Second Edition

RF and Microwave Handbook

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

Preface ix

Acknowledgments xi

Editors xiii

Advisory Board xv

Contributors xix

Introduction to Microwaves and RF
Patrick Fay **I**-1

1 Overview of Microwave Engineering
Mike Golio **1**-1

SECTION I Microwave Measurements

2 Linear Measurements
Ronald E. Ham **2**-1

3 Network Analyzer Calibration
Joseph Staudinger **3**-1

4 Absolute Magnitude and Phase Calibrations
Kate A. Remley, Paul D. Hale, and Dylan F. Williams **4**-1

5 Noise Measurements
Alfy Riddle **5**-1

6 Nonlinear Microwave Measurement and Characterization
J. Stevenson Kenney **6**-1

7 Theory of High-Power Load-Pull Characterization for RF and Microwave Transistors
John F. Sevic **7**-1

8 Pulsed Measurements
Anthony E. Parker, James G. Rathmell, and Jonathan B. Scott **8**-1

9 Microwave On-Wafer Test
Jean-Pierre Lanteri, Christopher Jones, and John R. Mahon **9**-1

10 High Volume Microwave Test
Jean-Pierre Lanteri, Christopher Jones, and John R. Mahon **10**-1

11 Large Signal Network Analysis/Waveform Measurements
Joseph M. Gering **11**-1

SECTION II Circuits

12 Receivers
Warren L. Seely **12**-1

13 Transmitters
Warren L. Seely **13**-1

14 Low Noise Amplifier Design
Jakub Kucera and Urs Lott **14**-1

15 Microwave Mixer Design
Anthony M. Pavio **15**-1

16 Modulation and Demodulation Circuitry
Charles Nelson **16**-1

17 Power Amplifier Fundamentals
Douglas A. Teeter and Edward T. Spears **17**-1

18 Handset Power Amplifier Design
Douglas A. Teeter and Edward T. Spears **18**-1

19 Class A Amplifiers
Warren L. Seely **19**-1

20 High Power Amplifiers
Brent Irvine, Todd Heckleman, and Aaron Radomski **20**-1

21 Oscillator Circuits
Alfy Riddle **21**-1

22 Phase Locked Loop Design
Robert Newgard **22**-1

23 Filters and Multiplexers
Richard V. Snyder **23**-1

24 RF Switches
Robert J. Trew **24**-1

25 Monolithic Microwave IC Technology
Lawrence P. Dunleavy **25**-1

26 Bringing RFICs to the Market
John C. Cowles **26**-1

SECTION III CAD, Simulation and Modeling

27 System Simulation
Joseph Staudinger **27**-1

28 Numerical Techniques for the Analysis and Design of RF/Microwave Structures
Manos M. Tentzeris **28**-1

29 Computer Aided Design of Passive Components
Daniel G. Swanson, Jr. **29**-1

30 Nonlinear RF and Microwave Circuit Analysis
Michael B. Steer and John F. Sevic **30**-1

31 Computer Aided Design (CAD) of Microwave Circuitry
Ron Kielmeyer **31**-1

32 Nonlinear Transistor Modeling for Circuit Simulation
Walter R. Curtice **32**-1

33 Behavioral Modeling of Microwave Power Amplifiers
Troels S. Nielsen **33**-1

34 Technology CAD
Peter A. Blakey **34**-1

Appendix A: Mathematics, Symbols, and Physical Constants **A**-1

Appendix B: Microwave Engineering Appendix
John P. Wendler **B**-1

Index **I**-1

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.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

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*.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

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

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

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

I.1 Introduction to Microwave and RF Engineering I-1
I.2 General Applications I-8
I.3 Frequency Band Definitions I-9
I.4 Overview of The RF and Microwave Handbook I-11
References I-12

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

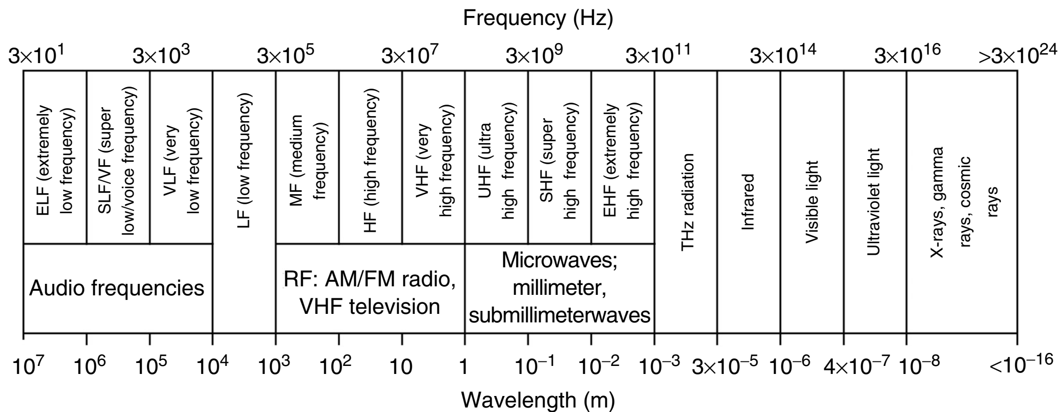


FIGURE I.1 Electromagnetic frequency spectrum and associated wavelengths.

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

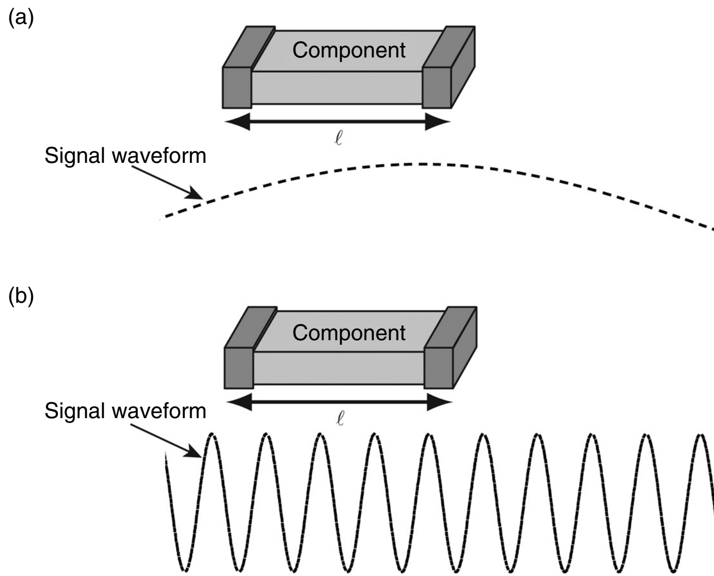


FIGURE I.2 Schematic representation of component dimensions relative to signal wavelengths. Conventional lumped-element analysis techniques are typically applicable for components for which $\ell < \lambda/10$ (a) since the phase change due to electromagnetic propagation across the component is small,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

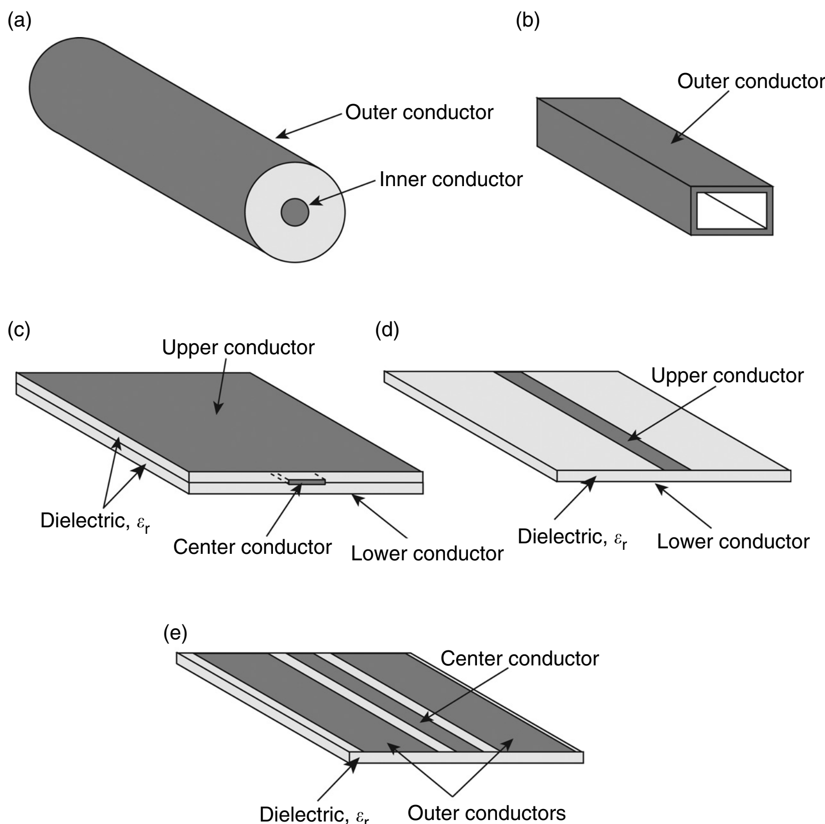


FIGURE I.3 Several common guided-wave structures. (a) coaxial cable, (b) rectangular waveguide, (c) stripline, (d) microstrip, and (e) coplanar waveguide.

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,

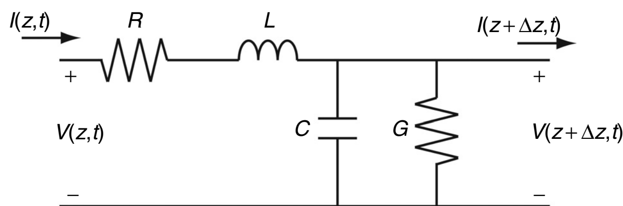


FIGURE I.4 Equivalent circuit for an incremental length of transmission line. A finite length of transmission line can be modeled as a series concatenation of sections of this form.

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

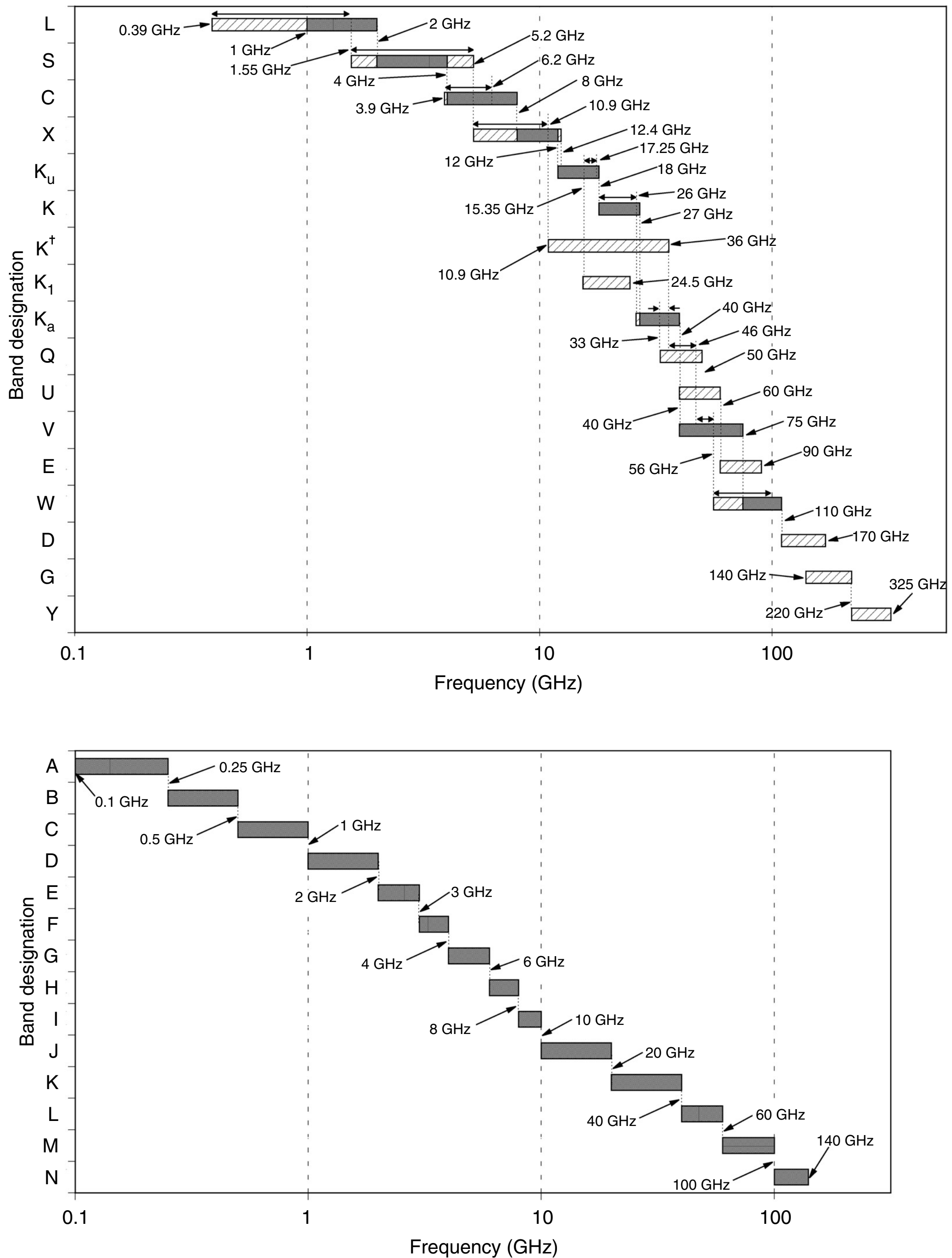


FIGURE I.5 Microwave and RF frequency band designations [1–7]. (a) Industrial and IEEE designations. Diagonal hashing indicates variation in the definitions found in literature; dark regions in the bars indicate the IEEE radar band definitions [8]. Double-ended arrows appearing above bands indicate alternative 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

1.1 Semiconductor Materials for RF and Microwave Applications **1**-1
1.2 Propagation and Attenuation in the Atmosphere **1**-3
1.3 Systems Applications **1**-5
Communications • Navigation • Sensors (Radar) • Heating
1.4 Measurements **1**-7
Small Signal • Large Signal • Noise • Pulsed $I-V$
1.5 Circuits and Circuit Technologies **1**-16
Low Noise Amplifier • Power Amplifier • Mixer • RF Switch • Filter • Oscillator
1.6 CAD, Simulation, and Modeling **1**-19
References **1**-20

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

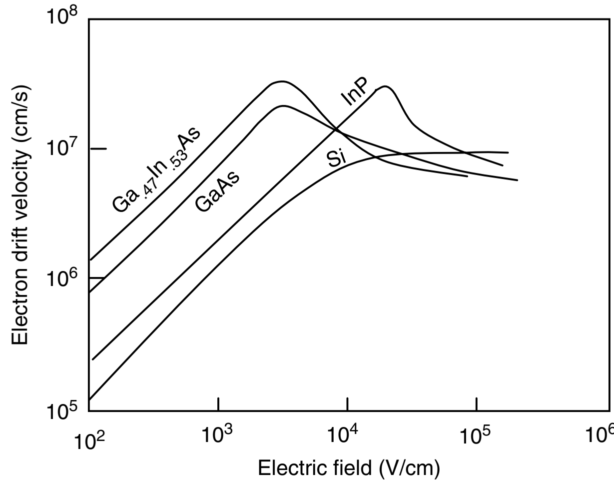


FIGURE 1.1 The electron velocity as a function of applied electric field for several semiconductor materials which are important for RF and microwave applications.

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

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

where v 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, $f_{\mathrm{III-V}}$, 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

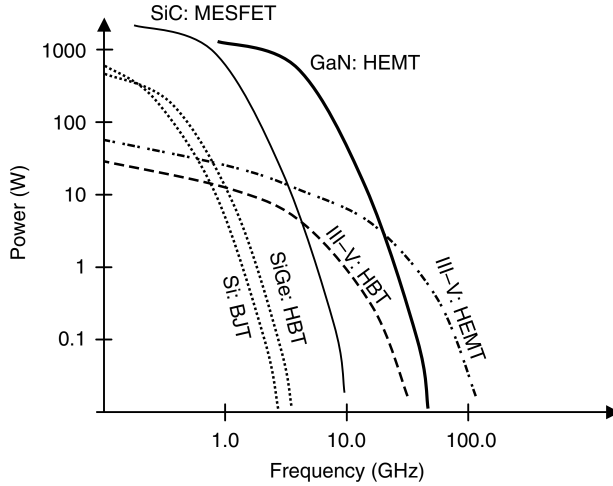


FIGURE 1.2 Semiconductor choices for RF applications are a strong function of the power and frequency required for the wireless application.

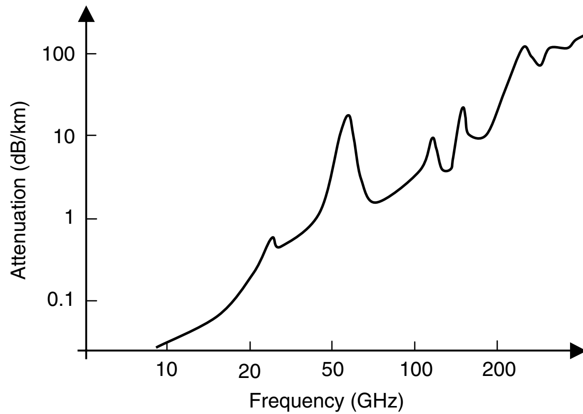


FIGURE 1.3 Attenuation of electromagnetic signals in the atmosphere as a function of frequency.

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

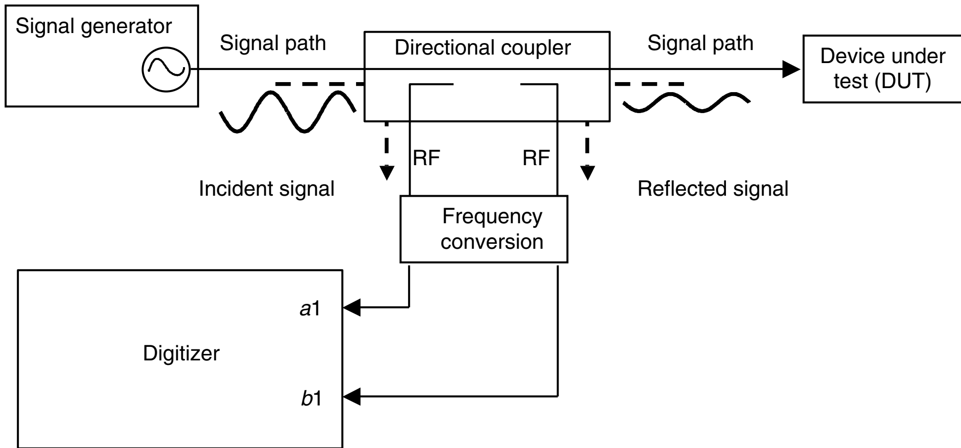


FIGURE 1.4 Vector network analyzer measurement configuration to determine *s*-parameters of a high-frequency device, component, or subsystem.

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.

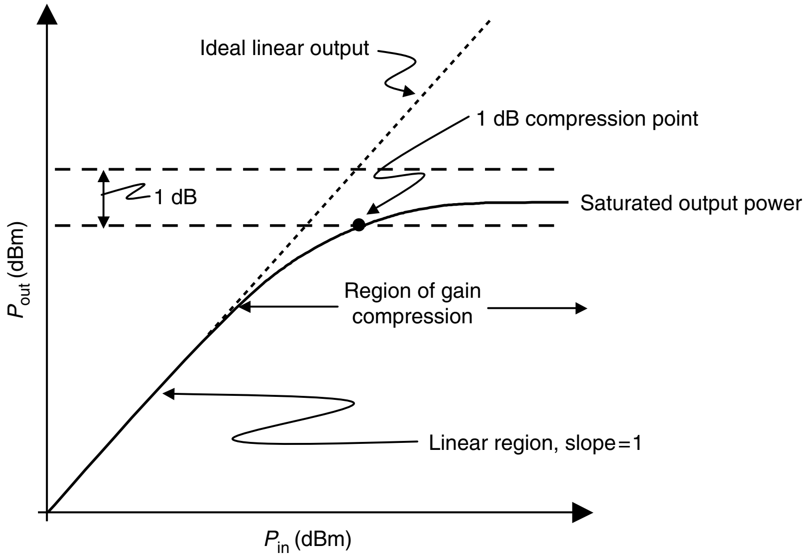


FIGURE 1.5 Output power versus input power at the fundamental frequency for a nonlinear circuit.

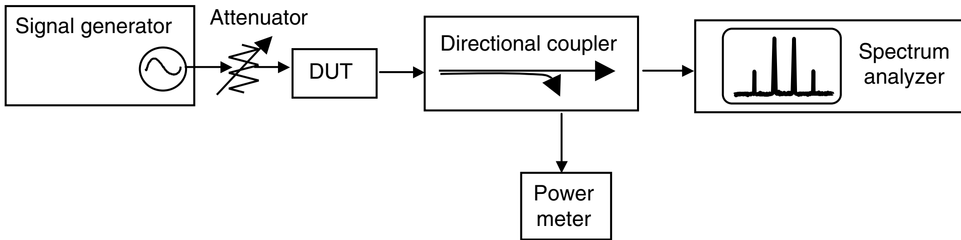


FIGURE 1.6 Measurement configuration to characterize gain compression and harmonic distortion. By replacing the signal generator with two combined signals at slightly offset frequencies, the configuration can also be used to measure intermodulation distortion.

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

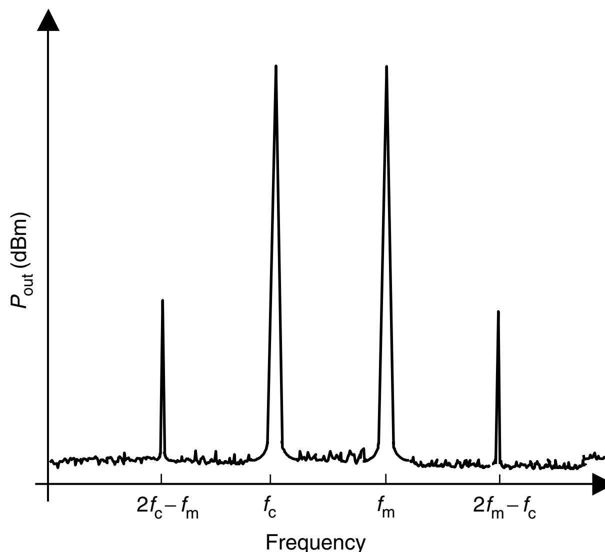


FIGURE 1.7 An illustration of signal spectrum due to intermodulation distortion from two signals at frequencies f_c and f_m.

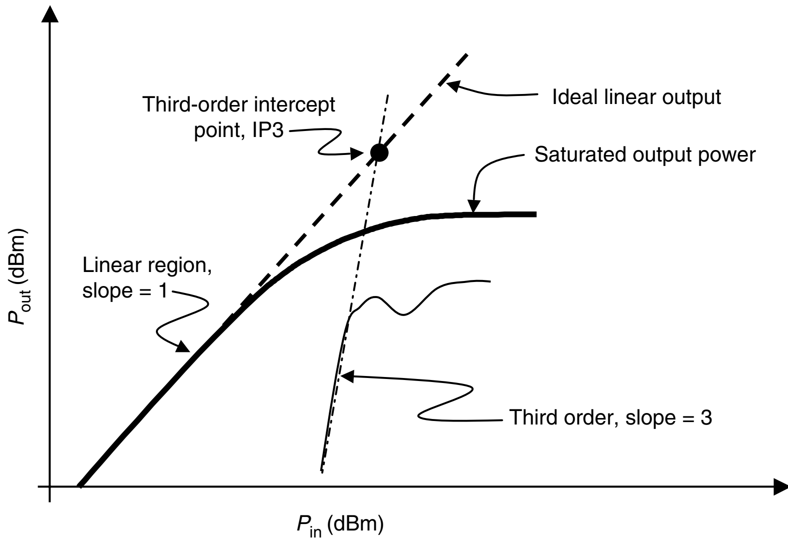


FIGURE 1.8 Relationship between signal output power and intermodulation distortion product levels.

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.

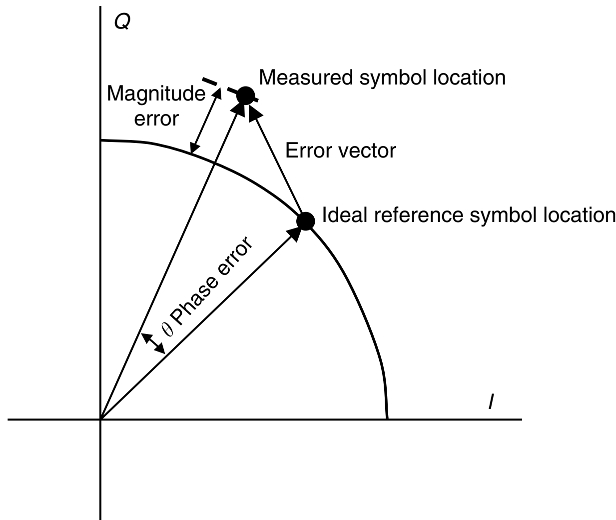


FIGURE 1.9 *I–Q* diagram indicating the error vector for EVM measurements.

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

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

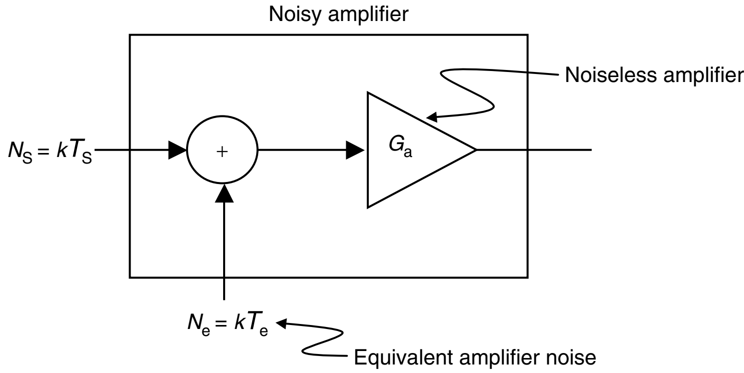


FIGURE 1.10 System view of amplifier noise.

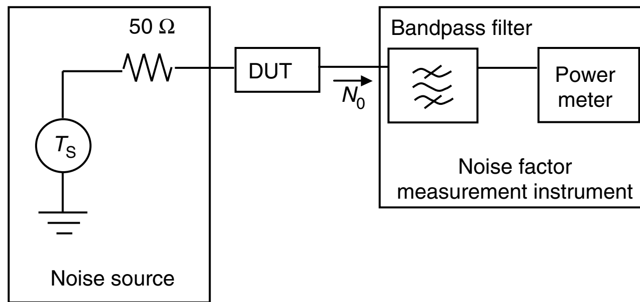


FIGURE 1.11 Measurement configuration for noise factor measurement.

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

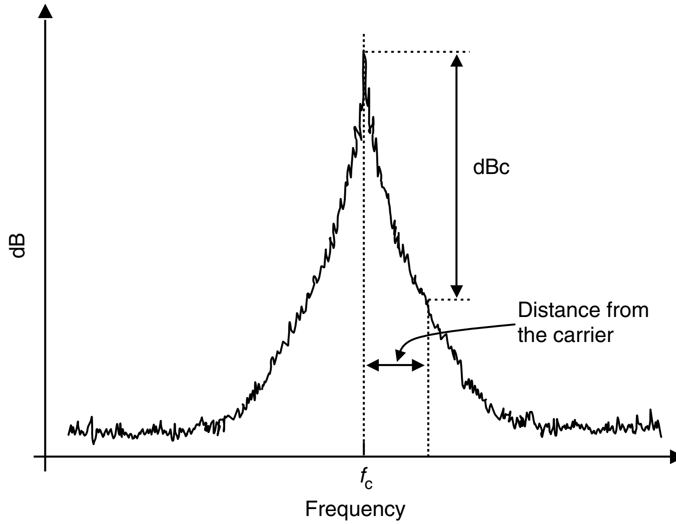


FIGURE 1.12 Typical phase noise spectrum observed on a spectrum analyzer.

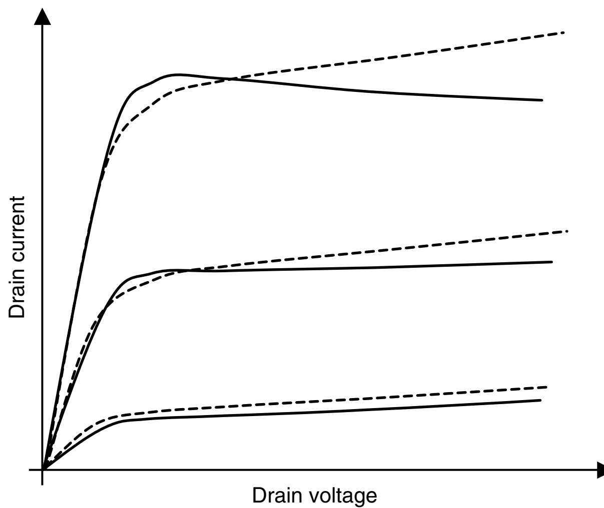


FIGURE 1.13 Pulsed *I–V* characteristics of a microwave FET. Solid lines are DC characteristics while dashed lines are pulsed.

