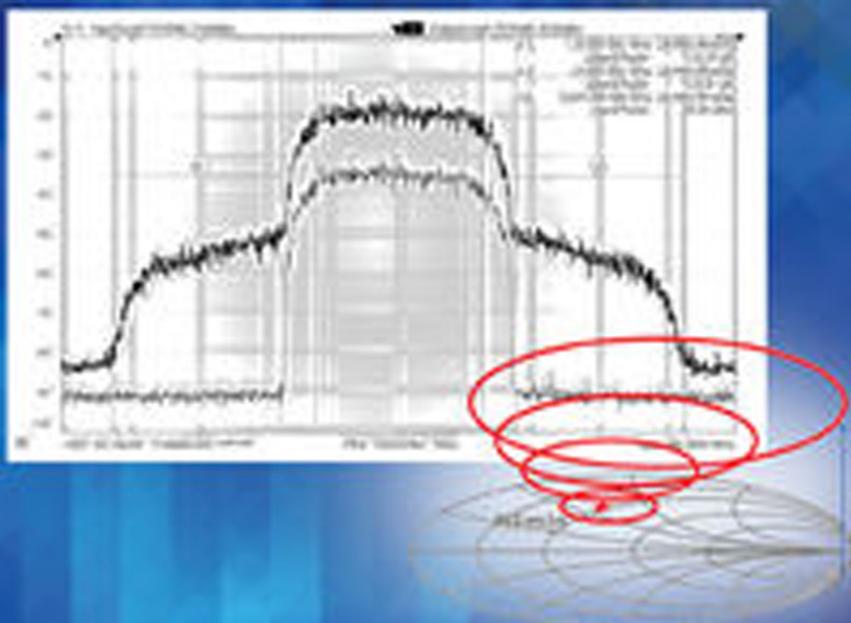


SECOND EDITION

HANDBOOK OF

MICROWAVE
COMPONENT
MEASUREMENTS

WITH ADVANCED VNA TECHNIQUES



JOEL P. DUNSMORE

WILEY

HANDBOOK OF MICROWAVE COMPONENT MEASUREMENTS

HANDBOOK OF MICROWAVE COMPONENT MEASUREMENTS WITH ADVANCED VNA TECHNIQUES

Second Edition

Joel P. Dunsmore

PhD., Sebastopol, CA, USA

WILEY

This edition first published 2020
© 2020 John Wiley & Sons Ltd

Edition History

John Wiley and Sons (1e, 2012)

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by law. Advice on how to obtain permission to reuse material from this title is available at <http://www.wiley.com/go/permissions>

The right of Joel P. Dunsmore to be identified as the author of this work has been asserted in accordance with law.

Registered Offices

John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA

John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial Office

The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Limit of Liability/Disclaimer of Warranty

MATLAB® is a trademark of The MathWorks, Inc. and is used with permission. The MathWorks does not warrant the accuracy of the text or exercises in this book. This work's use or discussion of MATLAB® software or related products does not constitute endorsement or sponsorship by The MathWorks of a particular pedagogical approach or particular use of the MATLAB® software.

While the publisher and authors have used their best efforts in preparing this work, they make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives, written sales materials or promotional statements for this work. The fact that an organization, website, or product is referred to in this work as a citation and/or potential source of further information does not mean that the publisher and authors endorse the information or services the organization, website, or product may provide or recommendations it may make. This work is sold with the understanding that the publisher is not engaged in rendering professional services. The advice and strategies contained herein may not be suitable for your situation. You should consult with a specialist where appropriate. Further, readers should be aware that websites listed in this work may have changed or disappeared between when this work was written and when it is read. Neither the publisher nor authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

Library of Congress Cataloging-in-Publication Data

Names: Dunsmore, Joel P., author.

Title: Handbook of microwave component measurements : with advanced VNA techniques / Joel P. Dunsmore.

Description: Second edition. | Hoboken, NJ : John Wiley & Sons, Inc.,

[2020] | Includes bibliographical references and index.

Identifiers: LCCN 2020003649 (print) | LCCN 2020003650 (ebook) | ISBN

9781119477136 (hardback) | ISBN 9781119477112 (adobe pdf) | ISBN 9781119477129 (epub)

Subjects: LCSH: Microwave devices—Testing.

Classification: LCC TK7876 .D84 2020 (print) | LCC TK7876 (ebook) | DDC 621.381/330287–dc23

LC record available at <https://lccn.loc.gov/2020003649>

LC ebook record available at <https://lccn.loc.gov/2020003650>

Cover design: Wiley

Cover image: Created by Joel Dunsmore, © Tuomas Lehtinen/Getty Images

Set in 10/12pt TimesLTStd by SPi Global, Chennai, India

Printed and bound by CPI Group (UK) Ltd, Croydon, CR0 4YY

To my dear wife Dana

Contents

Foreword to the Second Edition	xvii
Foreword to the First Edition	xix
Preface to the Second Edition	xxi
Preface to the First Edition	xxiii
Acknowledgments for the Second Edition	xxv
Acknowledgments from the First Edition	xxvii
1 Introduction to Microwave Measurements	1
1.1 Modern Measurement Process	2
1.2 A Practical Measurement Focus	3
1.3 Definition of Microwave Parameters	3
1.3.1 <i>S-Parameter Primer</i>	4
1.3.2 <i>Phase Response of Networks</i>	11
1.4 Power Parameters	13
1.4.1 <i>Incident and Reflected Power</i>	13
1.4.2 <i>Available Power</i>	13
1.4.3 <i>Delivered Power</i>	14
1.4.4 <i>Power Available from a Network</i>	14
1.4.5 <i>Available Gain</i>	15
1.5 Noise Figure and Noise Parameters	15
1.5.1 <i>Noise Temperature</i>	16
1.5.2 <i>Effective or Excess Input Noise Temperature</i>	17
1.5.3 <i>Excess Noise Power and Operating Temperature</i>	17
1.5.4 <i>Noise Power Density</i>	17
1.5.5 <i>Noise Parameters</i>	18
1.6 Distortion Parameters	19
1.6.1 <i>Harmonics</i>	19
1.6.2 <i>Second-Order Intercept</i>	19
1.6.3 <i>Two-Tone Intermodulation Distortion</i>	20
1.6.4 <i>Adjacent Channel Power and Adjacent Channel Level Ratio</i>	23
1.6.5 <i>Noise Power Ratio (NPR)</i>	24
1.6.6 <i>Error Vector Magnitude (EVM)</i>	25

