3.2	0.69183	0.4454	2.2450	22.3	112.2	168.4	33.4
3.4	0.67608	0.4791	2.0872	24.0	104.4	156.5	35.9
3.6	0.66069	0.5136	1.9472	25.7	97.4	146.0	38.5
3.8	0.64565	0.5488	1.8221	27.4	91.1	136.7	41.2
4.0	0.63096	0.5849	1.7097	29.2	85.5	128.2	43.9
4.2	0.61660	0.6218	1.6082	31.1	80.4	120.6	46.6
4.4	0.60256	0.6596	1.5161	33.0	75.8	113.7	49.5
4.6	0.58884	0.6982	1.4322	34.9	71.6	107.4	52.4
4.8	0.57544	0.7378	1.3554	36.9	67.8	101.7	55.3
5.0	0.56234	0.7783	1.2849	38.9	64.2	96.4	58.4
5.5	0.53088	0.8836	1.1317	44.2	56.6	84.9	66.3
6.0	0.50119	0.9953	1.0048	49.8	50.2	75.4	74.6
6.5	0.47315	1.1135	0.8981	55.7	44.9	67.4	83.5
7.0	0.44668	1.2387	0.8073	61.9	40.4	60.5	92.9
7.5	0.42170	1.3714	0.7292	68.6	36.5	54.7	102.9
8.0	0.39811	1.5119	0.6614	75.6	33.1	49.6	113.4
8.5	0.37584	1.6607	0.6021	83.0	30.1	45.2	124.6
9.0	0.35481	1.8184	0.5499	90.9	27.5	41.2	136.4
9.5	0.33497	1.9854	0.5037	99.3	25.2	37.8	148.9
10.0	0.31623	2.1623	0.4625	108.1	23.1	34.7	162.2
10.5	0.29854	2.3497	0.4256	117.5	21.3	31.9	176.2
11.0	0.28184	2.5481	0.3924	127.4	19.6	29.4	191.1
11.5	0.26607	2.7584	0.3625	137.9	18.1	27.2	206.9
12.0	0.25119	2.9811	0.3354	149.1	16.8	25.2	223.6
12.5	0.23714	3.2170	0.3109	160.8	15.5	23.3	241.3
13.0	0.22387	3.4668	0.2884	173.3	14.4	21.6	260.0
13.5	0.21135	3.7315	0.2680	186.6	13.4	20.1	279.9
14.0	0.19953	4.0119	0.2493	200.6	12.5	18.7	300.9
14.5	0.18836	4.3088	0.2321	215.4	11.6	17.4	323.2
15.0	0.17783	4.6234	0.2163	231.2	10.8	16.2	346.8
15.5	0.16788	4.9566	0.2018	247.8	10.1	15.1	371.7
16.0	0.15849	5.3096	0.1883	265.5	9.4	14.1	398.2
16.5	0.14962	5.6834	0.1759	284.2	8.8	13.2	426.3
17.0	0.14125	6.0795	0.1645	304.0	8.2	12.3	456.0
17.5	0.13335	6.4989	0.1539	324.9	7.7	11.5	487.4
18.0	0.12589	6.9433	0.1440	347.2	7.2	10.8	520.7
18.5	0.11885	7.4140	0.1349	370.7	6.7	10.1	556.0
19.0	0.11220	7.9125	0.1264	395.6	6.3	9.5	593.4
19.5	0.10593	8.4406	0.1185	422.0	5.9	8.9	633.0
20.0	0.10000	9.0000	0.1111	450.0	5.6	8.3	675.0
20.5	0.09441	9.5925	0.1042	479.6	5.2	7.8	719.4
21.0	0.08913	10.2202	0.0978	511.0	4.9	7.3	766.5
21.5	0.08414	10.8850	0.0919	544.3	4.6	6.9	816.4

TABLE 3 (continued)

Loss [dB]	Voltage Atten	Bridge Arm Z1 = Z2 = 1	Shunt Arm Z1 = Z2 = 1	Bridge Arm Z1 = Z2 = 50	Shunt Arm Z1 = Z2 = 50	Bridge Arm Z1 = Z2 = 75	Shunt Arm Z1 = Z2 = 75
22.0	0.07943	11.5893	0.0863	579.5	4.3	6.5	869.2
22.5	0.07499	12.3352	0.0811	616.8	4.1	6.1	925.1
23.0	0.07079	13.1254	0.0762	656.3	3.8	5.7	984.4
23.5	0.06683	13.9624	0.0716	698.1	3.6	5.4	1047.2
24.0	0.06310	14.8489	0.0673	742.4	3.4	5.1	1113.7
24.5	0.05957	15.7880	0.0633	789.4	3.2	4.8	1184.1
25.0	0.05623	16.7828	0.0596	839.1	3.0	4.5	1258.7
26.0	0.05012	18.9526	0.0528	947.6	2.6	4.0	1421.4
27.0	0.04467	21.3872	0.0468	1069.4	2.3	3.5	1604.0
28.0	0.03981	24.1189	0.0415	1205.9	2.1	3.1	1808.9
29.0	0.03548	27.1838	0.0368	1359.2	1.8	2.8	2038.8
30.0	0.03162	30.6228	0.0327	1531.1	1.6	2.4	2296.7
31.0	0.02818	34.4813	0.0290	1724.1	1.5	2.2	2586.1
32.0	0.02512	38.8107	0.0258	1940.5	1.3	1.9	2910.8
33.0	0.02239	43.6684	0.0229	2183.4	1.1	1.7	3275.1
34.0	0.01995	49.1187	0.0204	2455.9	1.0	1.5	3683.9
35.0	0.01778	55.2341	0.0181	2761.7	0.9	1.4	4142.6
36.0	0.01585	62.0957	0.0161	3104.8	0.8	1.2	4657.2
37.0	0.01413	69.7946	0.0143	3489.7	0.7	1.1	5234.6
38.0	0.01259	78.4328	0.0127	3921.6	0.6	1.0	5882.5
39.0	0.01122	88.1251	0.0113	4406.3	0.6	0.9	6609.4
40.0	0.01000	99.0000	0.0101	4950.0	0.5	0.8	7425.0
41.0	0.00891	111.2018	0.0090	5560.1	0.4	0.7	8340.1
42.0	0.00794	124.8925	0.0080	6244.6	0.4	0.6	9366.9
43.0	0.00708	140.2538	0.0071	7012.7	0.4	0.5	10519.0
44.0	0.00631	157.4893	0.0063	7874.5	0.3	0.5	11811.7
45.0	0.00562	176.8279	0.0057	8841.4	0.3	0.4	13262.1

Return Loss, Reflection Coefficient, VSWR, and Mismatch Loss

TABLE 4 Conversion Between Return Loss, Reflection Coefficient, VSWR, and Mismatch Loss

Return Loss [dB]	Reflection Coefficient (Rho)	VSWR ( ):1	Mismatch Loss [dB]	Return Loss [dB]	Reflection Coefficient (Rho)	VSWR ( ):1	Mismatch Loss [dB]
Infinite	0.0000	1.00	0.00	33.00	0.0224	1.05	0.00
50.00	0.0032	1.01	0.00	32.00	0.0251	1.05	0.00
49.00	0.0035	1.01	0.00	31.00	0.0282	1.06	0.00
48.00	0.0040	1.01	0.00	30.00	0.0316	1.07	0.00
47.00	0.0045	1.01	0.00	29.00	0.0355	1.07	0.01
46.00	0.0050	1.01	0.00	28.00	0.0398	1.08	0.01
45.00	0.0056	1.01	0.00	27.00	0.0447	1.09	0.01
44.00	0.0063	1.01	0.00	26.00	0.0501	1.11	0.01
43.00	0.0071	1.01	0.00	25.00	0.0562	1.12	0.01
42.00	0.0079	1.02	0.00	24.00	0.0631	1.13	0.02
41.00	0.0089	1.02	0.00	23.00	0.0708	1.15	0.02
40.00	0.0100	1.02	0.00	22.00	0.0794	1.17	0.03
39.00	0.0112	1.02	0.00	21.00	0.0891	1.20	0.03
38.00	0.0126	1.03	0.00	20.00	0.1000	1.22	0.04
37.00	0.0141	1.03	0.00	19.50	0.1059	1.24	0.05
36.00	0.0158	1.03	0.00	19.00	0.1122	1.25	0.06
35.00	0.0178	1.04	0.00	18.50	0.1189	1.27	0.06
34.00	0.0200	1.04	0.00	18.00	0.1259	1.29	0.07
17.50	0.1334	1.31	0.08	6.02	0.5000	3.00	1.25
17.00	0.1413	1.33	0.09	6.00	0.5012	3.01	1.26
16.50	0.1496	1.35	0.10	5.80	0.5129	3.11	1.33
16.00	0.1585	1.38	0.11	5.60	0.5248	3.21	1.40
15.50	0.1679	1.40	0.12	5.40	0.5370	3.32	1.48
15.00	0.1778	1.43	0.14	5.20	0.5495	3.44	1.56
14.50	0.1884	1.46	0.16	5.11	0.5556	3.50	1.60
14.00	0.1995	1.50	0.18	5.00	0.5623	3.57	1.65
13.50	0.2113	1.54	0.20	4.80	0.5754	3.71	1.75
13.00	0.2239	1.58	0.22	4.60	0.5888	3.86	1.85
12.50	0.2371	1.62	0.25	4.44	0.6000	4.00	1.94
12.00	0.2512	1.67	0.28	4.40	0.6026	4.03	1.96
11.50	0.2661	1.73	0.32	4.20	0.6166	4.22	2.08
11.00	0.2818	1.78	0.36	4.00	0.6310	4.42	2.20
10.50	0.2985	1.85	0.41	3.93	0.6364	4.50	2.25
10.00	0.3162	1.92	0.46	3.80	0.6457	4.64	2.34
9.80	0.3236	1.96	0.48	3.60	0.6607	4.89	2.49
9.60	0.3311	1.99	0.50	3.52	0.6667	5.00	2.55
9.54	0.3333	2.00	0.51	3.40	0.6761	5.17	2.65
9.40	0.3388	2.03	0.53	3.20	0.6918	5.49	2.83
9.20	0.3467	2.06	0.56	3.00	0.7079	5.85	3.02
9.00	0.3548	2.10	0.58	2.80	0.7244	6.26	3.23
8.80	0.3631	2.14	0.61	2.60	0.7413	6.73	3.46
8.60	0.3715	2.18	0.65	2.40	0.7586	7.28	3.72
8.40	0.3802	2.23	0.68	2.20	0.7762	7.94	4.01
8.20	0.3890	2.27	0.71	2.00	0.7943	8.72	4.33
8.00	0.3981	2.32	0.75	1.80	0.8128	9.69	4.69
7.80	0.4074	2.37	0.79	1.74	0.8182	10.00	4.81
7.60	0.4169	2.43	0.83	1.60	0.8318	10.89	5.11
7.40	0.4266	2.49	0.87	1.40	0.8511	12.44	5.60
7.36	0.4286	2.50	0.88	1.20	0.8710	14.50	6.17
7.20	0.4365	2.55	0.92	1.00	0.8913	17.39	6.87
7.00	0.4467	2.61	0.97	0.80	0.9120	21.73	7.74
6.80	0.4571	2.68	1.02	0.60	0.9333	28.96	8.89
6.60	0.4677	2.76	1.07	0.40	0.9550	43.44	10.56
6.40	0.4786	2.84	1.13	0.20	0.9772	86.86	13.47
6.20	0.4898	2.92	1.19	0.00	1.0000	Infinite	Infinite

Notes:

1. Return Loss = -20*log(|Rho|)
2. Mismatch Loss = -10*log(1-|Rho|^2)
3. VSWR = (1+|Rho|)/(1-|Rho|)

Waveguide Components

TABLE 5 Waveguide Performance and Dimensions

EIA WR-	Mil-W-85E RG( )/U	TE10 Cutoff Frequency [GHz]	Recommended Frequency Range		Theoretical Attenuation		Inside Dimensions		Tolerance	Outside Dimensions		Tolerance	Wall Thickness	Material	Contact Flange	Choke Flange	Cover Flange	Hole Pattern Figure
			Min [GHz]	Max [GHz]	Fmin [dB/100']	Fmax [dB/100']	a [inches]	b [Inches]	± [Inches]	[Inches]	[Inches]	± [Inches]	[Inches]					
3	139	173.5726	220.00	325.00	503.90	352.59	0.0340	0.0170	0.00020	4.156	(Diameter)	0.001		Silver				
4	137	137.2434	170.00	260.00	371.25	246.94	0.0430	0.0215	0.00020	3.156	(Diameter)	0.001		Silver				
5	135	115.7151	140.00	220.00	303.47	190.96	0.0510	0.0255	0.00025	2.156	(Diameter)	0.001		Silver				
7	136	90.7918	110.00	170.00	210.19	133.39	0.0650	0.0325	0.00025	1.156	(Diameter)	0.001		Silver				
8	138	73.7683	90.00	140.00	151.38	97.26	0.080	0.040	0.0003	0.156	(Diameter)	0.001		Silver				
10		59.0147	75.00	110.00			0.100	0.050	0.0005	0.180	0.130	0.002	0.040					1
12		48.3727	60.00	90.00	122.71	82.37	0.122	0.061	0.0005	0.202	0.141	0.002	0.040	Brass			387	1
12	99	48.3727	60.00	90.00	77.46	51.99	0.122	0.061	0.0005	0.202	0.141	0.002	0.040	Silver				1
15		39.8748	50.00	75.00	89.78	61.41	0.148	0.074	0.0010	0.228	0.154	0.002	0.040	Brass			385	1
15	98	39.8748	50.00	75.00	56.67	38.76	0.148	0.074	0.0010	0.228	0.154	0.002	0.040	Silver				1
19		31.3908	40.00	60.00			0.188	0.094	0.0010	0.268	0.174	0.002	0.040					1
22		26.3458	33.00	50.00	48.33	32.89	0.224	0.112	0.0010	0.304	0.192	0.002	0.040	Brass			383	1
22	97	26.3458	33.00	50.00	30.50	20.76	0.224	0.112	0.0010	0.304	0.192	0.002	0.040	Silver				1
28		21.0767	26.50	40.00	28.00	19.20	0.280	0.140	0.0015	0.360	0.220	0.002	0.040	Aluminum				1,2
28		21.0767	26.50	40.00	34.32	23.53	0.280	0.140	0.0015	0.360	0.220	0.002	0.040	Brass		600	599	1,2
28	96	21.0767	26.50	40.00	21.66	14.85	0.280	0.140	0.0015	0.360	0.220	0.002	0.040	Silver				1,2
34		17.3573	22.00	33.00			0.340	0.170	0.0020	0.420	0.250	0.003	0.040					
42	121	14.0511	18.00	26.50	16.86	12.40	0.420	0.170	0.0020	0.500	0.250	0.003	0.040	Aluminum	425	598	597	1,2
42	53	14.0511	18.00	26.50	20.66	15.19	0.420	0.170	0.0020	0.500	0.250	0.003	0.040	Brass	425	596	595	1,2
42	66	14.0511	18.00	26.50	13.04	9.59	0.420	0.170	0.0020	0.500	0.250	0.003	0.040	Silver	425	596A	595	1,2
51		11.5715	15.00	22.00			0.510	0.255	0.0025	0.590	0.335	0.003	0.040					2
62		9.4879	12.40	18.00	7.88	5.80	0.622	0.311	0.0025	0.702	0.391	0.003	0.040	Aluminum				2
62	91	9.4879	12.40	18.00	9.66	7.11	0.622	0.311	0.0025	0.702	0.391	0.003	0.040	Brass		541	419	2
62	107	9.4879	12.40	18.00	6.10	4.49	0.622	0.311	0.0025	0.702	0.391	0.003	0.040	Silver				2
75		7.8686	10.00	15.00			0.750	0.375	0.003	0.850	0.475	0.003	0.050					2
90	67	6.5572	8.20	12.40	5.29	3.66	0.900	0.400	0.003	1.000	0.500	0.003	0.050	Aluminum		136A	135	2
90	52	6.5572	8.20	12.40	6.48	4.49	0.900	0.400	0.003	1.000	0.500	0.003	0.050	Brass		40A	39	2
112	68	5.2598	7.05	10.00	3.39	2.63	1.122	0.497	0.004	1.250	0.625	0.004	0.064	Aluminum		137A	138	2
112	51	5.2598	7.05	10.00	4.15	3.23	1.122	0.497	0.004	1.250	0.625	0.004	0.064	Brass		52A	51	2
137	106	4.3014	5.85	8.20	2.42	1.91	1.372	0.622	0.004	1.500	0.750	0.004	0.064	Aluminum		440A	441	
137	50	4.3014	5.85	8.20	2.96	2.34	1.372	0.622	0.004	1.500	0.750	0.004	0.064	Brass		343A	344	
159		3.7116	4.90	7.05			1.590	0.795	0.004	1.718	0.923	0.004	0.064					

TABLE 5 (continued) Waveguide Performance and Dimensions

EIA WR-	Mil-W-85E RG( )/U	TE10 Cutoff Frequency [GHz]	Recommended Frequency Range Min [GHz]	Recommended Frequency Range Max [GHz]	Theoretical Attenuation Fmin [dB/100']	Theoretical Attenuation Fmax [dB/100']	Inside Dimensions a [inches]	Inside Dimensions b [Inches]	Tolerance ± [Inches]	Outside Dimensions [Inches]	Outside Dimensions [Inches]	Tolerance ± [Inches]	Wall Thickness [Inches]	Material	Contact Flange	Choke Flange	Cover Flange	Hole Pattern Figure
187	95	3.1525	3.95	5.85	1.70	1.18	1.872	0.872	0.005	2.000	1.000	0.005	0.064	Aluminum		406A	407	
187	49	3.1525	3.95	5.85	2.09	1.45	1.872	0.872	0.005	2.000	1.000	0.005	0.064	Brass		148B	149A	
229		2.5771	3.30	4.90			2.290	1.145	0.005	2.418	1.273	0.005	0.064					
284	75	2.0780	2.60	3.95	0.91	0.62	2.840	1.340	0.005	3.000	1.500	0.005	0.080	Aluminum		585	584	
284	48	2.0780	2.60	3.95	1.11	0.76	2.840	1.340	0.005	3.000	1.500	0.005	0.080	Brass		54B	53	
340	113	1.7357	2.20	3.30	0.65	0.45	3.400	1.700	0.005	3.560	1.860	0.005	0.080	Aluminum			554	
340	112	1.7357	2.20	3.30	0.80	0.56	3.400	1.700	0.005	3.560	1.860	0.005	0.080	Brass			553	
430	105	1.3724	1.70	2.60	0.48	0.32	4.300	2.150	0.005	4.460	2.310	0.005	0.080	Aluminum			437A	
430	104	1.3724	1.70	2.60	0.59	0.39	4.300	2.150	0.005	4.460	2.310	0.005	0.080	Brass			435A	
510		1.1572	1.45	2.20			5.100	2.550	0.005	5.260	2.710	0.005	0.080					
650	103	0.9079	1.12	1.70	0.26	0.17	6.500	3.250	0.005	6.660	3.410	0.005	0.080	Aluminum			418A	
650	69	0.9079	1.12	1.70	0.32	0.21	6.500	3.250	0.005	6.660	3.410	0.005	0.080	Brass			417A	
770	205	0.7664	0.96	1.45	0.20	0.13	7.700	3.850	0.005	7.950	4.100	0.005	0.125	Aluminum				
975	204	0.6053	0.75	1.12	0.14	0.09	9.750	4.875	0.010	10.000	5.125	0.010	0.125	Aluminum				
1150	203	0.5132	0.64	0.96	0.11	0.07	11.500	5.750	0.015	11.750	6.000	0.015	0.125	Aluminum				
1500	202	0.3934	0.49	0.75	0.07	0.05	15.000	7.500	0.015	15.250	7.750	0.015	0.125	Aluminum				
1800	201	0.3279	0.41	0.63	0.05	0.04	18.000	9.000	0.020	18.250	9.250	0.020	0.125	Aluminum				
2100		0.2810	0.35	0.53	0.04	0.03	21.000	10.500	0.020	21.250	10.750	0.020	0.125	Aluminum				
2300		0.2566	0.32	0.49	0.04	0.03	23.000	11.500	0.020	23.250	11.750	0.020	0.125	Aluminum				

Notes:

1. Conductivity of 63.0e6 Mhos/Meter used for Silver.
2. Conductivity of 37.7e6 Mhos/Meter used for Aluminum.
3. Conductivity of 25.1e6 Mhos/Meter used for Brass.
4. Loss is inversely proportional to the square root of conductivity.