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

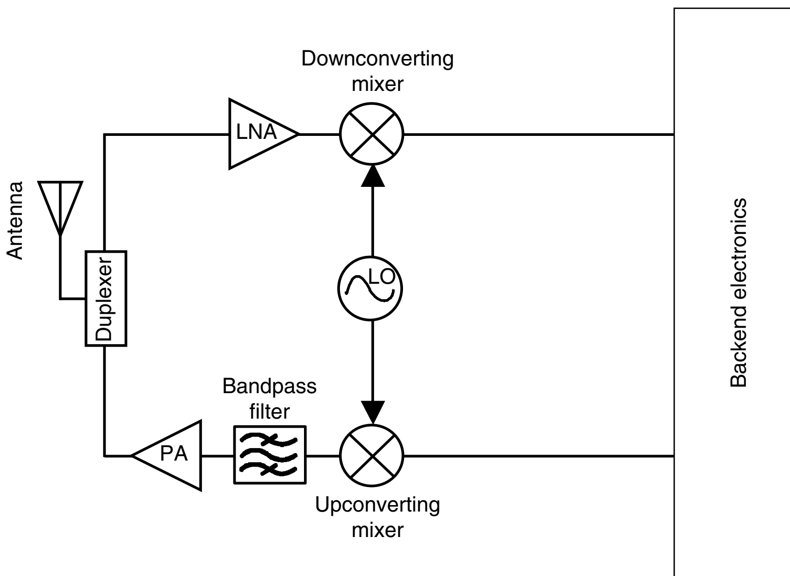


FIGURE 1.14 Generalized microwave radio architecture illustrating the microwave components in both the receiver and transmitter path.

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.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Microwave Measurements

2 Linear Measurements
Ronald E. Ham **2**-1
Signal Measurements • Network Measurements

3 Network Analyzer Calibration
Joseph Staudinger **3**-1
VNA Functionality • Sources of Measurement Uncertainties • Modeling VNA Systematic Errors • Calibration • Calibration Standards

4 Absolute Magnitude and Phase Calibrations
Kate A. Remley, Paul D. Hale, and Dylan F. Williams **4**-1
Absolute Calibration for Signal Measurements • Elements of Signal-Measurement Calibrations • Applications of Waveform Calibrations • Conclusion

5 Noise Measurements
Alfy Riddle **5**-1
Fundamentals of Noise • Detection • Noise Figure and Y-Factor Method • Phase Noise and Jitter • Summary

6 Nonlinear Microwave Measurement and Characterization
J. Stevenson Kenney **6**-1
Mathematical Characterization of Nonlinear Circuits • Harmonic Distortion • Gain Compression and Phase Distortion • Intermodulation Distortion • Multicarrier Intermodulation Distortion and Noise Power Ratio • Distortion of Digitally Modulated Signals • Summary

7 Theory of High-Power Load-Pull Characterization for RF and Microwave Transistors
John F. Sevic **7**-1
System Architecture for High-Power Load-Pull • Characterization of System Components • System Performance Verification • Summary

8 Pulsed Measurements
Anthony E. Parker, James G. Rathmell, and Jonathan B. Scott **8**-1
Introduction • Isothermal and Isodynamic Characteristics • Relevant Properties of Devices • Pulsed Measurement Equipment • Measurement Techniques • Data Processing • Summary • Defining Terms

9 Microwave On-Wafer Test
Jean-Pierre Lanteri, Christopher Jones, and John R. Mahon **9-**1
On-Wafer Test Capabilities and Applications • Test Accuracy Considerations • On-Wafer Test Interface • On-Wafer RF Test Benefits

10 High Volume Microwave Test
Jean-Pierre Lanteri, Christopher Jones, and John R. Mahon **10-**1
High Volume Microwave Component Needs • Test System Overview • High Volume Test Challenges • Data Analysis Overview • Conclusion

11 Large Signal Network Analysis/Waveform Measurements
Joseph M. Gering **11-**1
Introduction • Mathematical Background • Measurement Systems • Calibration • Measurement Extensions • Summary

2

Linear Measurements

Ronald E. Ham
MW and RF Consulting

2.1 Signal Measurements **2**-1
Time Domain • Frequency Domain • Modulation Domain
2.2 Network Measurements **2**-7
Power • Impedance • Network Analyzers
References **2**-22

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.

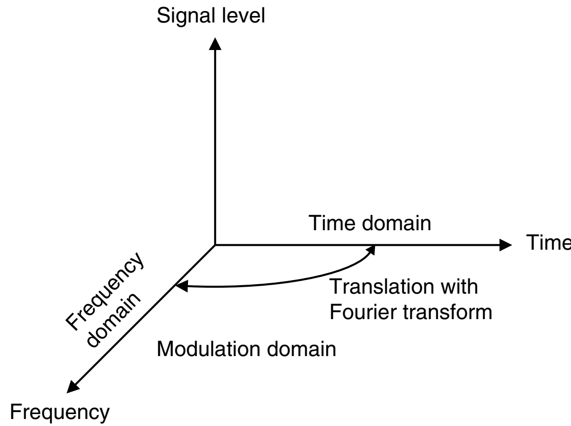


FIGURE 2.1 Three different types of measurements characterize signals.

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.

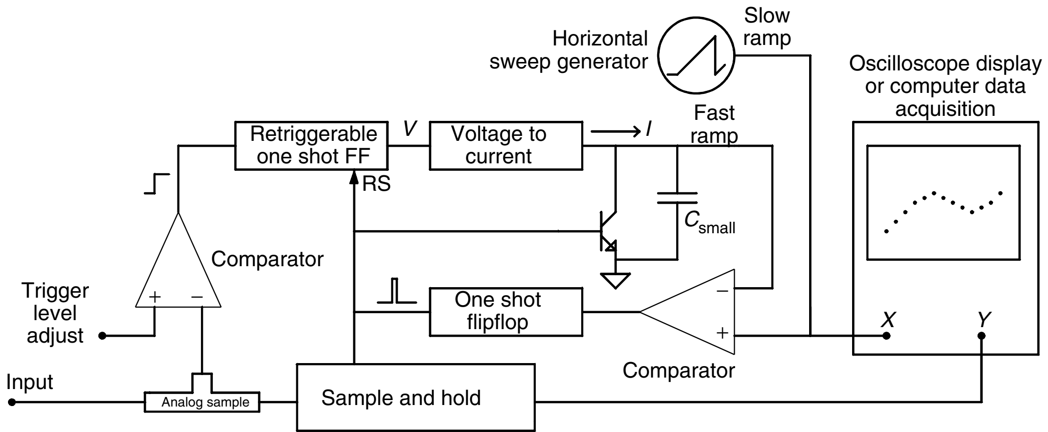


FIGURE 2.2 Simplified sampling oscilloscope.

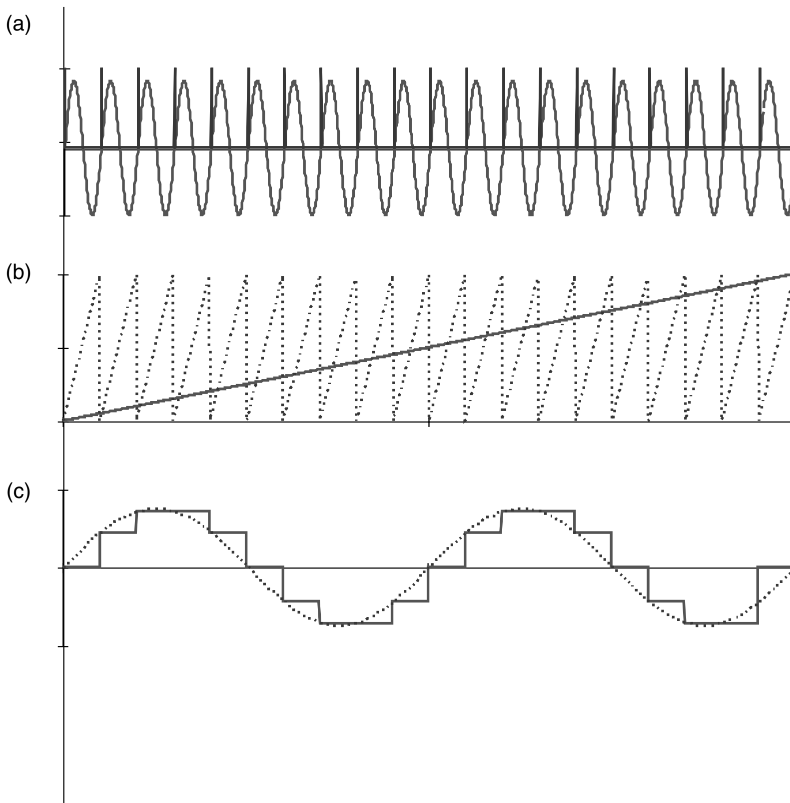
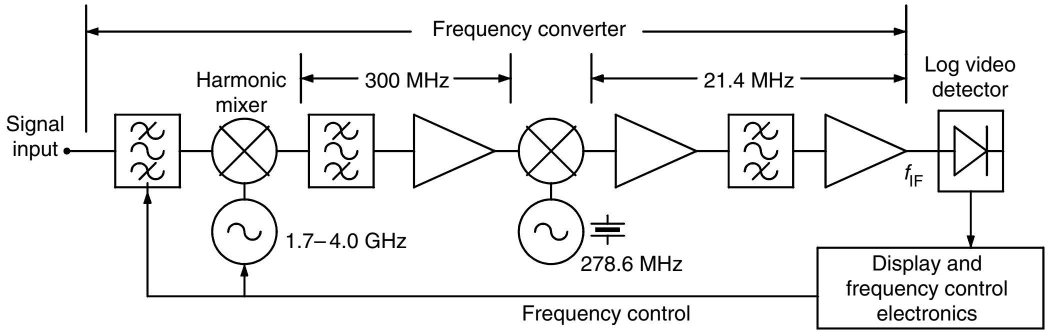


FIGURE 2.3 Primary waveforms in a sampling oscilloscope. (a) Input and zero crossing triggers, (b) fast display ramp and the slow sampling ramp, and (c) output waveform versus the input waveform.

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.

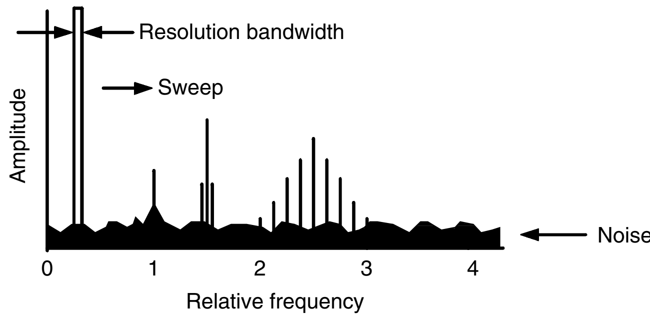


FIGURE 2.5 Swept frequency spectrum analyzer display.

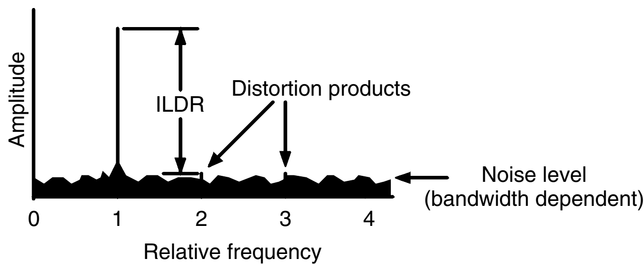


FIGURE 2.6 Instantaneous linear dynamic range.

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.

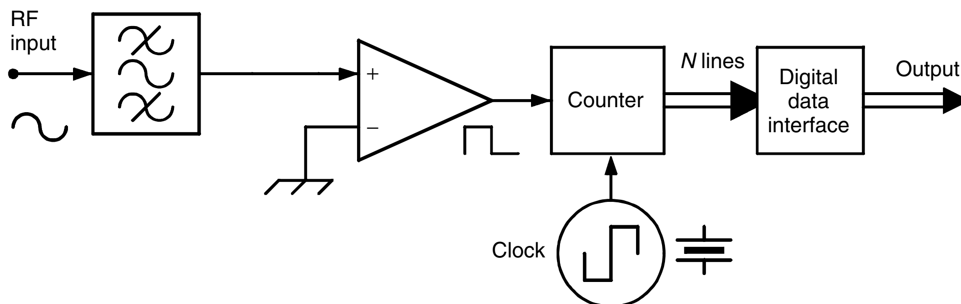


FIGURE 2.7 Frequency counter.

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

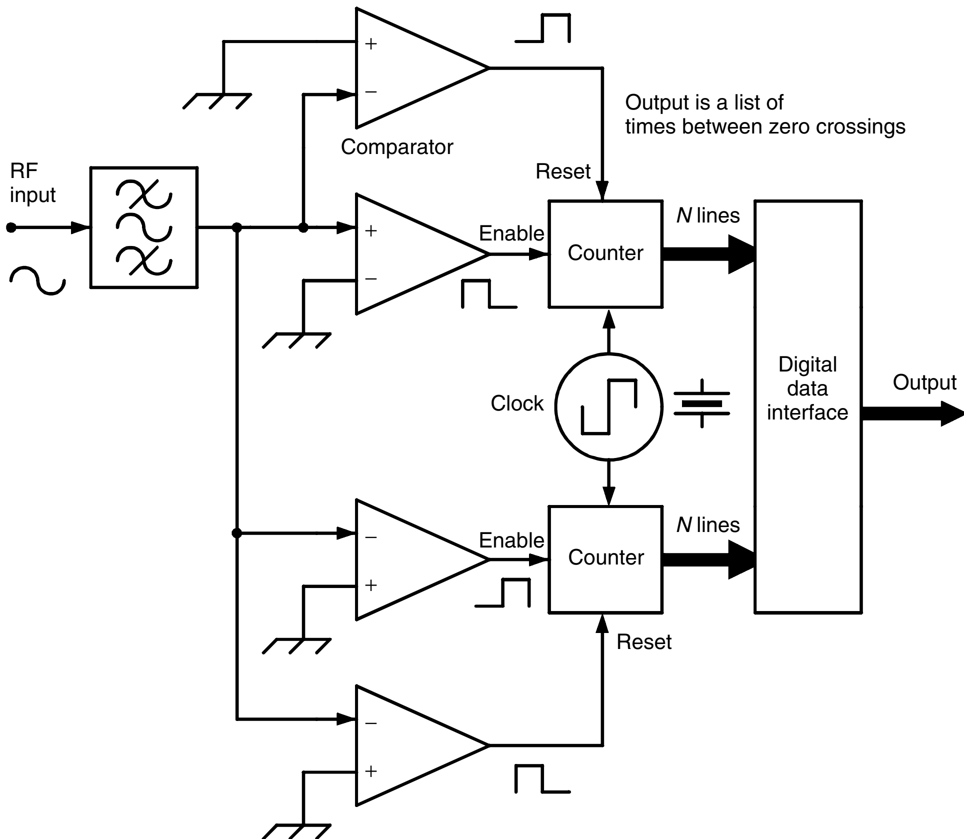


FIGURE 2.8 Modulation domain analyzer period measurements yield frequency as a function of time.

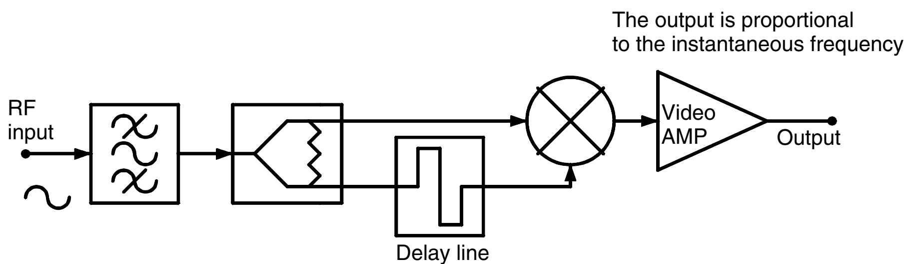


FIGURE 2.9 Instantaneous frequency and phase discriminator.

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

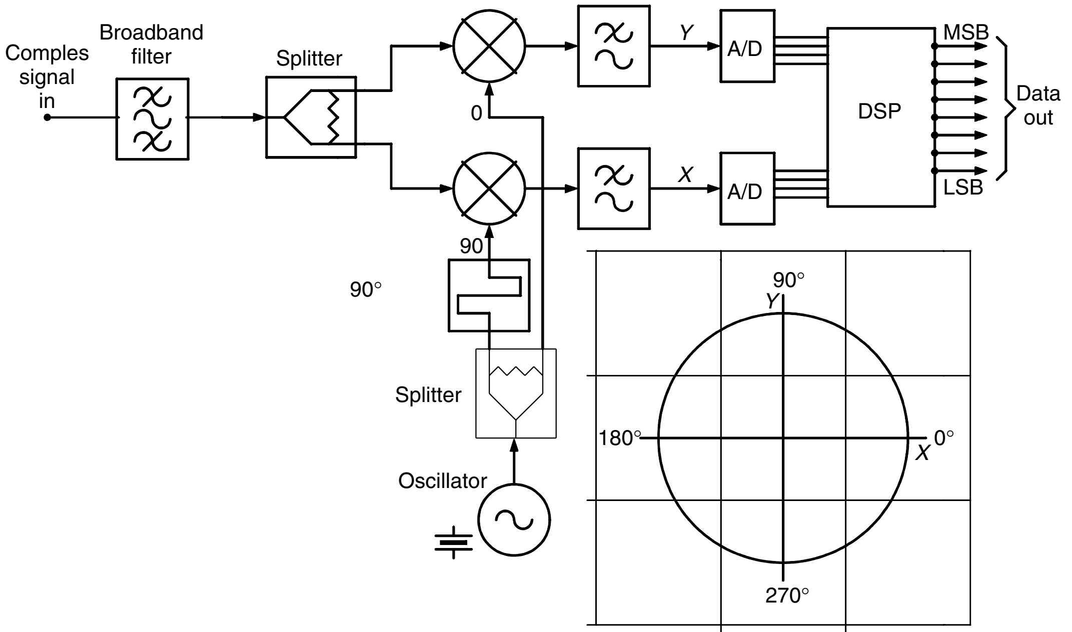


FIGURE 2.10 Modulation analyzer.

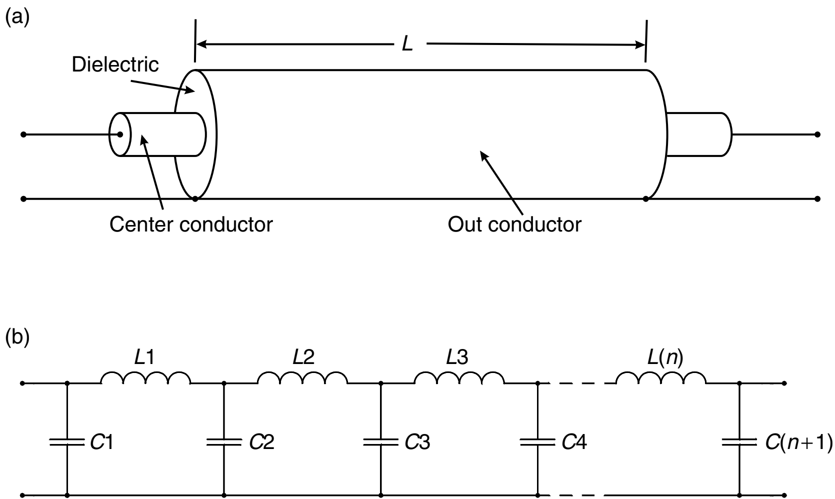


FIGURE 2.11 Examples of transmission lines. (a) Coaxial and (b) lumped element.

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

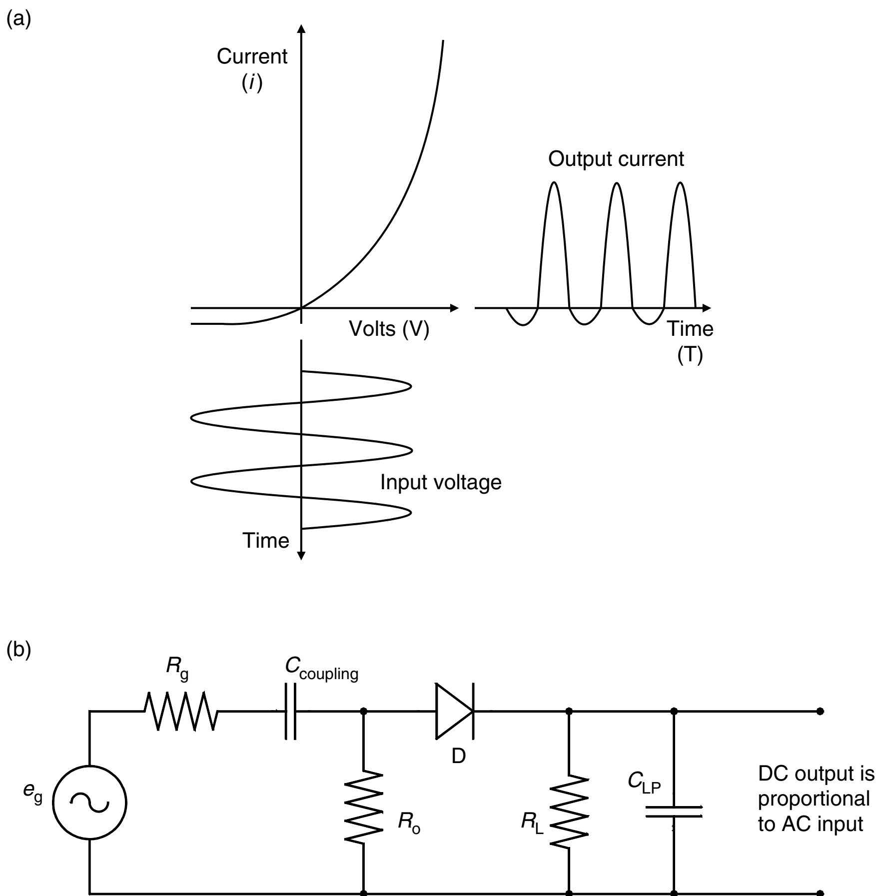


FIGURE 2.12 Diode detector. (a) Diode detector waveforms and (b) diode detector circuit.

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.

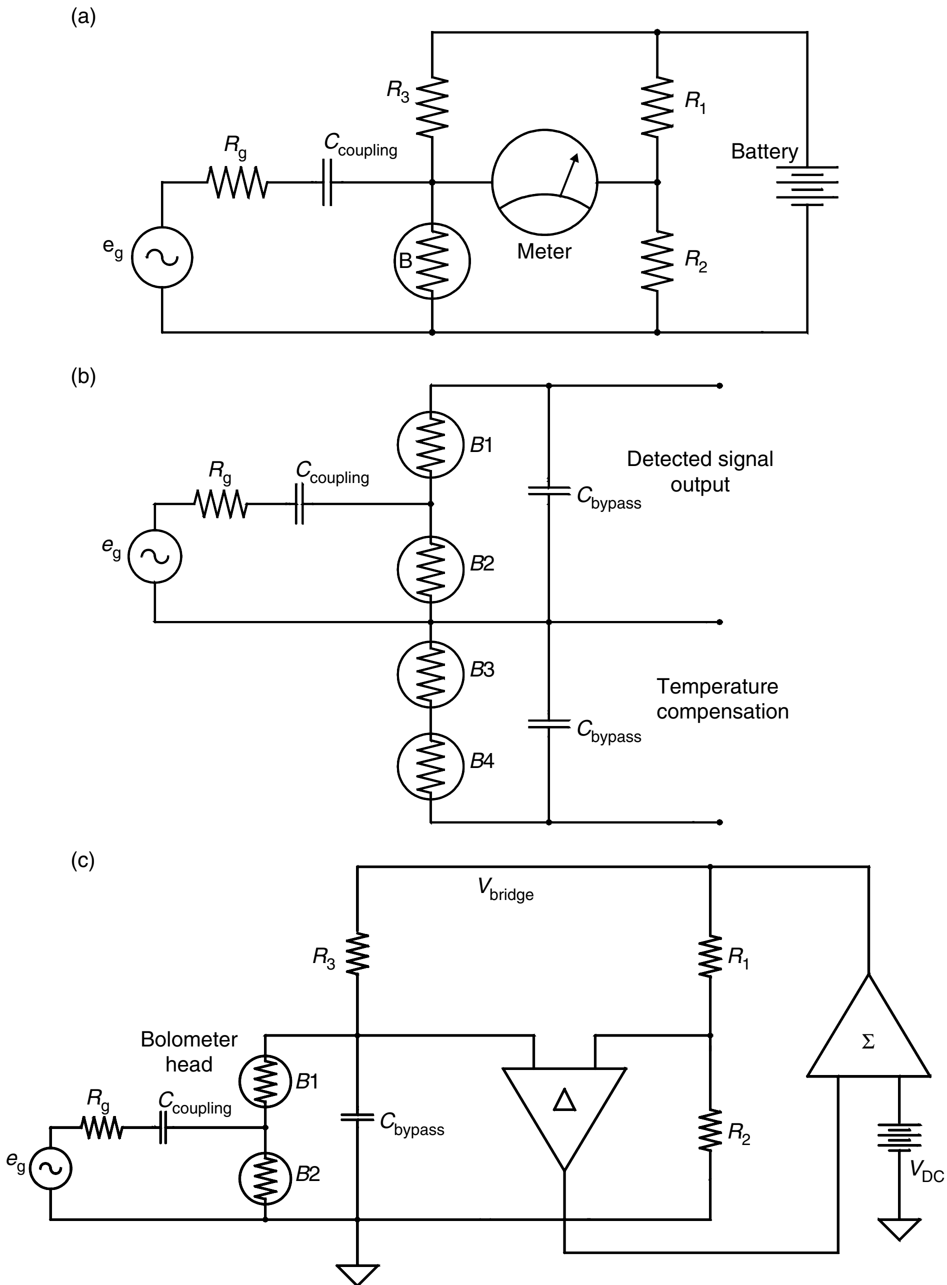


FIGURE 2.13 Thermally dependent detector circuits. (a) Bolometer in a bridge circuit, (b) temperature-compensated bolometer head, and (c) self-balancing bridge circuit.

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

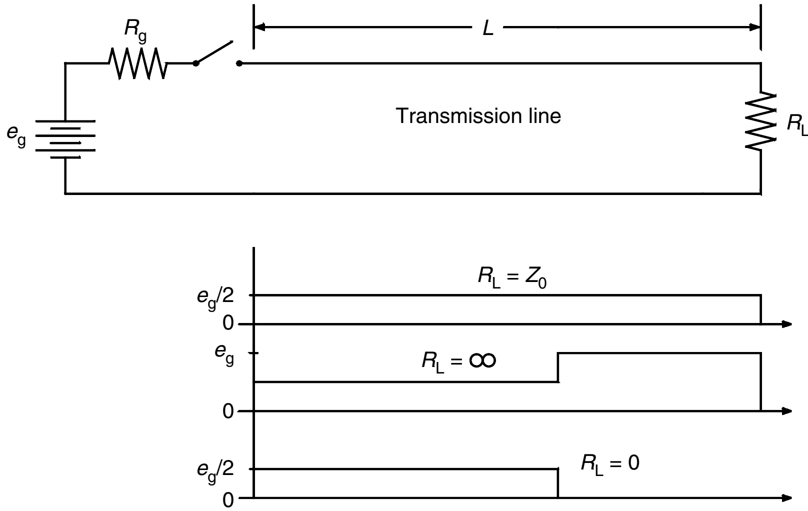


FIGURE 2.14 Switched DC line voltage at time > length/velocity for various impedances at the end of the line.

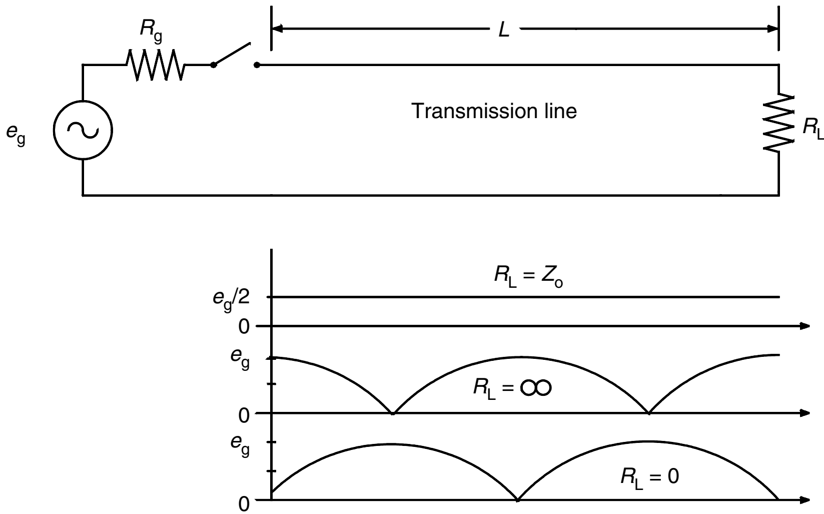


FIGURE 2.15 Waveforms on the line for a sinusoidal source and various impedances at the end of the line.