1.7	Characteristics of Microwave Components	26
1.8	Passive Microwave Components	27
1.8.1	<i>Cables, Connectors, and Transmission Lines</i>	27
1.8.2	<i>Connectors</i>	31
1.8.3	<i>Non-coaxial Transmission Lines</i>	44
1.9	Filters	47
1.10	Directional Couplers	49
1.11	Circulators and Isolators	51
1.12	Antennas	52
1.13	PC Board Components	53
1.13.1	<i>SMT Resistors</i>	53
1.13.2	<i>SMT Capacitors</i>	56
1.13.3	<i>SMT Inductors</i>	57
1.13.4	<i>PC Board Vias</i>	57
1.14	Active Microwave Components	58
1.14.1	<i>Linear and Non-linear</i>	58
1.14.2	<i>Amplifiers: System, Low-Noise, High Power</i>	58
1.14.3	<i>Mixers and Frequency Converters</i>	59
1.14.4	<i>Frequency Multiplier and Limiters and Dividers</i>	61
1.14.5	<i>Oscillators</i>	62
1.15	Measurement Instrumentation	63
1.15.1	<i>Power Meters</i>	63
1.15.2	<i>Signal Sources</i>	64
1.15.3	<i>Spectrum Analyzers</i>	65
1.15.4	<i>Vector Signal Analyzers</i>	66
1.15.5	<i>Noise Figure Analyzers</i>	67
1.15.6	<i>Network Analyzers</i>	67
	References	70
2	VNA Measurement Systems	71
2.1	Introduction	71
2.2	VNA Block Diagrams	72
2.2.1	<i>VNA Source</i>	73
2.2.2	<i>Understanding Source-Match</i>	76
2.2.3	<i>VNA Test Set</i>	82
2.2.4	<i>Directional Devices</i>	85
2.2.5	<i>VNA Receivers</i>	91
2.2.6	<i>IF and Data Processing</i>	95
2.2.7	<i>Multiport VNAs</i>	97
2.2.8	<i>High-Power Test Systems</i>	104
2.2.9	<i>VNA with mm-Wave Extenders</i>	105
2.3	VNA Measurement of Linear Microwave Parameters	107
2.3.1	<i>Measurement Limitations of the VNA</i>	107
2.3.2	<i>Limitations Due to External Components</i>	111
2.4	Measurements Derived from S-Parameters	112
2.4.1	<i>The Smith Chart</i>	112
2.4.2	<i>Transforming S-Parameters to Other Impedances</i>	117

2.4.3	<i>Concatenating Circuits and T-Parameters</i>	118
2.5	Modeling Circuits Using Y and Z Conversion	120
2.5.1	<i>Reflection Conversion</i>	120
2.5.2	<i>Transmission Conversion</i>	120
2.6	Other Linear Parameters	121
2.6.1	<i>Z-Parameters, or Open-Circuit Impedance Parameters</i>	122
2.6.2	<i>Y-Parameters, or Short-Circuit Admittance Parameters</i>	123
2.6.3	<i>ABCD Parameters</i>	124
2.6.4	<i>H-Parameters or Hybrid Parameters</i>	125
2.6.5	<i>Complex Conversions and Non-equal Reference Impedances</i>	126
	References	126
3	Calibration and Vector Error Correction	127
3.1	Introduction	127
3.1.1	<i>Error Correction and Linear Measurement Methods for S-Parameters</i>	128
3.1.2	<i>Power Measurements with a VNA</i>	131
3.2	Basic Error Correction for S-Parameters: Cal-Application	134
3.2.1	<i>12-Term Error Model</i>	134
3.2.2	<i>1-Port Error Model</i>	136
3.2.3	<i>8-Term Error Model</i>	136
3.3	Determining Error Terms: Cal-Acquisition for 12-Term Models	139
3.3.1	<i>1-Port Error Terms</i>	139
3.3.2	<i>1-Port Standards</i>	141
3.3.3	<i>2-Port Error Terms</i>	148
3.3.4	<i>12-Term to 11-Term Error Model</i>	153
3.4	Determining Error Terms: Cal-Acquisition for 8-Term Models	153
3.4.1	<i>TRL Standards and Raw Measurements</i>	153
3.4.2	<i>Special Cases for TRL Calibration</i>	157
3.4.3	<i>Unknown Thru or SOLR (Reciprocal Thru Calibration)</i>	158
3.4.4	<i>Applications of Unknown Thru Calibrations</i>	159
3.4.5	<i>QSOLT Calibration</i>	161
3.4.6	<i>Electronic Calibration (ECal™) or Automatic Calibration</i>	162
3.5	Waveguide Calibrations	166
3.6	Calibration for Source Power	167
3.6.1	<i>Calibrating Source Power for Source Frequency Response</i>	168
3.6.2	<i>Calibration for Power Sensor Mismatch</i>	169
3.6.3	<i>Calibration for Source Power Linearity</i>	171
3.7	Calibration for Receiver Power	173
3.7.1	<i>Some Historical Perspective</i>	173
3.7.2	<i>Modern Receiver Power Calibration</i>	173
3.7.3	<i>Response Correction for the Transmission Test Receiver</i>	178
3.7.4	<i>Power Waves vs. Actual Waves</i>	181
3.8	Calibrating Multiple Channels Simultaneously: Cal All	182
3.9	Multiport Calibration Strategies	186
3.9.1	<i>N × 2-Port Calibrations: Switching Test Sets</i>	186
3.9.2	<i>N-port Calibration: True Multiport</i>	188