Source:

1. Balanis, C.A., *Advanced Engineering Electromagnetics*, John Wiley & Sons, New York, 1990.
2. Catalog, Microwave Development Labs, Natick, MA.
3. Catalog, Aerowave Inc, Medford, MA.
4. Catalog, Formcraft Tool Co, Chicago, IL.
5. Catalog, Penn Engineering Components, No. Hollywood, CA.

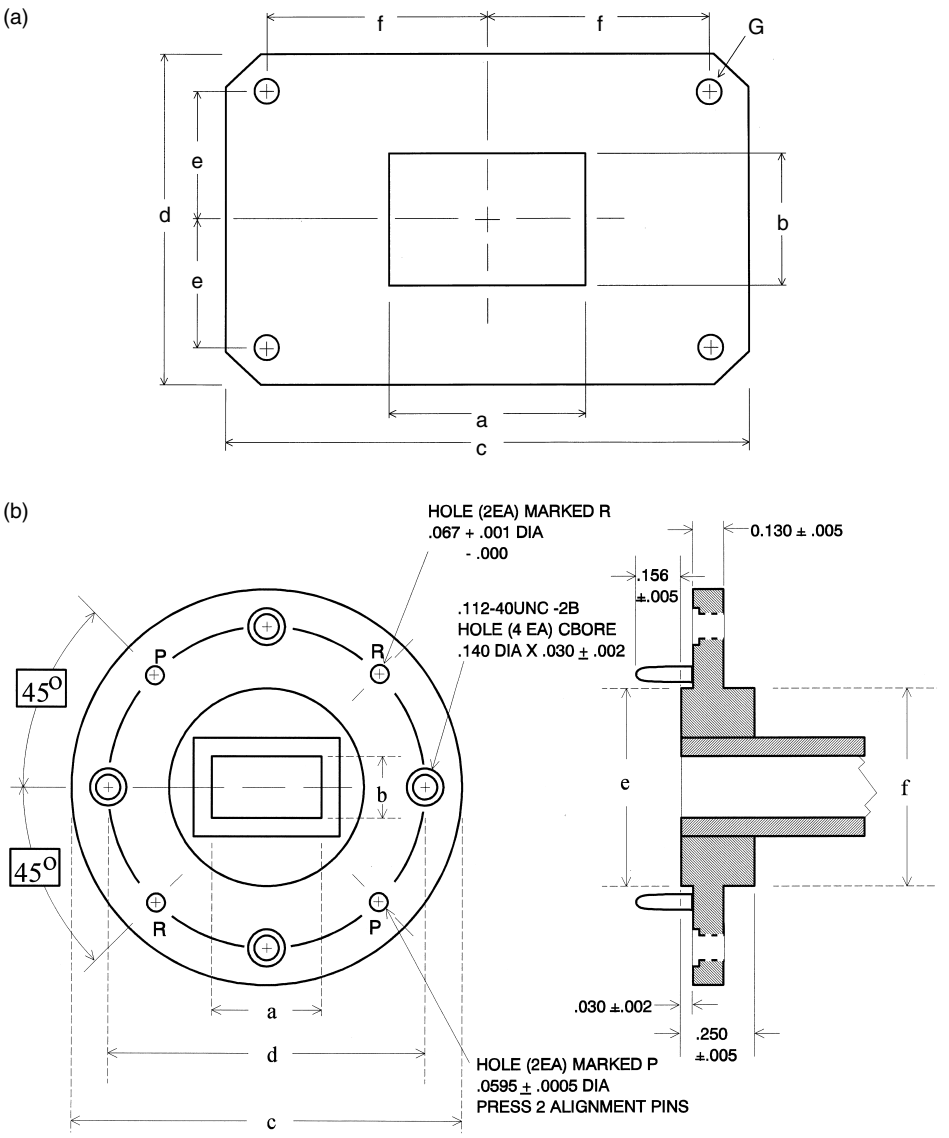


FIGURE B.4 (a) Standard waveguide flange dimensions (rectangular flange); (b) Standard waveguide flange dimensions (circular flange).

TABLE 6 Waveguide Flange Dimensions

Rectangular Flange Dimensions

WR-()	a [inches] (Ref.)	b [inches] (Ref.)	c [inches] ±0.015	d [inches] ±0.015	e [inches] ±0.005	f [inches] ±0.005	G [inches-dia] ±0.0015
28	0.280	0.140	0.750	1.750	0.265	0.250	0.1175
42	0.420	0.170	0.875	0.875	0.335	0.320	0.1175
51	0.510	0.255	1.313	1.313	0.497	0.478	0.1445
62	0.620	0.311	1.313	1.313	0.478	0.497	0.1445
75	0.750	0.375	1.500	2.500	0.560	0.520	0.1445
90	0.900	0.400	1.625	1.625	0.610	0.640	0.1705
112	1.122	0.497	1.875	1.875	0.676	0.737	0.1705

Circular Flange Dimensions

WR-()	a [inches] (Ref.)	b [inches] (Ref.)	c [inches] +0.000 –0.002	d [inches] BSC.	e [inches] ±0.005	f [inches] ±0.005
10	0.1	0.05	0.75	0.5625	0.375	0.312
12	0.122	0.061	0.75	0.5625	0.375	0.312
15	0.148	0.074	0.75	0.5625	0.375	0.312
19	0.188	0.094	1.125	0.9375	0.5	0.468
22	0.224	0.112	1.125	0.9375	0.5	0.468
28	0.28	0.14	1.125	0.9375	0.5	0.468
42	0.42	0.17	1.125	0.9375	0.625	0.625

Coaxial Cables

TABLE 7 Flexible Coax Specifications

RG-()/U	Mil-C-17/()	Zo	Loss dB/100' @1 MHz	Loss dB/100' @10 MHz	Loss dB/100' @100 MHz	Loss dB/100' @1000 MHz	Center Conductor	Outer Conductor	Jacket	Outside Diameter [Inches]	Dielectric	Velocity Factor [%]	Capacitance pF/foot	Dielectric Core Diameter [inches]	Maximum Voltage [RMS]
8		52	0.16	0.56	1.9	7.4	#13 Stranded Bare Copper .058 Dia	Bare Copper Braid, 97%	Black PVC	0.405	Polyethylene	66	29.5	0.285	3700
8A	[3] 163	52	0.16	0.56	1.9	7.4	#13 Stranded Bare Copper .058 Dia	Bare Copper Braid, 97%	Black PVC, Noncontaminating	0.405	Polyethylene	66	29.5	0.285	3700
9		51	0.18	0.62	2.1	8.2	#13 Stranded Silver Coated Copper .086 Dia	Double Braid, Silver Coated Inner, Bare Copper Outer, 97%	Gray PVC, Noncontaminating	0.42	Polyethylene	66	30	0.28	3700
9B	[4] 164	50					#13 Stranded Silver Coated Copper .089 Dia	Double Braid, Silver Coated Inner, Bare Copper Outer, 97%	PVC	0.36	Polyethylene	66		0.285	3701
11	[6]	75	0.19	0.66	2	7.1	#18 Stranded Tinned Copper .048	Bare Copper Braid, 97%	Black PVC	0.405	Flame Retardant Semi-Foam Polyethylene	66	20.5	0.285	300
11A		75	0.19	0.66	2	8.5	#18 Stranded Tinned Copper .048	Bare Copper Braid, 97%	Black PVC, Noncontaminating	0.405	Polyethylene	66	20.5	0.285	3700
58	[28]	50	0.42	1.5	5.4	22.8	#20 Tinned Copper .035 Dia	Tinned Copper Braid, 95%	Black PVC, Noncontaminating	0.193	Polyethylene	66	30.8	0.116	1400
58A		50	0.42	1.5	5.4	22.8	#20 Solid Bare Copper .035 Dia	Tinned Copper Braid, 95%	Black PVC	0.193	Polyethylene	66	30.8	0.116	1400
58C	155	50	0.42	1.5	5.4	22.8	#20 Tinned Copper .035 Dia	Tinned Copper Braid, 95%	Black PVC, Noncontaminating	0.193	Polyethylene	66	30.8	0.116	1400
59	[29]	75	0.6	1.1	3.4	12	#23 Solid Bare Copper Covered Steel .023 Dia	Bare Copper Braid, 95%	Black PVC, Noncontaminating	0.241	Polyethylene	66	20.5	0.146	1700
59B		75	0.6	1.1	3.4	12	#23 Solid Bare Copper Covered Steel .023 Dia	Bare Copper Braid, 95%	Black PVC, Noncontaminating	0.241	Polyethylene	66	20.5	0.146	1700
62A	[30]	93	0.25	0.85	2.7	8.7	#22 Solid Bare Copper Covered Steel .023 Dia	Bare Copper Braid, 95%	Black PVC, Noncontaminating	0.242	Semi-solid Polyethylene	84	13.5	0.146	750

TABLE 7 (continued)

RG-()/U	Mil-C-17/()	Zo	Loss dB/100' @1 MHz	Loss dB/100' @10 MHz	Loss dB/100' @100 MHz	Loss dB/100' @1000 MHz	Center Conductor	Outer Conductor	Jacket	Outside Diameter [Inches]	Dielectric	Velocity Factor [%]	Capacitance pF/foot	Dielectric Core Diameter [inches]	Maximum Voltage [RMS]
62B	[91] 97	93	0.31	0.9	2.9	11	#24 Solid Bare Copper Covered Steel .025 Dia	Bare Copper Braid, 95%	Black PVC, Noncontaminating	0.242	Semi-solid Polyethylene	84	13.5	0.146	750
63	[31]	125	0.19	0.52	1.5	5.8	#22 Solid Bare Copper Covered Steel .025 Dia	Bare Copper Braid, 97%	Black PVC, Noncontaminating	0.405	Semi-solid Polyethylene	84	9.7	0.285	750
71	[90]	93	0.25	0.85	2.7	8.7	#22 Solid Bare Copper Covered Steel .025 Dia	Double Braid, Tinned Copper Outer, Bare Copper Inner, 98%	Black Polyethylene	0.245	Semi-solid Polyethylene	84	13.5	0.146	750
122	[54] 157	50	0.4	1.7	7	29	#22 Stranded Tinned Copper .030 Dia	Tinned Copper Braid, 95%	Black PVC, Noncontaminating	0.16	Polyethylene	66	30.8	0.096	1400
141	[59] 170	50					#18 Solid Silver Coated Copper Covered Steel .037 Dia	Silver Coated Copper Braid, 94%	Fluorinated Ethylene-Propylene	0.17	TFE Teflon	69.5		0.116	1400
141A		50	0.34	1.1	3.9	13.5	#18 Solid Silver Coated Copper Covered Steel .037 Dia	Silver Coated Copper Braid, 94%	Tinted Brown Fiberglass	0.187	TFE Teflon	69.5	29.2	0.116	1400
142	[60] 158	50	0.34	1.1	3.9	13.5	#18 Solid Silver Coated Copper Covered Steel .037 Dia	Double Silver Coated Copper Braid, 94%	Tinted Brown Fluorinated Ethylene-Propylene	0.187	TFE Teflon	69.5	29.2	0.116	1400
174	[119] 173	50	1.9	3.3	8.4	34	#26 Stranded Bare Copper Covered Steel .019 Dia	Tinned Copper Braid, 90%	Black PVC Jacket	0.11	Polyethylene	66	30.8	0.059	1100
178	[93] 169	50					#30 Stranded Silver Coated Copper Covered Steel .012 Dia	Silver Coated Copper Braid, 96%	Fluorinated Ethylene-Propylene	0.071	TFE Teflon	69.5		0.033	750
178B		50	2.6	5.6	14	46	#30 Solid Silver Coated Copper Covered Steel .012 Dia	Silver Coated Copper Braid, 96%	White Fluorinated Ethylene-Propylene	0.071	TFE Teflon	69.5	29.2	0.033	750

179	[94]	75	3	5.3	10	24	#30 Solid Silver Coated Copper Covered Steel .012 Dia	Silver Coated Copper Braid, 95%	Tinted Brown Fluorinated Ethylene-Propylene	0.1	TFE Teflon	69.5	19.5	0.062	900
180	[95]	95	2.4	3.3	5.7	17	#30 Solid Silver Coated Copper Covered Steel .012 Dia	Silver Coated Copper Braid, 95%	Tinted Brown Fluorinated Ethylene-Propylene	0.141	TFE Teflon	69.5	15.4	0.102	1100
187	[68] 94	75					#30 Solid Silver Coated Copper Covered Steel .012 Dia	Silver Coated Copper Braid, 92.3%	Fluorinated Ethylene-Propylene	0.1	TFE Teflon	69.5		0.063	900
212	[73] 162	50	0.26	0.83	2.7	9.8	#15.5 Solid Silver Coated Copper .0556 Dia	Double Silver Coated Copper Braid, 95%	Black PVC Noncontaminating	0.332	Polyethylene	66	30.8	0.185	2200
213	[74] 163	50	0.18	0.62	2.1	8.2	#13 Stranded Bare Copper .089 Dia	Bare Copper Braid 97%	Black PVC Noncontaminating	0.405	Polyethylene	66	30.8	0.285	3700
214	[75] 164	50	0.17	0.55	1.9	8	#13 Stranded Silver Coated Copper .089 Dia	Double Silver Coated Copper Braid, 97%	Black PVC Noncontaminating	0.425	Polyethylene	66	30.8	0.285	3700
216	[77]	75	0.19	0.66	2	7.1	#18 Stranded Tinned Copper .048	Double Bare Copper Braid 95%	Black PVC Noncontaminating	0.425	Polyethylene	66	20.5	0.285	3700
223	[84]	50	0.35	1.2	4.1	14.5	#19 Solid Silver Coated Copper .034 Dia	Double Silver Coated Copper Braid, 95%	Black PVC Noncontaminating	0.212	Polyethylene	66	30.8	0.117	1700
303	[111] 170	50	0.34	1.1	3.9	13.5	#18 Solid Silver Coated Copper Covered Steel .037 Dia	Silver Coated Copper Braid, 95%	Tinted Brown Fluorinated Ethylene-Propylene	0.17	TFE Teflon	69.5	29.2	0.116	1400
316	[113] 172	50	1.2	2.7	8.3	29	#26 Stranded Silver Coated Copper Covered Steel .020 Dia	Silver Coated Copper Braid, 95%	White Fluorinated Ethylene-Propylene	0.098	TFE Teflon	69.5	29.2	0.06	900

Note: Mil-C-17/() part numbers were revised. Initial specification numbers are shown in brackets, current specification numbers are unbracketed.
Source:
Mil-C-17G
Mil-C-17G Supplement 1
Belden Master Catalog, Belden Wire & Cable Co, Richmond, IN.

TABLE 8 Semi-Rigid Coax Specifications

RG-()/U	Mil-C-17 Part Number M17/	Zo	Loss [dB/100 ft] Power [W] @500 MHz	Loss [dB/100 ft] Power [W] @1 GHz	Loss [dB/100 ft] Power [W] @3 GHz	Loss [dB/100 ft] Power [W] @5 GHz	Loss [dB/100 ft] Power [W] @10 GHz	Loss [dB/100 ft] Power [W] @18 GHz	Loss [dB/100 ft] Power [W] @20 GHz	Center Conductor Material	Center Conductor Diameter	Outer Conductor	Outside Diameter [Inches]	Dielectric	Dielectric Constant (1 GHz)	Max Capacitance pF/foot	Dielectric Diameter [inches]	Maximum Voltage (60Hz) [RMS]
	154-00001	50 ± 3.0	42 14	60 10	100 6	140 4.5	190 3.1		280 2	Silver Plated Copper Coated Steel	0.008 ±0.0005	Copper	0.034 ±0.001	Solid PTFE	2.03	29.9	0.026 ±0.001	750
	154-00002	50 ± 3.0	42 14	60 10	100 6	140 4.5	190 3.1		280 2	Silver Plated Copper Coated Steel	0.008 ±0.0005	Tin-Plated Copper	0.034 ±0.002	Solid PTFE	2.03	29.9	0.026 ±0.001	750
	151-00001	50 ± 2.5	28 45	40 32	70 18	90 13	130 9		190 65	Silver Plated Copper Coated Steel	0.0113 ±0.0005	Copper	0.047 ±0.001	Solid PTFE	2.03	29.9	0.037 ±0.001	1000
	151-00002	50 ± 2.5	28 45	40 32	70 18	90 13	130 9		190 6.5	Silver Plated Copper Coated Steel	0.0113 ±0.0005	Tin-Plated Copper	0.047 +0.002 –0.001	Solid PTFE	2.03	29.9	0.037 ±0.001	1000
405	133-RG-405	50 ± 1.5	15 180	22 130		50 54	80 35		130 20	Silver Plated Copper Coated Steel	0.0201 ±0.0005	Copper	0.0865 ±0.001	Solid PTFE	2.03	29.9	0.066 ±0.002	5000
	133-00001	50 ± 1.5	15 180	22 130		50 54	80 35		130 20	Silver Plated Copper Coated Steel	0.0201 ±0.0005	Tin-Plated Copper	0.0865 +0.002 –0.001	Solid PTFE	2.03	29.9	0.066 ±0.002	5000
	133-00002	50 ± 1.5	15 180	22 130		50 54	80 35		130 20	Silver Plated Copper	0.0201 ±0.0005	Copper	0.0865 ±0.001	Solid PTFE	2.03	29.9	0.066 ±0.002	5000
	133-00003	50 ± 1.5	15 180	22 130		50 54	80 35		130 20	Silver Plated Copper	0.0201 ±0.0005	Tin-Plated Copper	0.0865 +0.002 –0.001	Solid PTFE	2.03	29.9	0.066 ±0.002	5000
	133-00004	50 ± 1.5	15 180	22 130		50 54	80 35		130 20	Silver Plated Nickel Copper Coated Steel	0.0201 ±0.0005	Copper	0.0865 ±0.001	Solid PTFE	2.03	29.9	0.066 ±0.002	5000
	133-00005	50 ± 1.5	15 180	22 130		50 54	80 35		130 20	Silver Plated Nickel Copper Coated Steel	0.0201 ±0.0005	Tin-Plated Copper	0.0865 +0.002 –0.001	Solid PTFE	2.03	29.9	0.066 ±0.002	5000
	133-0006	50 ± 1.5	15 180	22 130		50 54	80 35		130 20	Silver Plated Copper Coated Steel	0.0201 ±0.0005	Copper	0.0865 ±0.001	Solid PTFE	2.03	29.9	0.066 ±0.002	5000