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.

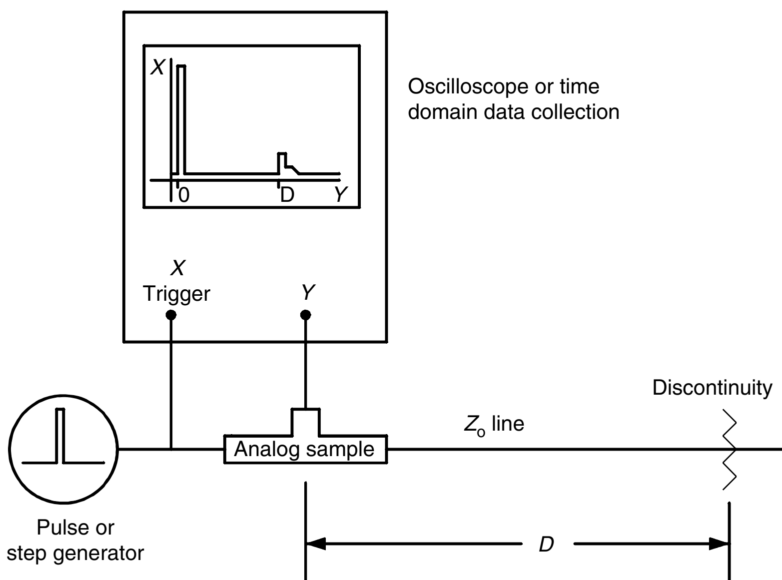


FIGURE 2.16 Time-domain reflectometer.

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

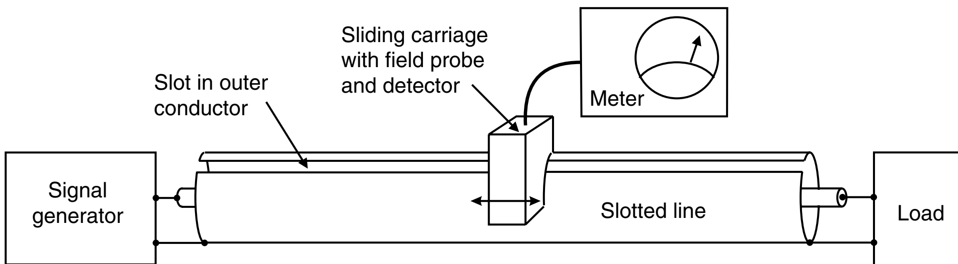


FIGURE 2.17 A slotted line is used to measure the impedance of an unknown load.

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

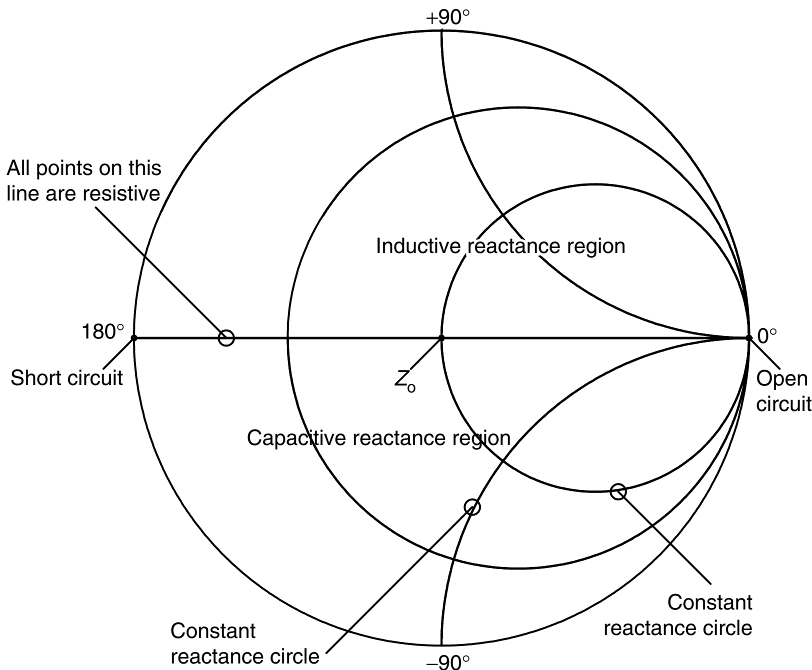


FIGURE 2.18 The Smith Chart is a plot of all nonnegative real impedances.

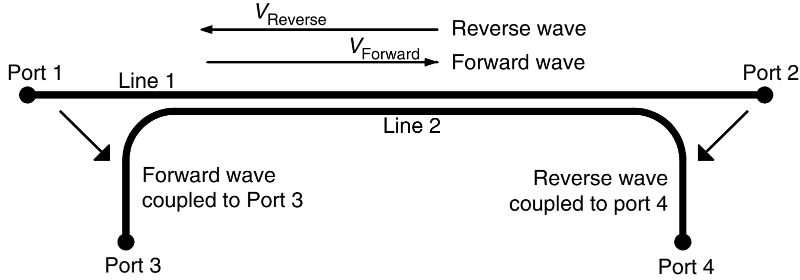


FIGURE 2.19 A directional coupler separates forward and reverse waves on a transmission line.

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.

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

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.

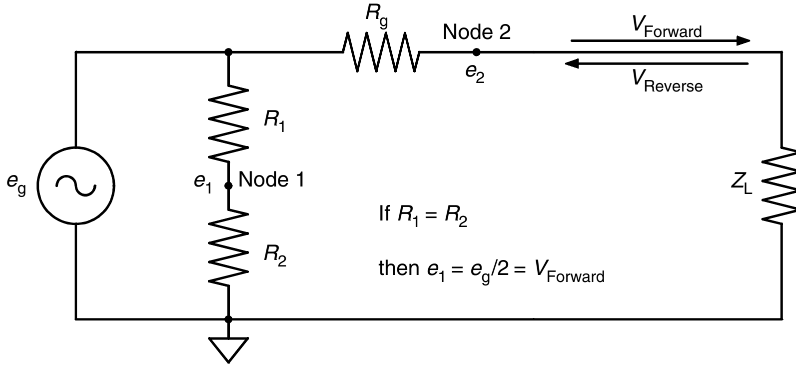


FIGURE 2.20 A resistive bridge can be used to measure the reverse wave on a transmission line.

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

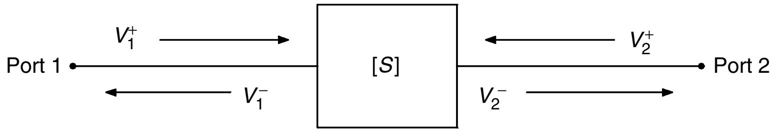


FIGURE 2.21 *s*-Parameters are defined by forward and reverse voltage waves.

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.

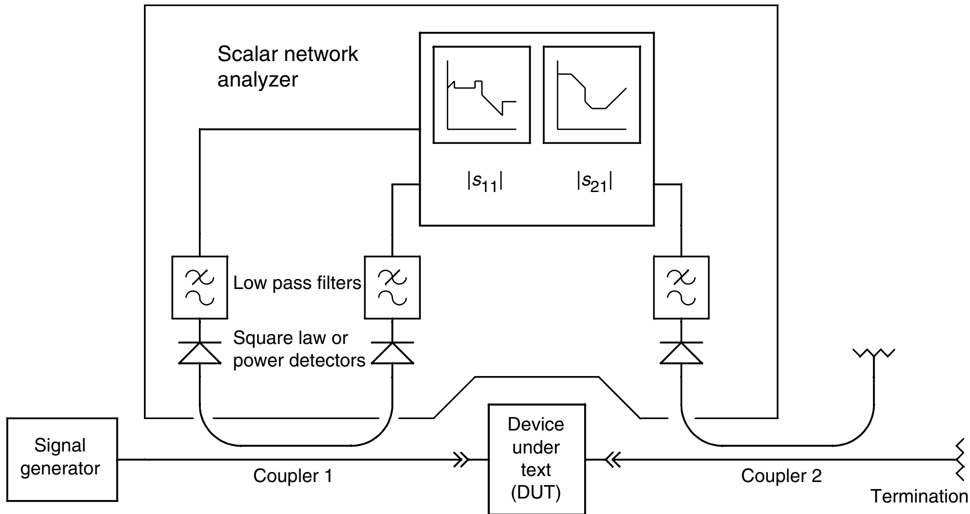


FIGURE 2.22 A scalar network analyzer can measure the magnitude of gain and return loss.

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.

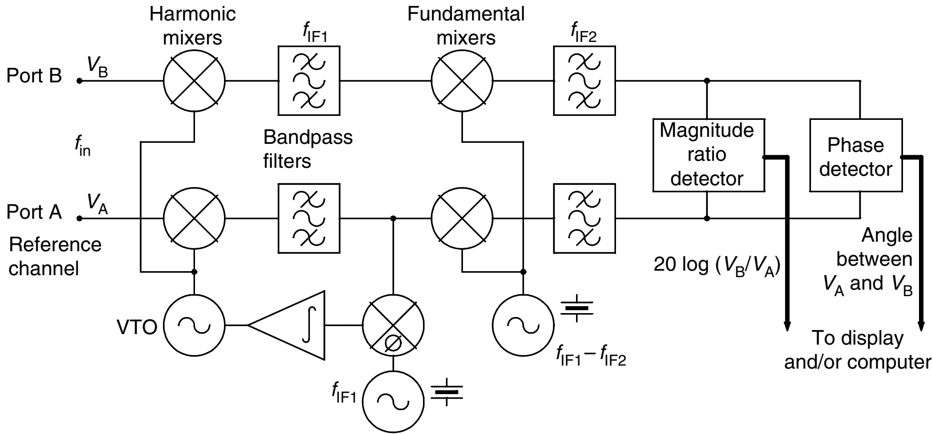


FIGURE 2.23 A vector network analyzer measures complex ratios.

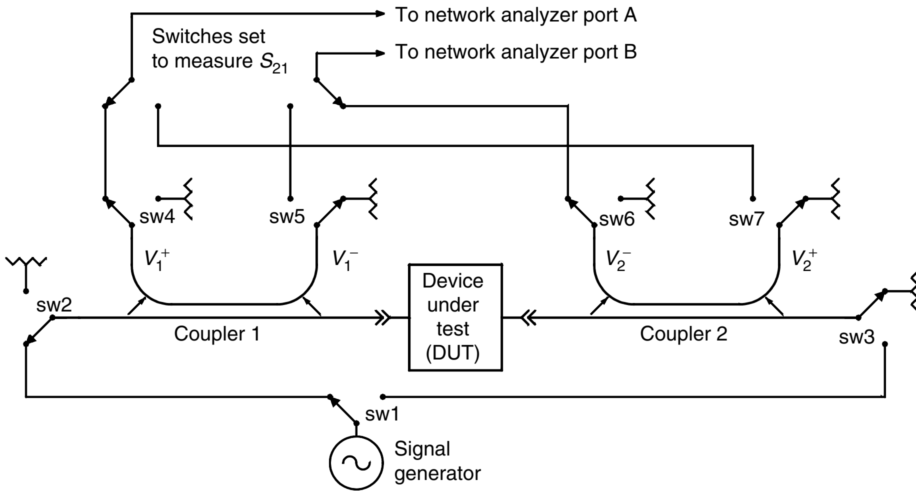


FIGURE 2.24 A two-port *s*-parameter test set can measure all four *s*-parameters without moving the DUT.

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.

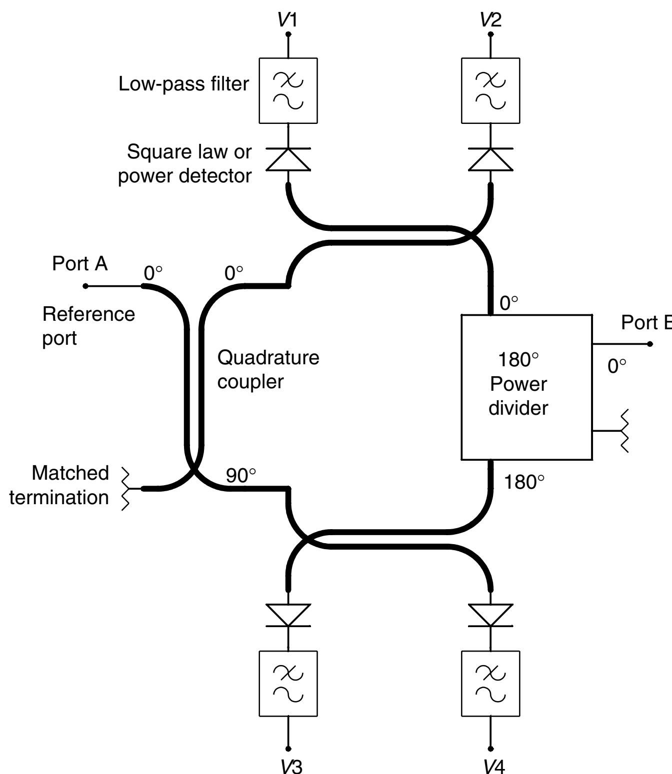


FIGURE 2.25 A six-port network can be used as a narrow-band vector analyzer.

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

3.1 VNA Functionality 3-1
3.2 Sources of Measurement Uncertainties 3-3
3.3 Modeling VNA Systematic Errors 3-3
3.4 Calibration 3-4
3.5 Calibration Standards 3-5
References 3-8

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

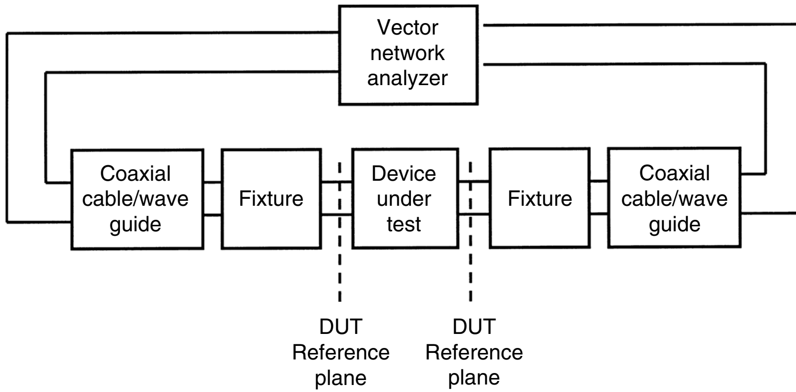


FIGURE 3.1 Typical measurement setup consisting of a device under test embedded in a fixture connected to the vector network analyzer with appropriate cables.

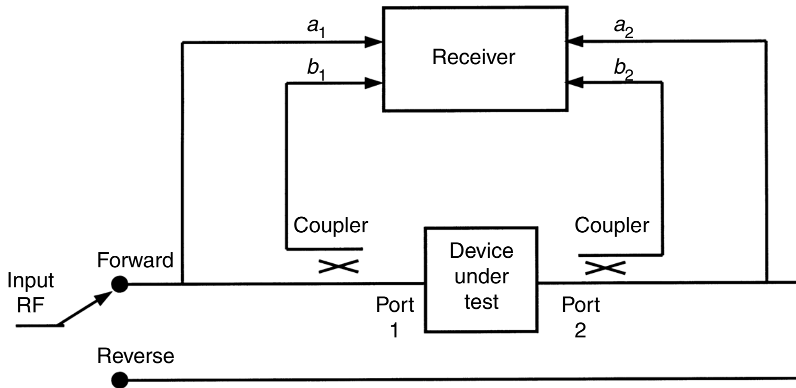


FIGURE 3.2 High simplified VNA block diagram.

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

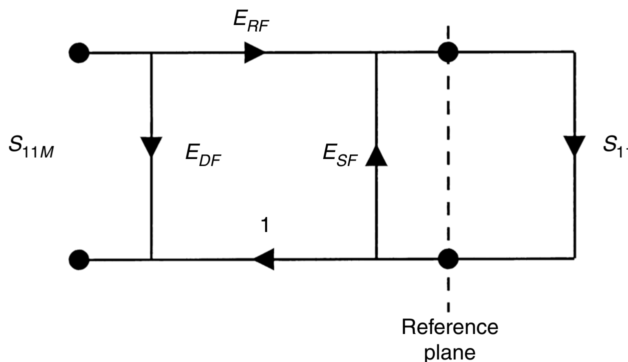


FIGURE 3.3 Typical one-port VNA error model for reflection coefficient measurements.

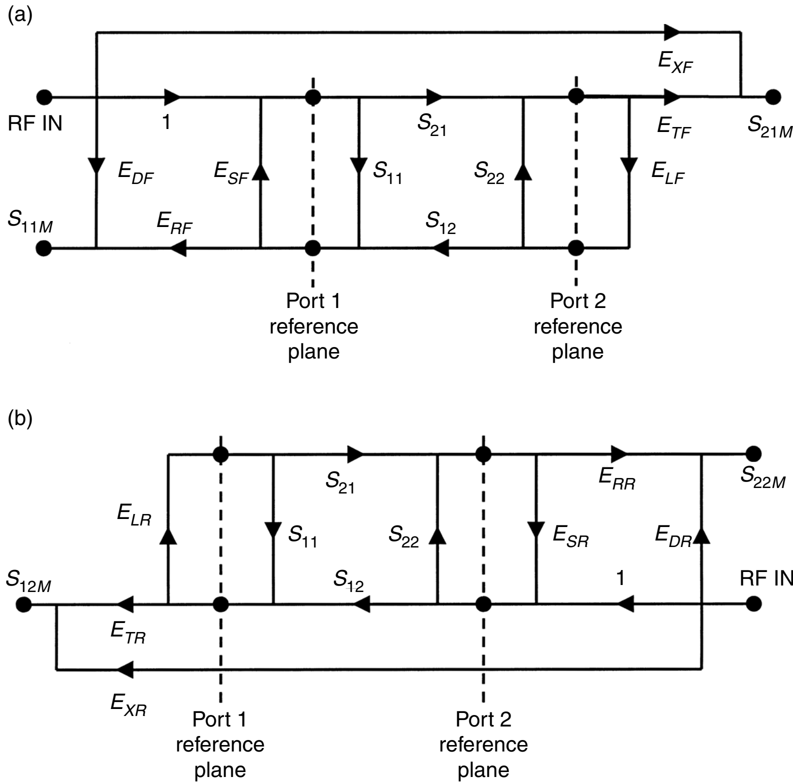


FIGURE 3.4 Typical two-port VNA error model: (a) forward model and (b) reverse model.

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:

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

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_{11M}, S_{21M}, S_{12M}, S_{22M} 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

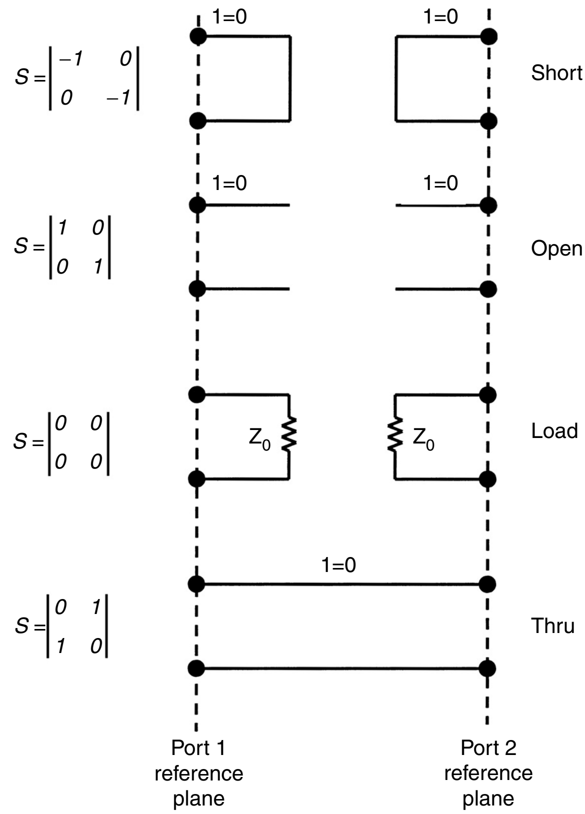


FIGURE 3.5 Electrical definition for lossless and ideal SOLT standards.

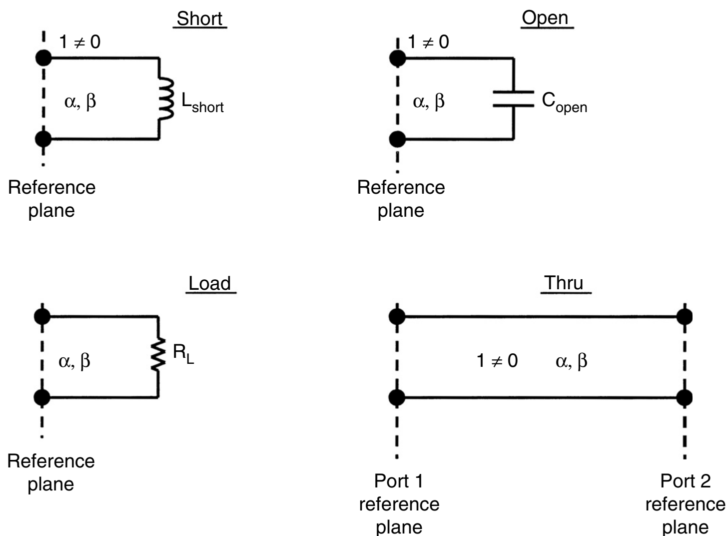


FIGURE 3.6 High frequency descriptions of SOLT standards generally consider nonzero length transmission lines, loss mechanisms, and fringing field effects associated with the open standard.

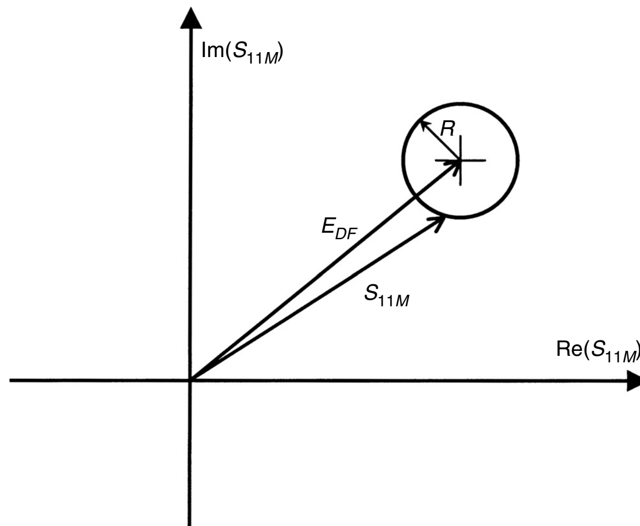


FIGURE 3.7 Characterizing directivity error terms using a sliding load termination.

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° or multiples of 180° (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*

4.1 Absolute Calibration for Signal Measurements **4**-1
4.2 Elements of Signal-Measurement Calibrations **4**-3
Instrument Impulse Response • Impedance Mismatch and Mismatch Correction • Other Considerations for Signal Measurements: Timebase Errors, Spectral Leakage, and Signal Level
4.3 Applications of Waveform Calibrations **4**-9
RF Harmonic Phase Calibration Example: LSNA Measurement of a Square Wave • Oscilloscope Impulse Response Calibration Example: Mixer Reciprocity Measurement
4.4 Conclusion **4**-15
Acknowledgments.. **4**-15
References **4**-15

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.

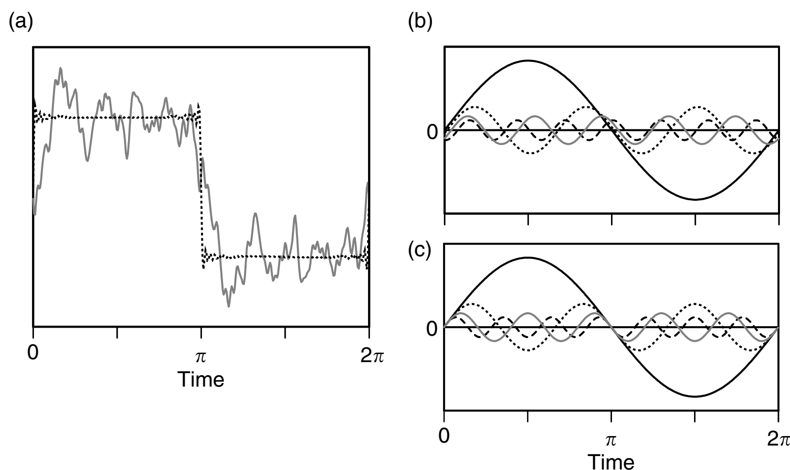


FIGURE 4.1 To illustrate the concept of absolute calibration, consider the measurement of a square wave and its Fourier series representation. The distorted (uncalibrated) waveform in (a) can be constructed from its Fourier series, the first few terms of which are shown in (b). When the measurement instrument 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

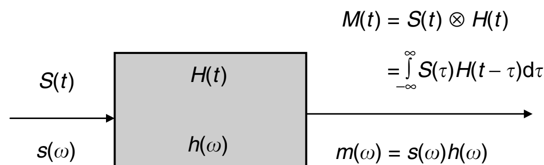


FIGURE 4.2 Absolute calibration for a linear, time-invariant measurement system. When determining the instrument response, a known calibration signal $s_c(\omega)$ is used as the input. The measured output of the instrument $m_c(\omega)$ is used to find the instrument response $h(\omega) = m_c(\omega)/s_c(\omega)$. 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 (∼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 $b = b/\sqrt{2}$.

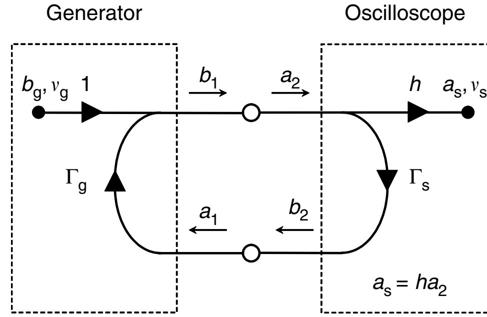


FIGURE 4.3 Microwave flow diagram describing the propagation of signals between a signal generator and a measurement instrument, for example, an oscilloscope. After [6].

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},

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

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

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

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

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

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

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

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.