3.10	Automatic In-Situ Calibrations: CalPod	191
3.10.1	<i>CalPod Initialization and Recorrection</i>	192
3.10.2	<i>CalPod-as-Ecal</i>	194
3.11	Devolved Calibrations	194
3.11.1	<i>Response Calibrations</i>	195
3.11.2	<i>Enhanced Response Calibration</i>	196
3.12	Determining Residual Errors	199
3.12.1	<i>Reflection Errors</i>	199
3.12.2	<i>Using Airlines to Determine Residual Errors</i>	199
3.13	Computing Measurement Uncertainties	210
3.13.1	<i>Uncertainty in Reflection Measurements</i>	210
3.13.2	<i>Uncertainty in Source Power</i>	211
3.13.3	<i>Uncertainty in Measuring Power (Receiver Uncertainty)</i>	212
3.14	S21 or Transmission Uncertainty	212
3.14.1	<i>General Uncertainty Equation for S21</i>	214
3.14.2	<i>Dynamic Uncertainty Computation</i>	215
3.15	Errors in Phase	218
3.16	Practical Calibration Limitations	219
3.16.1	<i>Cable Flexure</i>	220
3.16.2	<i>Changing Power after Calibration</i>	221
3.16.3	<i>Compensating for Changes in Step Attenuators</i>	223
3.16.4	<i>Connector Repeatability</i>	225
3.16.5	<i>Noise Effects</i>	226
3.16.6	<i>Drift: Short-Term and Long-Term</i>	227
3.16.7	<i>Interpolation of Error Terms</i>	229
3.16.8	<i>Calibration Quality: Electronic vs. Mechanical Kits</i>	231
	Reference	232
4	Time-Domain Transforms	235
4.1	Introduction	235
4.2	The Fourier Transform	236
4.2.1	<i>The Continuous Fourier Transform</i>	236
4.2.2	<i>Even and Odd Functions and the Fourier Transform</i>	236
4.2.3	<i>Modulation (Shift) Theorem</i>	237
4.3	The Discrete Fourier Transform	238
4.3.1	<i>Fast Fourier Transform (FFT) and Inverse Fast Fourier Transform (IFFT)</i>	238
4.3.2	<i>Discrete Fourier Transforms</i>	240
4.4	Fourier Transform (Analytic) vs. VNA Time Domain Transform	240
4.4.1	<i>Defining the Fourier Transform</i>	241
4.4.2	<i>Effects of Discrete Sampling</i>	242
4.4.3	<i>Effects of Truncated Frequency</i>	244
4.4.4	<i>Windowing to Reduce Effects of Truncation</i>	246
4.4.5	<i>Scaling and Renormalization</i>	248

4.5	Low-Pass Transforms	248
4.5.1	<i>Low-Pass Impulse Mode</i>	248
4.5.2	<i>DC Extrapolation</i>	249
4.5.3	<i>Low-Pass Step Mode</i>	249
4.5.4	<i>Band-Pass Mode</i>	251
4.6	Time-Domain Gating	252
4.6.1	<i>Gating Loss and Renormalization</i>	253
4.7	Examples of Time-Domain Transforms of Various Networks	256
4.7.1	<i>Time-Domain Response of Changes in Line Impedance</i>	256
4.7.2	<i>Time-Domain Response of Discrete Discontinuities</i>	257
4.7.3	<i>Time-Domain Responses of Various Circuits</i>	257
4.8	The Effects of Masking and Gating on Measurement Accuracy	259
4.8.1	<i>Compensation for Changes in Line Impedance</i>	259
4.8.2	<i>Compensation for Discrete Discontinuities</i>	260
4.8.3	<i>Time-Domain Gating</i>	260
4.8.4	<i>Estimating an Uncertainty Due to Masking</i>	265
4.9	Time-Domain Transmission Using VNA	265
4.10	Conclusions	269
	References	269
5	Measuring Linear Passive Devices	271
5.1	Transmission Lines, Cables, and Connectors	271
5.1.1	<i>Calibration for Low Loss Devices with Connectors</i>	271
5.1.2	<i>Measuring Electrically Long Devices</i>	273
5.1.3	<i>Attenuation Measurements</i>	278
5.1.4	<i>Return Loss Measurements</i>	295
5.1.5	<i>Cable Length and Delay</i>	305
5.2	Filters and Filter Measurements	306
5.2.1	<i>Filter Classes and Difficulties</i>	306
5.2.2	<i>Duplexer and Diplexers</i>	307
5.2.3	<i>Measuring Tunable High-Performance Filters</i>	308
5.2.4	<i>Measuring Transmission Response</i>	310
5.2.5	<i>High Speed vs. Dynamic Range</i>	315
5.2.6	<i>Extremely High Dynamic Range Measurements</i>	317
5.2.7	<i>Calibration Considerations</i>	326
5.3	Multiport Devices	327
5.3.1	<i>Differential Cables and Lines</i>	328
5.3.2	<i>Couplers</i>	328
5.3.3	<i>Hybrids, Splitters, and Dividers</i>	331
5.3.4	<i>Circulators and Isolators</i>	334
5.4	Resonators	336
5.4.1	<i>Resonator Responses on a Smith Chart</i>	336
5.5	Antenna Measurements	338
5.6	Conclusions	340
	References	341

6	Measuring Amplifiers	343
6.1	Amplifiers as a Linear Devices	343
6.1.1	<i>Pretesting an Amplifier</i>	344
6.1.2	<i>Optimizing VNA Settings for Calibration</i>	346
6.1.3	<i>Calibration for Amplifier Measurements</i>	347
6.1.4	<i>Amplifier Measurements</i>	351
6.1.5	<i>Analysis of Amplifier Measurements</i>	357
6.1.6	<i>Saving Amplifier Measurement Results</i>	367
6.2	Gain Compression Measurements	372
6.2.1	<i>Compression Definitions</i>	372
6.2.2	<i>AM-to-PM or Phase Compression</i>	376
6.2.3	<i>Swept Frequency Gain and Phase Compression</i>	377
6.2.4	<i>Gain Compression Application, Smart Sweep, and Safe-Sweep Mode</i>	378
6.3	Measuring High-Gain Amplifiers	384
6.3.1	<i>Setup for High-Gain Amplifiers</i>	386
6.3.2	<i>Calibration Considerations</i>	386
6.4	Measuring High-Power Amplifiers	389
6.4.1	<i>Configurations for Generating High Drive Power</i>	389
6.4.2	<i>Configurations for Receiving High-Power</i>	391
6.4.3	<i>Power Calibration and Pre/Post Leveling</i>	393
6.5	Making Pulsed-RF Measurements	394
6.5.1	<i>Wideband vs. Narrowband Measurements</i>	395
6.5.2	<i>Pulse Profile Measurements</i>	398
6.5.3	<i>Pulse-to-Pulse Measurements</i>	401
6.5.4	<i>DC Measurements for Pulsed RF Stimulus</i>	401
6.6	Distortion Measurements	403
6.6.1	<i>Harmonic Measurements on Amplifiers</i>	404
6.7	Measuring Doherty Amplifiers	410
6.8	X-Parameters, Load-Pull Measurements, Active Loads, and Hot S-Parameters	413
6.8.1	<i>Non-linear Responses and X-Parameters</i>	414
6.8.2	<i>Load-Pull, Source-Pull, and Load Contours</i>	417
6.8.3	<i>Hot S-Parameters and True Hot-S₂₂</i>	421
6.9	Conclusions on Amplifier Measurements	433
	References	434
7	Mixer and Frequency Converter Measurements	435
7.1	Mixer Characteristics	435
7.1.1	<i>Small Signal Model of Mixers</i>	438
7.1.2	<i>Reciprocity in Mixers</i>	442
7.1.3	<i>Scalar and Vector Responses</i>	444
7.2	Mixers vs. Frequency Converters	445
7.2.1	<i>Frequency Converter Design</i>	446
7.2.2	<i>Multiple Conversions and Spur Avoidance</i>	446
7.3	Mixers as a 12-Port Device	448
7.3.1	<i>Mixer Conversion Terms</i>	448