	133-00007	50 ± 1.5	15 180	22 130		50 54	80 35	130 20	Silver Plated Copper Coated Steel	0.0201 ±0.0005	Tin-Plated Copper	0.086 +0.0021 –0.001	Solid PTFE	2.03	29.9	0.066 ±0.002	5000
	133-00008	50 ± 1.5	15 180	22 130		50 54	80 35	130 20	Silver Plated Copper	0.0201 ±0.0005	Copper	0.0865 ±0.001	Solid PTFE	2.03	29.9	0.066 ±0.002	5000
	133-00009	50 ± 1.5	15 180	22 130		50 54	80 35	130 20	Silver Plated Copper	0.0201 ±0.0005	Tin-Plated Copper	0.0865 +0.002 –0.001	Solid PTFE	2.03	29.9	0.066 ±0.002	5000
	133-00010	50 ± 1.5	15 180	22 130		50 54	80 35	130 20	Silver Plated Nickel Copper Coated Steel	0.0201 ±0.0005	Copper	0.086 ±0.001	Solid PTFE	2.03	29.9	0.066 ±0.002	5000
	133-00011	50 ± 1.5	15 180	22 130		50 54	80 35	130 20	Silver Plated Nickel Copper Coated Steel	0.0201 ±0.0005	Tin-Plated Copper	0.086 +0.002 –0.001	Solid PTFE	2.03	29.9	0.066 ±0.002	5000
402	130-RG-402	50 ± 1.0	8 600	12 450	21 250	29 180	45 120	70 70	Silver Plated Copper Coated Steel	0.0362 ±0.0007	Copper	0.141 ±0.001	Solid PTFE	2.03	29.9	0.1175 ±0.001	5000
	130-00001	50 ± 1.0	8 600	12 450	21 250	29 180	45 120	70 70	Silver Plated Copper Coated Steel	0.0362 ±0.0007	Tin-Plated Copper	0.141 +0.002 –0.001	Solid PTFE	2.03	29.9	0.1175 ±0.001	5000
	130-00002	50 ± 1.0	8 601	12 451	21 251	29 181	45 121	70 71	Silver Plated Nickel Copper Coated Steel	0.0362 ±0.0007	Copper	0.141 ±0.001	Solid PTFE	2.03	29.9	0.1175 ±0.001	5000
	130-00003	50 ± 1.0	8 602	12 452	21 252	29 182	45 122	70 72	Silver Plated Nickel Copper Coated Steel	0.0362 ±0.0007	Tin-Plated Copper	0.141 +0.002 –0.001	Solid PTFE	2.03	29.9	0.1175 ±0.001	5000
	130-00004	50 ± 1.0	8 603	12 453	21 253	29 183	45 123	70 73	Silver Plated Copper Coated Steel	0.0362 ±0.0007	Copper	0.141 ±0.001	Solid PTFE	2.03	29.9	0.1175 ±0.001	5000
	130-00005	50 ± 1.0	8 604	12 454	21 254	29 184	45 124	70 74	Silver Plated Copper Coated Steel	0.0362 ±0.0007	Tin-Plated Copper	0.141 +0.002 –0.001	Solid PTFE	2.03	29.9	0.1175 ±0.001	5000
	130-00006	50 ± 1.0	8 605	12 455	21 255	29 185	45 125	70 75	Silver Plated Nickel Copper Coated Steel	0.0362 ±0.0007	Copper	0.141 ±0.001	Solid PTFE	2.03	29.9	0.1175 ±0.001	5000

TABLE 8 (continued)

RG-()/U	Mil-C-17 Part Number M17/	Zo	Loss [dB/100 ft] Power [W] @500 MHz	Loss [dB/100 ft] Power [W] @1 GHz	Loss [dB/100 ft] Power [W] @3 GHz	Loss [dB/100 ft] Power [W] @5 GHz	Loss [dB/100 ft] Power [W] @10 GHz	Loss [dB/100 ft] Power [W] @18 GHz	Loss [dB/100 ft] Power [W] @20 GHz	Center Conductor Material	Center Conductor Diameter	Outer Conductor	Outside Diameter [Inches]	Dielectric	Dielectric Constant (1 GHz)	Max Capacitance pF/foot	Dielectric Diameter [inches]	Maximum Voltage (60Hz) [RMS]
	130-00007	50 ± 1.0	8 606	12 456	21 256	29 186	45 126		70 76	Silver Plated Nickel Copper Coated Steel	0.0362 ±0.0007	Tin-Plated Copper	0.141 +0.002 –0.001	Solid PTFE	2.03	29.9	0.1175 ±0.001	5000
	130-00008	50 ± 1.0	8 607	12 457	21 257	29 187	45 127		70 77	Silver Plated Copper Coated Steel	0.0362 ±0.0007	Aluminum	0.141 ±0.001	Solid PTFE	2.03	29.9	0.1175 ±0.001	5000
	130-00009	50 ± 1.0	8 608	12 458	21 258	29 188	45 128		70 78	Silver Plated Copper Coated Steel	0.0362 ±0.0007	Aluminum	0.141 ±0.001	Solid PTFE	2.03	29.9	0.1175 ±0.001	5000
	130-00010	50 ± 1.0	8 609	12 459	21 259	29 189	45 129		70 79	Silver Plated Nickel Copper Coated Steel	0.0362 ±0.0007	Aluminum	0.141 ±0.001	Solid PTFE	2.03	29.9	0.1175 ±0.001	5000
	130-00011	50 ± 1.0	8 610	12 460	21 260	29 190	45 130		70 80	Silver Plated Nickel Copper Coated Steel	0.0362 ±0.0007	Aluminum	0.141 ±0.001	Solid PTFE	2.03	29.9	0.1175 ±0.001	5000
	130-00012	50 ± 1.0	8 611	12 461	21 261	29 191	45 131		70 81	Silver Plated Copper Coated Steel	0.0362 ±0.0007	Silver Plated Copper	0.141 +0.002 –0.001	Solid PTFE	2.03	29.9	0.1175 ±0.001	5000
	130-00013	50 ± 1.0	8 612	12 462	21 262	29 192	45 132		70 82	Silver Plated Nickel Copper Coated Steel	0.0362 ±0.0007	Silver Plated Copper	0.141 +0.002 –0.001	Solid PTFE	2.03	29.9	0.1175 ±0.001	5000
401	129-RG-401	50 ± 0.5	4.5 1900	7.5 1400	11 750		33 350	48 200	— —	Silver Plated Copper	0.0641 ±0.001	Copper	0.250 ±0.001	Solid PTFE	2.03	29.9	0.209 ±0.002	7500
	129-00001	50 ± 0.5	4.5 1900	7.5 1400	11 750		33 350	48 200	— —	Silver Plated Copper	0.0641 ±0.001	Tin-Plated Copper	0.250 +0.002 –0.001	Solid PTFE	2.03	29.9	0.209 ±0.002	7500

Notes: Attenuation/Power Ratings are maximum values for families.
Sources: Mil-C-17/130E
Semi-Rigid Coaxial Cable Catalog, Micro-Coax Components, Inc, Collegeville, PA.

Guide To Use of Dissimilar Metals In Sea Water, Marine Atmosphere, and Industrial Atmosphere

Code	Metals	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
A	Magnesium	G	I C C	I C C	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I
B	Zinc Zinc Coating		G	C C C	I C C	I C C	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I
C	Cadmium Beryllium			G	I C C	I C C	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I
D	Aluminum, Aluminum-Mg, Aluminum-Zn				G	C C C	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I
E	Aluminum-Copper					G	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I
F	Steels - Carbon, Low Alloy						G	I C	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I C C	I I I	I I I	I I I	I I I
G	Lead							G	C C C	I C C	I C C	I C C	I C C	I I I	I I C	I I C	I C C	I I C	I C C	I C C	I I C
H	Tin, Tin-Lead Indium								G	I I I	I I C	I I C	I C C	I C C	I C C	I C C	C C C	I C C	C C C	I I I	I I I
I	Stainless Steels - Martensitic, Ferritic									G	I I C	I I I	I I I	I I C	I I I	I I I	I I I	I I I	I I I	I I I	I I I
J	Chromium, Molybdenum, Tungsten										G	C C C	I C C	I I C	I I I	I C C	I C C	I C C	I C C	I C C	I C C
K	Stainless Steels - Austenitic, Type PH Super Strength Heat Resistant											G	I I I	I I I	I I I	I C C	I C C	I C C	I C C	I C C	I C C
L	Brass-Lead, Bronze												G	I C C	I C C	I I C	I C C	I I C	I C C	I C C	I C C
M	Brass - Low Copper, Bronze - Low Copper													G	I C C	I C C	I C C	I I C	I C C	I C C	I I C
N	Brass - High Copper, Bronze - High Copper														G	I I C	I C C	I I C	I C C	I C C	I I C
O	Copper - High Nickel, Monel															G	I C C	I C C	I C C	I C C	I C C
P	Nickel, Cobalt																G	I C C	I C C	I C C	I C C
Q	Titanium																	G	I C C	I C C	I C C
R	Silver																		G	C C C	I C C
S	Palladium, Rhodium, Gold, Platinum																			G	C C C
T	Graphite																				G

Sea Water

Marine Atmosphere | Industrial Atmosphere

Notes:
1. Metals are ranked from most anodic (A) to most cathodic (T).
2. "G" indicates bare identical metal couple compatible in sea water, marine atmosphere, and industrial atmosphere.
3. I indicates joined metals incompatible without appropriate protective coatings in the indicated environment.
4. C indicates joined metals are compatible in the indicated environment.
5. Reference: MIL-STD-889B.

Example: Can High-Copper Brass be put safely in direct contact with Gold?
High-Copper Brass has code letter M, and Gold has code letter T.
Find the intersection of Row M and Column T, to find that these metals are incompatible in salt water and marine atmosphere, but compatible in an industrial atmosphere.

Single Sideband and Image Reject Mixers

TABLE 10 Maximum Tolerable Phase Error (Degrees) for Single Sideband and Image Reject Mixers as a Function of Suppression and Amplitude Imbalance

Amplitude Imbalance	Suppression												
[dB]	−10 dBc	−13 dBc	−15 dBc	−17 dBc	−20 dBc	−23 dBc	−25 dBc	−27 dBc	−30 dBc	−33 dBc	−35 dBc	−37 dBc	−40 dBc
0.00	35.10	25.24	20.17	16.08	11.42	8.10	6.44	5.12	3.62	2.56	2.04	1.62	1.15
0.05	35.10	25.24	20.16	16.08	11.42	8.09	6.43	5.10	3.61	2.54	2.01	1.58	1.10
0.10	35.09	25.23	20.16	16.07	11.40	8.07	6.40	5.07	3.56	2.48	1.93	1.48	0.94
0.15	35.08	25.22	20.14	16.05	11.38	8.04	6.36	5.02	3.48	2.37	1.78	1.28	0.58
0.20	35.08	25.21	20.13	16.03	11.35	7.99	6.30	4.94	3.37	2.20	1.55	0.94	
0.25	35.06	25.19	20.10	16.00	11.30	7.93	6.22	4.84	3.23	1.97	1.20		
0.30	35.05	25.17	20.07	15.96	11.25	7.86	6.13	4.72	3.04	1.63	0.49		
0.35	35.03	25.14	20.04	15.92	11.19	7.77	6.01	4.57	2.79	1.12			
0.40	35.01	25.11	20.00	15.87	11.12	7.66	5.87	4.38	2.48				
0.45	34.99	25.07	19.96	15.81	11.03	7.54	5.72	4.17	2.08				
0.50	34.96	25.04	19.91	15.75	10.94	7.40	5.53	3.91	1.50				
0.55	34.93	24.99	19.85	15.68	10.84	7.25	5.32	3.61					
0.60	34.90	24.95	19.79	15.60	10.72	7.07	5.08	3.25					
0.65	34.87	24.90	19.73	15.51	10.60	6.88	4.81	2.80					
0.70	34.83	24.84	19.65	15.42	10.46	6.66	4.50	2.21					
0.75	34.79	24.78	19.58	15.32	10.31	6.42	4.13	1.32					
0.80	34.75	24.72	19.49	15.21	10.15	6.16	3.70						
0.85	34.70	24.65	19.41	15.10	9.97	5.86	3.18						
0.90	34.66	24.58	19.31	14.98	9.78	5.53	2.52						
0.95	34.61	24.50	19.21	14.84	9.58	5.16	1.53						
1.00	34.55	24.42	19.10	14.70	9.35	4.73							
1.10	34.44	24.24	18.87	14.40	8.86	3.65							
1.20	34.31	24.05	18.62	14.06	8.28	1.83							
1.30	34.17	23.84	18.34	13.67	7.61								
1.40	34.02	23.61	18.03	13.25	6.80								
1.50	33.86	23.35	17.69	12.78	5.82								
1.60	33.68	23.08	17.32	12.25	4.53								
1.70	33.50	22.79	16.92	11.67	2.52								
1.80	33.30	22.48	16.48	11.01									
1.90	33.09	22.14	16.00	10.28									
2.00	32.86	21.78	15.49	9.44									
2.10	32.63	21.39	14.93	8.47									
2.20	32.37	20.98	14.31	7.31									
2.30	32.11	20.54	13.64	5.87									
2.40	31.83	20.07	12.90	3.81									
2.50	31.54	19.56	12.08										
2.60	31.23	19.03	11.17										
2.70	30.90	18.45	10.13										
2.80	30.56	17.83	8.93										
2.90	30.21	17.17	7.48										
3.00	29.83	16.46	5.59										
3.10	29.44	15.68	2.41										
3.20	29.03	14.84											
3.30	28.60	13.92											
3.40	28.16	12.91											
3.50	27.69	11.77											
3.60	27.19	10.47											
3.70	26.68	8.94											
3.80	26.14	7.02											
3.90	25.57	4.24											
4.00	24.98												
4.10	24.35												
4.20	23.69												
4.30	23.00												
4.40	22.27												
4.50	21.49												
4.60	20.67												
4.70	19.79												
4.80	18.86												
4.90	17.85												
5.00	16.76												

Notes: Example: An image reject mixer requires 25 dB of unwanted image suppression, and a maximum amplitude imbalance of 1 dB is expected between the channels. Looking at the intersection of the 1 dB row and −25 dBc column shows that a maximum interchannel phase imbalance of 9.35 degrees is tolerable.
Suppression [dBc] = 10 log ((1 − 2 a cos(p) + a^2) / (1 + 2 a cos(p) + a^2)) where a is the voltage ratio amplitude imbalance and p is the phase imbalance.

Power, Voltage and Decibels

TABLE 11 Power, Voltage, dB Conversion Table

dB	Power Ratio	Voltage Ratio	dB	Power Ratio	Voltage Ratio	dB	Power Ratio	Voltage Ratio	dB	Power Ratio	Voltage Ratio
0	1	1	9	7.943282347	2.818382931	34	2511.886432	50.11872336	60	1000000	1000
0.1	1.023292992	1.011579454	9.5	8.912509381	2.985382619	35	3162.27766	56.23413252	61	1258925.412	1122.018454
0.2	1.047128548	1.023292992	10	10	3.16227766	36	3981.071706	63.09573445	62	1584893.192	1258.925412
0.3	1.071519305	1.035142167	11	12.58925412	3.548133892	37	5011.872336	70.79457844	63	1995262.315	1412.537545
0.4	1.096478196	1.047128548	12	15.84893192	3.981071706	38	6309.573445	79.43282347	64	2511886.432	1584.893192
0.5	1.122018454	1.059253725	13	19.95262315	4.466835922	39	7943.282347	89.12509381	65	3162277.66	1778.27941
0.6	1.148153621	1.071519305	14	25.11886432	5.011872336	40	10000	100	66	3981071.706	1995.262315
0.7	1.174897555	1.083926914	15	31.6227766	5.623413252	41	12589.25412	112.2018454	67	5011872.336	2238.721139
0.8	1.202264435	1.096478196	16	39.81071706	6.309573445	42	15848.93192	125.8925412	68	6309573.445	2511.886432
0.9	1.230268771	1.109174815	17	50.11872336	7.079457844	43	19952.62315	141.2537545	69	7943282.347	2818.382931
1	1.258925412	1.122018454	18	63.09573445	7.943282347	44	25118.86432	158.4893192	70	10000000	3162.27766
1.5	1.412537545	1.188502227	19	79.43282347	8.912509381	45	31622.7766	177.827941	71	12589254.12	3548.133892
2	1.584893192	1.258925412	20	100	10	46	39810.71706	199.5262315	72	15848931.92	3981.071706
2.5	1.77827941	1.333521432	21	125.8925412	11.22018454	47	50118.72336	223.8721139	73	19952623.15	4466.835922
3	1.995262315	1.412537545	22	158.4893192	12.58925412	48	63095.73445	251.1886432	74	25118864.32	5011.872336
3.5	2.238721139	1.496235656	23	199.5262315	14.12537545	49	79432.82347	281.8382931	75	31622776.6	5623.413252
4	2.511886432	1.584893192	24	251.1886432	15.84893192	50	100000	316.227766	76	39810717.06	6309.573445
4.5	2.818382931	1.678804018	25	316.227766	17.7827941	51	125892.5412	354.8133892	77	50118723.36	7079.457844
5	3.16227766	1.77827941	26	398.1071706	19.95262315	52	158489.3192	398.1071706	78	63095734.45	7943.282347
5.5	3.548133892	1.883649089	27	501.1872336	22.38721139	53	199526.2315	446.6835922	79	79432823.47	8912.509381
6	3.981071706	1.995262315	28	630.9573445	25.11886432	54	251188.6432	501.1872336	80	100000000	10000
6.5	4.466835922	2.11348904	29	794.3282347	28.18382931	55	316227.766	562.3413252	81	125892541.2	11220.18454
7	5.011872336	2.238721139	30	1000	31.6227766	56	398107.1706	630.9573445	82	158489319.2	12589.25412
7.5	5.623413252	2.371373706	31	1258.925412	35.48133892	57	501187.2336	707.9457844	83	199526231.5	14125.37545
8	6.309573445	2.511886432	32	1584.893192	39.81071706	58	630957.3445	794.3282347	84	251188643.2	15848.93192
8.5	7.079457844	2.66072506	33	1995.262315	44.66835922	59	794328.2347	891.2509381	85	316227766	17782.7941
86	398107170.6	19952.62315	−5.5	0.281838293	0.530884444	−35	0.000316228	0.017782794	−69	1.25893E-07	0.000354813
87	501187233.6	22387.21139	−6	0.251188643	0.501187234	−36	0.000251189	0.015848932	−70	0.0000001	0.000316228
88	630957344.5	25118.86432	−6.5	0.223872114	0.473151259	−37	0.000199526	0.014125375	−71	7.94328E-08	0.000281838
89	794328234.7	28183.82931	−7	0.199526231	0.446683592	−38	0.000158489	0.012589254	−72	6.30957E-08	0.000251189
90	1000000000	31622.7766	−7.5	0.177827941	0.421696503	−39	0.000125893	0.011220185	−73	5.01187E-08	0.000223872
91	1258925412	35481.33892	−8	0.158489319	0.398107171	−40	0.0001	0.01	−74	3.98107E-08	0.000199526
92	1584893192	39810.71706	−8.5	0.141253754	0.375837404	−41	7.94328E-05	0.008912509	−75	3.16228E-08	0.000177828
93	1995262315	44668.35922	−9	0.125892541	0.354813389	−42	6.30957E-05	0.007943282	−76	2.51189E-08	0.000158489
94	2511886432	50118.72336	−9.5	0.112201845	0.334965439	−43	5.01187E-05	0.007079458	−77	1.99526E-08	0.000141254
95	3162277660	56234.13252	−10	0.1	0.316227766	−44	3.98107E-05	0.006309573	−78	1.58489E-08	0.000125893
96	3981071706	63095.73445	−11	0.079432823	0.281838293	−45	3.16228E-05	0.005623413	−79	1.25893E-08	0.000112202
97	5011872336	70794.57844	−12	0.063095734	0.251188643	−46	2.51189E-05	0.005011872	−80	0.00000001	0.0001
98	6309573445	79432.82347	−13	0.050118723	0.223872114	−47	1.99526E-05	0.004466836	−81	7.94328E-09	8.91251E-05
99	7943282347	89125.09381	−14	0.039810717	0.199526231	−48	1.58489E-05	0.003981072	−82	6.30957E-09	7.94328E-05
100	10000000000	100000	−15	0.031622777	0.177827941	−49	1.25893E-05	0.003548134	−83	5.01187E-09	7.07946E-05
0	1	1	−16	0.025118864	0.158489319	−50	0.00001	0.003162278	−84	3.98107E-09	6.30957E-05
−0.1	0.977237221	0.988553095	−17	0.019952623	0.141253754	−51	7.94328E-06	0.002818383	−85	3.16228E-09	5.62341E-05
−0.2	0.954992586	0.977237221	−18	0.015848932	0.125892541	−52	6.30957E-06	0.002511886	−86	2.51189E-09	5.01187E-05
−0.3	0.933254301	0.966050879	−19	0.012589254	0.112201845	−53	5.01187E-06	0.002238721	−87	1.99526E-09	4.46684E-05
−0.4	0.912010839	0.954992586	−20	0.01	0.1	−54	3.98107E-06	0.001995262	−88	1.58489E-09	3.98107E-05
−0.5	0.891250938	0.944060876	−21	0.007943282	0.089125094	−55	3.16228E-06	0.001778279	−89	1.25893E-09	3.54813E-05
−0.6	0.87096359	0.933254301	−22	0.006309573	0.079432823	−56	2.51189E-06	0.001584893	−90	0.000000001	3.16228E-05
−0.7	0.851138038	0.922571427	−23	0.005011872	0.070794578	−57	1.99526E-06	0.001412538	−91	7.94328E-10	2.81838E-05
−0.8	0.831763771	0.912010839	−24	0.003981072	0.063095734	−58	1.58489E-06	0.001258925	−92	6.30957E-10	2.51189E-05
−0.9	0.812830516	0.901571138	−25	0.003162278	0.056234133	−59	1.25893E-06	0.001122018	−93	5.01187E-10	2.23872E-05
−1	0.794328235	0.891250938	−26	0.002511886	0.050118723	−60	0.000001	0.001	−94	3.98107E-10	1.99526E-05
−1.5	0.707945784	0.841395142	−27	0.001995262	0.044668359	−61	7.94328E-07	0.000891251	−95	3.16228E-10	1.77828E-05
−2	0.630957344	0.794328235	−28	0.001584893	0.039810717	−62	6.30957E-07	0.000794328	−96	2.51189E-10	1.58489E-05
−2.5	0.562341325	0.749894209	−29	0.001258925	0.035481339	−63	5.01187E-07	0.000707946	−97	1.99526E-10	1.41254E-05
−3	0.501187234	0.707945784	−30	0.001	0.031622777	−64	3.98107E-07	0.000630957	−98	1.58489E-10	1.25893E-05
−3.5	0.446683592	0.668343918	−31	0.000794328	0.028183829	−65	3.16228E-07	0.000562341	−99	1.25893E-10	1.12202E-05
−4	0.398107171	0.630957344	−32	0.000630957	0.025118864	−66	2.51189E-07	0.000501187	−100	1E-10	0.00001
−4.5	0.354813389	0.595662144	−33	0.000501187	0.022387211	−67	1.99526E-07	0.000446684			
−5	0.316227766	0.562341325	−34	0.000398107	0.019952623	−68	1.58489E-07	0.000398107			