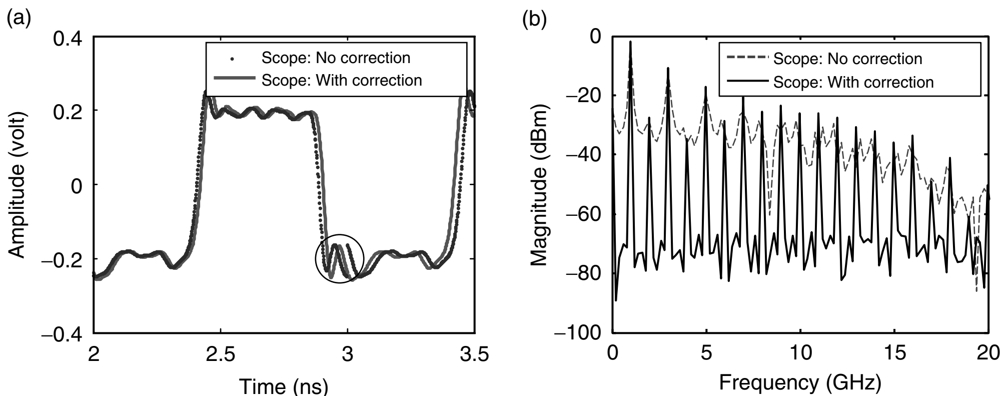


FIGURE 4.4 Oscilloscope measurement of a 1 GHz square-wave signal: (a) time domain and (b) frequency domain. The dotted lines are the raw data and the solid lines represent the calibrated oscilloscope measurement. The discontinuity in the uncorrected waveform circled in (a) is due to timebase distortion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

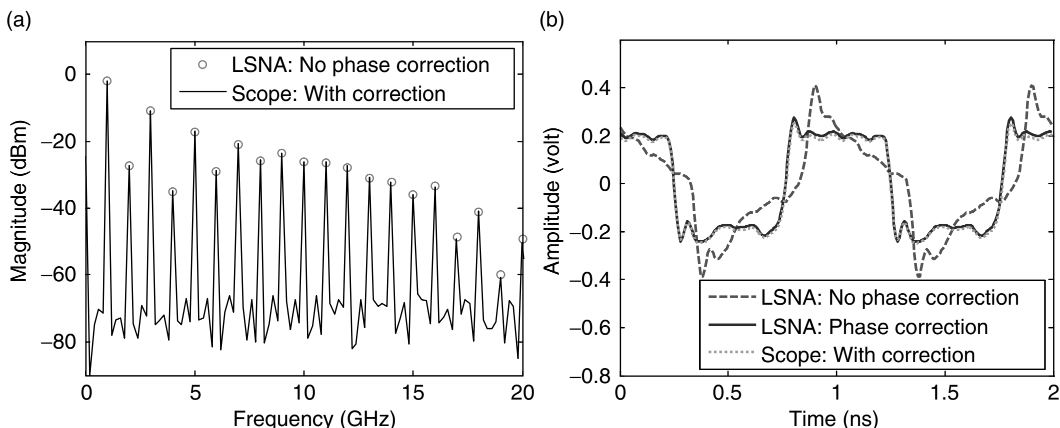


FIGURE 4.5 Magnitude-calibrated LSNA measurements compared to fully calibrated oscilloscope measurements. In (a), we see that the magnitudes compare well. In (b), LSNA measurements with and without phase calibration are compared to the oscilloscope measurement. The dashed line shows that the lack of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

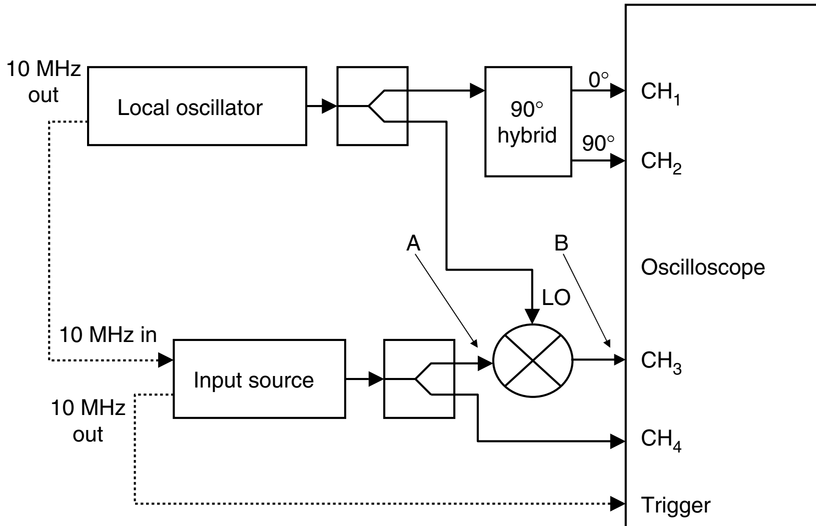


FIGURE 4.6 Set-up to measure the reciprocity of a microwave mixer by use of a calibrated oscilloscope. The phases of the signals at channels 3 and 4 are determined with respect to a fixed local oscillator phase, as measured on channels 1 and 2. From [57].

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.

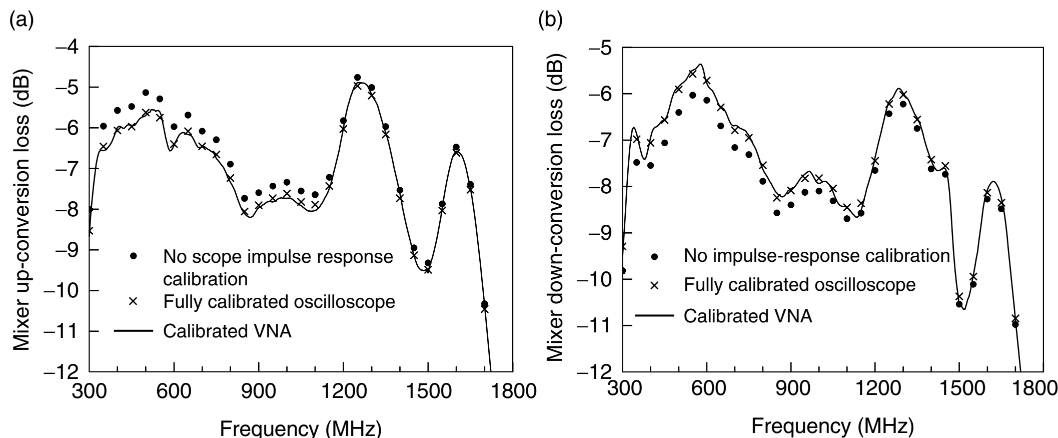


FIGURE 4.7 Measured magnitude response of a microwave mixer. (a) The up- and (b) the down-conversion processes. The solid lines represent power-meter calibrated VNA measurements. The symbols show measurements made with a sampling oscilloscope that was uncalibrated (dots) and calibrated (x's). Agreement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

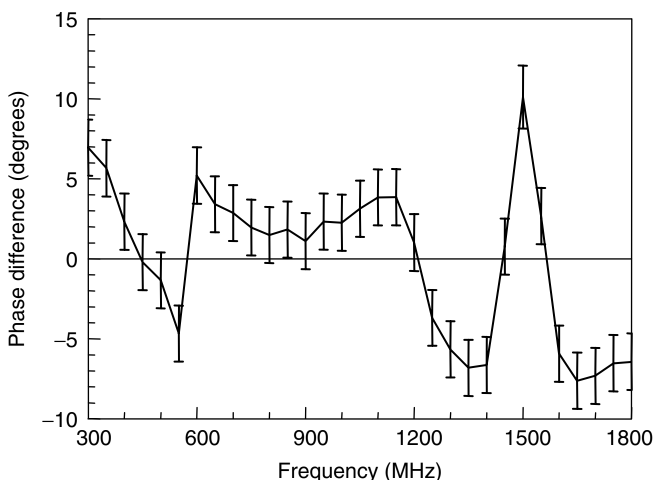


FIGURE 4.8 Phase difference of the mixer up- and down-conversion transfer functions measured with a digital sampling oscilloscope. The error bars represent the uncertainty in the measurement. The frequency axis is referred to the IF port. From [57].

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

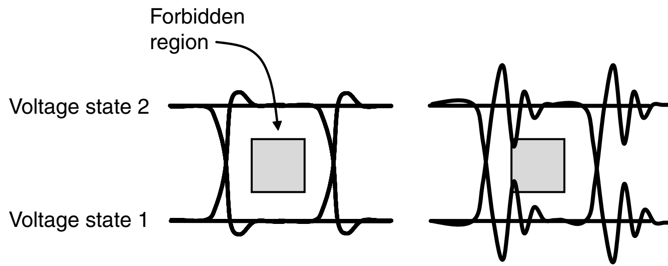


FIGURE 4.9 Schematic of a generic eye pattern measurement for a digital signal at baseband. Here, time is truncated to approximately one bit. The "Forbidden Region" is specified in a number of standards to determine pass/fail criteria for a particular component or to ensure error-free detection of a 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

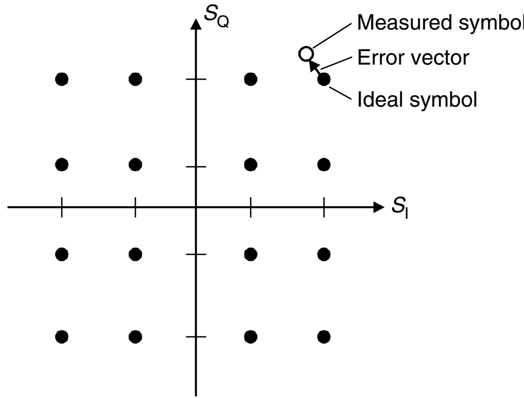


FIGURE 4.10 Constellation diagram for a 16-QAM signal. From [48].

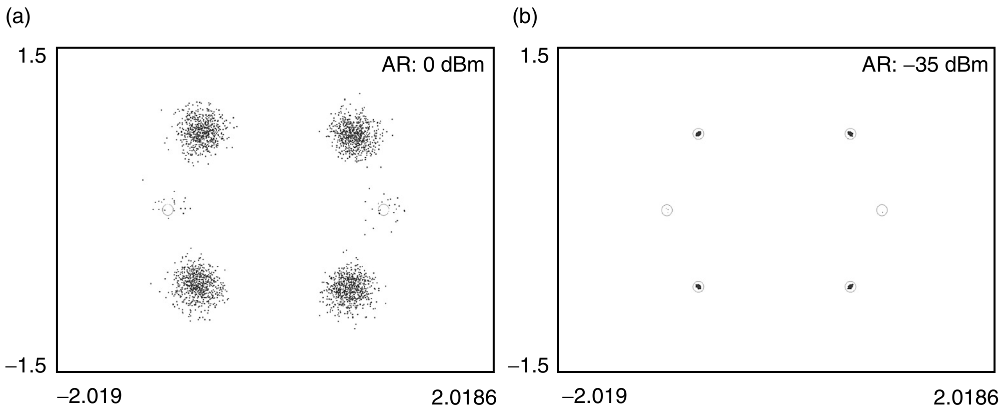


FIGURE 4.11 Illustration of uncorrected measurement-instrument-induced distortion on the constellation of a QPSK signal. Four QPSK information-payload symbol states are in quadrature with the graph's axes and two pilot symbol states are shown parallel to the horizontal axis. (a) The uncorrected measurement 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.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

5

Noise Measurements

5.1 Fundamentals of Noise 5-1
Statistics • Bandwidth
5.2 Detection 5-3
5.3 Noise Figure and *Y*-Factor Method 5-3
5.4 Phase Noise and Jitter 5-5
Introduction • Mathematical Basics • Phase Noise Measurements
5.5 Summary.. 5-10
References.. 5-10

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

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

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.

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

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

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.

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

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.

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

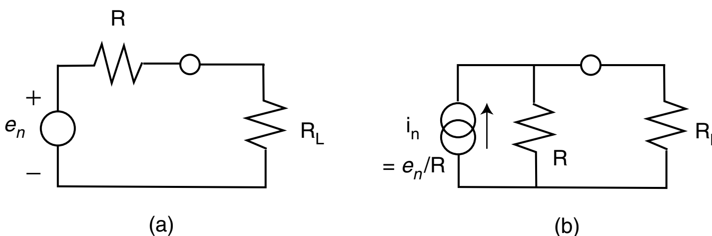


FIGURE 5.1 Equivalent thermal noise sources: (a) voltage and (b) current.

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

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

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

$$N_O\big(T_S\big) = \big(k\ T_S\ G_a + k\ T_e\ G_a\big)B \qquad (5.9)$$

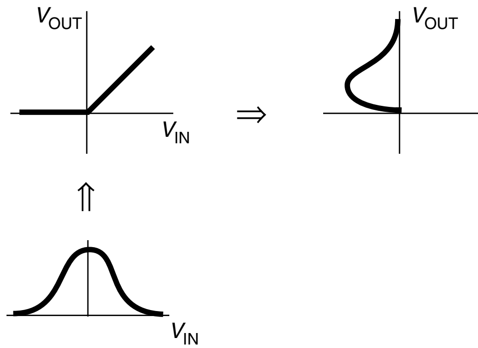


FIGURE 5.2 Nonlinear transformation of Gaussian noise.

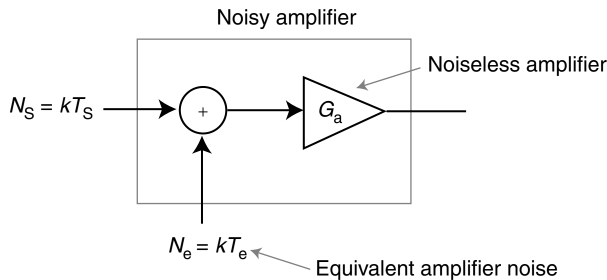


FIGURE 5.3 System view of amplifier noise.

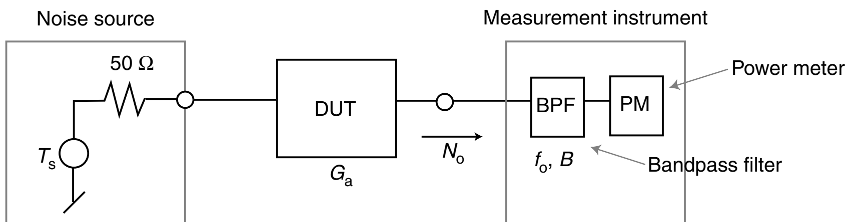


FIGURE 5.4 Test setup for Y-factor method.

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.

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

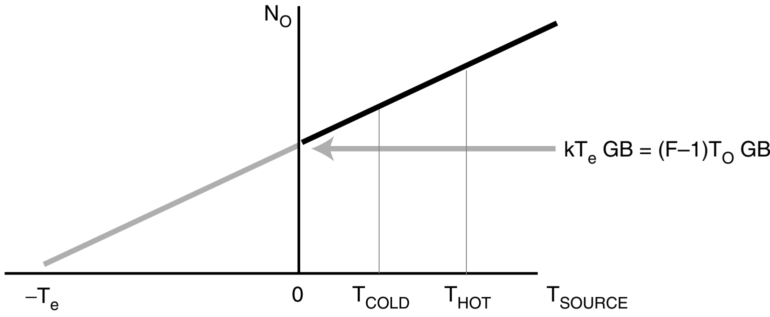


FIGURE 5.5 Output noise power versus source temperature.

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

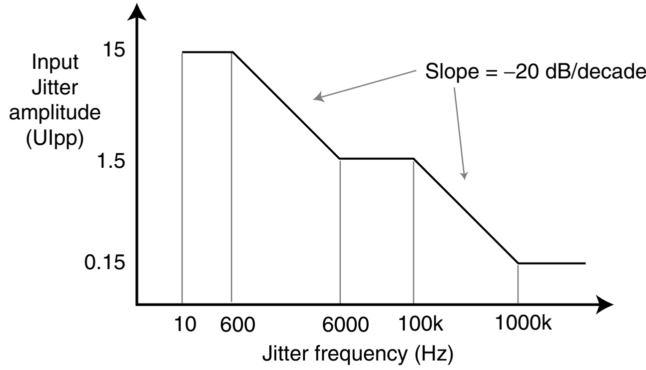


FIGURE 5.6 SONET category II jitter tolerance mask for OC-48.

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.

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

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:

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

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 radians^2/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.

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

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

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

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

$$\text{jitter} = \sqrt{\left({}_0\!\int^\infty S_\phi(f_m)\,df_m\right)} = \beta_1/\sqrt{2}\Big|_{\beta(t) = \beta_1\cos(\omega_m t)} \tag{5.20}$$

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].

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

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).

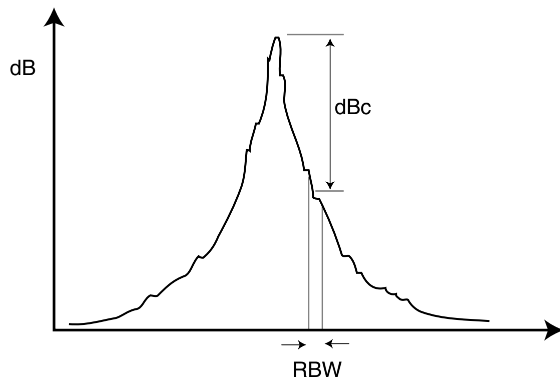


FIGURE 5.7 Typical measured spectrum on a spectrum analyzer.

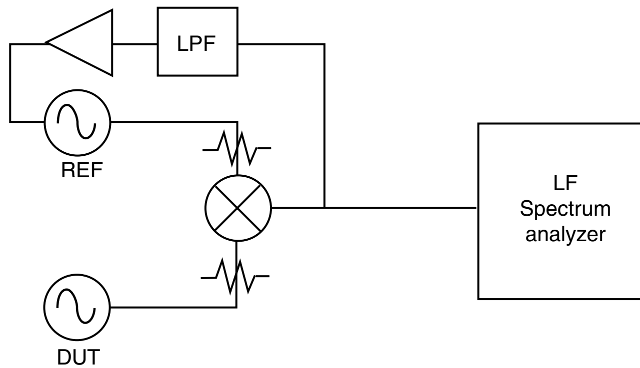


FIGURE 5.8 PLL phase noise measurement.

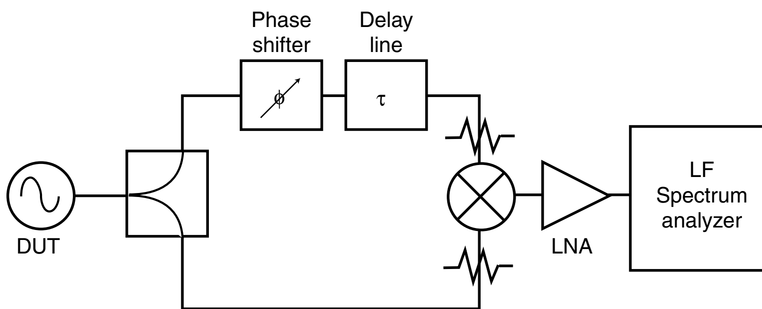


FIGURE 5.9 Transmission line discriminator.

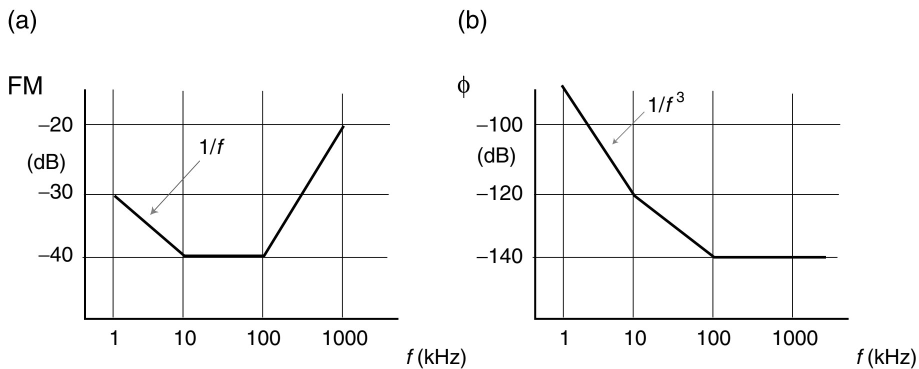


FIGURE 5.10 (a) FM and (b) phase noise spectral densities for the same device.

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

6.1 Mathematical Characterization of Nonlinear Circuits 6-2
Nonlinear Memoryless Circuits • Nonlinear Circuits with Memory
6.2 Harmonic Distortion 6-4
Harmonic Generation in Nonlinear Circuits • Measurement of Harmonic Distortion
6.3 Gain Compression and Phase Distortion 6-5
Gain Compression • Phase Distortion • Measurement of Gain Compression and Phase Deviation
6.4 Intermodulation Distortion 6-10
Two-Tone Intermodulation Distortion • Third-Order Intercept Point • Dynamic Range • Intermodulation Distortion of Cascaded Components • Measurement of Intermodulation Distortion
6.5 Multicarrier Intermodulation Distortion and Noise Power Ratio 6-13
Peak-to-Average Ratio of Multicarrier Signals • Noise–Power Ratio • Measurement of Multitone IMD and NPR
6.6 Distortion of Digitally Modulated Signals 6-16
Intermodulation Distortion of Digitally Modulated Signals • Measurement of ACPR, EVM, and ρ-Factor
6.7 Summary 6-21
References 6-21

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:*

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

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 *n*th order system will produce harmonics of *n*th 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

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

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.

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

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.

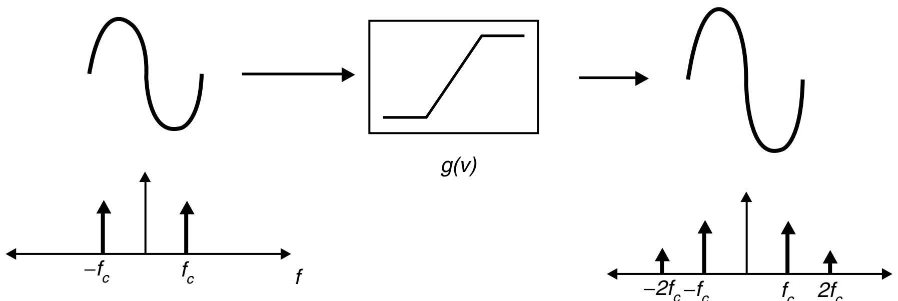


FIGURE 6.1 Effects of a nonlinear transfer characteristic on a sinusoidal input: harmonic distortion.

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

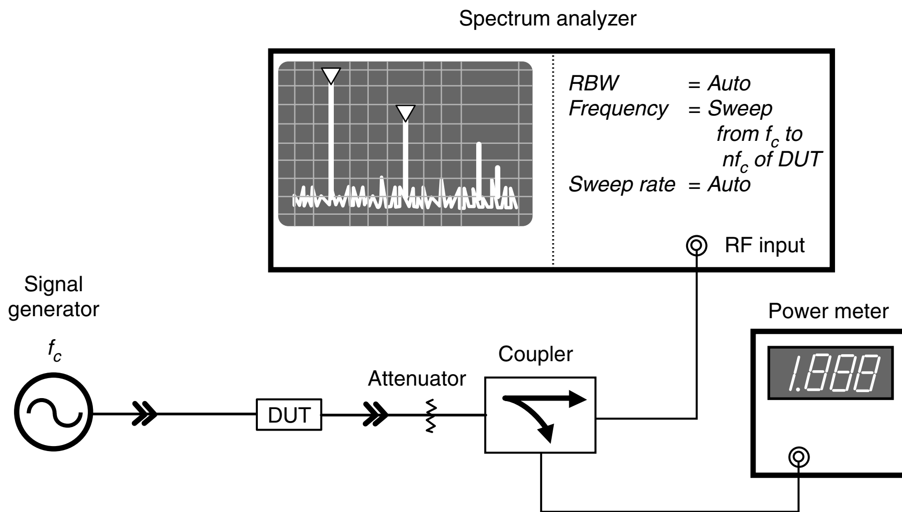


FIGURE 6.2 Setup used to measure harmonic distortion. Because harmonic levels are a function of output amplitude, a power meter is needed to accurately characterize the harmonic distortion properties.

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

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

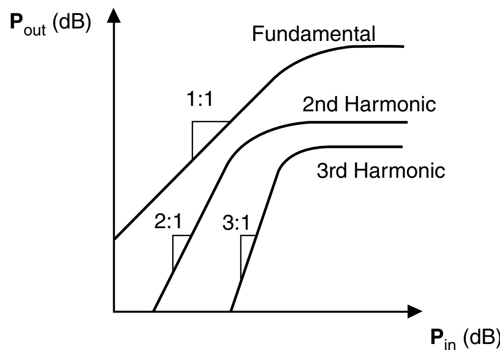


FIGURE 6.3 Output power vs. input power for a nonlinear circuit.

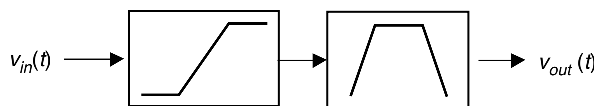


FIGURE 6.4 Bandpass nonlinear circuit.

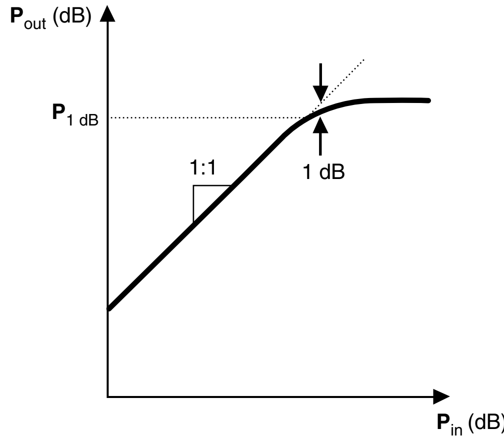


FIGURE 6.5 Gain compression of a bandpass nonlinear circuit. A figure of merit $P_{1\,\mathrm{dB}}$ is the output power at which the gain has been reduced by 1 dB.

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:

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

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.

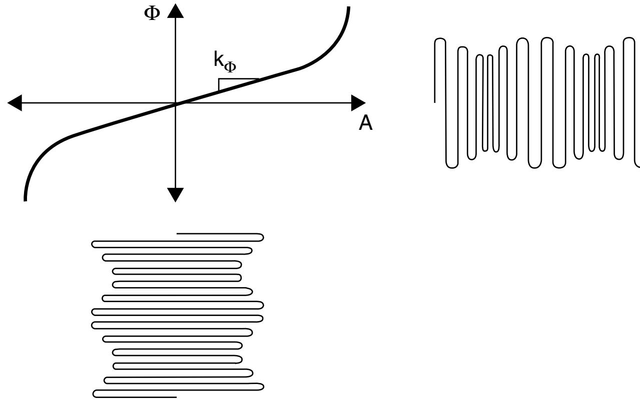


FIGURE 6.6 Effect of AM–PM distortion on a modulated signal. Input signal has AM component only. Output signal has interrelated AM and FM components due to the AM–PM distortion of the circuit.

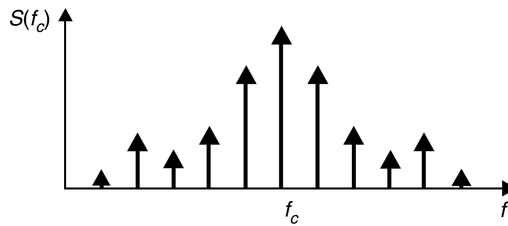


FIGURE 6.7 Output components of an amplitude-modulated signal distorted by AM–PM effects.

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. $P_{1\text{ dB}}$ 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 $\Delta\text{Mag}(S_{21}) = -1$ 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 $\Delta\Phi\,(P_{1\text{ dB}})$

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

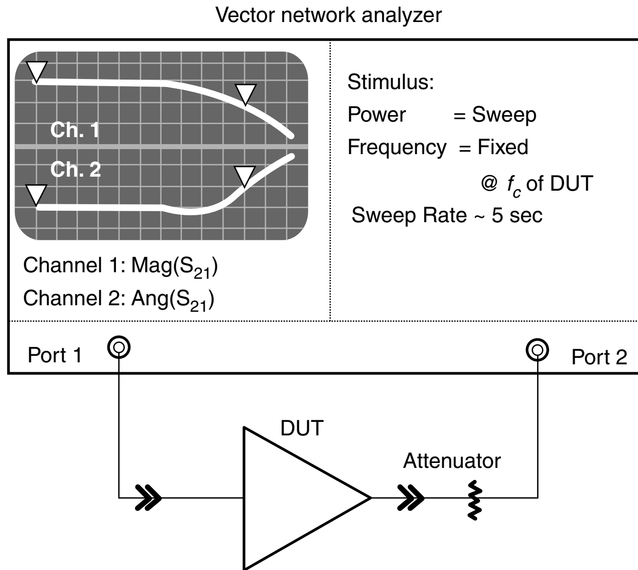


FIGURE 6.8 Setup used to measure gain compression and AM–PM.

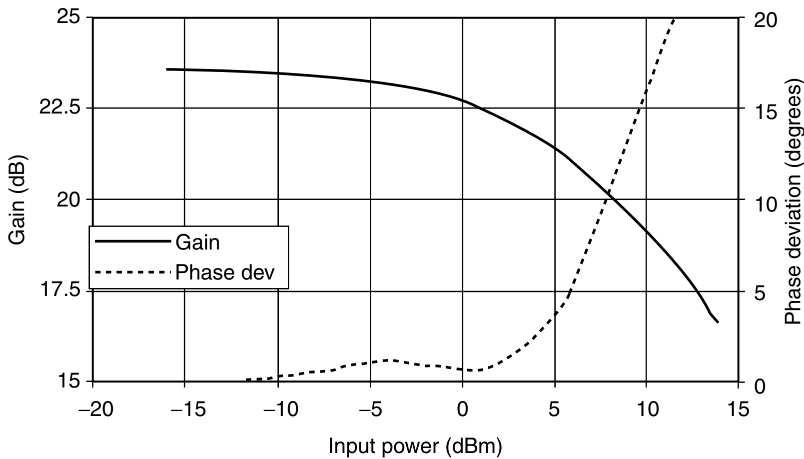


FIGURE 6.9 Measured gain compression and AM–PM of a 0.5 W 1960 MHz GaAs MESFET power amplifier IC using an HP8753C Vector Network Analyzer in power sweep mode.

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.

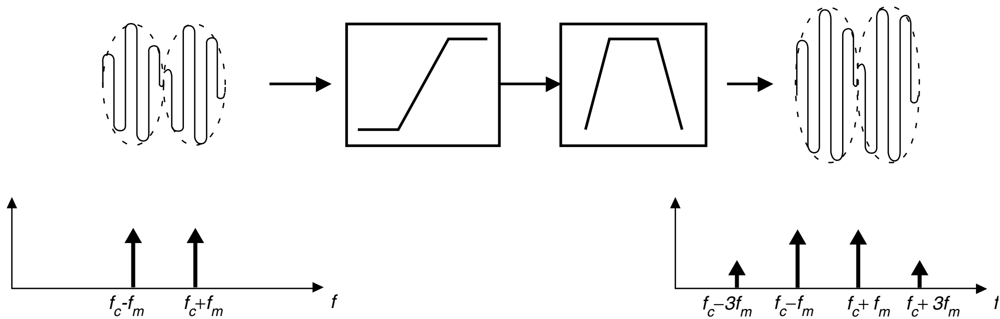


FIGURE 6.10 Intermodulation distortion of a two-tone signal. The output bandpass signal contains the original input signal as well as the harmonics of the envelope at the sum and difference frequencies.