7.4	Mixer Measurements: Frequency Response	451
7.4.1	<i>Introduction</i>	451
7.4.2	<i>Amplitude Response</i>	452
7.4.3	<i>Phase Response</i>	456
7.4.4	<i>Group Delay and Modulation Methods</i>	466
7.4.5	<i>Swept LO Measurements</i>	469
7.5	Calibration for Mixer Measurements	476
7.5.1	<i>Calibrating for Power</i>	476
7.5.2	<i>Calibrating for Phase</i>	479
7.5.3	<i>Determining the Phase and Delay of a Reciprocal Calibration Mixer</i>	482
7.6	Mixers Measurements vs. Drive Power	493
7.6.1	<i>Mixer Measurements vs. LO Drive</i>	493
7.6.2	<i>Mixer Measurements vs. RF Drive Level</i>	497
7.7	TOI and Mixers	501
7.7.1	<i>IMD vs. LO Drive Power</i>	502
7.7.2	<i>IMD vs. RF Power</i>	502
7.7.3	<i>IMD vs. Frequency Response</i>	505
7.8	Noise Figure in Mixers and Converters	507
7.9	Special Cases	507
7.9.1	<i>Mixers with RF or LO Multipliers</i>	507
7.9.2	<i>Segmented Sweeps</i>	509
7.9.3	<i>Measuring Higher-Order Products</i>	509
7.9.4	<i>Mixers with an Embedded LO</i>	515
7.9.5	<i>High-Gain and High-Power Converters</i>	517
7.10	I/Q Converters and Modulators	518
7.11	Conclusions on Mixer Measurements	530
	References	531
8	Spectrum Analysis: Distortion and Modulation Measurements	533
8.1	Spectrum Analysis in Vector Network Analyzers	534
8.1.1	<i>Spectrum Analysis Fundamentals</i>	534
8.1.2	<i>SA Block Diagrams: Image Rejection: Hardware vs. Software</i>	539
8.1.3	<i>Attributes of Repetitive Signals and Spectrum Measurements</i>	546
8.1.4	<i>Coherent Spectrum Analysis</i>	559
8.1.5	<i>Calibration of SA Results</i>	568
8.1.6	<i>Two-Tone Measurements, IMD, and TOI Definition</i>	571
8.1.7	<i>Measurement Techniques for Two-Tone TOI</i>	574
8.1.8	<i>Swept IMD</i>	576
8.1.9	<i>Optimizing Results</i>	579
8.1.10	<i>Error Correction</i>	582
8.2	Distortion Measurement of Complex Modulated Signals	583
8.2.1	<i>Adjacent Power Measurements</i>	584
8.2.2	<i>Noise Power Ratio (NPR) Measurements</i>	587
8.2.3	<i>NPR Signal Quality and Correction</i>	592
8.2.4	<i>EVM Derived from Distortion Measurements</i>	596

8.3	Measurements of Spurious Signals with VNA Spectrum Analyzer	605
8.3.1	<i>Spurious at Predictable Frequencies</i>	605
8.3.2	<i>Multiport Mixer Spurious Measurements</i>	607
8.3.3	<i>Spurious Oscillations</i>	608
8.4	Measurements of Pulsed Signals and Time-Gated Spectrum Analysis	611
8.4.1	<i>Understanding Pulsed Spectrum</i>	611
8.4.2	<i>Time-Gated Spectrum Analysis</i>	612
8.5	Summary	615
	Reference	615
9	Measuring Noise Figure and Noise Power	617
9.1	Noise-Figure Measurements for Amplifiers	617
9.1.1	<i>Definition of Noise Figure</i>	618
9.1.2	<i>Noise-Power Measurements</i>	619
9.1.3	<i>Computing Noise Figure from Noise Powers</i>	623
9.1.4	<i>Computing DUT Noise Figure from Y-Factor Measurements</i>	624
9.1.5	<i>Cold-Source Methods</i>	626
9.1.6	<i>Noise Parameters</i>	628
9.1.7	<i>Noise Parameter Measurement Results</i>	634
9.1.8	<i>Error Correction in Noise Figure Measurements</i>	637
9.2	Active Antenna Noise-Figure Measurements (G/T)	638
9.3	Noise Figure in Mixers and Converters	642
9.3.1	<i>Y-Factor Measurements on Mixers</i>	642
9.3.2	<i>Cold-Source Measurements on Mixers</i>	644
9.4	Other Noise-Related Measurements	650
9.4.1	<i>Noise Power Measurements with a VNA Spectrum Analyzer</i>	650
9.4.2	<i>Noise-Power Measurements</i>	650
9.4.3	<i>Noise Figure Measurements Using Spectrum Analysis</i>	653
9.4.4	<i>Carrier-to-Noise Measurements</i>	654
9.5	Uncertainty, Verification, and Improvement of Noise-Figure Measurements	655
9.5.1	<i>Uncertainty of Noise-Figure Measurements</i>	655
9.5.2	<i>Existing Methodologies</i>	656
9.5.3	<i>Techniques for Improving Noise-Figure Measurements</i>	665
9.6	Summary: Noise and Noise-Figure Measurements	668
	References	668
10	VNA Balanced Measurements	669
10.1	Differential and Balanced S-Parameters	669
10.2	3-Port Balanced Devices	674
10.3	Measurement Examples for Mixed-Mode Devices	675
10.3.1	<i>Passive Differential Devices: Balanced Transmission Lines</i>	675