Notes:
1. Multiply by appropriate factor.
2. Use voltage factor to convert S-parameters.

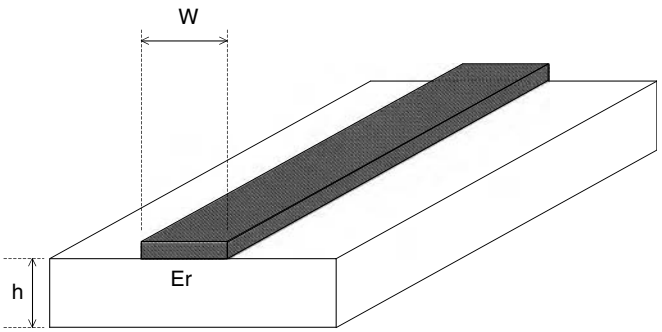


FIGURE B.5

TABLE 12a Zero Thickness Microstrip Dimensions, Effective Dielectric Constant, and PUL Capacitance and Inductance for Er = 2.2

Zo [Ohms]	W/h	Keff	pF/cm	nH/mm	Zo [Ohms]	W/h	Keff	pF/cm	nH/mm
1	250.3363	2.1861	49.3192	0.0493	57	2.5199	1.8500	0.7959	2.5860
2	123.5739	2.1728	24.5846	0.0983	58	2.4514	1.8471	0.7816	2.6294
3	81.4007	2.1601	16.3417	0.1471	59	2.3854	1.8443	0.7678	2.6727
4	60.3537	2.1480	12.2218	0.1955	60	2.3218	1.8416	0.7544	2.7160
5	47.7490	2.1364	9.7510	0.2438	61	2.2603	1.8389	0.7415	2.7592
6	39.3615	2.1253	8.1046	0.2918	62	2.2010	1.8362	0.7290	2.8024
7	33.3816	2.1146	6.9294	0.3395	63	2.1437	1.8336	0.7170	2.8456
8	28.9051	2.1044	6.0485	0.3871	64	2.0883	1.8310	0.7053	2.8887
9	25.4299	2.0946	5.3639	0.4345	65	2.0347	1.8285	0.6939	2.9318
10	22.6550	2.0851	4.8166	0.4817	66	1.9829	1.8259	0.6829	2.9749
11	20.3889	2.0760	4.3692	0.5287	67	1.9327	1.8235	0.6723	3.0179
12	18.5040	2.0673	3.9967	0.5755	68	1.8919	1.8214	0.6620	3.0612
13	16.9121	2.0589	3.6817	0.6222	69	1.8447	1.8190	0.6520	3.1042
14	15.5503	2.0508	3.4120	0.6688	70	1.7990	1.8166	0.6423	3.1471
15	14.3723	2.0429	3.1784	0.7152	71	1.7548	1.8143	0.6328	3.1900
16	13.3435	2.0354	2.9743	0.7614	72	1.7119	1.8120	0.6236	3.2329
17	12.4375	2.0280	2.7943	0.8075	73	1.6704	1.8097	0.6147	3.2757
18	11.6337	2.0210	2.6344	0.8536	74	1.6301	1.8075	0.6060	3.3186
19	10.9159	2.0141	2.4915	0.8994	75	1.5910	1.8053	0.5976	3.3614
20	10.2711	2.0075	2.3631	0.9452	76	1.5531	1.8031	0.5894	3.4041
21	9.6889	2.0010	2.2469	0.9909	77	1.5163	1.8010	0.5814	3.4469
22	9.1607	1.9948	2.1414	1.0365	78	1.4806	1.7988	0.5736	3.4896
23	8.6793	1.9887	2.0452	1.0819	79	1.4458	1.7968	0.5660	3.5322
24	8.2389	1.9828	1.9571	1.1273	80	1.4121	1.7947	0.5586	3.5749
25	7.8346	1.9771	1.8761	1.1726	81	1.3793	1.7926	0.5514	3.6175
26	7.4621	1.9715	1.8014	1.2177	82	1.3474	1.7906	0.5443	3.6601
27	7.1178	1.9661	1.7323	1.2628	83	1.3164	1.7886	0.5375	3.7027
28	6.7988	1.9608	1.6682	1.3078	84	1.2862	1.7867	0.5308	3.7453
29	6.5024	1.9557	1.6085	1.3528	85	1.2568	1.7847	0.5243	3.7878

TABLE 12a (continued)

Zo [Ohms]	W/h	Keff	pF/cm	nH/mm	Zo [Ohms]	W/h	Keff	pF/cm	nH/mm
30	6.2263	1.9507	1.5529	1.3976	86	1.2282	1.7828	0.5179	3.8303
31	5.9686	1.9458	1.5010	1.4424	87	1.2003	1.7809	0.5117	3.8728
32	5.7274	1.9410	1.4523	1.4871	88	1.1732	1.7791	0.5056	3.9152
33	5.5013	1.9364	1.4066	1.5318	89	1.1467	1.7772	0.4996	3.9577
34	5.2890	1.9319	1.3636	1.5763	90	1.1210	1.7754	0.4938	4.0001
35	5.0892	1.9274	1.3231	1.6208	91	1.0959	1.7736	0.4882	4.0424
36	4.9009	1.9231	1.2849	1.6653	92	1.0714	1.7718	0.4826	4.0848
37	4.7231	1.9189	1.2488	1.7096	93	1.0476	1.7700	0.4772	4.1272
38	4.5550	1.9147	1.2146	1.7539	94	1.0243	1.7683	0.4719	4.1695
39	4.3959	1.9107	1.1822	1.7982	95	1.0016	1.7665	0.4667	4.2118
40	4.2450	1.9067	1.1515	1.8424	96	0.9795	1.7658	0.4617	4.2552
41	4.1018	1.9028	1.1223	1.8865	97	0.9579	1.7651	0.4569	4.2987
42	3.9658	1.8990	1.0945	1.9306	98	0.9369	1.7644	0.4521	4.3421
43	3.8363	1.8953	1.0680	1.9746	99	0.9163	1.7637	0.4475	4.3855
44	3.7129	1.8917	1.0427	2.0186	100	0.8962	1.7629	0.4429	4.4289
45	3.5953	1.8881	1.0185	2.0625	101	0.8767	1.7621	0.4384	4.4722
46	3.4831	1.8846	0.9955	2.1064	102	0.8575	1.7613	0.4340	4.5154
47	3.3758	1.8811	0.9734	2.1502	103	0.8389	1.7605	0.4297	4.5586
48	3.2733	1.8778	0.9523	2.1940	104	0.8206	1.7596	0.4255	4.6018
49	3.1751	1.8745	0.9320	2.2378	105	0.8028	1.7588	0.4213	4.6449
50	3.0811	1.8712	0.9126	2.2814	106	0.7855	1.7579	0.4172	4.6880
51	2.9909	1.8680	0.8939	2.3251	107	0.7685	1.7570	0.4132	4.7310
52	2.9044	1.8649	0.8760	2.3687	108	0.7519	1.7561	0.4093	4.7740
53	2.8214	1.8618	0.8588	2.4122	109	0.7357	1.7552	0.4054	4.8169
54	2.7416	1.8587	0.8422	2.4557	110	0.7198	1.7543	0.4016	4.8599
55	2.6648	1.8558	0.8262	2.4992	111	0.7043	1.7534	0.3979	4.9027
56	2.5910	1.8528	0.8108	2.5426	112	0.6892	1.7524	0.3943	4.9456
113	0.6744	1.7515	0.3907	4.9884	132	0.4484	1.7330	0.3327	5.7964
114	0.6600	1.7505	0.3871	5.0312	133	0.4389	1.7321	0.3301	5.8387
115	0.6459	1.7496	0.3837	5.0739	134	0.4296	1.7311	0.3275	5.8810
116	0.6320	1.7486	0.3803	5.1166	135	0.4206	1.7302	0.3250	5.9232
117	0.6185	1.7477	0.3769	5.1593	136	0.4117	1.7292	0.3225	5.9654
118	0.6053	1.7467	0.3736	5.2020	137	0.4030	1.7283	0.3201	6.0076
119	0.5924	1.7457	0.3704	5.2446	138	0.3945	1.7273	0.3177	6.0498
120	0.5798	1.7447	0.3672	5.2872	139	0.3862	1.7264	0.3153	6.0920
121	0.5675	1.7438	0.3640	5.3298	140	0.3781	1.7254	0.3130	6.1342
122	0.5554	1.7428	0.3609	5.3723	141	0.3702	1.7245	0.3107	6.1763
123	0.5436	1.7418	0.3579	5.4148	142	0.3624	1.7236	0.3084	6.2185
124	0.5321	1.7408	0.3549	5.4573	143	0.3547	1.7226	0.3062	6.2606
125	0.5208	1.7399	0.3520	5.4998	144	0.3473	1.7217	0.3039	6.3027
126	0.5097	1.7389	0.3491	5.5422	145	0.3400	1.7208	0.3018	6.3448
127	0.4989	1.7379	0.3462	5.5846	146	0.3328	1.7199	0.2996	6.3868
128	0.4884	1.7369	0.3434	5.6270	147	0.3259	1.7190	0.2975	6.4289
129	0.4780	1.7360	0.3407	5.6694	148	0.3190	1.7181	0.2954	6.4709
130	0.4679	1.7350	0.3380	5.7118	149	0.3123	1.7172	0.2934	6.5130
131	0.4580	1.7340	0.3353	5.7541	150	0.3058	1.7163	0.2913	6.5550

Notes: Calculation of W/H has an error of less than 1%.

Source: Gupta, K.C., Garg, R., Bahl, I., Bhartia, P., *Microstrip Lines and Slotlines,* 2nd Ed., Artech House, Norwood, MA, 1996, 103.

TABLE 12b Zero Thickness Microstrip Dimensions, Effective Dielectric Constant, and PUL Capacitance and Inductance for Er =3.78

Zo [Ohms]	W/h	Keff	pF/cm	nH/mm	Zo [Ohms]	W/h	Keff	pF/cm	nH/mm
1	190.5772	3.7382	64.4927	0.0645	26	5.4498	3.1668	2.2831	1.5433
2	93.9049	3.6989	32.0763	0.1283	27	5.1890	3.1537	2.1940	1.5994
3	61.7511	3.6619	21.2770	0.1915	28	4.9474	3.1410	2.1113	1.6553
4	45.7086	3.6271	15.8817	0.2541	29	4.7230	3.1287	2.0345	1.7110
5	36.1035	3.5942	12.6477	0.3162	30	4.5140	3.1167	1.9629	1.7666
6	29.7137	3.5632	10.4941	0.3778	31	4.3190	3.1051	1.8961	1.8221
7	25.1593	3.5337	8.9578	0.4389	32	4.1365	3.0938	1.8335	1.8775
8	21.7508	3.5059	7.8070	0.4997	33	3.9655	3.0827	1.7747	1.9327
9	19.1053	3.4794	6.9133	0.5600	34	3.8050	3.0720	1.7195	1.9878
10	16.9935	3.4542	6.1994	0.6199	35	3.6539	3.0616	1.6676	2.0428
11	15.2694	3.4301	5.6162	0.6796	36	3.5116	3.0514	1.6185	2.0976
12	13.8357	3.4072	5.1309	0.7389	37	3.3773	3.0414	1.5722	2.1524
13	12.6252	3.3853	4.7210	0.7978	38	3.2504	3.0317	1.5284	2.2070
14	11.5899	3.3643	4.3702	0.8566	39	3.1302	3.0222	1.4869	2.2616
15	10.6946	3.3442	4.0666	0.9150	40	3.0164	3.0130	1.4475	2.3160
16	9.9129	3.3249	3.8014	0.9732	41	2.9083	3.0039	1.4101	2.3703
17	9.2247	3.3064	3.5678	1.0311	42	2.8056	2.9951	1.3745	2.4246
18	8.6143	3.2885	3.3605	1.0888	43	2.7080	2.9864	1.3406	2.4787
19	8.0694	3.2714	3.1754	1.1463	44	2.6150	2.9780	1.3082	2.5327
20	7.5800	3.2549	3.0090	1.2036	45	2.5263	2.9697	1.2774	2.5867
21	7.1382	3.2389	2.8586	1.2607	46	2.4417	2.9615	1.2479	2.6406
22	6.7375	3.2235	2.7222	1.3175	47	2.3609	2.9536	1.2197	2.6943
23	6.3724	3.2086	2.5978	1.3743	48	2.2836	2.9458	1.1927	2.7480
24	6.0386	3.1942	2.4840	1.4308	49	2.2097	2.9381	1.1669	2.8016
25	5.7321	3.1803	2.3794	1.4871	50	2.1389	2.9306	1.1421	2.8552
51	2.0711	2.9233	1.1183	2.9086	101	0.5140	2.7126	0.5439	5.5488
52	2.0060	2.9160	1.0954	2.9620	102	0.5008	2.7099	0.5383	5.6008
53	1.9435	2.9089	1.0734	3.0152	103	0.4878	2.7071	0.5328	5.6529
54	1.8840	2.9020	1.0523	3.0685	104	0.4752	2.7043	0.5274	5.7048
55	1.8268	2.8952	1.0320	3.1217	105	0.4630	2.7016	0.5222	5.7568
56	1.7719	2.8886	1.0124	3.1748	106	0.4511	2.6989	0.5170	5.8086
57	1.7190	2.8820	0.9935	3.2278	107	0.4394	2.6961	0.5119	5.8605
58	1.6682	2.8756	0.9752	3.2807	108	0.4281	2.6934	0.5069	5.9123
59	1.6192	2.8693	0.9577	3.3336	109	0.4171	2.6907	0.5020	5.9640
60	1.5720	2.8631	0.9407	3.3865	110	0.4064	2.6880	0.4972	6.0157
61	1.5265	2.8569	0.9243	3.4392	111	0.3960	2.6853	0.4924	6.0674
62	1.4826	2.8509	0.9084	3.4919	112	0.3858	2.6826	0.4878	6.1190
63	1.4402	2.8450	0.8931	3.5446	113	0.3759	2.6800	0.4832	6.1706
64	1.3993	2.8392	0.8782	3.5971	114	0.3663	2.6774	0.4788	6.2221
65	1.3597	2.8334	0.8638	3.6496	115	0.3569	2.6747	0.4744	6.2736
66	1.3215	2.8278	0.8499	3.7021	116	0.3477	2.6721	0.4701	6.3251
67	1.2845	2.8222	0.8364	3.7545	117	0.3388	2.6696	0.4658	6.3765
68	1.2488	2.8167	0.8233	3.8068	118	0.3301	2.6670	0.4616	6.4279
69	1.2142	2.8113	0.8106	3.8591	119	0.3217	2.6644	0.4575	6.4793
70	1.1807	2.8060	0.7982	3.9113	120	0.3134	2.6619	0.4535	6.5307
71	1.1482	2.8008	0.7862	3.9635	121	0.3054	2.6594	0.4496	6.5820
72	1.1168	2.7956	0.7746	4.0156	122	0.2976	2.6569	0.4457	6.6333
73	1.0863	2.7905	0.7633	4.0676	123	0.2900	2.6544	0.4418	6.6845
74	1.0568	2.7855	0.7523	4.1196	124	0.2826	2.6520	0.4381	6.7358
75	1.0282	2.7805	0.7416	4.1716	125	0.2754	2.6496	0.4344	6.7870
76	1.0004	2.7756	0.7312	4.2235	126	0.2683	2.6472	0.4307	6.8382
77	0.9735	2.7737	0.7215	4.2776	127	0.2615	2.6448	0.4271	6.8893
78	0.9474	2.7717	0.7120	4.3316	128	0.2548	2.6424	0.4236	6.9405
79	0.9220	2.7696	0.7027	4.3855	129	0.2483	2.6401	0.4201	6.9916

TABLE 12b (continued)

Zo [Ohms]	W/h	Keff	pF/cm	nH/mm	Zo [Ohms]	W/h	Keff	pF/cm	nH/mm
80	0.8974	2.7675	0.6936	4.4393	130	0.2420	2.6378	0.4167	7.0427
81	0.8735	2.7653	0.6848	4.4929	131	0.2358	2.6355	0.4134	7.0938
82	0.8503	2.7630	0.6762	4.5465	132	0.2298	2.6332	0.4101	7.1449
83	0.8277	2.7606	0.6677	4.6000	133	0.2239	2.6309	0.4068	7.1959
84	0.8058	2.7582	0.6595	4.6534	134	0.2182	2.6287	0.4036	7.2469
85	0.7845	2.7557	0.6514	4.7067	135	0.2126	2.6265	0.4004	7.2980
86	0.7639	2.7532	0.6436	4.7599	136	0.2072	2.6243	0.3973	7.3490
87	0.7438	2.7506	0.6359	4.8130	137	0.2019	2.6221	0.3943	7.3999
88	0.7242	2.7481	0.6284	4.8660	138	0.1968	2.6200	0.3912	7.4509
89	0.7052	2.7454	0.6210	4.9190	139	0.1917	2.6179	0.3883	7.5019
90	0.6868	2.7428	0.6138	4.9718	140	0.1869	2.6158	0.3853	7.5528
91	0.6688	2.7401	0.6068	5.0246	141	0.1821	2.6137	0.3825	7.6037
92	0.6513	2.7374	0.5999	5.0773	142	0.1775	2.6116	0.3796	7.6546
93	0.6344	2.7347	0.5931	5.1300	143	0.1729	2.6096	0.3768	7.7055
94	0.6178	2.7320	0.5865	5.1826	144	0.1685	2.6076	0.3741	7.7564
95	0.6018	2.7292	0.5801	5.2351	145	0.1642	2.6056	0.3713	7.8073
96	0.5861	2.7265	0.5737	5.2875	146	0.1600	2.6036	0.3687	7.8582
97	0.5709	2.7237	0.5675	5.3399	147	0.1560	2.6017	0.3660	7.9091
98	0.5561	2.7209	0.5615	5.3922	148	0.1520	2.5998	0.3634	7.9599
99	0.5417	2.7182	0.5555	5.4444	149	0.1481	2.5979	0.3608	8.0108
100	0.5277	2.7154	0.5497	5.4966	150	0.1444	2.5960	0.3583	8.0616

Notes: Calculation of W/H has an error of less than 1%.