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.

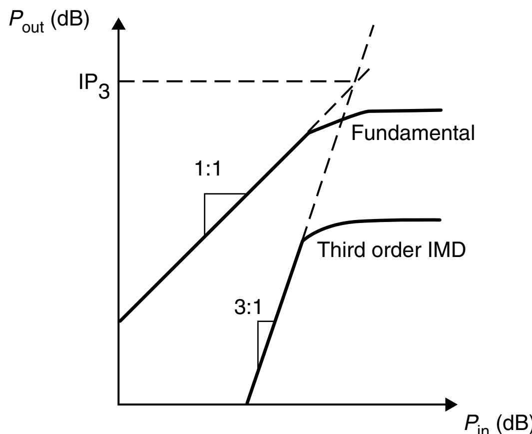


FIGURE 6.11 Relationship between signal output power and intermodulation distortion product power levels. Extrapolating the trends, a figure of merit called the *third-order intercept point* (IP_3) is obtained.

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.

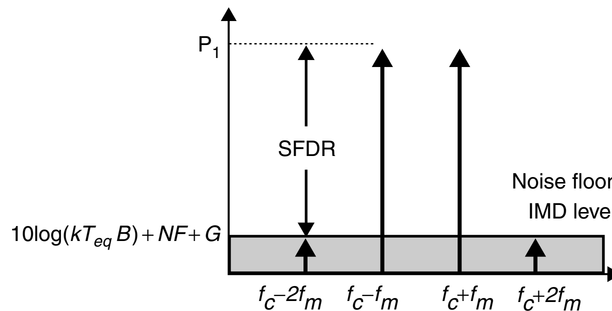


FIGURE 6.12 An illustration of spurious-free dynamic range, which defines the range of signal levels where the worst case signal-to-noise ratio is defined by the noise floor of the system, rather than the IMD level.

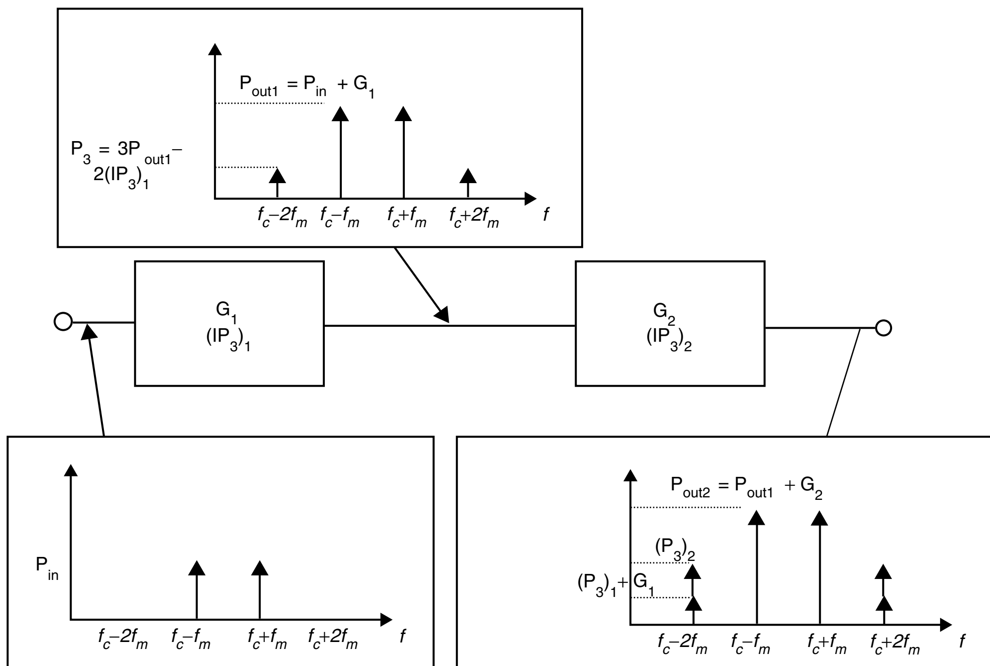


FIGURE 6.13 Effect of cascaded IMD levels. The IMD from the first stage may be power added to those of the second stage with levels adjusted for the gain of the stage.

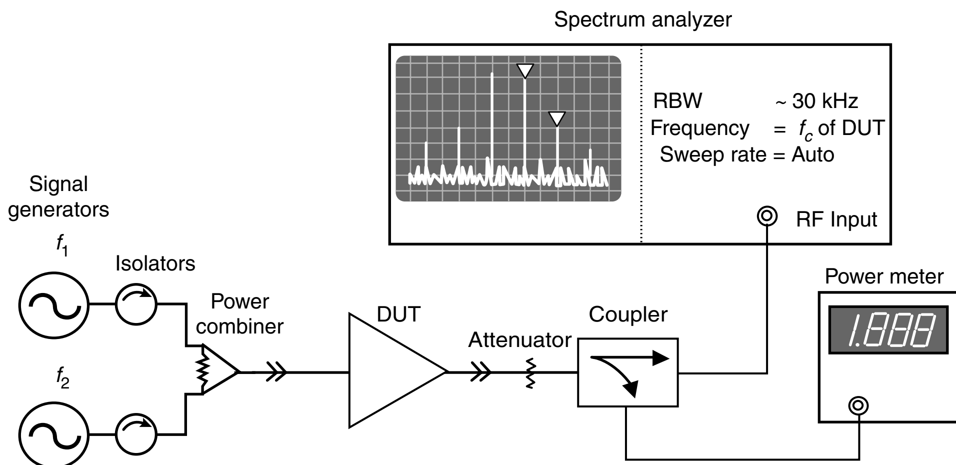


FIGURE 6.14 Setup used to measure two-tone intermodulation distortion.

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

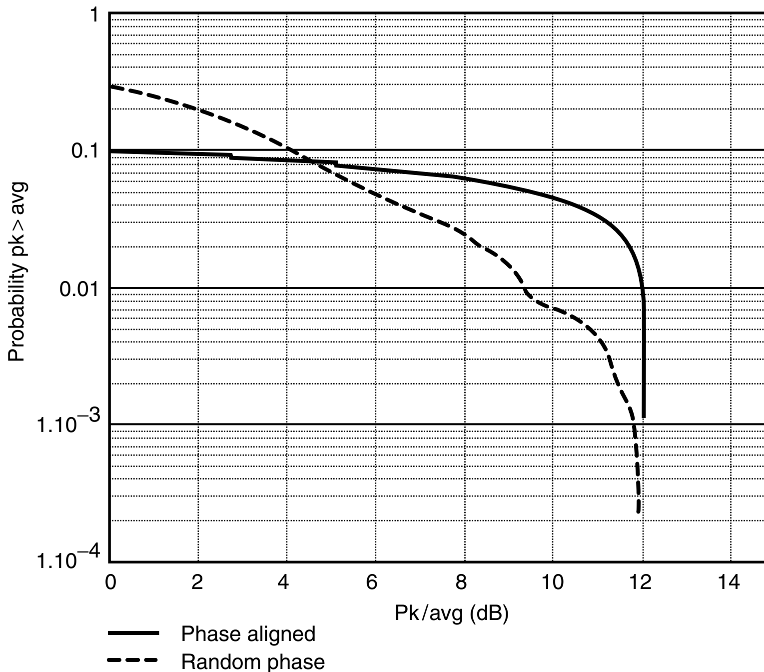


FIGURE 6.15 Distribution of *peak-to-average* ratio of a phase-aligned 16-carrier signal and a random-phase 16-carrier signal. The y-axis shows the probability that the signal exceeds a power level above average on the x-axis. While both signals ultimately have the same pk/avg ratio, their distributions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

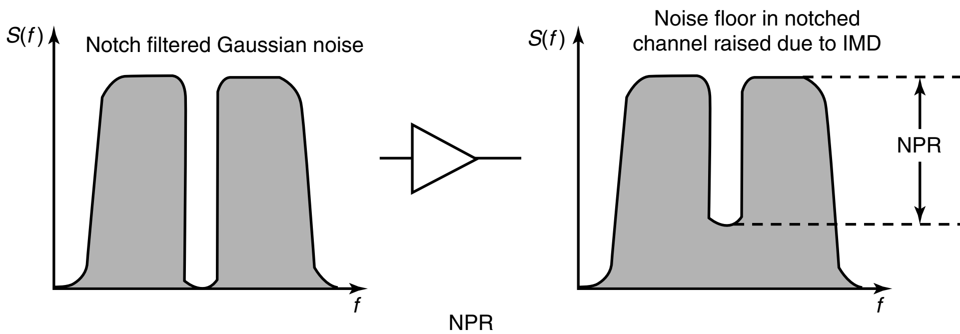


FIGURE 6.16 An illustration of NPR. NPR is essentially a measure of the carrier-to-interference level experienced by multiple carriers passing through a nonlinear component.

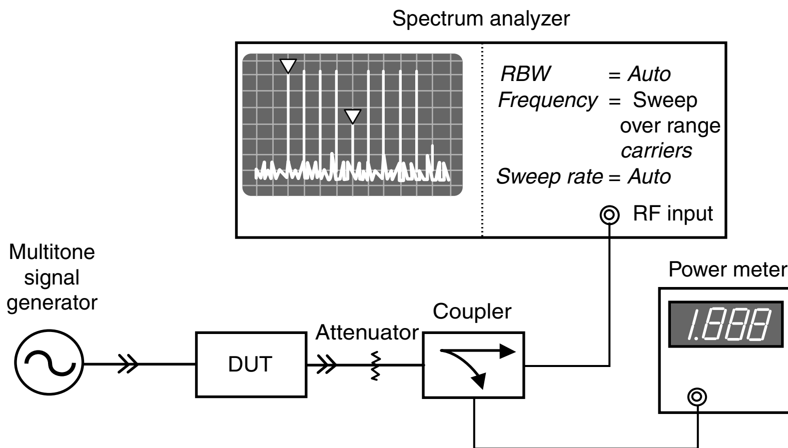


FIGURE 6.17 Measurement setup for multitone IMD. Tones are usually spaced equally, with the middle tone deleted to allow measurement of the worst-case IMD.

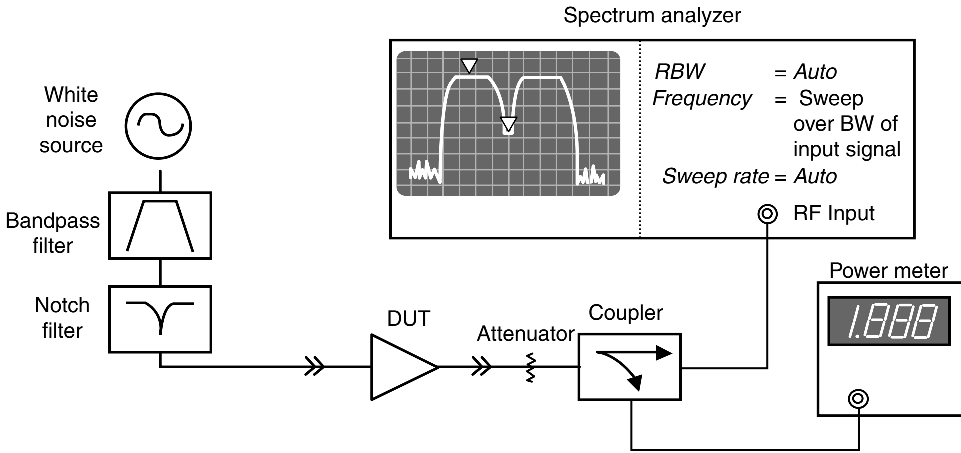


FIGURE 6.18 Noise Power Ratio measurement setup. The rejection of the notch filter should be at least 10 dB below the NPR level to avoid erroneous measurement.

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

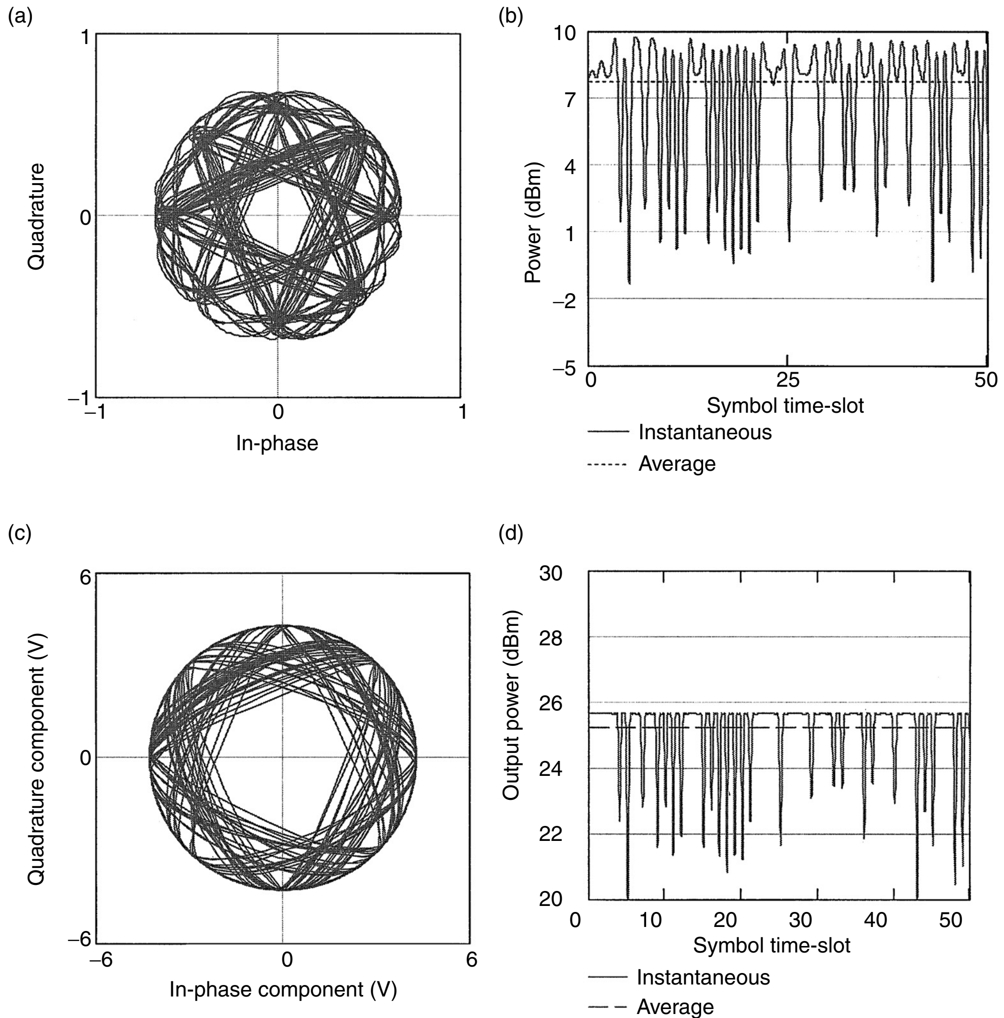


FIGURE 6.19 Effect of amplitude and phase distortion on digitally modulated signals. (a) Shows a ¼/4 DQPSK signal constellation, and its associated power envelope in (b). When such a signal is passed through a nonlinear amplifier, the resulting envelope is clipped (d), and portions of the constellation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

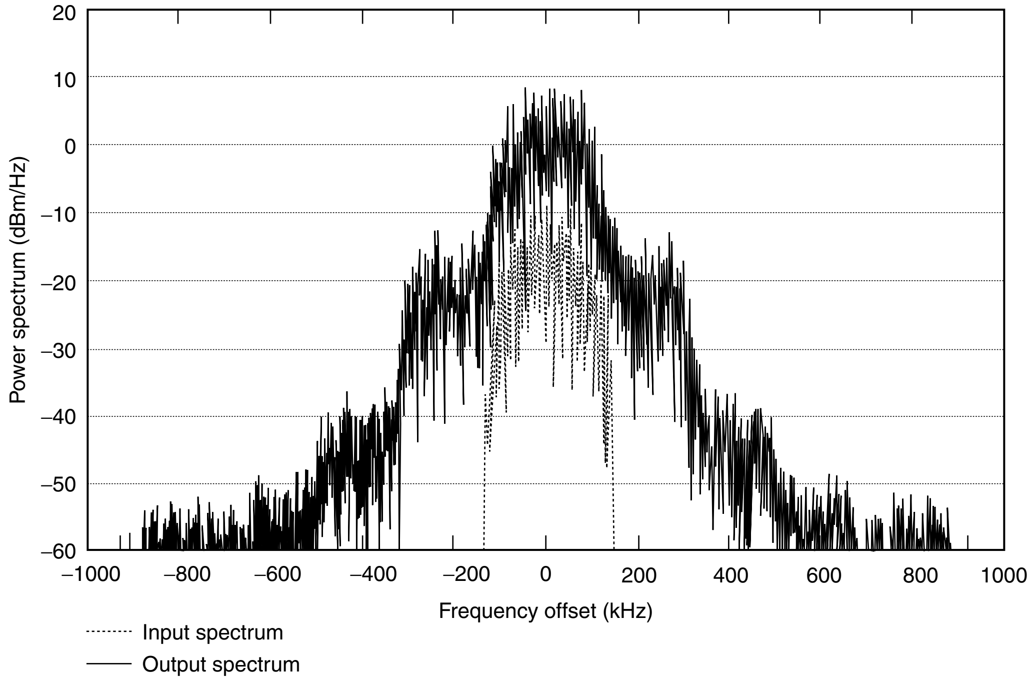


FIGURE 6.20 Effect of nonlinear distortion on a digitally modulated signal. The lower power input signal to a power amplifier has a frequency spectrum that is well contained within a specified channel bandwidth. IMD due to nonlinear distortion creates out-of-band products that may fall within the adjacent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

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

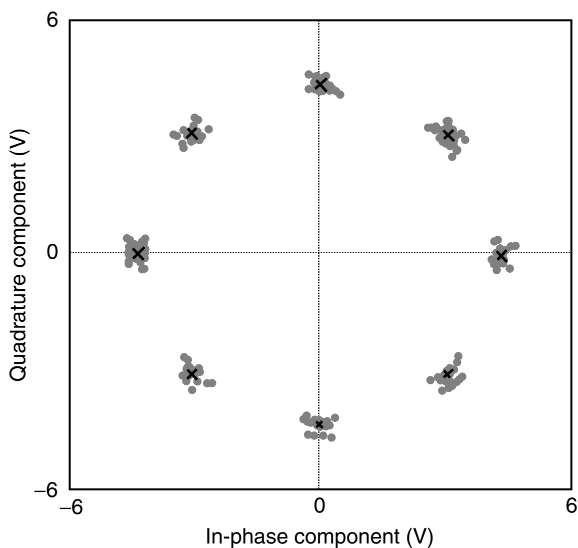


FIGURE 6.21 Errors in the demodulated *I–Q* constellation may result from the in-band IMD products. The *rms* summation of errors from the desired location (given by the × markers) give the *error vector magnitude* of the signal distortion.

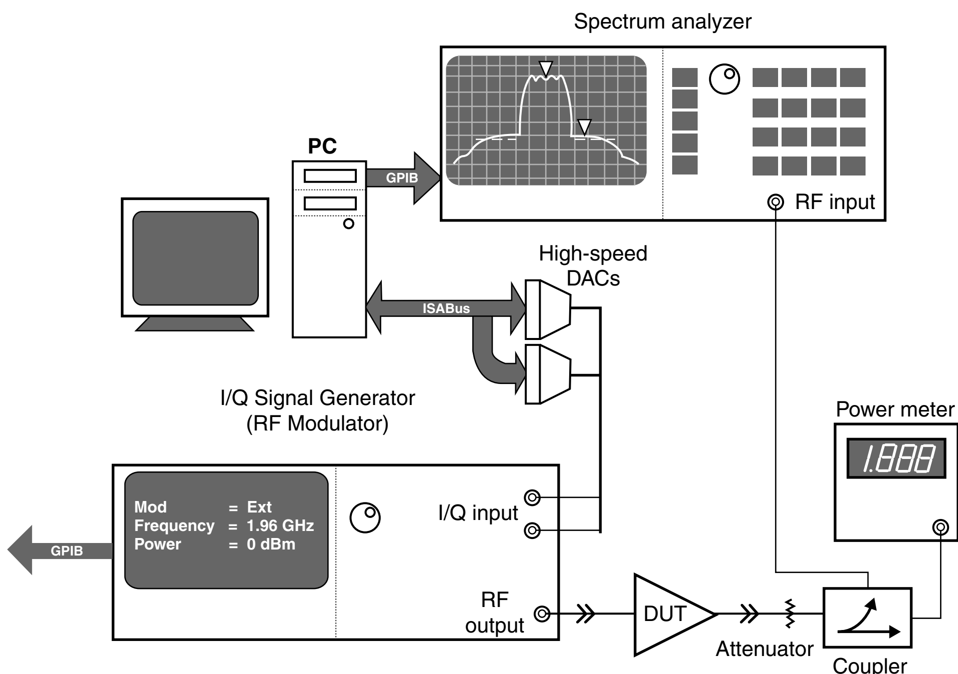


FIGURE 6.22 Measurement setup for ACPR. Waveforms are created using PCs or specialized arbitrary waveform generators. In either case, the baseband waveform must be upconverted to the center frequency of the DUT.

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.

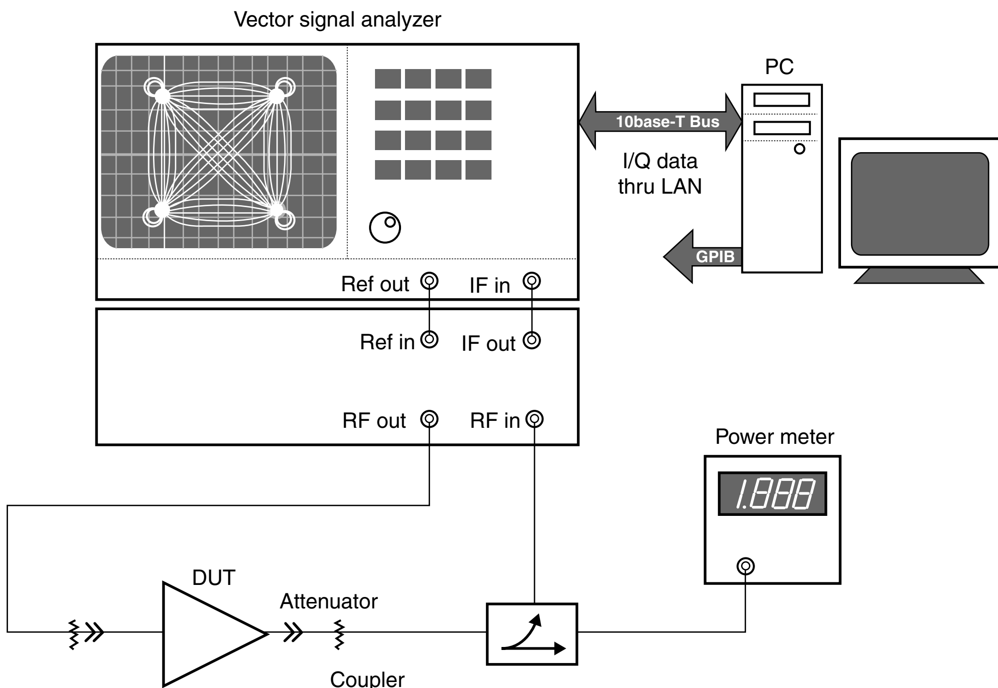


FIGURE 6.23 Setup for measuring EVM. The VSA demodulates the *I*–*Q* waveform and calculates the deviation from ideal to calculate EVM and ρ-factor as given in Equations 6.24 and 6.25, respectively.

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.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

7

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

John F. Sevic
Maury Microwave Corporation

7.1 System Architecture for High-Power Load-Pull 7-2
7.2 Characterization of System Components 7-5
Vector Network Analyzer Calibration Theory • S-Parameter Characterization of Tuners • S-Parameter Characterization of System Components • Fixture Characterization to Increase System VSWR
7.3 System Performance Verification 7-13
7.4 Summary 7-15
Acknowledgments 7-15
References 7-16

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

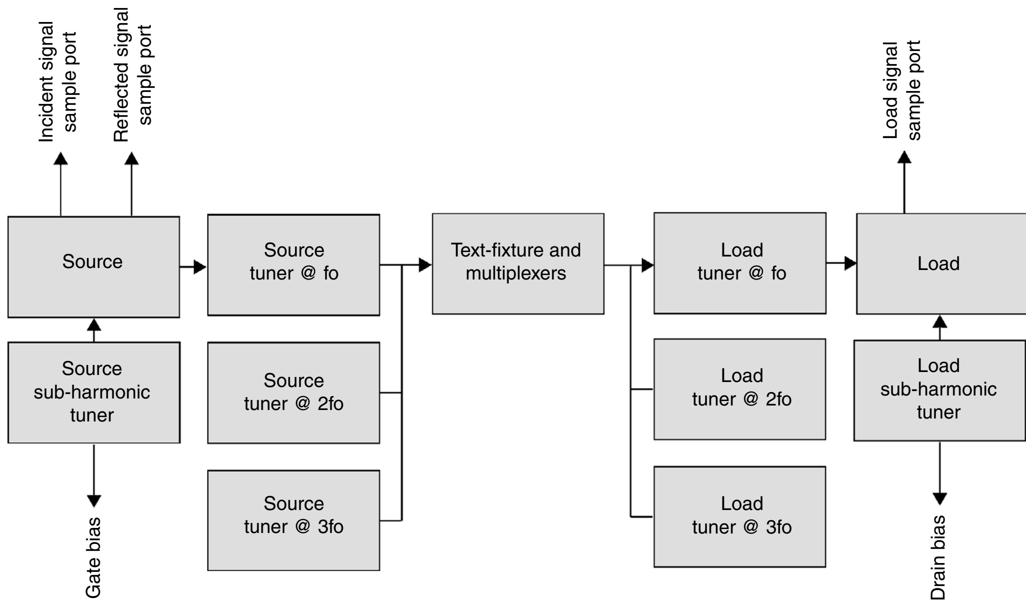


FIGURE 7.1 Block diagram of a generalized high-power load-pull system, illustrating the source, tuners, test-fixture, and load. The incident, reflected, and load signals are sampled at the three sampling points shown. Also shown, though not necessary, are harmonic and sub-harmonic tuners.

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].

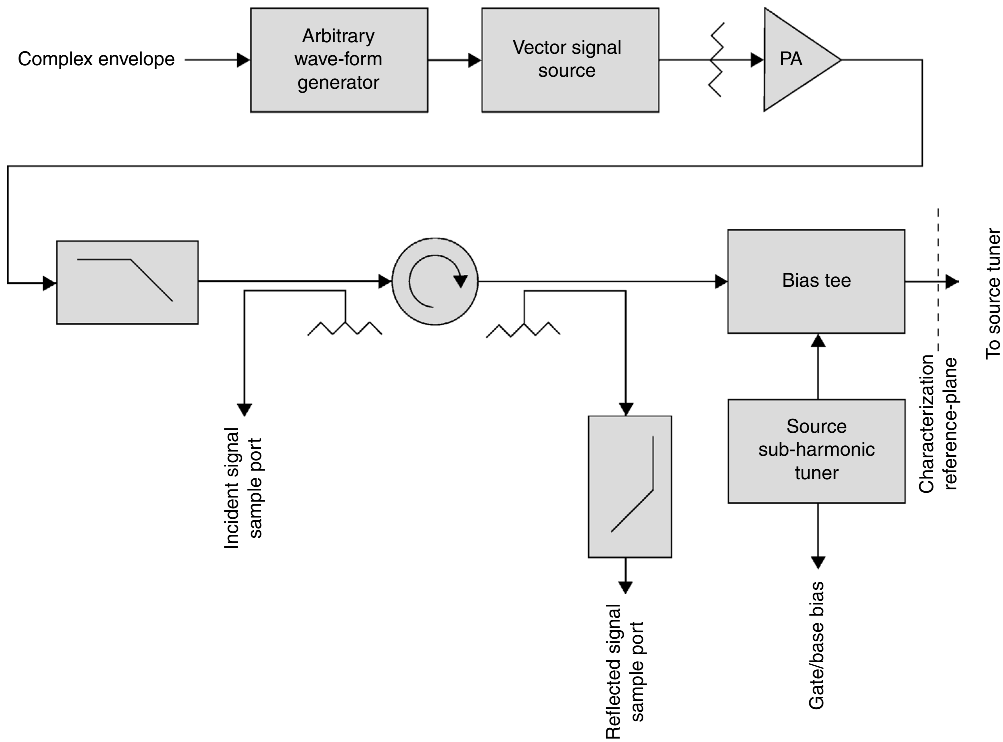


FIGURE 7.2 Detail of the source portion of Figure 7.1.