10.3.2	<i>Differential Amplifier Measurements</i>	680
10.3.3	<i>Differential Amplifiers and Non-linear Operation</i>	682
10.4	True-Mode VNA for Non-linear Testing	689
10.4.1	<i>True-Mode Instruments</i>	689
10.4.2	<i>True-Mode Measurements</i>	692
10.4.3	<i>Determining the Phase Skew of a Differential Device</i>	698
10.4.4	<i>Differential Harmonic Measurements</i>	700
10.5	Differential Testing Using Baluns, Hybrids, and Transformers	708
10.5.1	<i>Transformers vs. Hybrids</i>	708
10.5.2	<i>Using Hybrids and Baluns with a 2-Port VNA</i>	711
10.6	Distortion Measurements of Differential Devices	714
10.6.1	<i>Comparing Single-Ended IMD Measurement to True-Mode Measurements</i>	715
10.6.2	<i>Differential IMD without Baluns</i>	718
10.7	Noise Figure Measurements on Differential Devices	723
10.7.2	<i>Measurement Setup</i>	725
10.8	Conclusions on Differential Device Measurement	731
	References	732
11	Advanced Measurement Techniques	733
11.1	Creating Your Own Cal-Kits	733
11.1.1	<i>PC Board Example</i>	734
11.1.2	<i>Evaluating PC Board Fixtures</i>	735
11.2	Fixturing and De-embedding	750
11.2.1	<i>De-embedding Mathematics</i>	751
11.3	Determining S-Parameters for Fixtures	753
11.3.1	<i>Fixture Characterization Using 1-Port Calibrations</i>	753
11.4	Automatic Port Extensions (APE)	759
11.5	AFR: Fixture Removal Using Time Domain	764
11.5.1	<i>2-Port AFR</i>	764
11.5.2	<i>Fixture-Enhanced AFR</i>	768
11.5.3	<i>1-Port AFR</i>	770
11.6	Embedding Port-Matching Elements	772
11.7	Impedance Transformations	774
11.8	De-embedding High-Loss Devices	775
11.9	Understanding System Stability	778
11.9.1	<i>Determining Cable Transmission Stability</i>	778
11.9.2	<i>Determining Cable Mismatch Stability</i>	778
11.9.3	<i>Reflection Tracking Stability</i>	781
11.10	Some Final Comments on Advanced Techniques and Measurements	782
	References	783

Appendix A Physical Constants	785
Appendix B Common RF and Microwave Connectors	787
Appendix C Common Waveguides	789
Appendix D Some Definitions for Calibration Kit Opens and Shorts	791
Appendix E Frequency, Wavelength, and Period	795
Index	797

Foreword to the Second Edition

Our world today is perpetually changing due to advancements in technology. As we begin the new decade, we are on the verge of large-scale 5G deployments that will enable richer services across more devices while driving higher performance across the entire network, from the device to the edge to the core. Automotive capabilities are continuing the march toward progressively higher levels of autonomy. The number of connected devices being deployed in our homes, businesses, and cities, including devices we carry or wear, is growing exponentially. These are just a few of the changes that technology is enabling, and they will lead to further changes and innovations. The pace of change and the associated level of technological complexity are only accelerating. As has long been true with technology innovation, practical advancements in technology are possible only with corresponding advancements in measurement. Measurement helps confirm the creation of a new technology in the first place and later helps in applying and disseminating the technology with the required levels of performance, quality, and cost.

The measurement challenges with today's new technologies increasingly require total solutions. Traditional hardware measurement instruments are no longer enough. Of course, state-of-the-art hardware is still required, but significant software capabilities that build on the core capabilities of the hardware are equally vital. Complex calibrations, advanced measurement algorithms, and sophisticated data analysis and visualization allow the raw measurement data to be turned into useful information that yields actionable insights. This is especially the case with microwave measurements, including measurements of microwave components based on vector network analyzer solutions.

It's also the case that no one measurement solution fits all. Different form factors are preferred for different use cases. Benchtop instruments are often used in laboratory or product development settings. Modular solutions, with their density and configuration flexibility, can be appropriate for manufacturing. Portable, handheld solutions are required in field environments. Still other formats are useful for experimental or educational purposes. Sometimes the software is best executed within the measurement instrument, and sometimes it is preferred that it run elsewhere, on a desktop computer or even on computing resources located in the cloud.

In this second edition of his comprehensive book, Dr. Joel Dunsmore has refined the already strong original presentation of the foundational concepts of microwave component measurements, while adding important new material on modulated measurements and on the use of modular instruments. Other additions cover the increased importance of calibration and metrology in the context of 5G measurement challenges. Throughout, the value of understanding the measurement science – how the calculations are made and why – is conveyed in

a compelling way. Dr. Dunsmore can do all this effectively because of his now nearly four decades of experience as an innovative practitioner and educator, making advanced measurement ideas real through new products and solutions and sharing those ideas with others. This book is the latest example of applying his talents to help move forward the technologies that connect and secure today's perpetually changing world.

Ron Nersesian

Chairman and CEO, Keysight Technologies

Foreword to the First Edition

The electronics industry has undergone revolutionary changes in the past 20 years. System performance has significantly advanced, physical size of hardware has shrunk, quality and reliability have greatly improved and manufacturing costs have dramatically decreased. Underlying these advances has been the phenomenal growth in RF test and measurement capability. Modern-day RF test equipment has progressed to the point where it is not uncommon to measure signals below -100 dBm at milliseconds speed. Even more astounding is the ability to marry RF test capability with analysis software whereby test equipment can produce linear and non-linear models of the device under test to significantly improve the life of the design engineer, using this capability.

RF and microwave components have played an important role in this revolutionary change. Component size has shrunk, parasitics have been reduced, quality standards have greatly improved and costs have reduced ten-fold. At the same time, test fixtures and interconnects have improved to enable a higher level of precision during characterization and production measurement. In parallel with these advances, test equipment has improved to an extent where there has been a revolution in the capabilities to make precise and fast measurements of RF and microwave components. The success of a manufacturer of RF and microwave components is directly linked to the quality and capability of measuring component performance during the design, qualification and production phase of the product life cycle. From a practical point of view, the testing must be fast (1–2 seconds), the accuracy very precise (hundredths of a dB), with a high degree of repeatability. Each phase of the life cycle imposes its unique requirements for measurement accuracy and data collection.

During the design phase, full characterization of performance, including amplitude and phase, is a must in order to establish a reference for future production runs. So it becomes a necessary requirement to characterize and de-embed the test setup and test fixtures to isolate actual device performance. Fortunately, the modern vector network analyzer provides support in this regard. Consequently, the performance data obtained during the design phase becomes the gold standard for evaluating statistical variation obtained from future production lots; and these lots are accepted or rejected based on the statistical results, with sigma generally being the statistic most closely watched by the test or QA engineer to evaluate production lot acceptance.