Source: Gupta, K.C., Garg, R., Bahl, I., Bhartia, P., *Microstrip Lines and Slotlines,* 2nd Ed., Artech House, Norwood, MA, 1996, 103.

TABLE 12c Zero Thickness Microstrip Dimensions, Effective Dielectric Constant, and PUL Capacitance and Inductance for Er =5.75

Zo [Ohms]	W/h	Keff	pF/cm	nH/mm	Zo [Ohms]	W/h	Keff	pF/cm	nH/mm
1	154.1545	5.6626	79.3758	0.0794	55	1.3098	4.1200	1.2310	3.7239
2	75.8057	5.5818	39.4035	0.1576	56	1.2664	4.1088	1.2074	3.7864
3	49.7547	5.5068	26.0920	0.2348	57	1.2248	4.0978	1.1846	3.8488
4	36.7611	5.4372	19.4449	0.3111	58	1.1847	4.0869	1.1627	3.9112
5	28.9839	5.3723	15.4628	0.3866	59	1.1461	4.0763	1.1415	3.9734
6	23.8117	5.3116	12.8127	0.4613	60	1.1089	4.0658	1.1210	4.0355
7	20.1263	5.2548	10.9235	0.5352	61	1.0731	4.0555	1.1012	4.0976
8	17.3689	5.2014	9.5094	0.6086	62	1.0386	4.0453	1.0821	4.1596
9	15.2296	5.1512	8.4118	0.6814	63	1.0054	4.0353	1.0636	4.2214
10	13.5223	5.1037	7.5357	0.7536	64	0.9733	4.0305	1.0464	4.2859
11	12.1289	5.0589	6.8204	0.8253	65	0.9423	4.0265	1.0297	4.3507
12	10.9706	5.0163	6.2257	0.8965	66	0.9124	4.0222	1.0136	4.4153
13	9.9929	4.9759	5.7236	0.9673	67	0.8836	4.0178	0.9979	4.4797
14	9.1570	4.9375	5.2942	1.0377	68	0.8557	4.0132	0.9827	4.5439
15	8.4343	4.9008	4.9229	1.1077	69	0.8288	4.0084	0.9679	4.6080
16	7.8035	4.8659	4.5987	1.1773	70	0.8028	4.0035	0.9535	4.6719
17	7.2484	4.8324	4.3133	1.2466	71	0.7777	3.9985	0.9394	4.7357
18	6.7562	4.8004	4.0602	1.3155	72	0.7534	3.9933	0.9258	4.7993
19	6.3169	4.7697	3.8342	1.3841	73	0.7299	3.9881	0.9125	4.8628
20	5.9225	4.7403	3.6312	1.4525	74	0.7072	3.9828	0.8996	4.9261
21	5.5666	4.7119	3.4479	1.5205	75	0.6853	3.9774	0.8870	4.9893
22	5.2438	4.6847	3.2817	1.5883	76	0.6641	3.9720	0.8747	5.0524
23	4.9499	4.6584	3.1302	1.6559	77	0.6435	3.9665	0.8628	5.1153
24	4.6812	4.6331	2.9916	1.7232	78	0.6236	3.9609	0.8511	5.1781
25	4.4346	4.6087	2.8644	1.7902	79	0.6044	3.9554	0.8397	5.2408

TABLE 12c (continued)

Zo [Ohms]	W/h	Keff	pF/cm	nH/mm	Zo [Ohms]	W/h	Keff	pF/cm	nH/mm
26	4.2075	4.5851	2.7471	1.8571	80	0.5858	3.9498	0.8287	5.3034
27	3.9979	4.5623	2.6388	1.9237	81	0.5678	3.9442	0.8178	5.3659
28	3.8037	4.5402	2.5384	1.9901	82	0.5503	3.9386	0.8073	5.4283
29	3.6233	4.5187	2.4451	2.0563	83	0.5334	3.9329	0.7970	5.4905
30	3.4554	4.4980	2.3581	2.1223	84	0.5171	3.9273	0.7870	5.5527
31	3.2988	4.4778	2.2769	2.1881	85	0.5012	3.9217	0.7771	5.6148
32	3.1523	4.4583	2.2010	2.2538	86	0.4859	3.9161	0.7676	5.6768
33	3.0151	4.4393	2.1297	2.3193	87	0.4710	3.9105	0.7582	5.7387
34	2.8863	4.4208	2.0628	2.3846	88	0.4566	3.9049	0.7490	5.8005
35	2.7652	4.4028	1.9997	2.4497	89	0.4427	3.8994	0.7401	5.8623
36	2.6511	4.3853	1.9403	2.5147	90	0.4292	3.8938	0.7314	5.9239
37	2.5434	4.3682	1.8842	2.5795	91	0.4161	3.8883	0.7228	5.9855
38	2.4417	4.3516	1.8311	2.6441	92	0.4034	3.8829	0.7144	6.0470
39	2.3455	4.3353	1.7808	2.7087	93	0.3912	3.8774	0.7063	6.1085
40	2.2543	4.3195	1.7331	2.7730	94	0.3793	3.8720	0.6983	6.1699
41	2.1678	4.3040	1.6878	2.8373	95	0.3677	3.8667	0.6904	6.2312
42	2.0856	4.2889	1.6448	2.9014	96	0.3565	3.8614	0.6828	6.2925
43	2.0075	4.2741	1.6037	2.9653	97	0.3457	3.8561	0.6753	6.3537
44	1.9289	4.2588	1.5645	3.0288	98	0.3352	3.8508	0.6679	6.4148
45	1.8591	4.2449	1.5272	3.0926	99	0.3250	3.8456	0.6607	6.4759
46	1.7926	4.2312	1.4916	3.1562	100	0.3152	3.8405	0.6537	6.5369
47	1.7291	4.2178	1.4576	3.2198	101	0.3056	3.8354	0.6468	6.5979
48	1.6684	4.2048	1.4250	3.2832	102	0.2963	3.8304	0.6400	6.6588
49	1.6104	4.1919	1.3938	3.3464	103	0.2874	3.8254	0.6334	6.7197
50	1.5549	4.1794	1.3638	3.4096	104	0.2787	3.8204	0.6269	6.7806
51	1.5017	4.1671	1.3351	3.4727	105	0.2702	3.8155	0.6205	6.8414
52	1.4507	4.1550	1.3076	3.5356	106	0.2620	3.8107	0.6143	6.9022
53	1.4018	4.1431	1.2811	3.5985	107	0.2541	3.8059	0.6082	6.9629
54	1.3549	4.1315	1.2556	3.6612	108	0.2464	3.8011	0.6022	7.0236
109	0.2389	3.7964	0.5963	7.0842	130	0.1255	3.7101	0.4942	8.3525
110	0.2317	3.7918	0.5905	7.1449	131	0.1217	3.7065	0.4902	8.4127
111	0.2247	3.7872	0.5848	7.2055	132	0.1180	3.7030	0.4863	8.4729
112	0.2179	3.7827	0.5792	7.2660	133	0.1144	3.6996	0.4824	8.5331
113	0.2113	3.7782	0.5738	7.3265	134	0.1110	3.6962	0.4786	8.5933
114	0.2049	3.7738	0.5684	7.3871	135	0.1076	3.6928	0.4748	8.6535
115	0.1987	3.7694	0.5631	7.4475	136	0.1044	3.6895	0.4711	8.7137
116	0.1927	3.7651	0.5580	7.5080	137	0.1012	3.6863	0.4675	8.7739
117	0.1869	3.7608	0.5529	7.5684	138	0.0982	3.6830	0.4639	8.8341
118	0.1813	3.7566	0.5479	7.6288	139	0.0952	3.6799	0.4603	8.8943
119	0.1758	3.7524	0.5430	7.6892	140	0.0923	3.6767	0.4569	8.9544
120	0.1705	3.7483	0.5382	7.7496	141	0.0896	3.6737	0.4534	9.0146
121	0.1653	3.7443	0.5334	7.8099	142	0.0869	3.6706	0.4500	9.0748
122	0.1603	3.7403	0.5288	7.8703	143	0.0842	3.6676	0.4467	9.1350
123	0.1555	3.7363	0.5242	7.9306	144	0.0817	3.6647	0.4434	9.1952
124	0.1508	3.7324	0.5197	7.9909	145	0.0792	3.6618	0.4402	9.2553
125	0.1462	3.7286	0.5153	8.0512	146	0.0768	3.6589	0.4370	9.3155
126	0.1418	3.7248	0.5109	8.1115	147	0.0745	3.6561	0.4339	9.3757
127	0.1375	3.7210	0.5066	8.1717	148	0.0723	3.6533	0.4308	9.4359
128	0.1334	3.7173	0.5024	8.2320	149	0.0701	3.6505	0.4277	9.4961
129	0.1294	3.7137	0.4983	8.2922	150	0.0680	3.6478	0.4247	9.5562

Notes: Calculation of W/H has an error of less than 1%.

Source: Gupta, K.C., Garg, R., Bahl, I., Bhartia, P., *Microstrip Lines and Slotlines,* 2nd Ed., Artech House, Norwood, MA, 1996, 103.

TABLE 12d Zero Thickness Microstrip Dimensions, Effective Dielectric Constant, and PUL Capacitance and Inductance for Er = 9.4

Zo [Ohms]	W/h	Keff	pF/cm	nH/mm	Zo [Ohms]	W/h	Keff	pF/cm	nH/mm
1	120.1221	9.2047	101.2010	0.1012	24	3.4066	7.1750	3.7229	2.1444
2	58.8859	9.0280	50.1125	0.2004	25	3.2161	7.1309	3.5630	2.2269
3	38.5353	8.8676	33.1101	0.2980	26	3.0409	7.0885	3.4157	2.3090
4	28.3901	8.7212	24.6268	0.3940	27	2.8791	7.0475	3.2797	2.3909
5	22.3208	8.5871	19.5493	0.4887	28	2.7293	7.0079	3.1537	2.4725
6	18.2864	8.4635	16.1735	0.5822	29	2.5902	6.9697	3.0366	2.5538
7	15.4132	8.3493	13.7691	0.6747	30	2.4609	6.9326	2.9276	2.6348
8	13.2647	8.2433	11.9712	0.7662	31	2.3403	6.8967	2.8258	2.7156
9	11.5985	8.1445	10.5771	0.8567	32	2.2275	6.8619	2.7306	2.7961
10	10.2695	8.0521	9.4653	0.9465	33	2.1220	6.8281	2.6413	2.8764
11	9.1854	7.9655	8.5584	1.0356	34	2.0229	6.7952	2.5574	2.9564
12	8.2846	7.8841	7.8050	1.1239	35	1.9221	6.7606	2.4780	3.0356
13	7.5248	7.8074	7.1695	1.2116	36	1.8364	6.7301	2.4037	3.1152
14	6.8754	7.7348	6.6264	1.2988	37	1.7556	6.7004	2.3336	3.1947
15	6.3143	7.6661	6.1571	1.3853	38	1.6792	6.6715	2.2673	3.2740
16	5.8248	7.6009	5.7477	1.4714	39	1.6070	6.6434	2.2045	3.3530
17	5.3942	7.5389	5.3875	1.5570	40	1.5386	6.6159	2.1449	3.4319
18	5.0126	7.4798	5.0682	1.6421	41	1.4738	6.5891	2.0884	3.5105
19	4.6722	7.4234	4.7833	1.7268	42	1.4122	6.5629	2.0346	3.5890
20	4.3668	7.3694	4.5276	1.8110	43	1.3537	6.5373	1.9834	3.6673
21	4.0913	7.3178	4.2968	1.8949	44	1.2981	6.5122	1.9346	3.7454
22	3.8416	7.2683	4.0876	1.9784	45	1.2451	6.4877	1.8880	3.8233
23	3.6143	7.2207	3.8971	2.0616	46	1.1946	6.4638	1.8436	3.9010
47	1.1465	6.4403	1.8011	3.9786	99	0.1530	5.8352	0.8139	7.9771
48	1.1005	6.4173	1.7604	4.0560	100	0.1472	5.8268	0.8052	8.0518
49	1.0567	6.3948	1.7215	4.1332	101	0.1417	5.8184	0.7966	8.1265
50	1.0148	6.3728	1.6841	4.2103	102	0.1365	5.8102	0.7883	8.2012
51	0.9747	6.3596	1.6494	4.2901	103	0.1314	5.8021	0.7801	8.2758
52	0.9364	6.3507	1.6165	4.3711	104	0.1265	5.7942	0.7720	8.3504
53	0.8997	6.3412	1.5849	4.4519	105	0.1217	5.7864	0.7642	8.4251
54	0.8646	6.3312	1.5543	4.5323	106	0.1172	5.7787	0.7565	8.4997
55	0.8309	6.3208	1.5248	4.6124	107	0.1128	5.7712	0.7489	8.5742
56	0.7987	6.3100	1.4963	4.6923	108	0.1086	5.7638	0.7415	8.6488
57	0.7678	6.2989	1.4687	4.7719	109	0.1046	5.7565	0.7342	8.7234
58	0.7382	6.2875	1.4421	4.8512	110	0.1006	5.7493	0.7271	8.7979
59	0.7098	6.2759	1.4163	4.9303	111	0.0969	5.7423	0.7201	8.8725
60	0.6826	6.2641	1.3914	5.0091	112	0.0933	5.7354	0.7133	8.9470
61	0.6564	6.2521	1.3673	5.0877	113	0.0898	5.7286	0.7065	9.0216
62	0.6313	6.2400	1.3439	5.1661	114	0.0864	5.7219	0.6999	9.0961
63	0.6072	6.2278	1.3213	5.2443	115	0.0832	5.7154	0.6934	9.1707
64	0.5841	6.2156	1.2994	5.3223	116	0.0801	5.7090	0.6871	9.2452
65	0.5619	6.2032	1.2781	5.4001	117	0.0771	5.7026	0.6808	9.3197
66	0.5406	6.1909	1.2575	5.4777	118	0.0743	5.6964	0.6747	9.3943
67	0.5201	6.1786	1.2375	5.5552	119	0.0715	5.6903	0.6687	9.4688
68	0.5004	6.1662	1.2181	5.6325	120	0.0688	5.6843	0.6627	9.5434
69	0.4814	6.1539	1.1992	5.7096	121	0.0662	5.6785	0.6569	9.6179
70	0.4632	6.1417	1.1809	5.7866	122	0.0638	5.6727	0.6512	9.6925
71	0.4457	6.1295	1.1631	5.8634	123	0.0614	5.6670	0.6456	9.7670
72	0.4289	6.1173	1.1458	5.9401	124	0.0591	5.6615	0.6401	9.8416
73	0.4127	6.1053	1.1290	6.0166	125	0.0569	5.6560	0.6346	9.9162
74	0.3972	6.0933	1.1127	6.0931	126	0.0548	5.6506	0.6293	9.9907
75	0.3822	6.0814	1.0968	6.1694	127	0.0527	5.6454	0.6241	10.0653
76	0.3679	6.0696	1.0813	6.2456	128	0.0508	5.6402	0.6189	10.1399
77	0.3540	6.0579	1.0662	6.3217	129	0.0489	5.6351	0.6138	10.2146
78	0.3407	6.0464	1.0516	6.3977	130	0.0471	5.6301	0.6088	10.2892
79	0.3279	6.0350	1.0373	6.4736	131	0.0453	5.6252	0.6039	10.3638

TABLE 12d (continued)

Zo [Ohms]	W/h	Keff	pF/cm	nH/mm	Zo [Ohms]	W/h	Keff	pF/cm	nH/mm
80	0.3156	6.0236	1.0233	6.5494	132	0.0436	5.6204	0.5991	10.4385
81	0.3038	6.0125	1.0098	6.6251	133	0.0420	5.6157	0.5943	10.5131
82	0.2924	6.0014	0.9965	6.7007	134	0.0404	5.6111	0.5897	10.5878
83	0.2815	5.9905	0.9836	6.7762	135	0.0389	5.6065	0.5850	10.6625
84	0.2709	5.9797	0.9710	6.8517	136	0.0375	5.6021	0.5805	10.7372
85	0.2608	5.9691	0.9588	6.9271	137	0.0361	5.5977	0.5761	10.8119
86	0.2510	5.9586	0.9468	7.0024	138	0.0347	5.5934	0.5717	10.8867
87	0.2416	5.9482	0.9351	7.0777	139	0.0334	5.5892	0.5673	10.9614
88	0.2326	5.9380	0.9237	7.1529	140	0.0322	5.5850	0.5631	11.0362
89	0.2239	5.9280	0.9125	7.2281	141	0.0310	5.5809	0.5589	11.1110
90	0.2155	5.9180	0.9016	7.3032	142	0.0298	5.5770	0.5547	11.1858
91	0.2074	5.9083	0.8910	7.3782	143	0.0287	5.5730	0.5507	11.2606
92	0.1997	5.8986	0.8806	7.4532	144	0.0276	5.5692	0.5467	11.3354
93	0.1922	5.8891	0.8704	7.5281	145	0.0266	5.5654	0.5427	11.4103
94	0.1850	5.8798	0.8605	7.6031	146	0.0256	5.5617	0.5388	11.4852
95	0.1781	5.8706	0.8507	7.6779	147	0.0247	5.5581	0.5350	11.5600
96	0.1715	5.8616	0.8412	7.7528	148	0.0237	5.5545	0.5312	11.6350
97	0.1651	5.8526	0.8319	7.8276	149	0.0229	5.5510	0.5274	11.7099
98	0.1589	5.8439	0.8228	7.9023	150	0.0220	5.5476	0.5238	11.7848

Notes: Calculation of W/H has an error of less than 1%.