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.

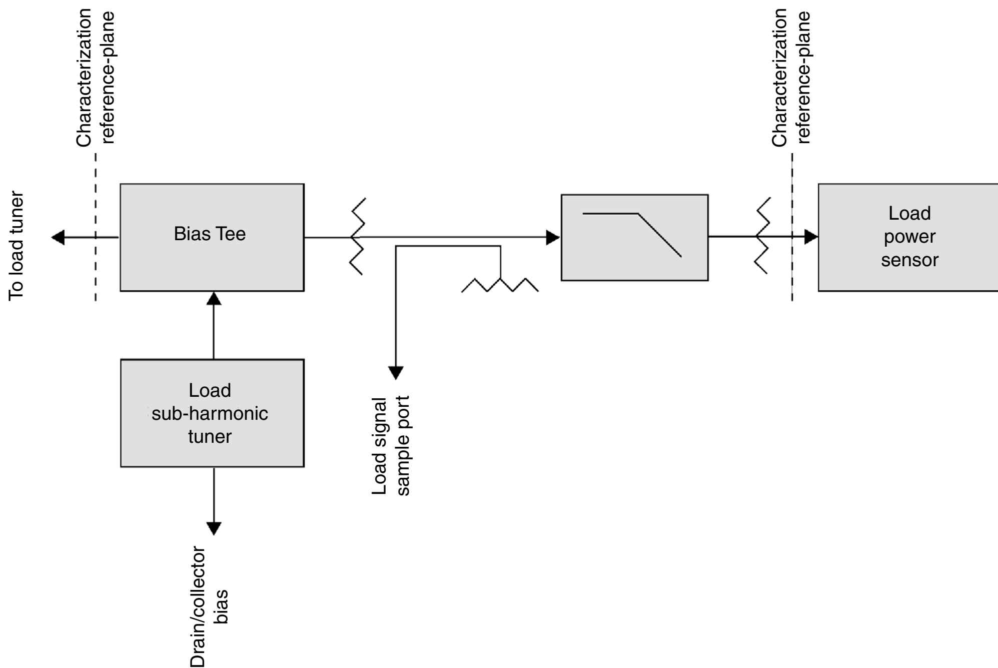


FIGURE 7.3 Detail of the load portion of Figure 7.1.

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].

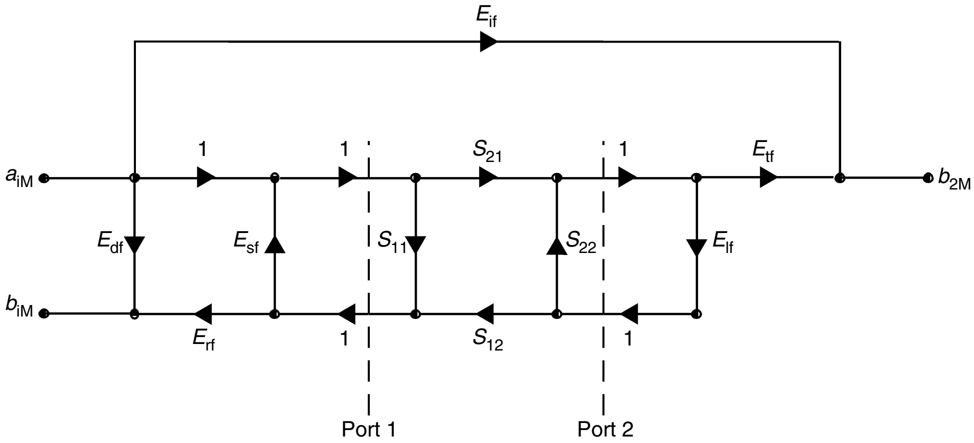


FIGURE 7.4 Signal-flow graph of the forward direction of a typical VNA.

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

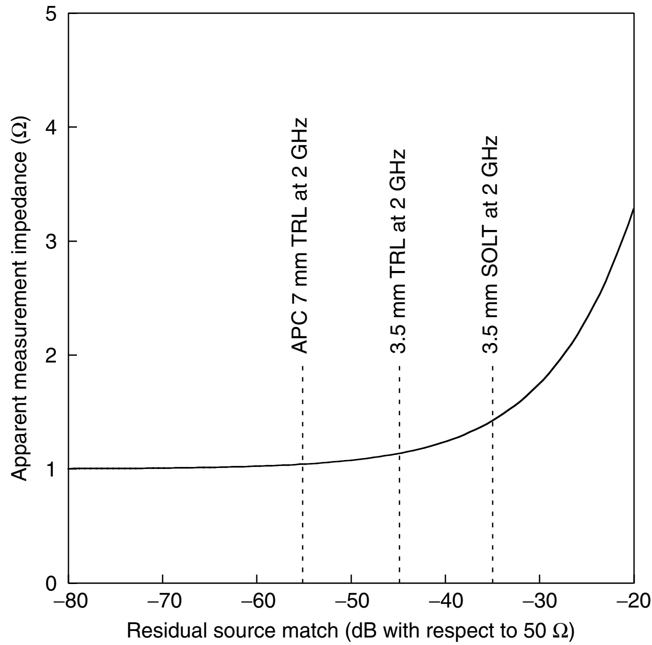


FIGURE 7.5 The influence of residual source match on the ability of a VNA to resolve a 1 Ω impedance with a 50 Ω reference impedance. The calibration performance numbers are typical for an HP 8510C with an 8514B test-set operating a 2 GHz.

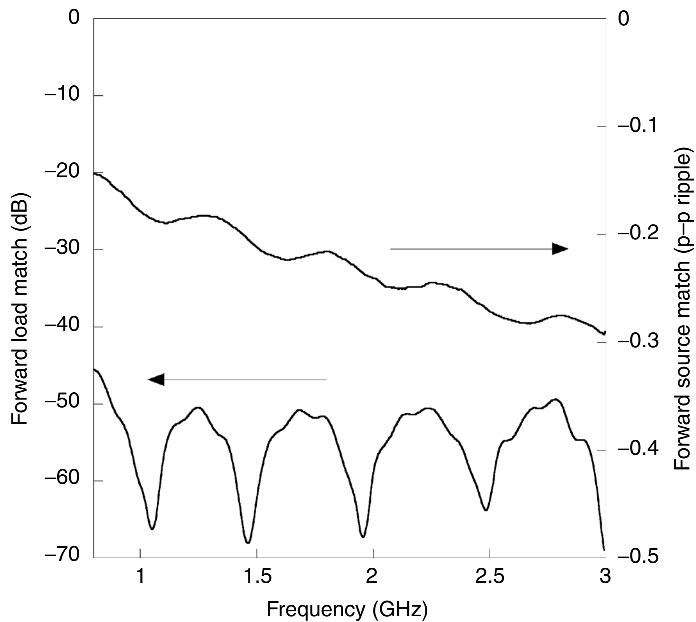


FIGURE 7.6 Typical response of an APC 7 mm TRL calibration using an offset-short and delay-line to extract source match and load match, respectively. The data were taken from an HP 8510C with an HP 8514B test set.

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.

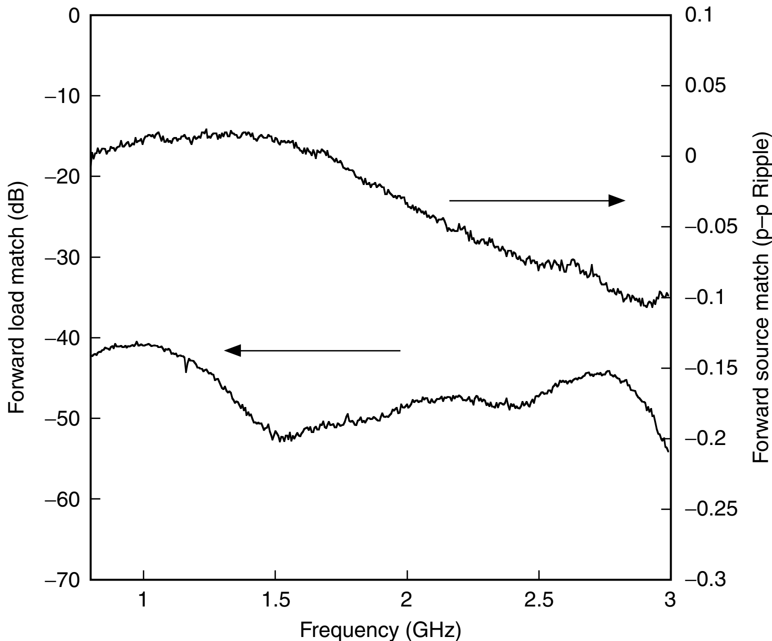


FIGURE 7.7 Typical response of a 3.5 mm TRL calibration using an offset-short and delay-line to extract source match and load match, respectively. The data were taken from an HP 8510C with an HP 8514B test set.

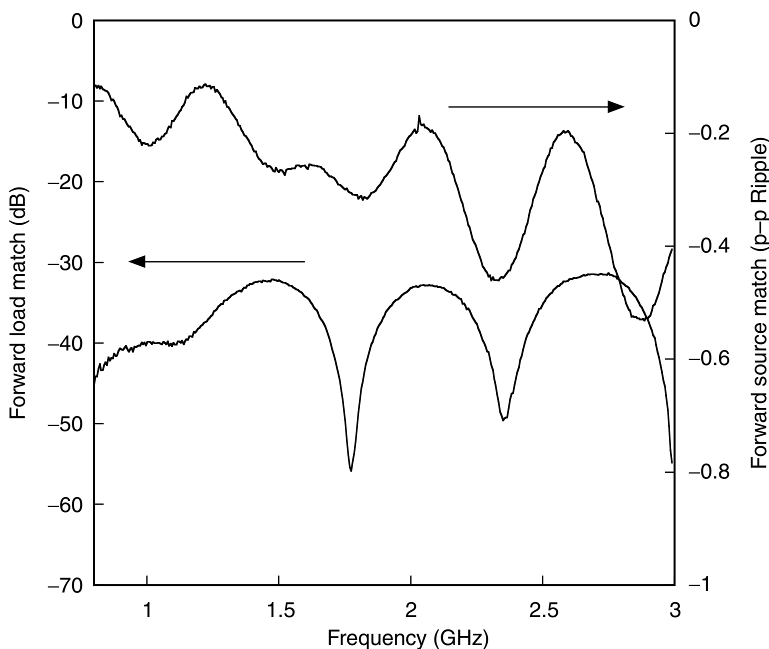


FIGURE 7.8 Typical response of a 3.5 mm SOLT calibration using an offset-short and delay-line to extract source match and load match, respectively. The data were taken from an HP 8510C with an HP 8514B test set.

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

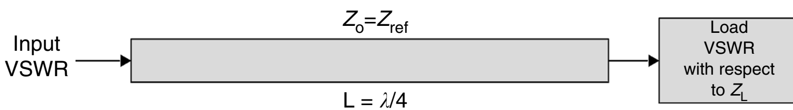


FIGURE 7.9 Network to describe the mismatch invariance property of lossless networks.

*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

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

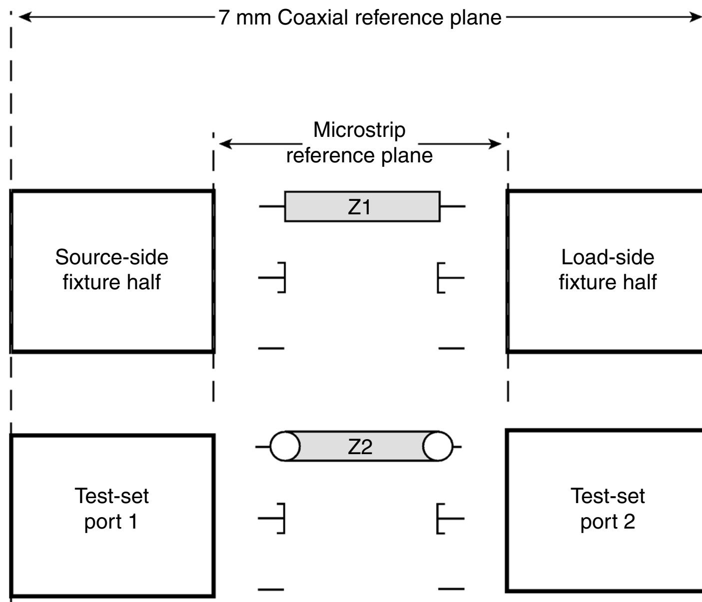


FIGURE 7.10 Reference-plane definitions for a two-tier calibration used for fixture characterization. The first tier is based on a TRL APC 7 mm calibration and the second tier is based on a microstrip TRL calibration.

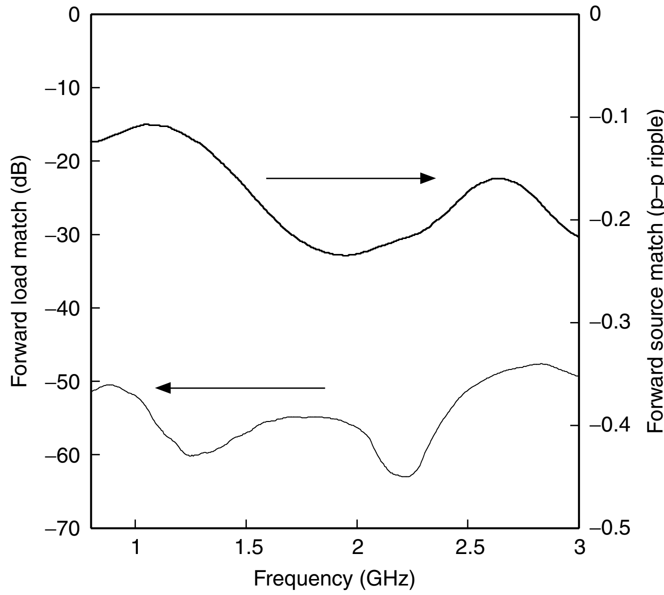


FIGURE 7.11 Microstrip TRL calibration using an offset-short and delay-line to extract source match and load match, respectively. The data were taken from an HP 8510C with an HP 8514B test set.

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

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

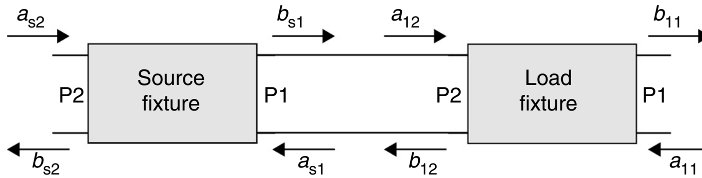


FIGURE 7.12 Port and traveling-wave definitions for cascading the source-fixture and load-fixture to examine the accuracy of the two-tier calibration fixture characterization.

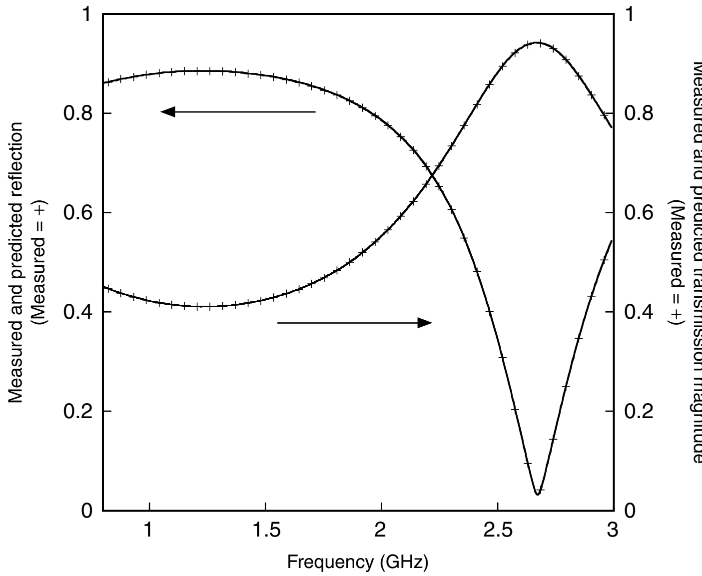


FIGURE 7.13 Forward reflection and transmission magnitude comparison of measured and cascaded fixture response. The error is so small the curves sit on top of each other.

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.

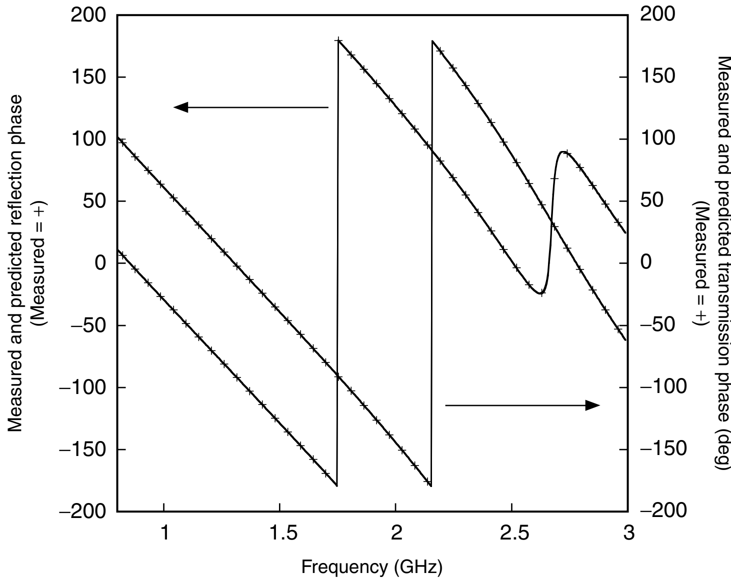


FIGURE 7.14 Forward reflection and transmission phase comparison of measured and cascaded fixture response. The error is so small the curves sit on top of each other.

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

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

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.

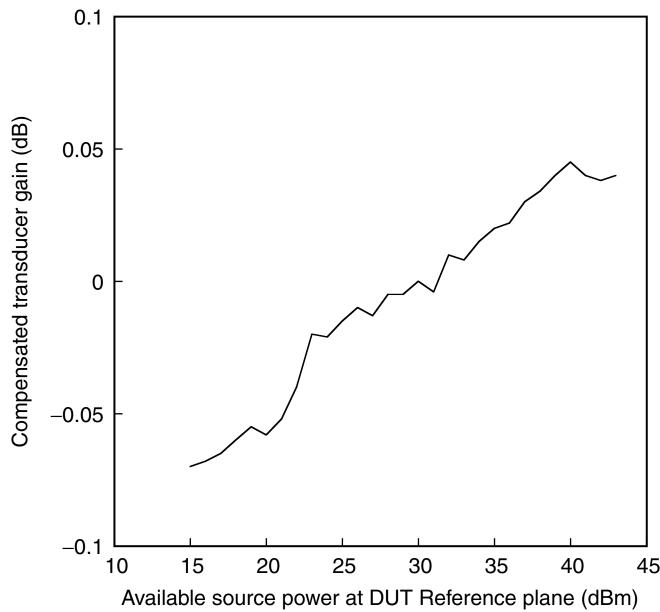


FIGURE 7.15 Measured transducer gain under the condition of conjugate match with mismatch loss compensation included.

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

8.1 Introduction **8**-1
8.2 Isothermal and Isodynamic Characteristics **8**-2
Small-Signal Conditions • Device Dispersion • Large-Signal Conditions • Pulsed Measurements
8.3 Relevant Properties of Devices **8**-7
Safe-Operating Area • Thermal Dispersion • Charge-Trapping • Time Constants • Pulsed-*I*/*V* and Pulsed-RF Characteristics
8.4 Pulsed Measurement Equipment **8**-11
System Architecture • Technical Considerations • Commercial Measurement Systems
8.5 Measurement Techniques **8**-19
The Pulse-Domain Paradigm and Timing • General Techniques
8.6 Data Processing **8**-26
Interpolation and Gridding
8.7 Summary **8**-28
Defining Terms **8**-29
References **8**-29

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

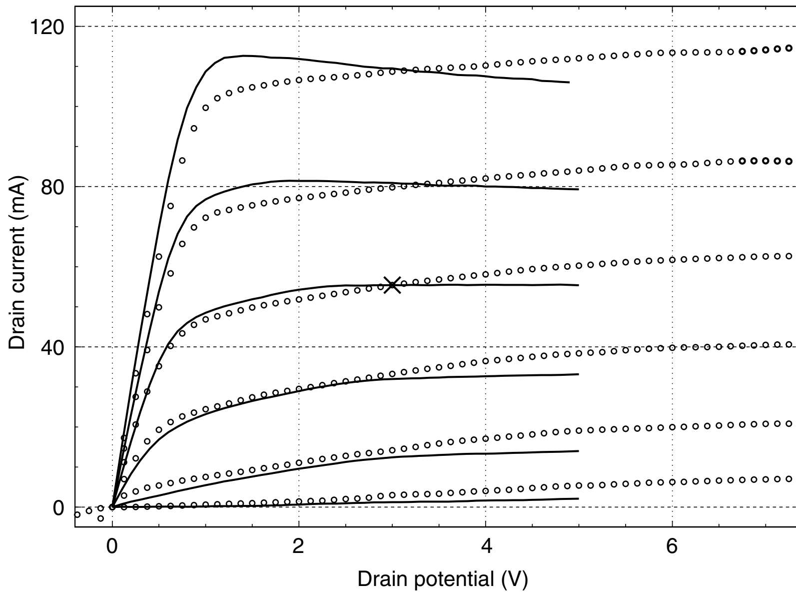


FIGURE 8.1 Characteristic curves for a MESFET. Shown are the DC characteristics (—) and the pulsed characteristics (∘), with 300 ns pulses separated by 200 ms quiescent periods, for the quiescent point $V_{DS} = 3.0$ V, $I_D = 55.4$ mA (×). Gate-source potential from −2.0 to +0.5 V in 0.5 V steps is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

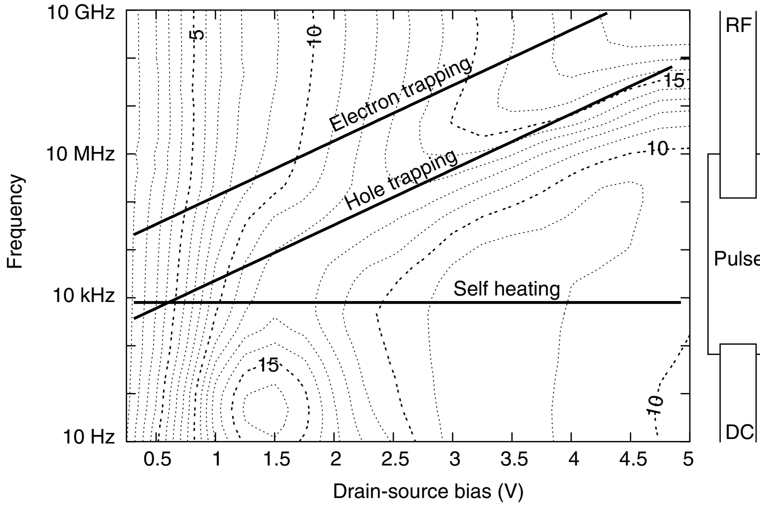


FIGURE 8.2 Frequency and drain-bias variation of intrinsic gain for a HEMT at a gate-bias of $V_{GS} = -0.8$ V. The characteristic frequencies of heating and charge trapping effects are indicated by the labeled lines. Also indicated are the typical frequency ranges applicable to DC, pulsed, and RF measurements.

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.

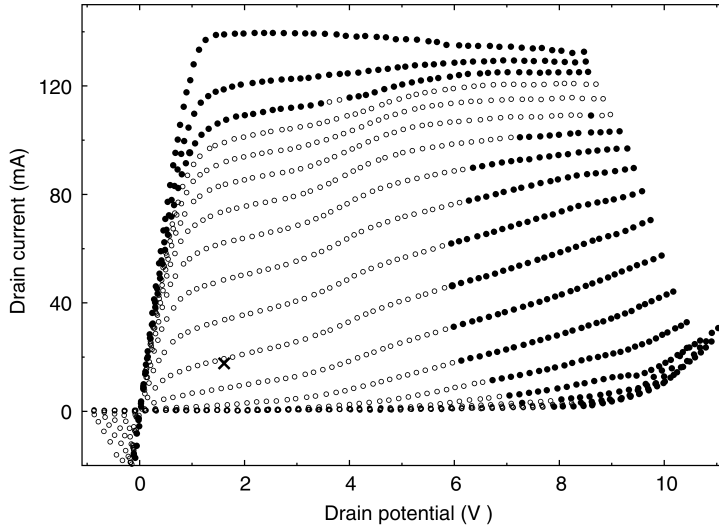


FIGURE 8.3 An example of the extended characteristic curves for an HEMT obtained with 500 ns pulses at 10 ms intervals for the quiescent point $V_{DS} = 1.6$ V, $I_D = 17.7$ mA (×). The solid points (•) are those for which the magnitude of gate current is greater than 1 mA. Gate-source potential from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

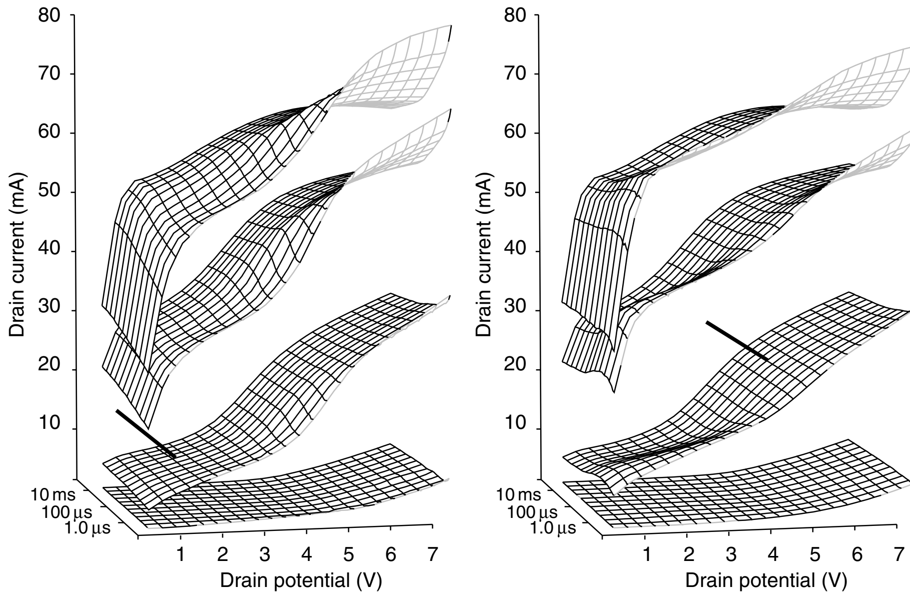


FIGURE 8.4 Time-evolution characteristics showing transients after a step change from the initial condition shown by solid lines (—). Each graph shows surfaces of drain current with gate-source potential from −1.5 to 0.0 V in 0.5 V steps as the parameter.

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

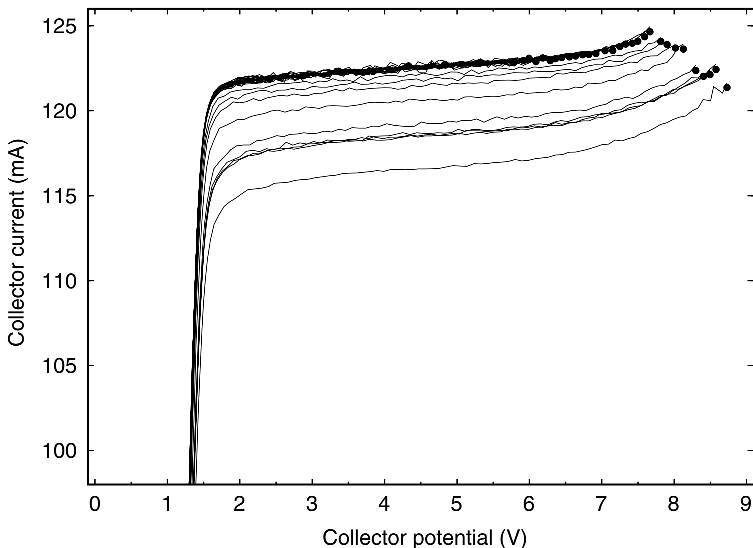


FIGURE 8.5 A single collector characteristic measured on a 140 μm^2 III–V HBT with sequentially increasing maximum voltage (shown by •) applied in 1 μs pulses. Note the progressive deterioration above a certain instantaneous dissipation level.