When published specs are provided for a component, it is very important to understand that these specifications are simply markers which allow for a first impression or summary review of the component performance. However, only when performance data and performance graphs are provided for all parameters, including both amplitude and phase, would the “real” performance of the component be known. Furthermore, in characterizing components for use

in customer evaluations, it is important to provide this data both within and outside the specified bandwidth. In non-linear components such as a frequency mixer, higher-order harmonics of the RF and LO signals are generated, and depending on the load impedance outside the specified frequency range, these higher-order harmonics can get reflected back into the mixer, causing an interaction between the desired signals and the unwanted harmonics. Fortunately, modern-day analyzers can make these harmonic measurements relatively quickly and easy.

Dr. Dunsmore has captured the essence of modern-day measurements. He provides a practical understanding of measurement capabilities and limitations. He provides a means for the test engineer to not only make measurements, but also to understand test concepts, anticipate measurement results and learn how to isolate and characterize the performance of the DUT, independent of potential errors inherent in the test environment. I am confident that this book will serve as a reference for understanding measurement methods, test block diagrams and measurement limitations so that correlation between the manufacturer and user would take place by using a common reference. This book has the potential to be an invaluable source to further the progress of the RF and microwave world.

Harvey Kaylie
President and Founder of Mini-Circuits

Preface to the Second Edition

It has been seven years since the first edition of this book, and while the fundamentals of RF and microwave component measurement (and thus the fundamental chapters) are largely unchanged, the advent of new methods and technologies has greatly increased the capabilities of the vector network analyzers, furthering the trend to fully integrated test solutions for complete component characterization.

Chapter 1 has been updated with better measurements of high-frequency connectors (3.5, 2.4, 1.85, and 1 mm) showing their response and moding characteristics, as well as adding some discussion on new modulation measurements such as noise power ratio (NPR) and adjacent channel power ratio (ACPR).

Chapter 2 adds new material on multiport network analyzers with very large port counts and new broadband mm-wave network analyzers.

Chapter 3 has been somewhat simplified with respect to details of power calibration (deleting some obsolete methods). Added were details on in-situ calibration modules (CalPods), a new concept of calibrating multiple channels and measurements with a single calibration (i.e. cal-all), and a new discussion of real-time uncertainty capability.

Chapter 4's theoretical material is unchanged (the Fourier transform remains as it was), but a new section on simulated time-domain reflectometer (TDR) measurements including eye diagrams has been added.

Chapter 5 is essentially unchanged, with just some minor updates.

Chapter 6 keeps the basics of amplifier test unchanged but adds several new topics including improved harmonic measurement methods, a method for testing dual input Doherty-type amplifiers, an updated discussion of X-parameters, and a new discussion of active or Hot S-parameters. The topic of distortion measurements (2-tone intermodulation distortion [IMD]) has moved to the new Chapter 8. The discussion of noise figure measurements has been moved to its own chapter, the new Chapter 9.

Chapter 7 has additions in the area of multichannel mixer test with a new method for measuring phase difference between multichannel mixers, as well as a new method for absolute phase measurements of a mixer. Also added is a new method for swept higher-order products measurements. An entirely new section on measuring I/Q mixers has been added. As with amplifiers in Chapter 6, the topic of noise figure measurements of mixers has been consolidated in the new Chapter 9.

Chapter 8 is almost entirely new and adds the new topic of measuring components with modulated signals to the realm of modern VNA testing. First, I give detailed discussion of how VNAs can perform spectrum analysis and details of modulated signal characteristics particularly related to when they are generated as a waveform played back on an arbitrary waveform

generator. Then new methods for extremely accurate power measurements on modulated signals are introduced. Methods for measuring distortion in the form of adjacent channel power is covered, along with noise power ratio. A new method for using a VNA to directly measure error vector magnitude (EVM) on power amplifiers is introduced. Finally, new methods for pulsed spectrum measurements are discussed. This chapter contains a lot of detail on spectrum analysis that was previously not well known and fully explains the differences one sees in the appearance of a spectrum when applying different measurement and detection methods.

Chapter 9 collects the topics of noise figure measurements for both amplifiers and frequency converters into a single chapter. New material on noise parameter measurements is included, as well as a new section on measuring the noise figure of an active antenna. Further, there is a detailed discussion on a new method for noise-figure verification.

Chapter 10 now contains all the information on differential measurements. New material on measuring the differential harmonics of differential amplifiers has been added, as well as making IMD measurements without using baluns. A new technique is shown for determining the phase skew of a differential amplifier.

Chapter 11 contains the material on advanced techniques for fixturing and creating calibration kits, with various other topics (previously Chapter 9). This contains new material on the one-port (open only) automatic fixture removal (AFR) method.

Several of these new capabilities are introduced that dramatically change the way components may be tested. These include the integration of full spectrum analysis capabilities in a VNA, and the ability to synchronously control the frequency, power, and phase of multiple sources while measuring at multiple different frequencies. These capabilities are generally implemented in the form a purpose-built application function. While details of how these functions operate can vary between instrument manufacturers (and frankly, between different versions of firmware from a single manufacturer), the key concepts are explained to allow the reader to fully understand the benefits of these new methods.

I would like to thank my colleagues from Keysight Technologies for their assistance in reviewing this material and for their help in developing the new material. As always, any mistakes are mine alone.

July 2019

Joel P. Dunsmore
Sebastopol, CA

Preface to the First Edition

This book is a bit of mixture between basic and advanced, and between theoretical and practical. Unfortunately, the dividing lines are not particularly clear and depend considerably upon the training and experience of the reader. While primarily a text about measurements techniques, there is considerable information about device attributes that will be useful to both a designer and a test engineer, as one purpose of device-test is to ascertain attributes of devices that do not follow the simplified models commonly associated with these devices. In practice, it is the unexpected responses that consume a majority of the time spent in test and troubleshooting designs, particularly related to active devices such as amplifiers and mixers.

The principle instrument for testing microwave components is the vector network analyzer (VNA), and recent advances have increased the test capabilities of this instrument to cover far more than simple gain and match measurements. As a designer of VNAs for more than 30 years, I have been involved in consulting on the widest range of microwave test needs from cell phone components to satellite multiplexers. The genesis and goal of this book are to provide to the reader a distillation of that experience to improve the quality and efficiency of the R&D and production test engineer. The focus is on modern test methods; the best practices have changed with changing instrument capability and occasionally the difference between legacy methods and new techniques are sufficiently great as to be particularly highlighted.