TABLE 12e Zero Thickness Microstrip Dimensions, Effective Dielectric Constant, and PUL Capacitance and Inductance for Er = 9.8

Zo [Ohms]	W/h	Keff	pF/cm	nH/mm	Zo [Ohms]	W/h	Keff	pF/cm	nH/mm
1	117.6023	9.5914	103.3045	0.1033	57	0.7347	6.5379	1.4963	4.8615
2	57.6329	9.4030	51.1424	0.2046	58	0.7059	6.5254	1.4691	4.9421
3	37.7044	9.2322	33.7840	0.3041	59	0.6783	6.5128	1.4428	5.0224
4	27.7700	9.0767	25.1237	0.4020	60	0.6519	6.5000	1.4174	5.1025
5	21.8272	8.9344	19.9408	0.4985	61	0.6265	6.4870	1.3927	5.1824
6	17.8771	8.8035	16.4952	0.5938	62	0.6022	6.4740	1.3689	5.2621
7	15.0641	8.6827	14.0413	0.6880	63	0.5788	6.4609	1.3458	5.3415
8	12.9606	8.5706	12.2066	0.7812	64	0.5564	6.4477	1.3234	5.4208
9	11.3295	8.4662	10.7841	0.8735	65	0.5349	6.4346	1.3017	5.4999
10	10.0285	8.3688	9.6496	0.9650	66	0.5142	6.4214	1.2807	5.5788
11	8.9673	8.2775	8.7244	1.0557	67	0.4944	6.4082	1.2603	5.6575
12	8.0857	8.1917	7.9558	1.1456	68	0.4753	6.3951	1.2405	5.7360
13	7.3419	8.1109	7.3075	1.2350	69	0.4570	6.3820	1.2213	5.8144
14	6.7063	8.0345	6.7535	1.3237	70	0.4395	6.3690	1.2026	5.8927
15	6.1572	7.9622	6.2749	1.4118	71	0.4226	6.3561	1.1844	5.9708
16	5.6782	7.8937	5.8573	1.4995	72	0.4064	6.3432	1.1668	6.0488
17	5.2568	7.8285	5.4900	1.5866	73	0.3908	6.3305	1.1497	6.1266
18	4.8834	7.7664	5.1644	1.6733	74	0.3758	6.3179	1.1330	6.2043
19	4.5503	7.7071	4.8738	1.7595	75	0.3614	6.3053	1.1168	6.2820
20	4.2515	7.6505	4.6131	1.8452	76	0.3476	6.2929	1.1010	6.3595
21	3.9820	7.5963	4.3778	1.9306	77	0.3343	6.2807	1.0857	6.4368
22	3.7377	7.5443	4.1645	2.0156	78	0.3215	6.2685	1.0707	6.5141
23	3.5154	7.4944	3.9703	2.1003	79	0.3092	6.2565	1.0561	6.5913
24	3.3122	7.4464	3.7926	2.1846	80	0.2974	6.2447	1.0419	6.6684
25	3.1259	7.4002	3.6296	2.2685	81	0.2860	6.2329	1.0281	6.7455

TABLE 12e (continued)

Zo [Ohms]	W/h	Keff	pF/cm	nH/mm	Zo [Ohms]	W/h	Keff	pF/cm	nH/mm
26	2.9545	7.3557	3.4795	2.3522	82	0.2751	6.2214	1.0146	6.8224
27	2.7962	7.3128	3.3408	2.4355	83	0.2646	6.2100	1.0015	6.8993
28	2.6497	7.2713	3.2124	2.5185	84	0.2545	6.1987	0.9887	6.9760
29	2.5138	7.2312	3.0930	2.6012	85	0.2448	6.1876	0.9762	7.0528
30	2.3873	7.1923	2.9819	2.6837	86	0.2355	6.1766	0.9640	7.1294
31	2.2693	7.1547	2.8781	2.7659	87	0.2265	6.1658	0.9520	7.2060
32	2.1591	7.1182	2.7811	2.8478	88	0.2179	6.1552	0.9404	7.2825
33	2.0558	7.0827	2.6901	2.9295	89	0.2096	6.1447	0.9291	7.3590
34	1.9590	7.0483	2.6046	3.0109	90	0.2016	6.1344	0.9180	7.4355
35	1.8614	7.0124	2.5237	3.0916	91	0.1939	6.1242	0.9071	7.5118
36	1.7777	6.9805	2.4480	3.1727	92	0.1865	6.1142	0.8965	7.5882
37	1.6987	6.9494	2.3766	3.2535	93	0.1794	6.1044	0.8862	7.6645
38	1.6241	6.9191	2.3090	3.3342	94	0.1726	6.0947	0.8760	7.7408
39	1.5535	6.8896	2.2450	3.4146	95	0.1660	6.0852	0.8661	7.8170
40	1.4867	6.8609	2.1843	3.4948	96	0.1597	6.0758	0.8565	7.8932
41	1.4233	6.8328	2.1266	3.5749	97	0.1536	6.0666	0.8470	7.9693
42	1.3632	6.8053	2.0718	3.6547	98	0.1478	6.0575	0.8377	8.0455
43	1.3060	6.7785	2.0197	3.7343	99	0.1422	6.0486	0.8286	8.1216
44	1.2517	6.7523	1.9699	3.8138	100	0.1368	6.0398	0.8198	8.1977
45	1.1999	6.7266	1.9225	3.8931	101	0.1316	6.0312	0.8111	8.2737
46	1.1507	6.7015	1.8772	3.9721	102	0.1266	6.0227	0.8026	8.3498
47	1.1037	6.6770	1.8339	4.0510	103	0.1217	6.0143	0.7942	8.4258
48	1.0588	6.6529	1.7924	4.1298	104	0.1171	6.0062	0.7860	8.5018
49	1.0161	6.6293	1.7527	4.2083	105	0.1127	5.9981	0.7780	8.5778
50	0.9752	6.6149	1.7158	4.2895	106	0.1084	5.9902	0.7702	8.6538
51	0.9361	6.6054	1.6810	4.3722	107	0.1043	5.9824	0.7625	8.7298
52	0.8988	6.5953	1.6474	4.4545	108	0.1003	5.9748	0.7549	8.8057
53	0.8630	6.5846	1.6150	4.5365	109	0.0965	5.9673	0.7476	8.8817
54	0.8289	6.5735	1.5837	4.6182	110	0.0928	5.9600	0.7403	8.9576
55	0.7961	6.5619	1.5536	4.6996	111	0.0893	5.9527	0.7332	9.0336
56	0.7648	6.5501	1.5245	4.7807	112	0.0859	5.9456	0.7262	9.1095
113	0.0826	5.9387	0.7194	9.1855	132	0.0396	5.8280	0.6101	10.6295
114	0.0795	5.9318	0.7126	9.2614	133	0.0381	5.8232	0.6052	10.7056
115	0.0765	5.9251	0.7060	9.3374	134	0.0366	5.8185	0.6005	10.7818
116	0.0736	5.9185	0.6996	9.4133	135	0.0352	5.8139	0.5958	10.8579
117	0.0708	5.9120	0.6932	9.4893	136	0.0339	5.8093	0.5912	10.9341
118	0.0681	5.9057	0.6870	9.5652	137	0.0326	5.8049	0.5866	11.0102
119	0.0655	5.8994	0.6808	9.6412	138	0.0314	5.8005	0.5821	11.0864
120	0.0630	5.8933	0.6748	9.7172	139	0.0302	5.7962	0.5777	11.1626
121	0.0606	5.8873	0.6689	9.7931	140	0.0290	5.7920	0.5734	11.2389
122	0.0583	5.8814	0.6631	9.8691	141	0.0279	5.7879	0.5691	11.3151
123	0.0561	5.8756	0.6574	9.9451	142	0.0269	5.7838	0.5649	11.3914
124	0.0540	5.8699	0.6517	10.0211	143	0.0258	5.7799	0.5608	11.4677
125	0.0519	5.8643	0.6462	10.0971	144	0.0249	5.7760	0.5567	11.5440
126	0.0499	5.8588	0.6408	10.1731	145	0.0239	5.7722	0.5527	11.6203
127	0.0480	5.8534	0.6354	10.2492	146	0.0230	5.7684	0.5487	11.6966
128	0.0462	5.8482	0.6302	10.3252	147	0.0221	5.7647	0.5448	11.7730
129	0.0445	5.8430	0.6250	10.4013	148	0.0213	5.7611	0.5410	11.8494
130	0.0428	5.8379	0.6200	10.4773	149	0.0205	5.7576	0.5372	11.9258
131	0.0411	5.8329	0.6150	10.5534	150	0.0197	5.7541	0.5334	12.0022

Notes: Calculation of W/H has an error of less than 1%.

Source: Gupta, K.C., Garg, R., Bahl, I., Bhartia, P., *Microstrip Lines and Slotlines,* 2nd Ed., Artech House, Norwood, MA, 1996, 103.

TABLE 12f Zero Thickness Microstrip Dimensions, Effective Dielectric Constant, and PUL Capacitance and Inductance for Er = 11.6

Zo [Ohms]	W/h	Keff	pF/cm	nH/mm	Zo [Ohms]	W/h	Keff	pF/cm	nH/mm
1	107.9253	11.3278	112.2672	0.1123	26	2.6229	8.5446	3.7502	2.5351
2	52.8209	11.0843	55.5270	0.2221	27	2.4782	8.4928	3.6003	2.6246
3	34.5132	10.8654	36.6506	0.3299	28	2.3444	8.4426	3.4615	2.7138
4	25.3888	10.6674	27.2364	0.4358	29	2.2202	8.3942	3.3325	2.8026
5	19.9316	10.4873	21.6044	0.5401	30	2.1047	8.3473	3.2124	2.8912
6	16.3052	10.3226	17.8617	0.6430	31	1.9970	8.3019	3.1003	2.9794
7	13.7232	10.1712	15.1973	0.7447	32	1.8884	8.2543	2.9948	3.0667
8	11.7929	10.0313	13.2059	0.8452	33	1.7964	8.2125	2.8967	3.1545
9	10.2964	9.9016	11.6625	0.9447	34	1.7101	8.1718	2.8045	3.2420
10	9.1031	9.7809	10.4321	1.0432	35	1.6291	8.1324	2.7178	3.3293
11	8.1299	9.6682	9.4289	1.1409	36	1.5527	8.0939	2.6361	3.4163
12	7.3215	9.5625	8.5958	1.2378	37	1.4807	8.0565	2.5589	3.5031
13	6.6397	9.4632	7.8932	1.3340	38	1.4128	8.0201	2.4859	3.5897
14	6.0573	9.3696	7.2931	1.4294	39	1.3485	7.9846	2.4168	3.6759
15	5.5541	9.2812	6.7747	1.5243	40	1.2877	7.9499	2.3513	3.7620
16	5.1153	9.1975	6.3226	1.6186	41	1.2301	7.9161	2.2890	3.8478
17	4.7294	9.1180	5.9249	1.7123	42	1.1754	7.8830	2.2299	3.9335
18	4.3875	9.0424	5.5725	1.8055	43	1.1235	7.8507	2.1735	4.0189
19	4.0826	8.9703	5.2581	1.8982	44	1.0742	7.8192	2.1199	4.1041
20	3.8091	8.9016	4.9760	1.9904	45	1.0273	7.7883	2.0687	4.1890
21	3.5625	8.8358	4.7215	2.0822	46	0.9827	7.7654	2.0207	4.2758
22	3.3390	8.7728	4.4908	2.1736	47	0.9402	7.7532	1.9762	4.3653
23	3.1357	8.7124	4.2807	2.2645	48	0.8998	7.7401	1.9334	4.4545
24	2.9499	8.6543	4.0887	2.3551	49	0.8612	7.7262	1.8922	4.5432
25	2.7796	8.5984	3.9125	2.4453	50	0.8244	7.7117	1.8526	4.6315
51	0.7893	7.6966	1.8145	4.7195	101	0.0957	6.9814	0.8726	8.9017
52	0.7558	7.6810	1.7778	4.8072	102	0.0917	6.9719	0.8635	8.9837
53	0.7238	7.6650	1.7425	4.8945	103	0.0880	6.9625	0.8545	9.0657
54	0.6933	7.6487	1.7084	4.9816	104	0.0844	6.9534	0.8458	9.1477
55	0.6641	7.6322	1.6755	5.0683	105	0.0809	6.9444	0.8372	9.2297
56	0.6362	7.6154	1.6438	5.1548	106	0.0776	6.9356	0.8287	9.3116
57	0.6095	7.5985	1.6131	5.2410	107	0.0744	6.9269	0.8205	9.3936
58	0.5840	7.5815	1.5835	5.3270	108	0.0714	6.9185	0.8124	9.4756
59	0.5596	7.5643	1.5549	5.4127	109	0.0684	6.9102	0.8044	9.5576
60	0.5363	7.5472	1.5273	5.4982	110	0.0656	6.9020	0.7967	9.6396
61	0.5139	7.5301	1.5005	5.5835	111	0.0630	6.8940	0.7890	9.7216
62	0.4925	7.5129	1.4747	5.6686	112	0.0604	6.8862	0.7815	9.8036
63	0.4721	7.4959	1.4496	5.7535	113	0.0579	6.8786	0.7742	9.8857
64	0.4525	7.4789	1.4253	5.8382	114	0.0555	6.8710	0.7670	9.9677
65	0.4337	7.4620	1.4018	5.9227	115	0.0533	6.8637	0.7599	10.0498
66	0.4158	7.4452	1.3790	6.0071	116	0.0511	6.8565	0.7530	10.1318
67	0.3986	7.4286	1.3569	6.0913	117	0.0490	6.8494	0.7461	10.2139
68	0.3821	7.4121	1.3355	6.1753	118	0.0470	6.8425	0.7394	10.2960
69	0.3663	7.3957	1.3147	6.2592	119	0.0450	6.8357	0.7329	10.3781
70	0.3512	7.3795	1.2945	6.3430	120	0.0432	6.8290	0.7264	10.4602
71	0.3367	7.3635	1.2749	6.4266	121	0.0414	6.8225	0.7201	10.5423
72	0.3228	7.3477	1.2558	6.5101	122	0.0397	6.8161	0.7138	10.6245
73	0.3095	7.3321	1.2373	6.5935	123	0.0381	6.8099	0.7077	10.7067
74	0.2967	7.3166	1.2193	6.6768	124	0.0365	6.8037	0.7017	10.7888
75	0.2845	7.3014	1.2018	6.7600	125	0.0350	6.7977	0.6957	10.8711
76	0.2728	7.2864	1.1847	6.8430	126	0.0336	6.7919	0.6899	10.9533
77	0.2616	7.2716	1.1682	6.9260	127	0.0322	6.7861	0.6842	11.0355
78	0.2508	7.2570	1.1520	7.0089	128	0.0309	6.7804	0.6786	11.1178
79	0.2405	7.2426	1.1363	7.0918	129	0.0296	6.7749	0.6730	11.2001
80	0.2306	7.2285	1.1210	7.1745	130	0.0284	6.7695	0.6676	11.2824

TABLE 12f (continued)

Zo [Ohms]	W/h	Keff	pF/cm	nH/mm	Zo [Ohms]	W/h	Keff	pF/cm	nH/mm
81	0.2211	7.2146	1.1061	7.2572	131	0.0273	6.7642	0.6622	11.3647
82	0.2121	7.2009	1.0916	7.3398	132	0.0261	6.7590	0.6570	11.4471
83	0.2033	7.1874	1.0774	7.4224	133	0.0251	6.7539	0.6518	11.5294
84	0.1950	7.1741	1.0636	7.5049	134	0.0240	6.7489	0.6467	11.6118
85	0.1870	7.1611	1.0501	7.5873	135	0.0231	6.7440	0.6417	11.6943
86	0.1793	7.1483	1.0370	7.6697	136	0.0221	6.7392	0.6367	11.7767
87	0.1719	7.1357	1.0242	7.7520	137	0.0212	6.7345	0.6318	11.8592
88	0.1649	7.1233	1.0117	7.8343	138	0.0203	6.7300	0.6271	11.9417
89	0.1581	7.1111	0.9994	7.9166	139	0.0195	6.7255	0.6223	12.0242
90	0.1516	7.0992	0.9875	7.9988	140	0.0187	6.7210	0.6177	12.1067
91	0.1454	7.0874	0.9758	8.0810	141	0.0179	6.7167	0.6131	12.1893
92	0.1394	7.0759	0.9645	8.1632	142	0.0172	6.7125	0.6086	12.2718
93	0.1337	7.0646	0.9533	8.2453	143	0.0165	6.7084	0.6042	12.3545
94	0.1282	7.0535	0.9424	8.3274	144	0.0158	6.7043	0.5998	12.4371
95	0.1230	7.0426	0.9318	8.4095	145	0.0152	6.7003	0.5955	12.5197
96	0.1179	7.0319	0.9214	8.4916	146	0.0146	6.6964	0.5912	12.6024
97	0.1131	7.0214	0.9112	8.5736	147	0.0140	6.6926	0.5870	12.6851
98	0.1085	7.0111	0.9013	8.6556	148	0.0134	6.6889	0.5829	12.7679
99	0.1040	7.0010	0.8915	8.7377	149	0.0128	6.6852	0.5788	12.8506
100	0.0998	6.9911	0.8820	8.8197	150	0.0123	6.6817	0.5748	12.9334

Notes: Calculation of W/H has an error of less than 1%.

Source: Gupta, K.C., Garg, R., Bahl, I., Bhartia, P., *Microstrip Lines and Slotlines,* 2nd Ed., Artech House, Norwood, MA, 1996, 103.