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

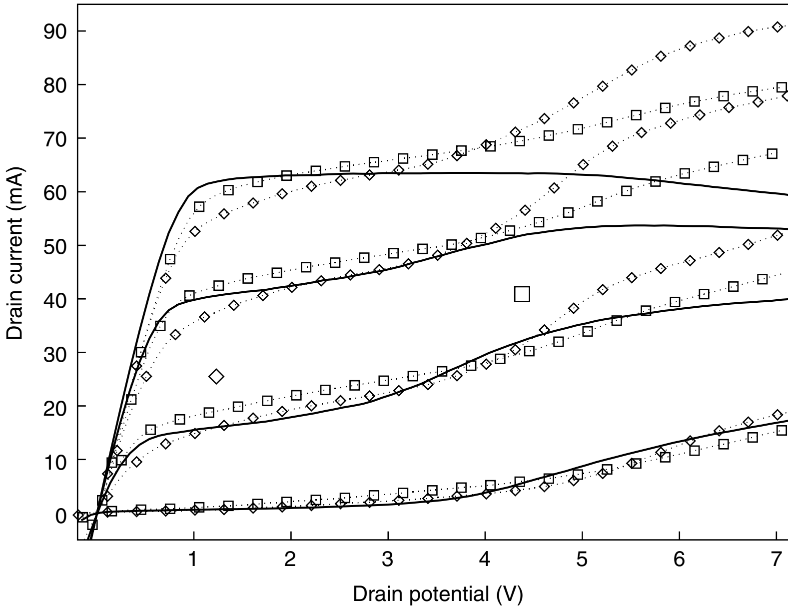


FIGURE 8.6 DC and pulsed characteristics at 100 ns for two initial operating conditions, with gate-source potential from −1.5 to 0.0 V in 0.5 V steps as the parameter. The graph shows two sets of pulse data covering the same terminal potentials as the DC data (—). Shown is pulse data (◇) from initial 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.

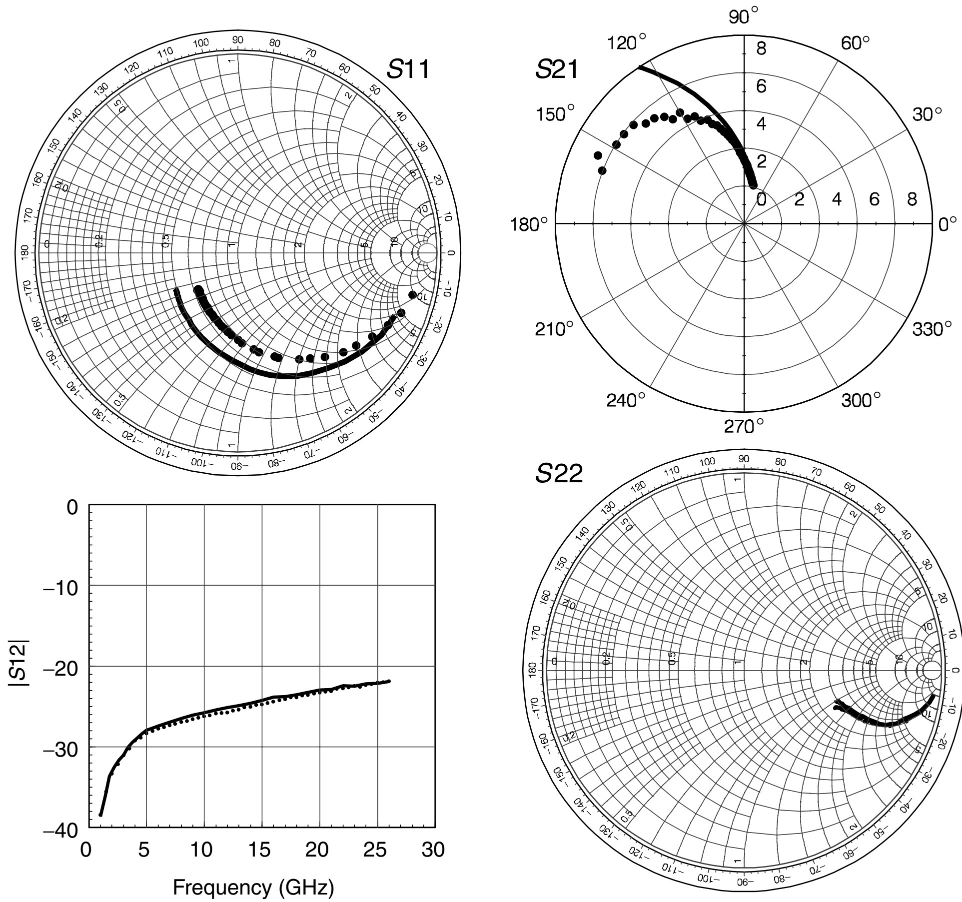


FIGURE 8.7 S-parameters measured at the same bias point with off-state and on-state quiescent conditions. The on-state parameters are from static, or DC measurements (—) and the off-state parameters are from measurements in a pulsed bias at the same point with off-state quiescent periods (•).

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.

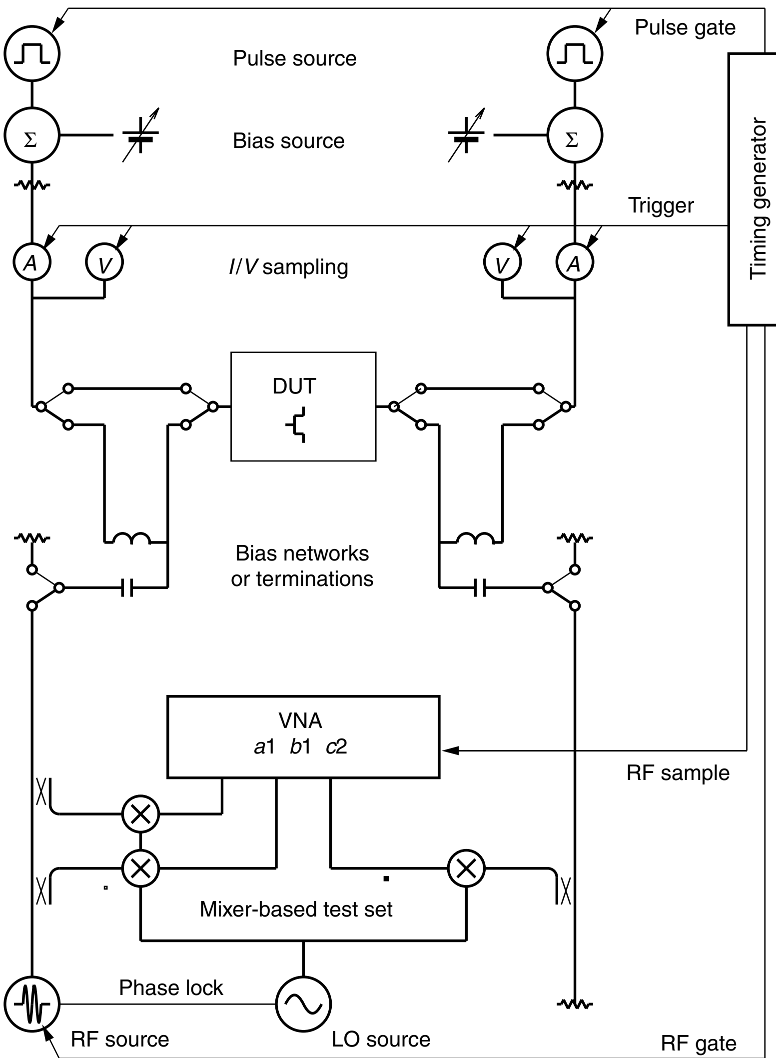


FIGURE 8.8 Simplified diagram of a generic pulsed measurement system. Alternative connections provide load terminations when there is no pulsed-RF test set or directly connect the pulsed-I/V subsystem to the DUT.

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.

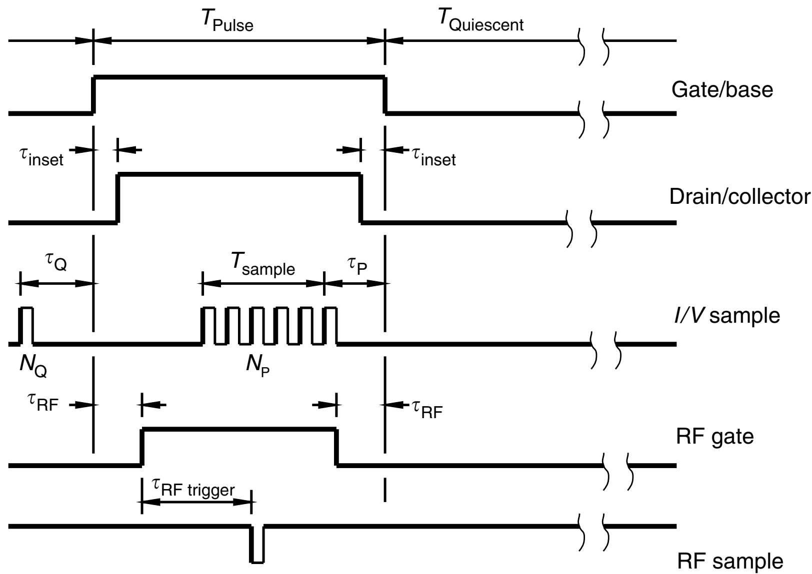


FIGURE 8.9 Generic timing diagram for each pulsed measurement pulse event.

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.

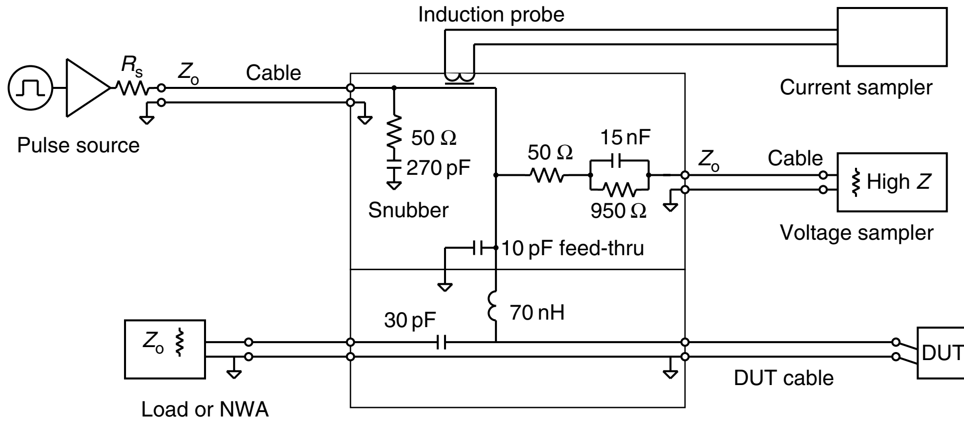


FIGURE 8.10 Schematic of a bias network that provides RF termination and pulsed bias feed with voltage and current measuring points.

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.

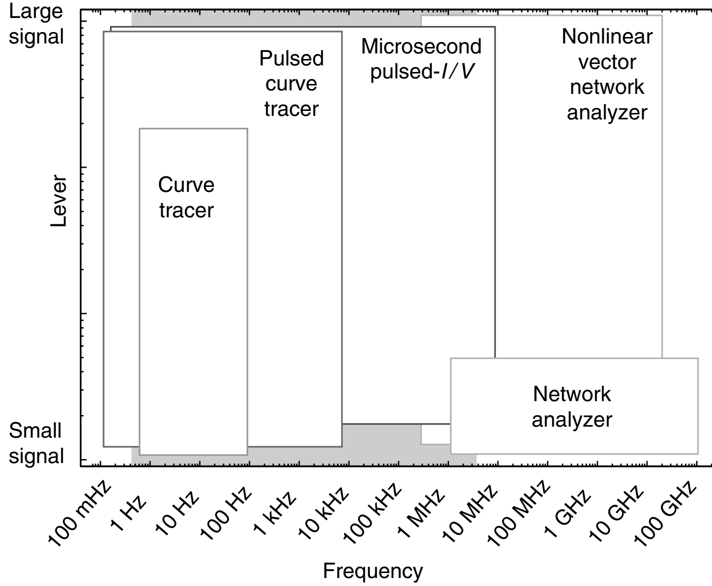


FIGURE 8.11 Relative position of various types of measurement equipment, including pulsed-I/V systems, in term of measurement frequency and signal level. The shaded area indicates the frequency range of typical dispersion effects in low-power FETs at moderate bias.

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

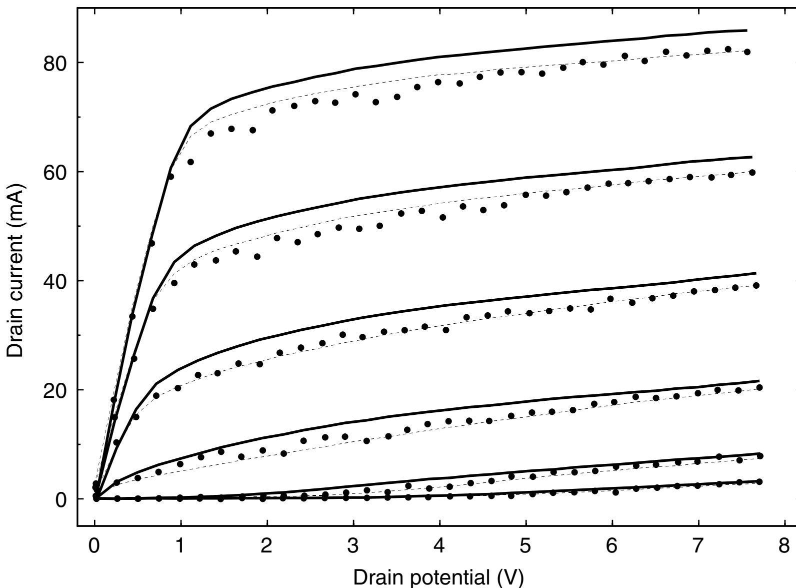


FIGURE 8.12 Characteristic curves for a MESFET measured with three different pulse sequences: a step-and-sweep with 1 μs pulses and 1 μs quiescent periods (- - - -), the same pulses sequenced in pseudorandom order (•), and an isodynamic measurement. The latter used 800 ns pulses with 1ms quiescent periods. 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

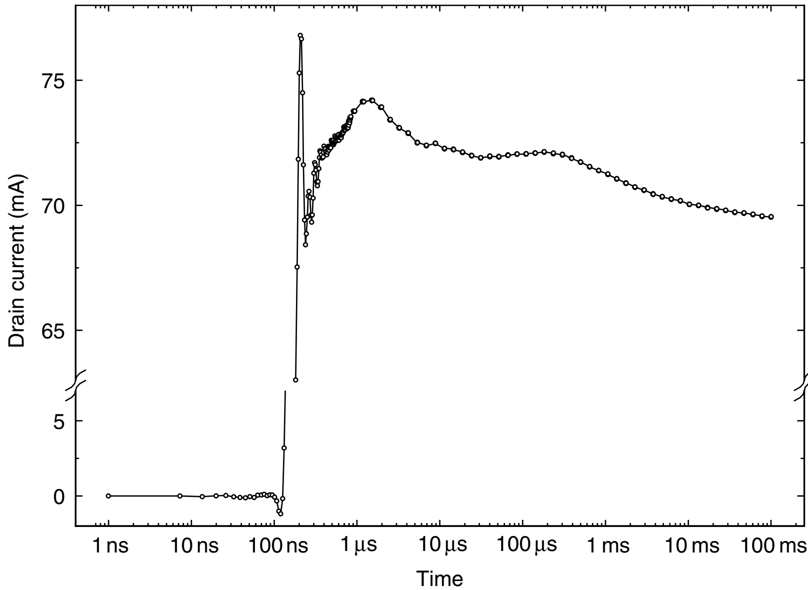


FIGURE 8.13 Transient response measured with eight repeated measurements at 50 ns intervals. Each repetition is shifted by 6.25 ns to give the composite response shown.

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

$$v'_{DS} = v_{DS} - i_D R_D - (i_D + i_G) R_S,$$

$$v'_{GS} = v_{GS} - i_G R_G - (i_D + i_G) R_S. \quad (8.2)$$

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 $\mathrm{d}i_D/\mathrm{d}v_{GS}$ and $\mathrm{d}i_D/\mathrm{d}v_{DS}$ 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

9.1 On-Wafer Test Capabilities and Applications 9-1
Fixtured Test Limitations • On-Wafer Test Enabler: Coplanar Probes • On-Wafer Test Capabilities • On-Wafer RF Test Applications

9.2 Test Accuracy Considerations 9-6
Test Equipment Manufacturer • System Integration • Calibration Technique • Dynamic Range

9.3 On-Wafer Test Interface 9-11

9.4 On-Wafer RF Test Benefits 9-15

References 9-16

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

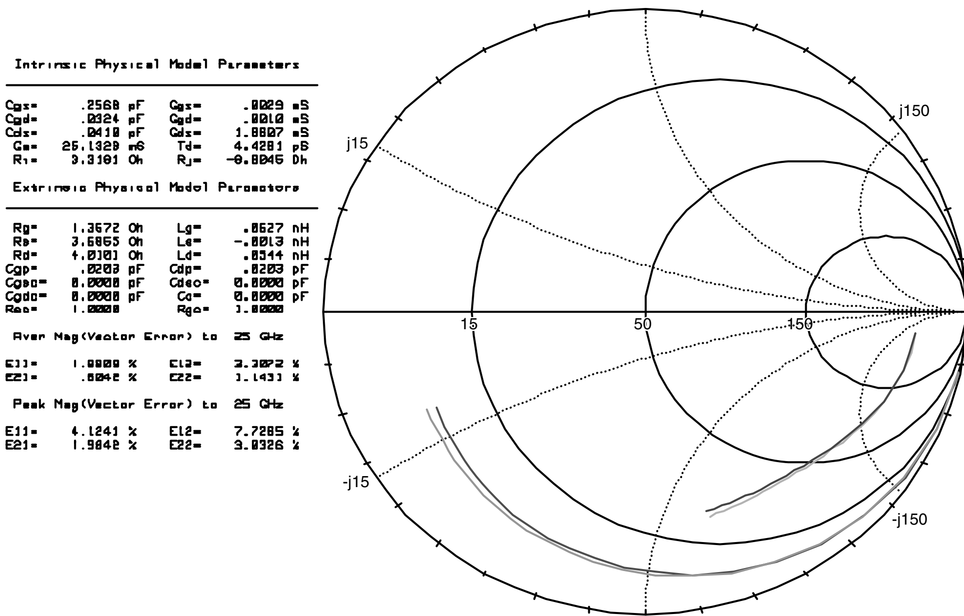


FIGURE 9.1 Equivalent circuit FET model extraction and fit with measurement.

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.

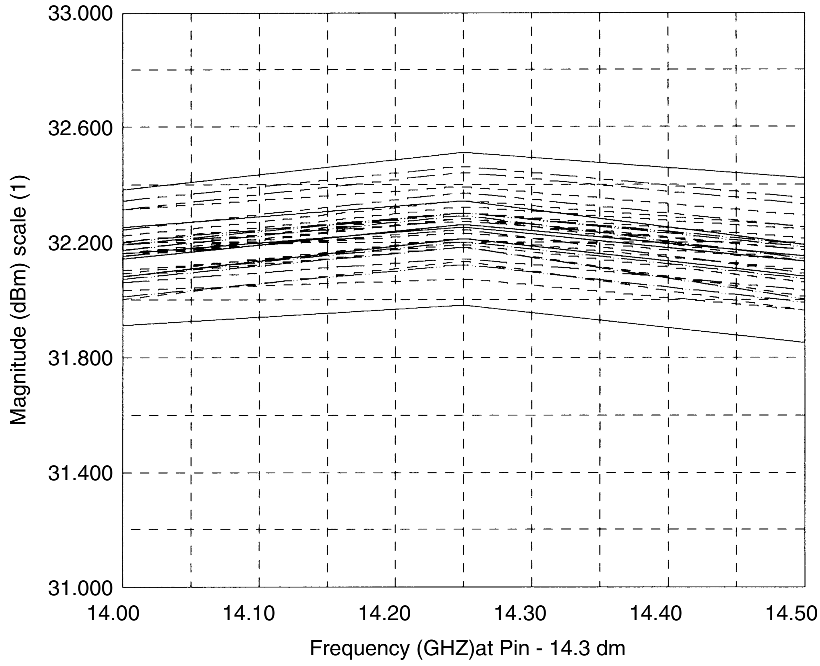


FIGURE 9.2 Pout response of Ku band PAs across a wafer.

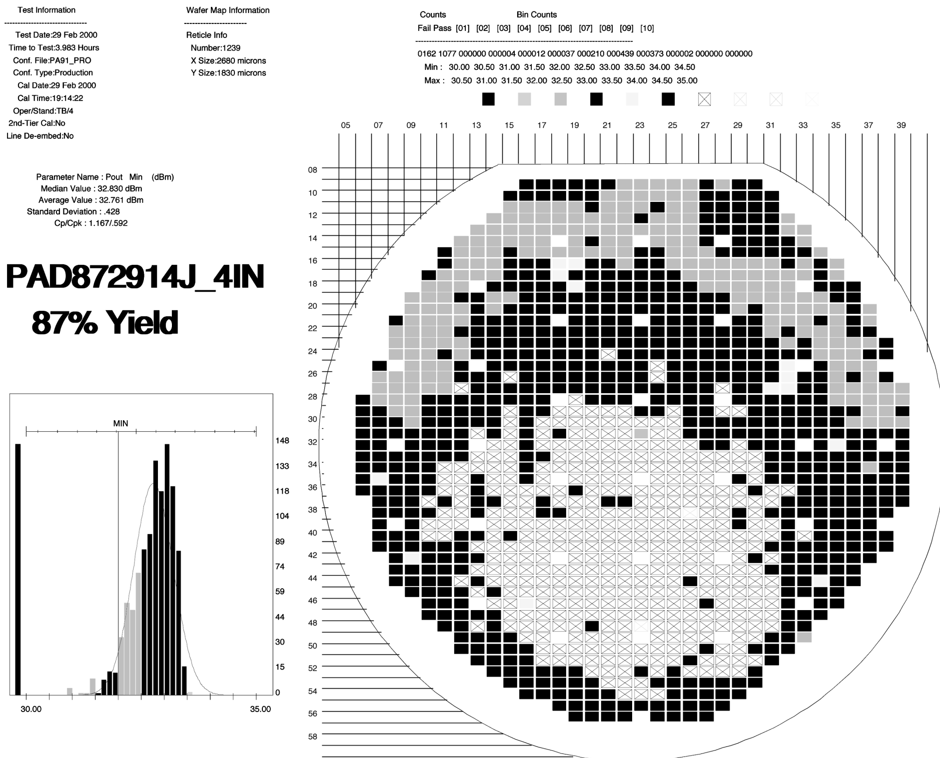


FIGURE 9.3 Wafter map of known good dies from on-wafer test.

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.

$$C_{NF} = R_{NF} + 10LOG\left(\frac{\left(\left|E_{10}^{2}\right|\right)\left(1-\left|G_{ns}^{2}\right|\right)}{\left(\left(1-\left|E_{sf}+\left(E_{10}^{2}G_{ns}/\left(1-G_{ns}E_{df}\right)\right)^{2}\right|\right)\right)\left(\left|\left(1-E_{df}G_{ns}\right)^{2}\right|\right)}\right) \tag{9.3}$$

$$C_{NF} = R_{NF} + 10LOG\left(\left|E_{10}^{2}\right|\right) \tag{9.4}$$

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

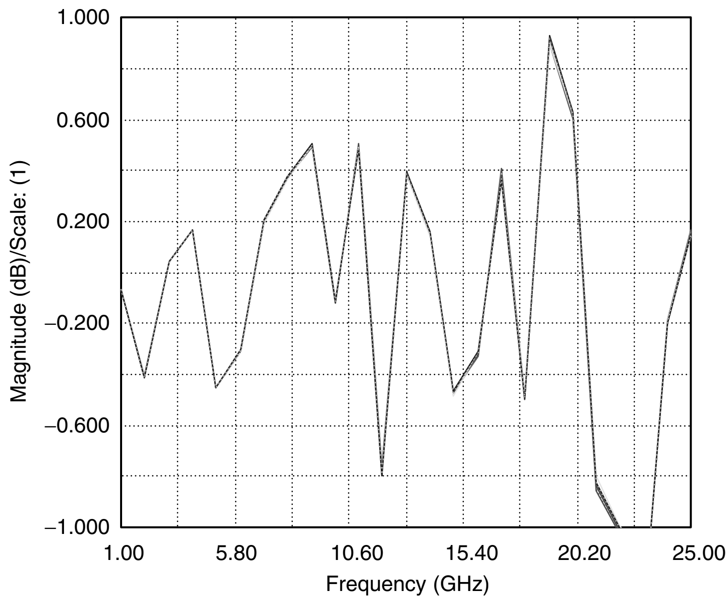


FIGURE 9.4 S_{21} vector to scalar measurement comparison of discrete FET (mismatched).

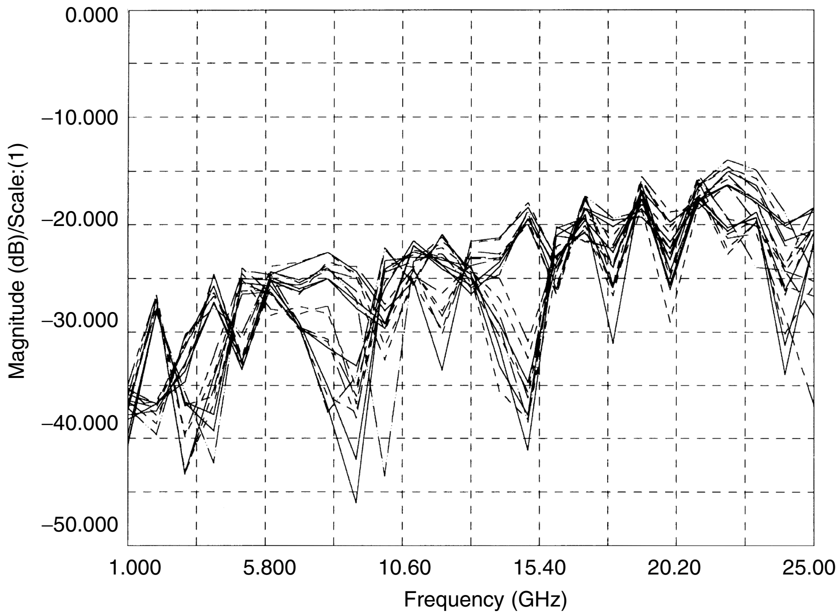


FIGURE 9.5 E_{lf} and E_{lr} error terms over a 5-month period.

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.

$$S_{21a} = \frac{\left(\left(S_{21m} - E_{xf}\right)/E_{tf}\right)\left(1 + \left(S_{22m} - E_{dr}\right)\left(E_{sr} - E_{lf}\right)/E_{rr}\right)}{\left(1 + \left(\left(S_{11m} - E_{df}\right)E_{sf}/E_{rf}\right)\right)\left(1 + \left(\left(S_{22m} - E_{dr}\right)E_{sr}/E_{rr}\right)\right) - \left(\left(S_{21m} - E_{xf}\right)\left(S_{12m} - E_{xr}\right)E_{lf}E_{lr}/E_{tf}E_{tr}\right)} \tag{9.5}$$

$$S_{21a} = \frac{S_{21m}\big|}{E_{tf}} \quad E_{lf}, E_{sf}, E_{sr}, E_{lr}, E_{df} \rightarrow 0 \tag{9.6}$$

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

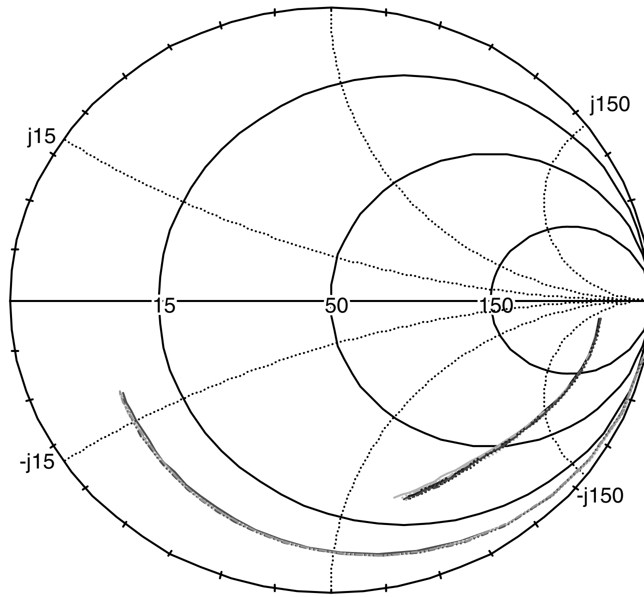


FIGURE 9.6 S_{11} and S_{22} of PCM FETs (mismatched) across a wafer.