Chapter 1 is intended as an introduction to microwave theory and microwave components. The first half introduces characterization concepts common to RF and microwave work. Some important mathematical results are presented that are useful in understanding the results of subsequent chapters. The second half of Chapter 1 introduces some common microwave connectors, transmission lines, and components, as well as some discussion of the basic microwave test instrumentation. This chapter is especially useful to engineers new to RF and microwave testing.

Chapter 2 provides a detailed look into the composition of common VNA designs along with their limitations. While this level of detail is not normally needed by the casual user, test engineers trying to understand measurement results at a very precise level will find it useful to understand overall results are affected by VNA test configuration. While the modern VNA can make a wide range of measurements, including distortion, power, and noise figure measurements, still the principal use is in measuring S-parameters. The second half of Chapter 2 illustrates many useful parameters derived from basic S-parameters.

Perhaps the most arcane aspect of using VNAs for test is the calibration and error correction process. Chapter 3 is a comprehensive discussion of the error models for VNAs, calibration methods, uncertainty analysis, and evaluation of calibration residuals. This chapter also introduces the idea of source and receiver power calibrations, about which, excluding this

book, very little formal information is currently available. The chapter concludes with many practical aspects of VNAs that affect the quality of calibrated measurements.

Chapter 4 may be the most mathematically rigorous, covering the very useful topic of time-domain transforms used in VNAs. The topic of gating, its effects, and compensation methods is examined in particular. These first four chapters comprise the introductory material to microwave component measurements.

The remaining chapters are focused on describing particular cases for microwave component measurements. Chapter 5 is devoted to passive microwave components such as cables and connectors, transmission lines, filters, isolators, and couplers. Best practices and methods for dealing with common problems are discussed for each component.

Chapter 6 is all about amplifier measurements and provides the understanding needed for complete characterization. In particular, difficulties with measuring high-gain and high-power amplifiers are discussed, including pulsed RF measurements. Non-linear measurements such as harmonics and 2-tone intermodulation are introduced, and many of the concepts for distortion and noise measurements are equally valid whether using a spectrum analyzer or a modern VNA for the test receiver.

Chapter 7 extends the discussion of active device test to that of mixers. Because few engineers have experience with mixers and they are often only superficially covered in engineering courses, the chapter starts with a detailed discussion of the modeling and characteristics of mixers and frequency converters. Measurement methods for mixers can be quite complicated, especially for the phase or delay response. Several key methods are discussed, with a new method of calibrating, using a phase reference, presented in detail for the first time. Besides the magnitude and phase frequency response, methods for measuring mixer characteristics versus RF and LO power are presented, along with distortion and noise measurements. This chapter is required reading for any test engineer dealing with mixers or frequency converters.

Chapter 8 brings in the concept of differential and balanced devices and provides complete details on the analysis and measurement methods for differential devices including non-linear responses, noise figure, and distortion.

Chapter 9 provides a collection of very useful techniques and concepts for the test engineer particularly with respect to test fixturing including complete discussion of creating in-fixture calibration kits.

February 2012

Joel P. Dunsmore
Sebastopol, CA

Acknowledgments for the Second Edition

For the new material in Chapter 6 (Hot parameters), I'd like to thank David Root and Jan Verspecht for their theoretical work, and Xin Chen for the implementation. Chapter 7 (differential and IQ mixer measurements) was implemented in firmware by Xin Chen as well. Many thanks for his great efforts.

My thanks for the implementation work in Chapter 8, which was realized in software by Jean-Pierre Teyssier, along with his expert consulting on the concepts of coherent spectrum analysis.

Acknowledgments from the First Edition

Many of my colleagues assisted in the development and review of this book, and I would like to acknowledge their help here. Henri Komrij, my R&D manager, has been a great supporter from the initial concept, as well as Greg Peters, V.P. and general manager of the Components Test Division. Many R&D engineers in our lab contributed to the review of the manuscript, and their expertise in each field is sincerely appreciated: Keith Anderson, Dara Sariaslani, Dave Blackham, Ken Wong, Shinya Goto, Bob Shoulders, Dave Ballo, Clive Barnett, Cheng Ning, Xin Chen, Mihai Marcu, and Loren Betts. They did an excellent job, and any remaining errors are entirely and regrettably my own.

Many of the new methods and techniques presented here rely on the difficult and precise implementation of measurement methods and algorithms, and I'd like to thank our software design team, Johan Ericsson, Sue Wood, Jim Kerr, Phil Hoard, Jade Hughes, Brad Hokkanen, Niels Jensen, Raymond Taylor, Dennis McCarthy, Andy Cannon, Wil Stark, Yu-Chen Hu, Zhi-Wen Wong, and Yang Yang, as well as their managers, Sean Hubert, Qi Gao, and Dexter Yamaguchi for all their help over the years in implementing in our products many of the functions described here.

Finally, I would like to remember here Dr. Roger Pollard, who as my PhD adviser at University of Leeds and as a colleague during his sabbaticals at HP and Agilent Technologies, provided advice, mentoring, and friendship; he will be greatly missed.

1

Introduction to Microwave Measurements

“To measure is to know.”¹ This is a book about the art and science of measuring microwave components. While this work is based entirely on science, there is some art in the process, and the terms *skilled-in-the-art* and *state-of-the-art* take on particular significance when viewing the task of measuring microwave components. The goal of this work is to provide the latest, state-of-the-art methods and techniques for acquiring the optimum measurements of the myriad of microwave components. This goal naturally leads to the use of the vector network analyzer (VNA) as the principal test equipment, supported by the use of power meters, spectrum analyzers (SAs), signal sources and noise sources, impedance tuners, and other accessories.

Note here the careful use of the word *optimum*; this implies there are trade-offs between the cost and complexity of the measurement system, the time or duration of the measurement, the analytically computed uncertainty and traceability, and some heretofore unknown intangibles that all affect the overall measurement. For the best possible measurement, ignoring any consequence of time or cost, one can often go to national standards laboratories to find these best methods, but they would not suit a practical or commercial application. Thus, here the attempt is to strike an optimum balance between minimal errors in the measurement and practical consequences of the measurement techniques. The true value of this book is in providing insight into the wide range of issues and troubles that one encounters in trying to carefully and correctly ascertain the characteristics of one’s microwave component. The details have been gathered from decades of experience in hundreds of direct interactions with actual measurements; some problems are obvious and common, and others are subtle and rare. It is hoped that the reader can use this handbook to avoid many hours of unproductive test time.