TABLE 12g Zero Thickness Microstrip Dimensions, Effective Dielectric Constant, and PUL Capacitance and Inductance for Er =11.9

Zo [Ohms]	W/h	Keff	pF/cm	nH/mm	Zo [Ohms]	W/h	Keff	pF/cm	nH/mm
1	106.5298	11.6168	113.6899	0.1137	55	0.6453	7.8084	1.6947	5.1265
2	52.1270	11.3637	56.2223	0.2249	56	0.6179	7.7908	1.6626	5.2139
3	34.0530	11.1365	37.1049	0.3339	57	0.5918	7.7731	1.6316	5.3009
4	25.0454	10.9312	27.5710	0.4411	58	0.5667	7.7554	1.6016	5.3878
5	19.6583	10.7446	21.8678	0.5467	59	0.5428	7.7375	1.5726	5.4744
6	16.0785	10.5741	18.0780	0.6508	60	0.5199	7.7197	1.5446	5.5607
7	13.5298	10.4175	15.3802	0.7536	61	0.4980	7.7019	1.5176	5.6469
8	11.6245	10.2730	13.3640	0.8553	62	0.4771	7.6841	1.4914	5.7328
9	10.1474	10.1390	11.8014	0.9559	63	0.4571	7.6664	1.4660	5.8186
10	8.9696	10.0144	10.5558	1.0556	64	0.4379	7.6488	1.4414	5.9041
11	8.0091	9.8981	9.5403	1.1544	65	0.4195	7.6313	1.4176	5.9895
12	7.2113	9.7891	8.6970	1.2524	66	0.4020	7.6140	1.3946	6.0748
13	6.5385	9.6867	7.9859	1.3496	67	0.3852	7.5968	1.3722	6.1598
14	5.9637	9.5902	7.3784	1.4462	68	0.3691	7.5797	1.3505	6.2447
15	5.4672	9.4991	6.8538	1.5421	69	0.3536	7.5628	1.3295	6.3295
16	5.0342	9.4128	6.3961	1.6374	70	0.3389	7.5462	1.3090	6.4142
17	4.6534	9.3309	5.9937	1.7322	71	0.3247	7.5297	1.2892	6.4987
18	4.3160	9.2531	5.6370	1.8264	72	0.3112	7.5134	1.2699	6.5831
19	4.0152	9.1789	5.3189	1.9201	73	0.2982	7.4973	1.2511	6.6674
20	3.7454	9.1081	5.0334	2.0134	74	0.2858	7.4814	1.2329	6.7516
21	3.5020	9.0404	4.7759	2.1062	75	0.2739	7.4658	1.2152	6.8356
22	3.2816	8.9755	4.5424	2.1985	76	0.2625	7.4504	1.1980	6.9196
23	3.0810	8.9134	4.3298	2.2905	77	0.2516	7.4352	1.1812	7.0035
24	2.8977	8.8536	4.1355	2.3820	78	0.2411	7.4202	1.1649	7.0873
25	2.7297	8.7961	3.9572	2.4732	79	0.2311	7.4055	1.1490	7.1711

TABLE 12g (continued)

Zo [Ohms]	W/h	Keff	pF/cm	nH/mm	Zo [Ohms]	W/h	Keff	pF/cm	nH/mm
26	2.5751	8.7408	3.7930	2.5641	80	0.2215	7.3910	1.1335	7.2547
27	2.4324	8.6874	3.6413	2.6545	81	0.2123	7.3767	1.1185	7.3383
28	2.3004	8.6359	3.5009	2.7447	82	0.2035	7.3627	1.1038	7.4218
29	2.1779	8.5861	3.3704	2.8345	83	0.1950	7.3489	1.0895	7.5053
30	2.0640	8.5378	3.2489	2.9240	84	0.1869	7.3353	1.0755	7.5887
31	1.9578	8.4911	3.1355	3.0132	85	0.1791	7.3220	1.0619	7.6721
32	1.8513	8.4425	3.0288	3.1014	86	0.1717	7.3089	1.0486	7.7554
33	1.7607	8.3995	2.9295	3.1902	87	0.1646	7.2960	1.0356	7.8386
34	1.6756	8.3577	2.8362	3.2787	88	0.1577	7.2834	1.0230	7.9219
35	1.5956	8.3171	2.7485	3.3669	89	0.1512	7.2709	1.0106	8.0051
36	1.5204	8.2776	2.6658	3.4549	90	0.1449	7.2588	0.9985	8.0882
37	1.4494	8.2391	2.5877	3.5426	91	0.1389	7.2468	0.9868	8.1713
38	1.3824	8.2016	2.5139	3.6301	92	0.1331	7.2350	0.9752	8.2544
39	1.3190	8.1651	2.4440	3.7173	93	0.1276	7.2235	0.9640	8.3375
40	1.2591	8.1295	2.3777	3.8043	94	0.1223	7.2122	0.9530	8.4206
41	1.2023	8.0947	2.3147	3.8910	95	0.1172	7.2011	0.9422	8.5036
42	1.1484	8.0607	2.2548	3.9775	96	0.1124	7.1902	0.9317	8.5866
43	1.0973	8.0275	2.1979	4.0639	97	0.1077	7.1795	0.9214	8.6696
44	1.0487	7.9951	2.1436	4.1499	98	0.1033	7.1691	0.9113	8.7526
45	1.0026	7.9633	2.0918	4.2358	99	0.0990	7.1588	0.9015	8.8356
46	0.9586	7.9500	2.0446	4.3263	100	0.0949	7.1487	0.8919	8.9185
47	0.9168	7.9367	1.9994	4.4167	101	0.0909	7.1388	0.8824	9.0015
48	0.8770	7.9226	1.9560	4.5067	102	0.0872	7.1292	0.8732	9.0845
49	0.8391	7.9078	1.9143	4.5962	103	0.0836	7.1197	0.8641	9.1674
50	0.8029	7.8923	1.8742	4.6854	104	0.0801	7.1104	0.8552	9.2504
51	0.7684	7.8762	1.8356	4.7743	105	0.0768	7.1012	0.8466	9.3333
52	0.7354	7.8598	1.7984	4.8628	106	0.0736	7.0923	0.8380	9.4163
53	0.7040	7.8429	1.7626	4.9510	107	0.0705	7.0835	0.8297	9.4992
54	0.6740	7.8258	1.7280	5.0389	108	0.0676	7.0749	0.8215	9.5822
109	0.0648	7.0665	0.8135	9.6652	130	0.0266	6.9243	0.6752	11.4107
110	0.0621	7.0583	0.8056	9.7481	131	0.0255	6.9190	0.6698	11.4940
111	0.0595	7.0502	0.7979	9.8311	132	0.0245	6.9138	0.6645	11.5774
112	0.0571	7.0423	0.7903	9.9141	133	0.0235	6.9086	0.6592	11.6608
113	0.0547	7.0345	0.7829	9.9971	134	0.0225	6.9036	0.6541	11.7442
114	0.0524	7.0269	0.7756	10.0801	135	0.0216	6.8987	0.6490	11.8276
115	0.0503	7.0195	0.7685	10.1632	136	0.0207	6.8939	0.6440	11.9110
116	0.0482	7.0122	0.7615	10.2462	137	0.0198	6.8892	0.6391	11.9945
117	0.0462	7.0050	0.7546	10.3293	138	0.0190	6.8845	0.6342	12.0780
118	0.0443	6.9980	0.7478	10.4123	139	0.0182	6.8800	0.6294	12.1615
119	0.0424	6.9911	0.7412	10.4954	140	0.0174	6.8756	0.6247	12.2451
120	0.0407	6.9844	0.7346	10.5785	141	0.0167	6.8713	0.6201	12.3287
121	0.0390	6.9778	0.7282	10.6617	142	0.0160	6.8670	0.6156	12.4123
122	0.0374	6.9714	0.7219	10.7448	143	0.0154	6.8628	0.6111	12.4959
123	0.0358	6.9651	0.7157	10.8280	144	0.0147	6.8588	0.6067	12.5795
124	0.0343	6.9589	0.7096	10.9112	145	0.0141	6.8548	0.6023	12.6632
125	0.0329	6.9528	0.7036	10.9944	146	0.0135	6.8509	0.5980	12.7469
126	0.0316	6.9469	0.6978	11.0776	147	0.0130	6.8471	0.5938	12.8306
127	0.0302	6.9411	0.6920	11.1608	148	0.0124	6.8433	0.5896	12.9144
128	0.0290	6.9354	0.6863	11.2441	149	0.0119	6.8396	0.5855	12.9982
129	0.0278	6.9298	0.6807	11.3274	150	0.0114	6.8361	0.5814	13.0820

Notes: Calculation of W/H has an error of less than 1%.

Source: Gupta, K.C., Garg, R., Bahl, I., Bhartia, P., *Microstrip Lines and Slotlines,* 2nd Ed., Artech House, Norwood, MA, 1996, 103.

TABLE 12h Zero Thickness Microstrip Dimensions, Effective Dielectric Constant, and PUL Capacitance and Inductance for Er =12.88

Zo [Ohms]	W/h	Keff	pF/cm	nH/mm	Zo [Ohms]	W/h	Keff	pF/cm	nH/mm
1	102.3159	12.5596	118.2135	0.1182	24	2.7402	9.5011	4.2841	2.4676
2	50.0315	12.2746	58.4323	0.2337	25	2.5791	9.4384	4.0991	2.5619
3	32.6633	12.0197	38.5483	0.3469	26	2.4309	9.3779	3.9288	2.6559
4	24.0084	11.7903	28.6339	0.4581	27	2.2942	9.3197	3.7715	2.7494
5	18.8329	11.5823	22.7043	0.5676	28	2.1677	9.2635	3.6258	2.8427
6	15.3941	11.3928	18.7648	0.6755	29	2.0503	9.2091	3.4905	2.9355
7	12.9460	11.2191	15.9610	0.7821	30	1.9316	9.1518	3.3637	3.0273
8	11.1161	11.0591	13.8660	0.8874	31	1.8324	9.1020	3.2463	3.1197
9	9.6977	10.9111	12.2425	0.9916	32	1.7397	9.0536	3.1365	3.2117
10	8.5668	10.7736	10.9487	1.0949	33	1.6528	9.0068	3.0335	3.3035
11	7.6446	10.6455	9.8939	1.1972	34	1.5714	8.9612	2.9369	3.3950
12	6.8787	10.5255	9.0182	1.2986	35	1.4948	8.9170	2.8459	3.4862
13	6.2329	10.4130	8.2799	1.3993	36	1.4228	8.8739	2.7602	3.5772
14	5.6812	10.3071	7.6492	1.4993	37	1.3549	8.8320	2.6792	3.6678
15	5.2048	10.2071	7.1046	1.5985	38	1.2908	8.7911	2.6027	3.7582
16	4.7893	10.1125	6.6296	1.6972	39	1.2302	8.7513	2.5302	3.8484
17	4.4239	10.0228	6.2119	1.7952	40	1.1729	8.7125	2.4614	3.9383
18	4.1003	9.9376	5.8418	1.8927	41	1.1186	8.6745	2.3962	4.0280
19	3.8117	9.8565	5.5117	1.9897	42	1.0672	8.6375	2.3341	4.1174
20	3.5529	9.7791	5.2155	2.0862	43	1.0184	8.6014	2.2751	4.2066
21	3.3196	9.7051	4.9483	2.1822	44	0.9721	8.5791	2.2205	4.2989
22	3.1082	9.6342	4.7061	2.2778	45	0.9281	8.5645	2.1693	4.3928
23	2.9159	9.5663	4.4856	2.3729	46	0.8863	8.5487	2.1202	4.4863
47	0.8465	8.5321	2.0731	4.5794	99	0.0845	7.6726	0.9333	9.1471
48	0.8086	8.5147	2.0278	4.6720	100	0.0808	7.6620	0.9233	9.2331
49	0.7726	8.4967	1.9843	4.7643	101	0.0774	7.6516	0.9136	9.3192
50	0.7383	8.4781	1.9425	4.8562	102	0.0740	7.6415	0.9040	9.4052
51	0.7056	8.4591	1.9023	4.9478	103	0.0709	7.6315	0.8946	9.4912
52	0.6744	8.4397	1.8635	5.0390	104	0.0678	7.6218	0.8855	9.5773
53	0.6447	8.4201	1.8263	5.1299	105	0.0649	7.6123	0.8765	9.6633
54	0.6163	8.4002	1.7903	5.2206	106	0.0621	7.6030	0.8677	9.7494
55	0.5893	8.3802	1.7557	5.3109	107	0.0595	7.5938	0.8591	9.8354
56	0.5634	8.3601	1.7223	5.4010	108	0.0569	7.5849	0.8506	9.9215
57	0.5388	8.3399	1.6900	5.4908	109	0.0545	7.5761	0.8423	10.0076
58	0.5152	8.3198	1.6588	5.5804	110	0.0521	7.5675	0.8342	10.0937
59	0.4928	8.2996	1.6288	5.6697	111	0.0499	7.5592	0.8262	10.1798
60	0.4713	8.2796	1.5997	5.7588	112	0.0477	7.5509	0.8184	10.2659
61	0.4508	8.2596	1.5716	5.8477	113	0.0457	7.5429	0.8107	10.3521
62	0.4312	8.2397	1.5443	5.9365	114	0.0437	7.5350	0.8032	10.4382
63	0.4125	8.2200	1.5180	6.0250	115	0.0418	7.5273	0.7958	10.5244
64	0.3946	8.2005	1.4925	6.1133	116	0.0400	7.5198	0.7885	10.6106
65	0.3775	8.1811	1.4678	6.2015	117	0.0383	7.5124	0.7814	10.6968
66	0.3611	8.1619	1.4439	6.2895	118	0.0367	7.5052	0.7744	10.7830
67	0.3455	8.1429	1.4207	6.3774	119	0.0351	7.4981	0.7676	10.8693
68	0.3306	8.1242	1.3982	6.4652	120	0.0336	7.4912	0.7608	10.9556
69	0.3163	8.1057	1.3763	6.5528	121	0.0321	7.4844	0.7542	11.0419
70	0.3026	8.0874	1.3551	6.6402	122	0.0308	7.4778	0.7477	11.1282
71	0.2895	8.0694	1.3346	6.7276	123	0.0294	7.4713	0.7413	11.2146
72	0.2770	8.0517	1.3146	6.8148	124	0.0282	7.4649	0.7350	11.3009
73	0.2651	8.0342	1.2952	6.9020	125	0.0270	7.4587	0.7288	11.3873
74	0.2536	8.0169	1.2763	6.9890	126	0.0258	7.4526	0.7227	11.4737
75	0.2427	8.0000	1.2579	7.0759	127	0.0247	7.4467	0.7167	11.5602
76	0.2323	7.9833	1.2401	7.1628	128	0.0236	7.4409	0.7109	11.6466
77	0.2222	7.9669	1.2227	7.2496	129	0.0226	7.4352	0.7051	11.7331
78	0.2127	7.9507	1.2058	7.3363	130	0.0217	7.4296	0.6994	11.8197
79	0.2035	7.9348	1.1894	7.4229	131	0.0207	7.4241	0.6938	11.9062

TABLE 12h (continued)

Zo [Ohms]	W/h	Keff	pF/cm	nH/mm	Zo [Ohms]	W/h	Keff	pF/cm	nH/mm
80	0.1947	7.9192	1.1734	7.5095	132	0.0198	7.4188	0.6883	11.9928
81	0.1864	7.9039	1.1578	7.5960	133	0.0190	7.4136	0.6829	12.0794
82	0.1783	7.8889	1.1425	7.6825	134	0.0182	7.4085	0.6775	12.1660
83	0.1707	7.8741	1.1277	7.7689	135	0.0174	7.4035	0.6723	12.2527
84	0.1633	7.8596	1.1133	7.8552	136	0.0166	7.3986	0.6671	12.3393
85	0.1563	7.8453	1.0992	7.9415	137	0.0159	7.3938	0.6621	12.4261
86	0.1496	7.8314	1.0854	8.0278	138	0.0152	7.3891	0.6570	12.5128
87	0.1431	7.8176	1.0720	8.1140	139	0.0146	7.3845	0.6521	12.5996
88	0.1370	7.8042	1.0589	8.2002	140	0.0140	7.3800	0.6473	12.6863
89	0.1311	7.7910	1.0461	8.2864	141	0.0134	7.3756	0.6425	12.7732
90	0.1255	7.7781	1.0336	8.3725	142	0.0128	7.3713	0.6378	12.8600
91	0.1201	7.7654	1.0215	8.4587	143	0.0122	7.3671	0.6331	12.9469
92	0.1149	7.7529	1.0095	8.5448	144	0.0117	7.3630	0.6286	13.0338
93	0.1100	7.7407	0.9979	8.6309	145	0.0112	7.3590	0.6241	13.1207
94	0.1052	7.7288	0.9865	8.7169	146	0.0107	7.3551	0.6196	13.2077
95	0.1007	7.7171	0.9754	8.8030	147	0.0103	7.3512	0.6152	13.2946
96	0.0964	7.7056	0.9645	8.8890	148	0.0098	7.3475	0.6109	13.3817
97	0.0922	7.6944	0.9539	8.9751	149	0.0094	7.3438	0.6067	13.4687
98	0.0883	7.6833	0.9435	9.0611	150	0.0090	7.3402	0.6025	13.5558

Notes: Calculation of W/H has an error of less than 1%.

Source: Gupta, K.C., Garg, R., Bahl, I., Bhartia, P., *Microstrip Lines and Slotlines,* 2nd Ed., Artech House, Norwood, MA, 1996, 103.

TABLE 12i Zero Thickness Microstrip Dimensions, Effective Dielectric Constant, and PUL Capacitance and Inductance for Er =35

Zo [Ohms]	W/h	Keff	pF/cm	nH/mm	Zo [Ohms]	W/h	Keff	pF/cm	nH/mm
1	61.2525	33.5453	193.1947	0.1932	57	0.1141	20.3209	2.6380	8.5709
2	29.6179	32.3412	94.8478	0.3794	58	0.1063	20.2652	2.5890	8.7093
3	19.1302	31.3265	62.2321	0.5601	59	0.0990	20.2113	2.5417	8.8477
4	13.9141	30.4569	46.0216	0.7363	60	0.0923	20.1591	2.4961	8.9860
5	10.8011	29.7005	36.3573	0.9089	61	0.0860	20.1087	2.4521	9.1243
6	8.7367	29.0345	29.9561	1.0784	62	0.0801	20.0598	2.4096	9.2626
7	7.2699	28.4417	25.4132	1.2452	63	0.0746	20.0126	2.3686	9.4010
8	6.1757	27.9094	22.0275	1.4098	64	0.0695	19.9670	2.3289	9.5393
9	5.3291	27.4273	19.4101	1.5722	65	0.0648	19.9228	2.2906	9.6776
10	4.6555	26.9878	17.3286	1.7329	66	0.0604	19.8801	2.2534	9.8160
11	4.1073	26.5845	15.6351	1.8918	67	0.0562	19.8389	2.2175	9.9543
12	3.6529	26.2124	14.2315	2.0493	68	0.0524	19.7990	2.1827	10.0928
13	3.2705	25.8674	13.0501	2.2055	69	0.0488	19.7604	2.1490	10.2312
14	2.9445	25.5460	12.0424	2.3603	70	0.0455	19.7232	2.1163	10.3697
15	2.6635	25.2453	11.1732	2.5140	71	0.0424	19.6872	2.0846	10.5082
16	2.4189	24.9629	10.4161	2.6665	72	0.0395	19.6525	2.0538	10.6468
17	2.2042	24.6968	9.7510	2.8180	73	0.0368	19.6189	2.0239	10.7855
18	2.0144	24.4452	9.1623	2.9686	74	0.0343	19.5865	1.9949	10.9242
19	1.8358	24.1924	8.6351	3.1173	75	0.0319	19.5551	1.9667	11.0630
20	1.6891	23.9716	8.1658	3.2663	76	0.0298	19.5249	1.9394	11.2018
21	1.5569	23.7610	7.7427	3.4145	77	0.0277	19.4957	1.9127	11.3407
22	1.4372	23.5597	7.3594	3.5619	78	0.0258	19.4674	1.8869	11.4796
23	1.3284	23.3669	7.0106	3.7086	79	0.0241	19.4402	1.8617	11.6187
24	1.2292	23.1820	6.6918	3.8545	80	0.0224	19.4139	1.8372	11.7578
25	1.1385	23.0043	6.3995	3.9997	81	0.0209	19.3884	1.8133	11.8970

TABLE 12i (continued)