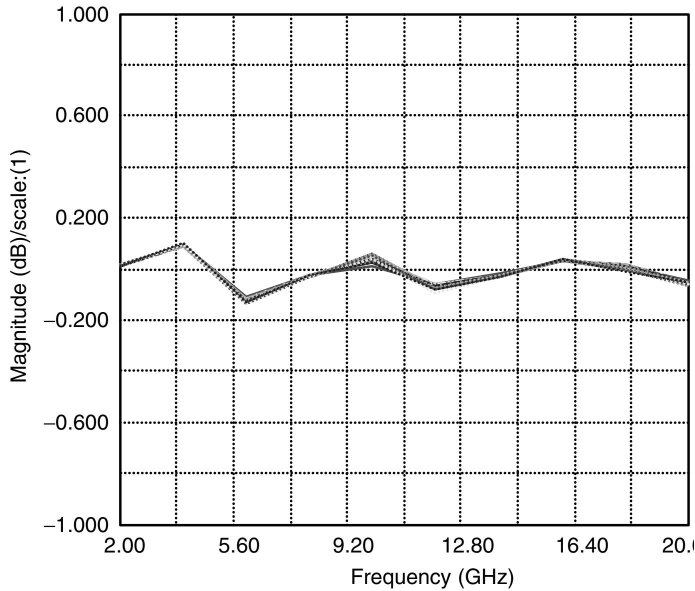


FIGURE 9.7 S_{21} vector to scalar measurement comparison for matched SPDT switch.

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.

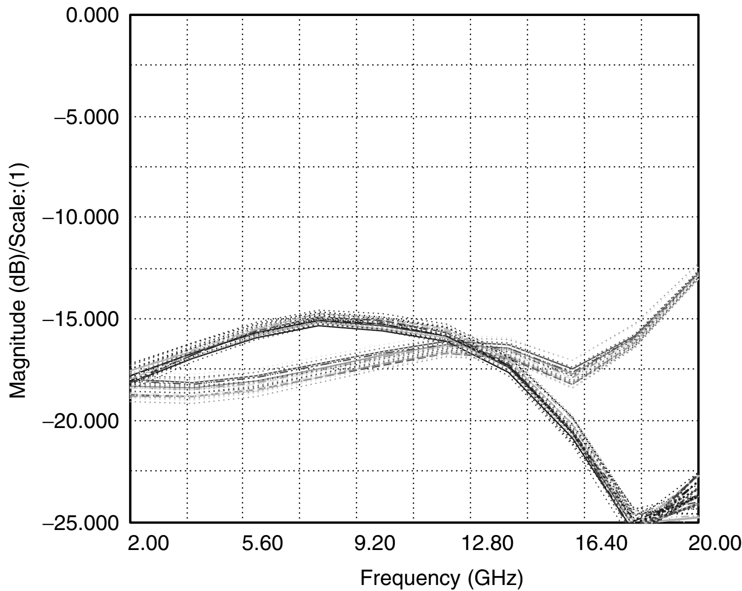


FIGURE 9.8 S_{11} and S_{22} of matched SPDT switch.

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

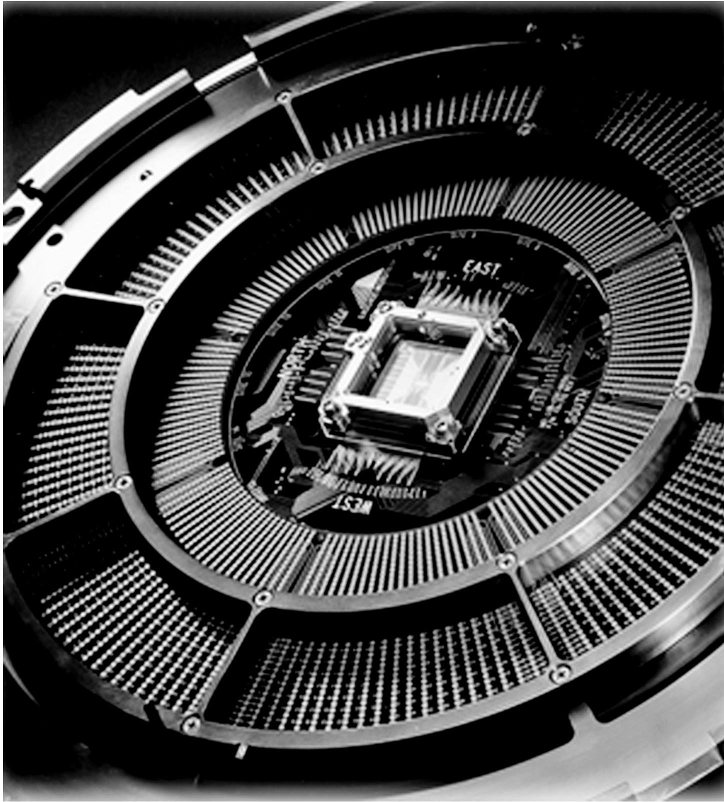


FIGURE 9.10 Ring-type RF probe card.

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

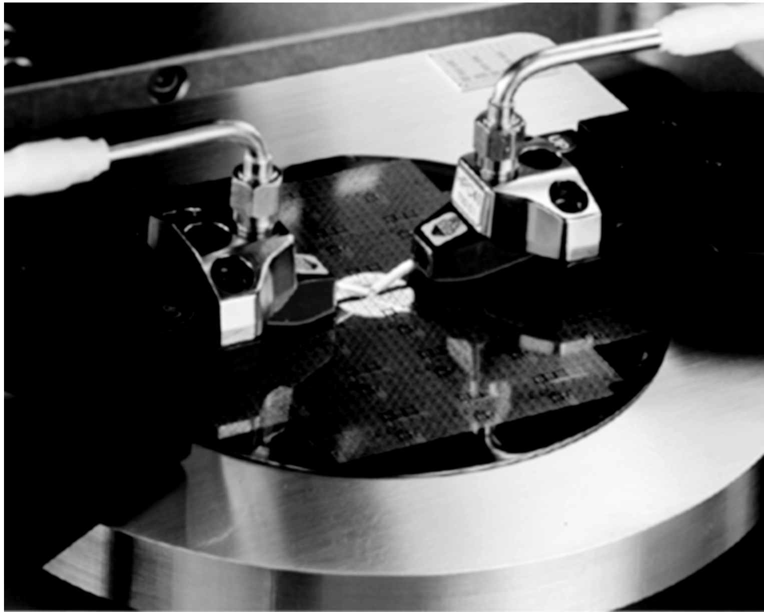


FIGURE 9.11 RF probes mounted on manipulators.

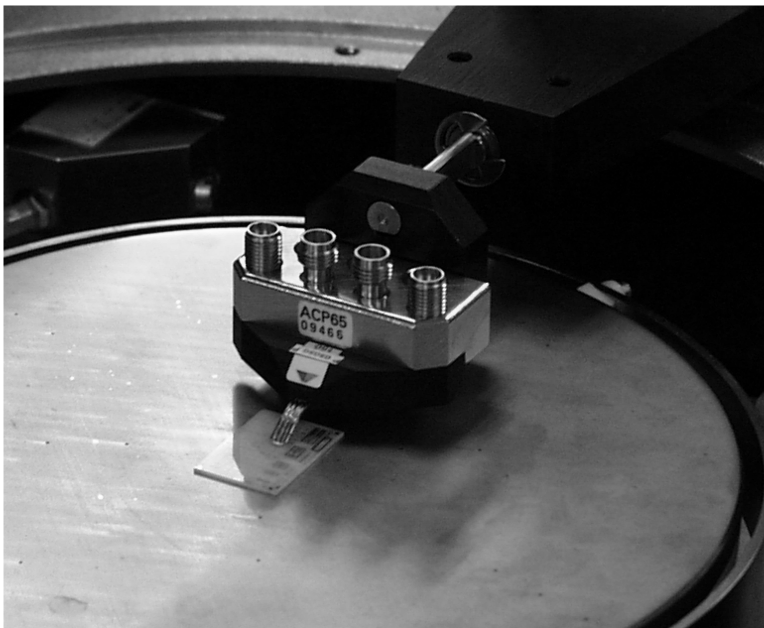


FIGURE 9.12 Four-part RF probe.

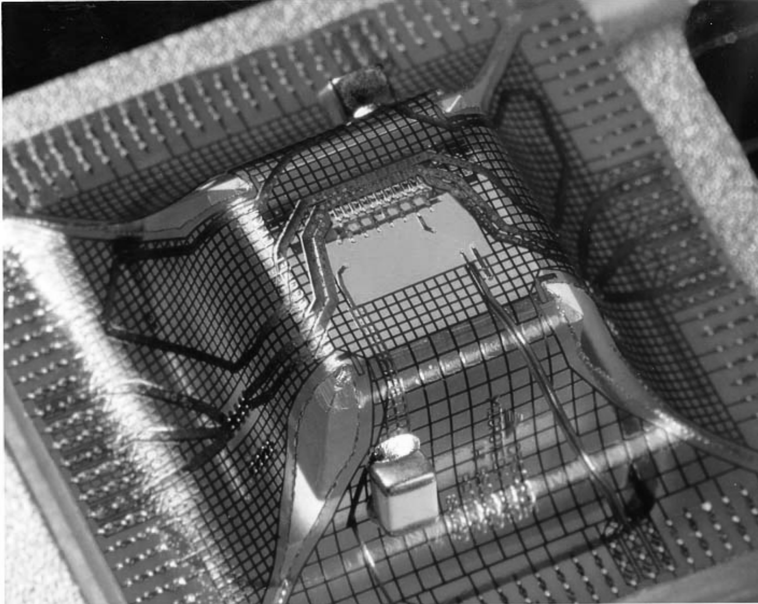


FIGURE 9.13 Cascade Microtech pyramid probe.

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

10.1 High Volume Microwave Component Needs 10-1
Cellular Phone Market Impact • High Volume RF Component Functions and Test Specifications • High Volume Test Success Factors

10.2 Test System Overview 10-4
Hardware: Rack and Stack vs. High Speed IC Testers • System Software Integration • RFIC Test Handlers • Contact Interface and Test Board

10.3 High Volume Test Challenges 10-8
Required Infrastructure • Accuracy and Repeatability Challenges • Volume and Cost Relationship • Product Mix Impact

10.4 Data Analysis Overview 10-13
Product Data Requirements and Database • Database Tools • Test Operation Data

10.5 Conclusion 10-17

References 10-17

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

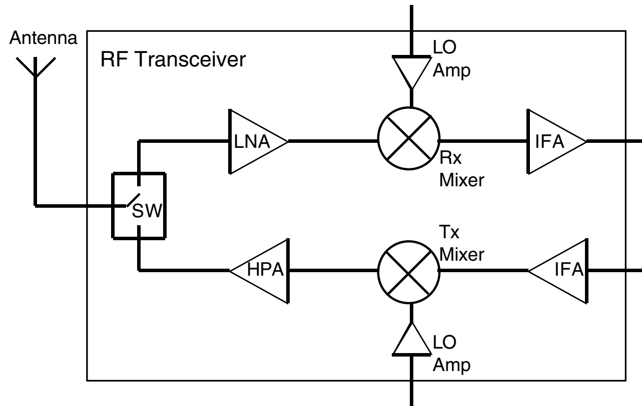


FIGURE 10.1 Typical RF transceiver building blocks.

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.

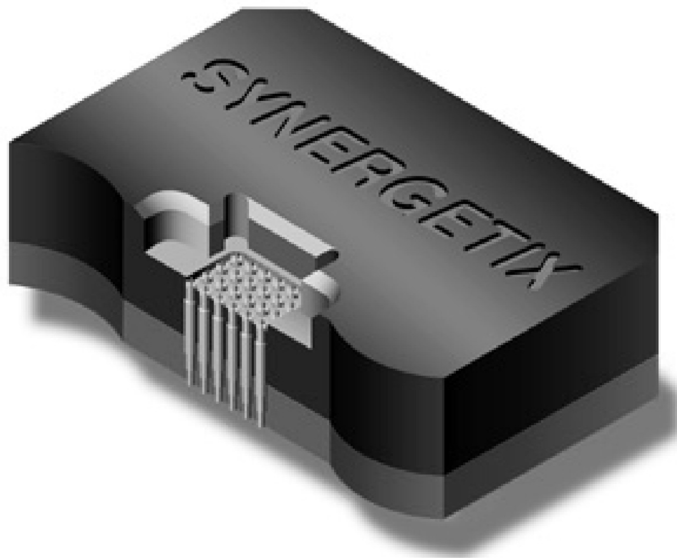


FIGURE 10.2 Pogo pin contactor.

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:

$$\sigma^2 \text{ gauge} = \sigma^2 \text{ repeatability} + \sigma^2 \text{ reproducibility}$$

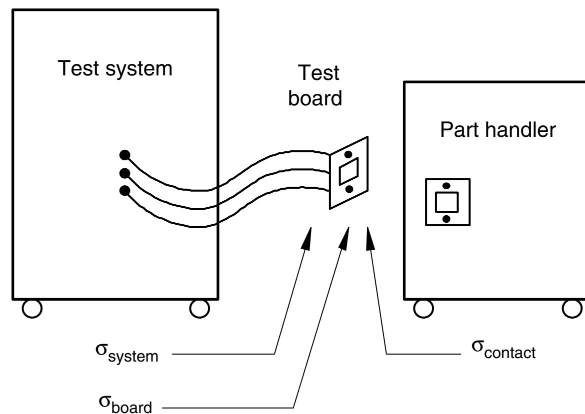


FIGURE 10.3 Sources of variability in an automated test setup.

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

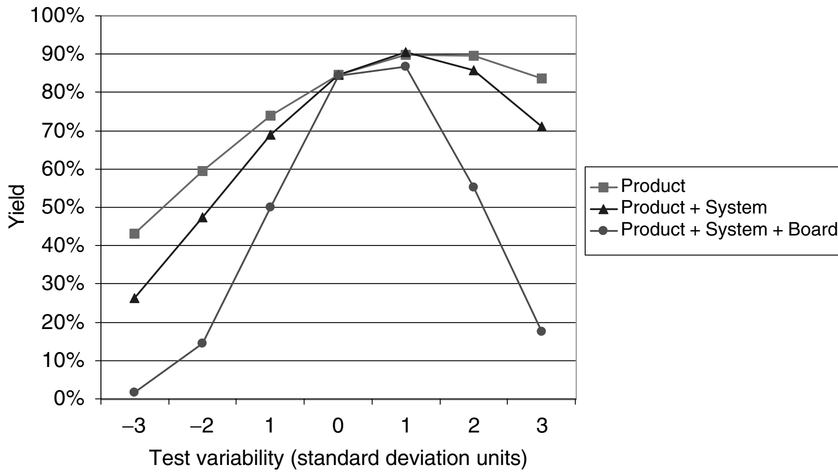


FIGURE 10.4 Yield versus variability for test system elements.

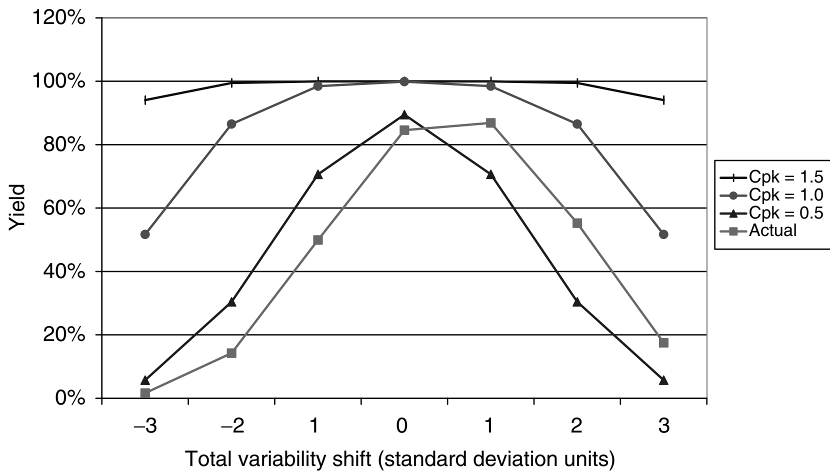


FIGURE 10.5 Yield versus variability as function of Cpk.

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:

$$Capacity = \frac{(Time\ Available)(Efficiency)}{Test\ Time + Handling\ Time} \tag{10.1}$$

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.

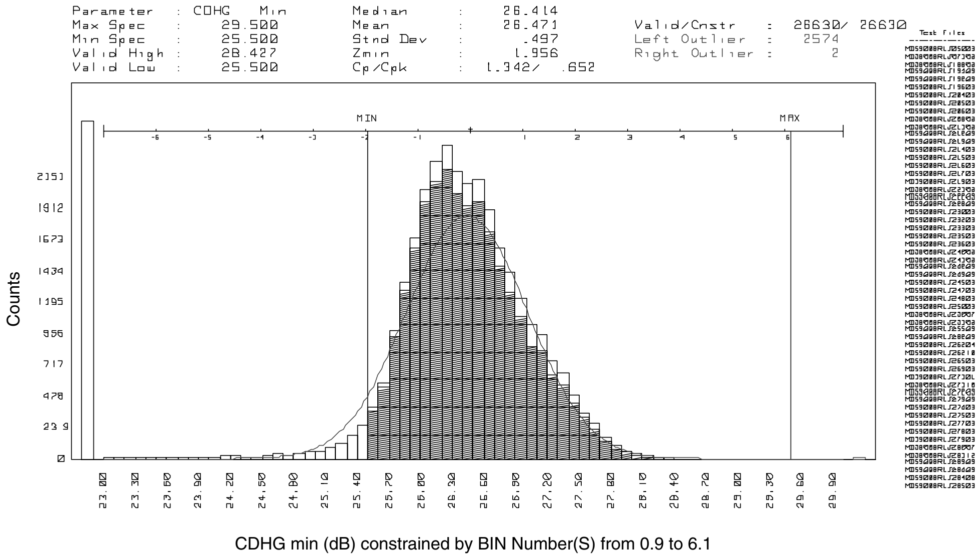


FIGURE 10.6 Histogram for distribution analysis.

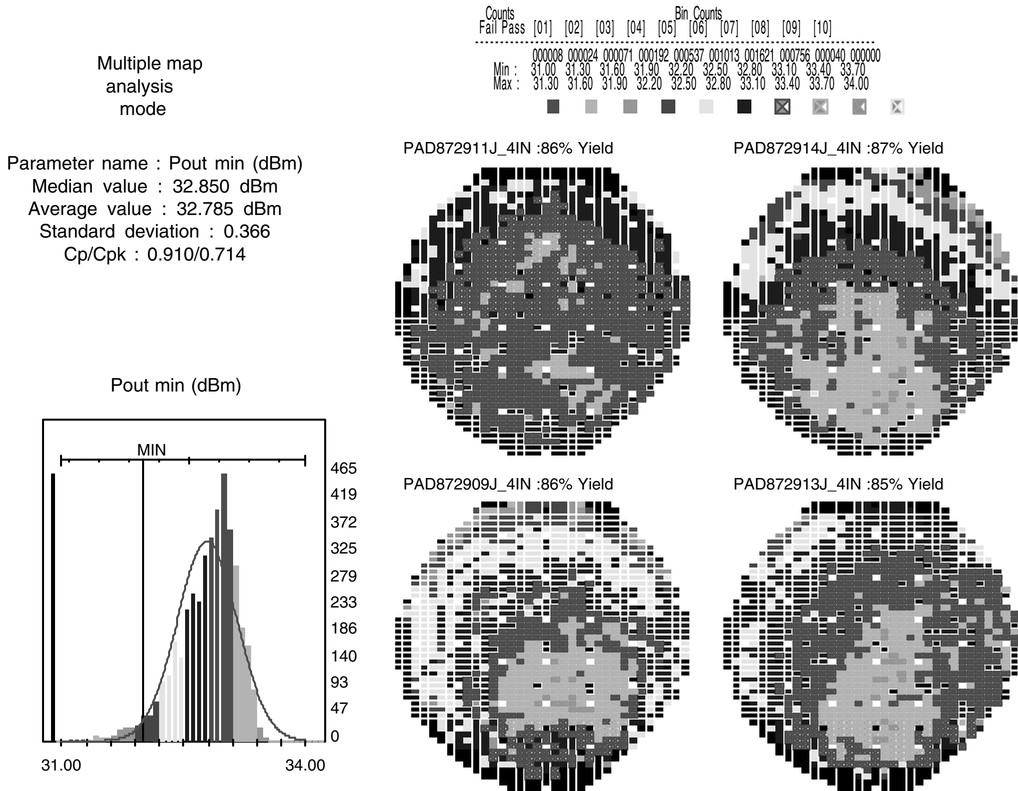


FIGURE 10.7 Wafer maps for yield pattern analysis.

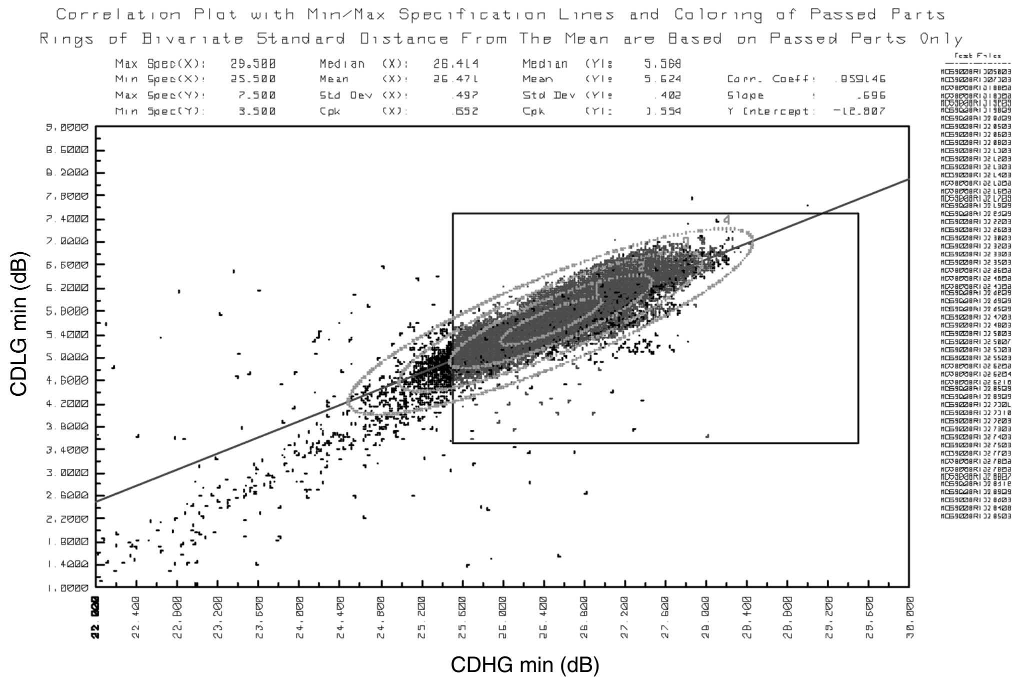


FIGURE 10.8 Scatter plot for parameter correlation analysis.

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

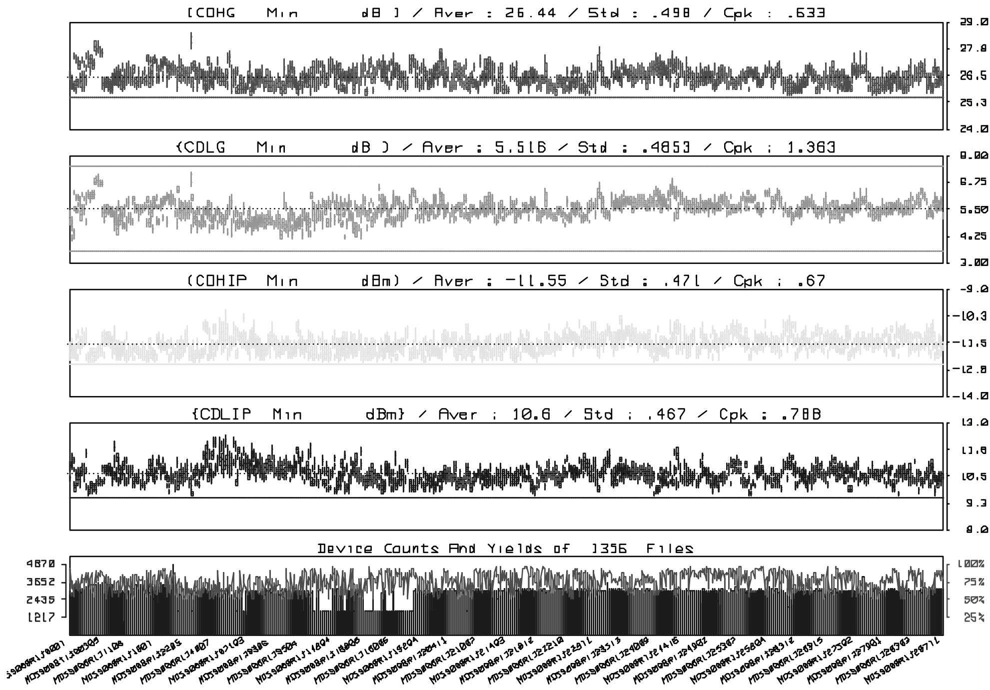


FIGURE 10.9 Multiple parameter control charts for product performance analysis.

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

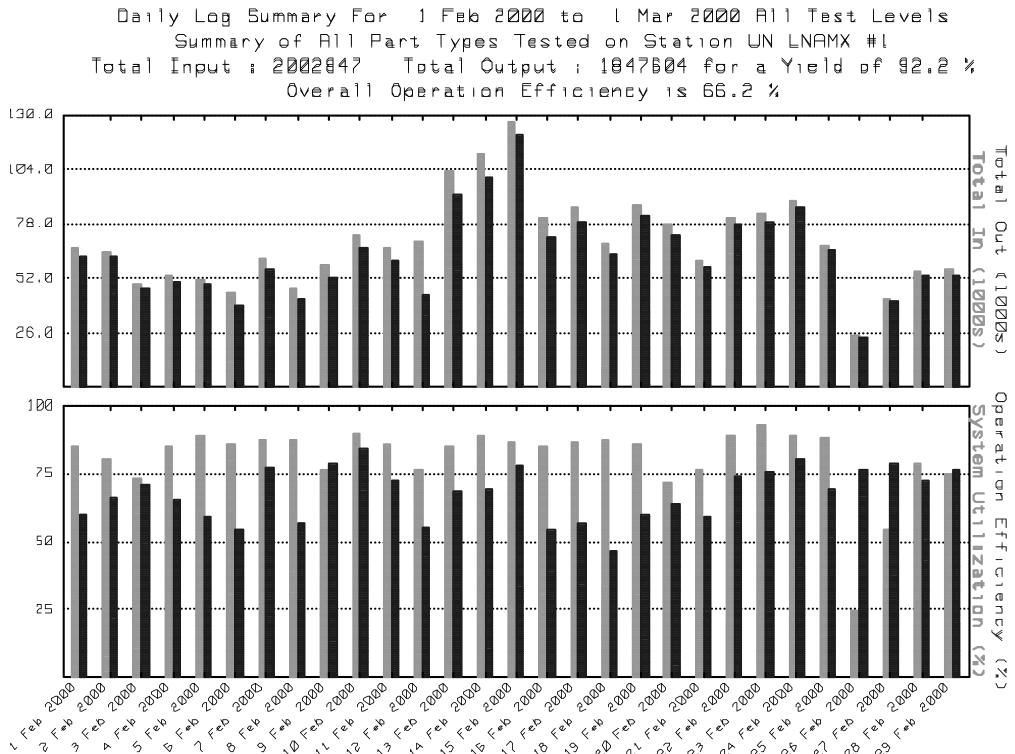


FIGURE 10.10 Yield and operation efficiency analysis tool.

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.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

11

Large Signal Network Analysis/Waveform Measurements

11.1 Introduction **11**-1
11.2 Mathematical Background **11**-2
11.3 Measurement Systems **11**-6
Sampling Down-Converter Based LSNA • Mixer Down-Converter Based LSNA • Comparison of Sampling and Mixer-Based LSNAs
11.4 Calibration **11**-9
s-Parameter Calibration • Power Calibration • Phase Calibration
11.5 Measurement Extensions **11**-13
Source-Pull, Load-Pull, and Active Injection • Multisine (Multitone) Signals • Hot *s*-Parameters
11.6 Summary **11**-14
References **11**-14

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]:

$$f(t) = \sum_{n=-\infty}^{\infty} C_n e^{jn\omega t}, \tag{11.1}$$

where

$$C_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) e^{-jn\theta} d\theta \tag{11.2}$$

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

$$f(t) = C_0 + \sum_{n=1}^{\infty} \text{Re}\left\{2C_n e^{jn\omega t}\right\}. \tag{11.3}$$

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

$$f(t) = \sum_{n=0}^{\infty} \text{Re}\left\{K_n e^{jn\omega t}\right\} = \text{Re}\left\{\sum_{n=0}^{\infty} K_n e^{jn\omega t}\right\}. \tag{11.5}$$

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,

$$f(t) = \text{Re}\left\{\sum_{n=0}^{N} K_n e^{j\omega_n t}\right\}, \tag{11.6}$$

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	$2\,\omega_c + 5\,\omega_m$