Zo [Ohms]	W/h	Keff	pF/cm	nH/mm	Zo [Ohms]	W/h	Keff	pF/cm	nH/mm
26	1.0554	22.8334	6.1304	4.1442	82	0.0195	19.3639	1.7900	12.0362
27	0.9790	22.6972	5.8857	4.2907	83	0.0181	19.3402	1.7674	12.1755
28	0.9087	22.6289	5.6670	4.4429	84	0.0169	19.3173	1.7453	12.3149
29	0.8438	22.5532	5.4624	4.5939	85	0.0157	19.2952	1.7238	12.4544
30	0.7840	22.4719	5.2708	4.7437	86	0.0147	19.2739	1.7028	12.5940
31	0.7286	22.3863	5.0911	4.8925	87	0.0137	19.2533	1.6823	12.7336
32	0.6774	22.2977	4.9222	5.0403	88	0.0127	19.2334	1.6624	12.8733
33	0.6300	22.2069	4.7633	5.1873	89	0.0119	19.2142	1.6429	13.0131
34	0.5861	22.1148	4.6136	5.3333	90	0.0111	19.1957	1.6238	13.1530
35	0.5453	22.0220	4.4724	5.4787	91	0.0103	19.1778	1.6052	13.2929
36	0.5074	21.9291	4.3390	5.6233	92	0.0096	19.1605	1.5871	13.4329
37	0.4723	21.8365	4.2128	5.7673	93	0.0089	19.1439	1.5693	13.5730
38	0.4397	21.7445	4.0933	5.9107	94	0.0083	19.1278	1.5520	13.7132
39	0.4093	21.6536	3.9800	6.0535	95	0.0078	19.1122	1.5350	13.8535
40	0.3811	21.5638	3.8724	6.1959	96	0.0072	19.0972	1.5184	13.9938
41	0.3549	21.4755	3.7702	6.3378	97	0.0067	19.0828	1.5022	14.1342
42	0.3305	21.3889	3.6730	6.4792	98	0.0063	19.0688	1.4863	14.2747
43	0.3078	21.3039	3.5805	6.6203	99	0.0059	19.0553	1.4708	14.4152
44	0.2867	21.2208	3.4923	6.7610	100	0.0055	19.0422	1.4556	14.5559
45	0.2670	21.1396	3.4081	6.9014	101	0.0051	19.0297	1.4407	14.6966
46	0.2487	21.0603	3.3278	7.0416	102	0.0047	19.0175	1.4261	14.8374
47	0.2317	20.9831	3.2510	7.1814	103	0.0044	19.0058	1.4118	14.9782
48	0.2158	20.9078	3.1776	7.3211	104	0.0041	18.9945	1.3978	15.1191
49	0.2010	20.8346	3.1073	7.4605	105	0.0038	18.9836	1.3841	15.2601
50	0.1873	20.7635	3.0399	7.5998	106	0.0036	18.9730	1.3707	15.4012
51	0.1745	20.6944	2.9753	7.7388	107	0.0033	18.9629	1.3575	15.5423
52	0.1625	20.6272	2.9134	7.8778	108	0.0031	18.9530	1.3446	15.6835
53	0.1514	20.5621	2.8539	8.0166	109	0.0029	18.9436	1.3319	15.8247
54	0.1411	20.4989	2.7967	8.1553	110	0.0027	18.9344	1.3195	15.9661
55	0.1314	20.4377	2.7418	8.2939	111	0.0025	18.9256	1.3073	16.1074
56	0.1224	20.3784	2.6889	8.4324	112	0.0023	18.9170	1.2954	16.2489
113	0.0022	18.9088	1.2836	16.3904	132	0.0006	18.7969	1.0956	19.0896
114	0.0020	18.9009	1.2721	16.5320	133	0.0005	18.7928	1.0872	19.2321
115	0.0019	18.8932	1.2608	16.6736	134	0.0005	18.7889	1.0790	19.3747
116	0.0018	18.8858	1.2497	16.8153	135	0.0005	18.7851	1.0709	19.5173
117	0.0016	18.8786	1.2387	16.9571	136	0.0004	18.7815	1.0629	19.6600
118	0.0015	18.8717	1.2280	17.0989	137	0.0004	18.7779	1.0551	19.8027
119	0.0014	18.8651	1.2175	17.2407	138	0.0004	18.7745	1.0473	19.9454
120	0.0013	18.8587	1.2071	17.3826	139	0.0003	18.7713	1.0397	20.0882
121	0.0012	18.8524	1.1970	17.5246	140	0.0003	18.7681	1.0322	20.2310
122	0.0012	18.8465	1.1870	17.6666	141	0.0003	18.7650	1.0248	20.3738
123	0.0011	18.8407	1.1771	17.8087	142	0.0003	18.7621	1.0175	20.5167
124	0.0010	18.8351	1.1675	17.9508	143	0.0003	18.7592	1.0103	20.6596
125	0.0009	18.8297	1.1580	18.0930	144	0.0002	18.7565	1.0032	20.8026
126	0.0009	18.8245	1.1486	18.2352	145	0.0002	18.7538	0.9962	20.9456
127	0.0008	18.8195	1.1394	18.3775	146	0.0002	18.7513	0.9893	21.0886
128	0.0008	18.8146	1.1304	18.5198	147	0.0002	18.7488	0.9825	21.2316
129	0.0007	18.8100	1.1215	18.6622	148	0.0002	18.7464	0.9758	21.3747
130	0.0007	18.8055	1.1127	18.8046	149	0.0002	18.7441	0.9692	21.5178
131	0.0006	18.8011	1.1041	18.9471	150	0.0002	18.7419	0.9627	21.6609

Notes: Calculation of W/H has an error of less than 1%.

Source: Gupta, K.C., Garg, R., Bahl, I., Bhartia, P., *Microstrip Lines and Slotlines,* 2nd Ed., Artech House, Norwood, MA, 1996, 103.

TABLE 12j Zero Thickness Microstrip Dimensions, Effective Dielectric Constant, and PUL Capacitance and Inductance for Er =85

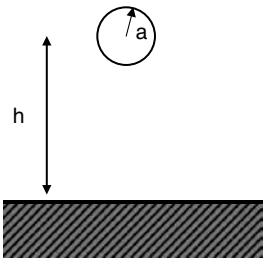
Zo [Ohms]	W/h	Keff	pF/cm	nH/mm	Zo [Ohms]	W/h	Keff	pF/cm	nH/mm
1	38.5925	79.6824	297.7560	0.2978	26	0.3739	51.7462	9.2288	6.2387
2	18.3705	75.6651	145.0765	0.5803	27	0.3349	51.4124	8.8583	6.4577
3	11.6859	72.5009	94.6738	0.8521	28	0.3000	51.0886	8.5150	6.6757
4	8.3710	69.9235	69.7318	1.1157	29	0.2688	50.7757	8.1961	6.8930
5	6.3982	67.7680	54.9189	1.3730	30	0.2409	50.4745	7.8994	7.1095
6	5.0937	65.9270	45.1398	1.6250	31	0.2159	50.1850	7.6226	7.3253
7	4.1695	64.3277	38.2191	1.8727	32	0.1935	49.9076	7.3640	7.5407
8	3.4821	62.9184	33.0733	2.1167	33	0.1734	49.6422	7.1218	7.7557
9	2.9517	61.6613	29.1034	2.3574	34	0.1554	49.3886	6.8947	7.9702
10	2.5309	60.5285	25.9513	2.5951	35	0.1393	49.1467	6.6813	8.1845
11	2.1895	59.4981	23.3904	2.8302	36	0.1249	48.9161	6.4804	8.3986
12	1.8946	58.5089	21.2622	3.0618	37	0.1119	48.6966	6.2911	8.6125
13	1.6651	57.6611	19.4840	3.2928	38	0.1003	48.4877	6.1124	8.8263
14	1.4696	56.8730	17.9682	3.5218	39	0.0899	48.2891	5.9435	9.0400
15	1.3012	56.1365	16.6614	3.7488	40	0.0806	48.1003	5.7835	9.2537
16	1.1551	55.4455	15.5236	3.9740	41	0.0723	47.9210	5.6319	9.4673
17	1.0274	54.7949	14.5245	4.1976	42	0.0648	47.7509	5.4881	9.6810
18	0.9153	54.4535	13.6748	4.4306	43	0.0581	47.5893	5.3514	9.8947
19	0.8165	54.1607	12.9202	4.6642	44	0.0521	47.4361	5.2213	10.1085
20	0.7290	53.8383	12.2376	4.8950	45	0.0467	47.2907	5.0975	10.3224
21	0.6514	53.4976	11.6179	5.1235	46	0.0418	47.1529	4.9794	10.5364
22	0.5825	53.1467	11.0534	5.3498	47	0.0375	47.0222	4.8667	10.7505
23	0.5211	52.7921	10.5375	5.5743	48	0.0336	46.8984	4.7590	10.9648
24	0.4664	52.4383	10.0645	5.7972	49	0.0301	46.7810	4.6561	11.1792
25	0.4176	52.0888	9.6297	6.0185	50	0.0270	46.6697	4.5575	11.3938
51	0.0242	46.5643	4.4631	11.6085	101	0.0001	44.8028	2.2106	22.5503
52	0.0217	46.4644	4.3726	11.8234	102	0.0001	44.7963	2.1888	22.7720
53	0.0195	46.3698	4.2857	12.0385	103	0.0001	44.7901	2.1674	22.9936
54	0.0175	46.2801	4.2023	12.2538	104	0.0001	44.7842	2.1464	23.2153
55	0.0156	46.1952	4.1221	12.4693	105	0.0001	44.7787	2.1258	23.4371
56	0.0140	46.1148	4.0449	12.6849	106	0.0001	44.7734	2.1056	23.6589
57	0.0126	46.0386	3.9707	12.9008	107	0.0001	44.7685	2.0858	23.8808
58	0.0113	45.9665	3.8992	13.1168	108	0.0000	44.7638	2.0664	24.1027
59	0.0101	45.8982	3.8302	13.3330	109	0.0000	44.7593	2.0474	24.3247
60	0.0091	45.8335	3.7637	13.5495	110	0.0000	44.7551	2.0287	24.5467
61	0.0081	45.7722	3.6996	13.7661	111	0.0000	44.7511	2.0103	24.7688
62	0.0073	45.7142	3.6376	13.9829	112	0.0000	44.7473	1.9923	24.9908
63	0.0065	45.6592	3.5777	14.1999	113	0.0000	44.7437	1.9745	25.2130
64	0.0059	45.6072	3.5198	14.4170	114	0.0000	44.7404	1.9572	25.4351
65	0.0052	45.5579	3.4638	14.6344	115	0.0000	44.7371	1.9401	25.6573
66	0.0047	45.5112	3.4095	14.8519	116	0.0000	44.7341	1.9233	25.8795
67	0.0042	45.4670	3.3570	15.0696	117	0.0000	44.7312	1.9068	26.1018
68	0.0038	45.4252	3.3061	15.2875	118	0.0000	44.7285	1.8906	26.3241
69	0.0034	45.3856	3.2568	15.5056	119	0.0000	44.7259	1.8746	26.5464
70	0.0030	45.3481	3.2089	15.7238	120	0.0000	44.7235	1.8589	26.7688
71	0.0027	45.3126	3.1625	15.9422	121	0.0000	44.7212	1.8435	26.9911
72	0.0024	45.2789	3.1174	16.1607	122	0.0000	44.7190	1.8284	27.2135
73	0.0022	45.2471	3.0736	16.3794	123	0.0000	44.7169	1.8135	27.4360
74	0.0020	45.2169	3.0311	16.5982	124	0.0000	44.7149	1.7988	27.6584
75	0.0018	45.1884	2.9897	16.8172	125	0.0000	44.7131	1.7844	27.8809
76	0.0016	45.1614	2.9495	17.0363	126	0.0000	44.7113	1.7702	28.1034
77	0.0014	45.1358	2.9104	17.2556	127	0.0000	44.7097	1.7562	28.3259
78	0.0013	45.1115	2.8723	17.4750	128	0.0000	44.7081	1.7425	28.5484
79	0.0011	45.0886	2.8352	17.6946	129	0.0000	44.7066	1.7289	28.7710
80	0.0010	45.0669	2.7991	17.9142	130	0.0000	44.7052	1.7156	28.9936

TABLE 12j (continued)

Zo [Ohms]	W/h	Keff	pF/cm	nH/mm	Zo [Ohms]	W/h	Keff	pF/cm	nH/mm
81	0.0009	45.0463	2.7639	18.1340	131	0.0000	44.7038	1.7025	29.2162
82	0.0008	45.0268	2.7296	18.3539	132	0.0000	44.7026	1.6896	29.4388
83	0.0007	45.0084	2.6962	18.5739	133	0.0000	44.7014	1.6768	29.6614
84	0.0007	44.9909	2.6636	18.7941	134	0.0000	44.7002	1.6643	29.8840
85	0.0006	44.9744	2.6317	19.0143	135	0.0000	44.6992	1.6519	30.1067
86	0.0005	44.9587	2.6007	19.2347	136	0.0000	44.6981	1.6398	30.3293
87	0.0005	44.9439	2.5704	19.4551	137	0.0000	44.6972	1.6278	30.5520
88	0.0004	44.9299	2.5408	19.6757	138	0.0000	44.6963	1.6160	30.7747
89	0.0004	44.9166	2.5118	19.8963	139	0.0000	44.6954	1.6043	30.9974
90	0.0003	44.9040	2.4836	20.1171	140	0.0000	44.6946	1.5929	31.2201
91	0.0003	44.8921	2.4560	20.3379	141	0.0000	44.6938	1.5816	31.4429
92	0.0003	44.8808	2.4290	20.5588	142	0.0000	44.6931	1.5704	31.6656
93	0.0002	44.8701	2.4026	20.7798	143	0.0000	44.6924	1.5594	31.8883
94	0.0002	44.8600	2.3767	21.0009	144	0.0000	44.6917	1.5486	32.1111
95	0.0002	44.8504	2.3515	21.2220	145	0.0000	44.6911	1.5379	32.3339
96	0.0002	44.8414	2.3267	21.4432	146	0.0000	44.6905	1.5273	32.5567
97	0.0002	44.8328	2.3025	21.6645	147	0.0000	44.6899	1.5169	32.7794
98	0.0001	44.8247	2.2788	21.8859	148	0.0000	44.6894	1.5067	33.0022
99	0.0001	44.8170	2.2556	22.1073	149	0.0000	44.6889	1.4966	33.2250
100	0.0001	44.8097	2.2329	22.3288	150	0.0000	44.6884	1.4866	33.4478

Notes: Calculation of W/H has an error of less than 1%.

Source: Gupta, K.C., Garg, R., Bahl, I., Bhartia, P., *Microstrip Lines and Slotlines,* 2nd Ed., Artech House, Norwood, MA, 1996, 103.



$$L_{WIRE\ over\ GND} = \frac{\mu_0 \mu_r}{2\pi} \cosh^{-1} \frac{h}{a} \text{ (nH/m)} \qquad \text{for } h \gg a$$

FIGURE B.6 Inductance of a wire over a ground plane. When consistent units are used for the wire radius, a, and the height over the ground plane, h, the formula provides an estimate of the inductance per unit length in (nH/m).

Bondwires, Ribbons, Mesh

The assembly of semiconductor and hybrid integrated circuits for microwave and millimeter-wave frequencies generally requires the use of gold bondwires, ribbons, or mesh. The impedance of the interconnection must be accounted for in a good design. Unfortunately, there is no single accepted electrical model. The complexity required of the model will depend on the frequency of operation and the general impedance levels of the circuits being connected. At low frequencies and moderate to high impedances, the connection is frequently modeled as an inductor (sometimes in series with a resistor); at high frequencies, a full 3-D electromagnetic simulation may be required for accurate results. At intermediate points it may be modeled as a high impedance transmission line or as a lumped LC circuit. Note that the resistances of the rf interconnects should be included in the design of extremely low-noise circuits as they will affect the noise figure. In connecting a semiconductor die to package leads, it may also be necessary to model

the mutual inductances and interlead capacitances in addition to the usual self inductances and shunt capacitance. Figure 13 illustrates one method of modeling bond wire inductance that has been shown adequate for many microwave applications. More sophisticated methods of modeling bond wires, ribbon or mesh are described in the references.

References

1. Grover, F.W., *Inductance Calculations*, Dover Publications, New York, NY Available through Instrument Society of America, Research Triangle Park, NC.
2. Caulton, M., Lumped Elements in Microwave Circuits, from Advances in Microwaves 1974, Academic Press, NY 1974, pp. 143–202.
3. Wadell, B. C., Transmission Line Design Handbook, Artech House, Boston, MA 1991, pp. 151–155.
4. Terman, F.E., *Radio Engineer's Handbook*, McGraw-Hill, 1943 pp. 48–51.
5. Caverly, R. H., "Characteristic Impedance of Integrated Circuit Bond Wires," IEEE Transactions on Microwave Theory and Techniques, Vol MTT-34, No. 9, September, 1986, pp. 982–984.
6. Kuester, E. F., Chang, D. C., "Propagating Modes Along a Thin Wire Located Above a Grounded Dielectric Slab," IEEE Transactions on Microwave Theory and Techniques, Vol MTT-25, No. 12, December, 1977, pp. 1065–1069.
7. Mondal, J. P. "Octagonal spiral inductor measurement and modelling for MMIC applications," Int. J. Electronics, 1990, Vol 68, No. 1, pp. 113–125.
8. MIL-STD-883E, "Test Method Standard, Microcircuits" US Department of Defense.
9. Gupta, K. C., Garg, E., Bahl, I., Bhartia, P., *Microstrip Lines and Slothlines*, 2nd Ed., Artech House, Norwood, MA, 1996, 103.

Index

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

Y

Electrical Engineering

RF AND MICROWAVE CIRCUITS, MEASUREMENTS, AND MODELING

The unique behavior and requirements associated with RF and microwave systems establishes a need for unique and complex models and simulation tools. The required toolset for a microwave circuit designer includes unique device models, both 2D and 3D electromagnetic simulators, as well as frequency domain based small signal and large signal circuit and system simulators. This unique suite of tools requires a design procedure that is also distinctive. With 532 illustrations, 71 tables, and 573 equations, **RF and Microwave Circuits, Measurements, and Modeling** highlights the challenges RF and microwave circuit designers face in their day-to-day tasks. Discussing the technologies and applications at the heart of modern wireless communications, the book examines not only the distinct design tools of the microwave circuit designer, but also the design procedures that must be followed to use them effectively.

New chapters discuss handset and high-power PAs, magnitude and phase calibration, large signal network analysis as well as behavior modeling of power amplifiers. Other chapters feature new and updated material that reflects recent progress in such areas as linear and pulsed measurements, transmitters and receivers, and CAD tools.

Features

- Details small-signal, phase, pulsed, waveform and noise measurements and calibration procedures for these measurements
- Explores RF and microwave circuit designs in terms of performance and critical design specifications
- Discusses transmitters and receivers first in terms of functional circuit block, then examines each block individually
- Examines fundamental amplifier issues, low noise amplifiers, power amplifiers for handset applications and high power amplifiers
- Covers other circuit functions including oscillators, mixers, modulators, phase locked loops, filters and multiplexers

7218
ISBN 0-8493-7218-6
90000
9 780849 372186
www.crcpress.com

CRC Press
Taylor & Francis Group
an informa business
www.taylorandfrancisgroup.com

6000 Broken Sound Parkway, NW
Suite 300, Boca Raton, FL 33487
270 Madison Avenue
New York, NY 10016
2 Park Square, Milton Park
Abingdon, Oxon OX14 4RN, UK