For the most part, the mathematical derivations in this book are intended to provide the reader with a straightforward connection between the derived values and the underlying characteristics. In some cases, the derivation will be provided in full if it is not accessible from

¹ Lord Kelvin, “On Measurement.”

existing literature; in other cases, a reference to the derivation will be provided. There are extensive tables and figures, with key sections providing many of the important formulas. The mathematical level of this handbook is geared to a college senior or working engineer with the intention of providing the most useful formulas in an approachable way. As such, sums will be preferred to integrals; finite differences will be preferred to derivatives; and divs, grads, and curls will be entirely eschewed.

The chapters are intended to self-standing for the most part. In many cases, there will be common material to many measurement types, such as the mathematical derivation of the parameters or the calibration and error-correction methods, and these will be gathered in the introductory chapters, though well referenced in the measurement chapters. In some cases, older methods of historical interest are given (there are many volumes on these older techniques), but by and large only the most modern techniques are presented. The focus here is on the practical microwave engineer facing modern, practical problems.

1.1 Modern Measurement Process

Throughout the discussion of measurements, a six-step procedure will be followed that applies to most measurement problems. When approaching a measurement, these steps are as follows:

Pretest: This important first step is often ignored, resulting in meaningless measurements and wasted time. During the pretest, measurements of the device-under-test (DUT) are performed to coarsely determine some of its attributes. During pretest, it is also determined if the DUT is plugged in, turned on, and operating as expected. Many times the gain, match, or power handling is discovered to be different than expected, and much time and effort can be saved by finding this out early.

Optimize: Once the coarse attributes of the device have been determined, the measurement parameters and measurement system can be optimized to give the best results for that particular device. This might include adding an attenuator to the measurement receivers, adding booster amplifiers to the source, or just changing the number of points in a measurement to capture the true response of the DUT. Depending upon the device's particular characteristic response relative to the system errors, different choices for calibration methods or calibration standards might be required.

Calibrate: Many users will skip to this step, only to find that something in the setup does not provide the needed conditions and they must go back to the first step, retest, and optimize before recalibration. Calibration is the process of characterizing the measurement system so that systematic errors can be removed from the measurement result. This is not the same as obtaining a calibration sticker for an instrument but really is the first step, the *acquisition* step of the error correction process that allows improved measurement results.

Measure: Finally, some stimulus is applied to the DUT, and its response to the stimulus is measured. During the measurement, many aspects of the stimulus must be considered, as well as the order of testing and other testing conditions. These include not only the specific test conditions but also pre-conditions such as previous power states to account for non-linear responses of the DUT.

Analyze: Once the raw data is taken, error correction factors (the *application* step of error correction) are applied to produce a corrected result. Further mathematical manipulations

on the measurement result can be performed to create more useful figures of merit, and the data from one set of conditions can be correlated with other conditions to provide useful insight into the DUT.

Save data: The final step is saving the results in a useful form. Sometimes this can be as simple as capturing a screen dump, but often it means saving results in such a way that they can be used in follow-up simulations and analysis.

1.2 A Practical Measurement Focus

The techniques used for component measurements in the microwave world change dramatically depending upon the attributes of the components; thus, the first step in describing the optimum measurement methods is understanding the expected behavior of the DUT. In describing the attributes and measurements of microwave components it is tempting to go back to first principles and derive all the underlying mathematics for each component and measurement described, but such an endeavor would require several volumes to complete. One could literally write a book on the all the attributes of almost any *single* component, so for this book the focus will be on only those final results useful for describing practical attributes of the components to be characterized, with quotes and references of many results without the underlying derivation.

There have been examples of books on microwave measurements that focus on the metrology kind of measurements (Collier and Skinner 2007) made in national laboratories such as the National Institute for Standards and Technology (NIST, USA) or the National Physical Laboratory (NPL, UK), but the methods used there don't transfer well or at all to the commercial market. For the most part, the focus of this book will be on practical measurement examples of components found in commercial and aerospace/defense industries. The measurements focus will be commercial characterization rather than the kinds of metrology found in standards labs.

Also, while there has been a great deal written about components in general or ideal terms, as well as much academic analysis of these idealized components, in practice these components contain significant parasitic effects that cause their behavior to differ dramatically from that described in many textbooks. Unfortunately, these effects are often not well understood, or difficult to consider in an analytic sense, and so are revealed only during an actual measurement of a physical device. In this chapter, the idealized analysis of many components is described, but the descriptions are extended to some of the real-world detriments that cause these components' behavior to vary from the expected analytical response.

1.3 Definition of Microwave Parameters

In this section, many of the relevant parameters used in microwave components are derived from the fundamental measurements of voltage and current on the ports. For simplicity, the derivations will focus on measurements made under the conditions of termination in real valued impedances, with the goal of providing mathematical derivations that are straightforward to follow and readily applicable to practical cases.

In microwave measurements, the fundamental parameter of measurement is power. One of the key goals of microwave circuit design is to optimize the power transfer from one circuit to another such as from an amplifier to an antenna. In the microwave world, power is almost

always referred to as either an incident power or a reflected power, in the context of power traveling along a transmission structure. The concept of traveling waves is of fundamental importance to understanding microwave measurements, and to engineers who haven't had a course on transmission lines and traveling waves, and even some who have, the concept of power flow and traveling waves can be confusing.

1.3.1 S-Parameter Primer

S-parameters have been developed in the context of microwave measurements but have a clear relationship to voltages and currents that are the common reference for most electrical engineers. This section will develop the definition of traveling waves and from that the definition of S-parameters, in a way that is both rigorous and ideally intuitive; the development will be incremental, rather than just quoting results, in hopes of engendering an intuitive understanding.

This signal traveling along a transmission line is known as a *traveling wave* (Marks and Williams 1992) and has a forward component and a reverse component. Figure 1.1 shows the schematic of a two-wire transmission structure with a source and a load.

If the voltage from the source is sinusoidal, it is represented by the phasor notation

$$v_s(t) = \text{Re}(|V_s|e^{j(\omega t+\phi)}), \text{ or } V_s = |V_s|e^{j(\omega t+\phi)} \quad (1.1)$$

The voltage and current at the load are

$$V_L = |V_L|e^{j\phi_L^V}, I_L = |I_L|e^{j\phi_L^I} \quad (1.2)$$

The voltage along the line is defined as $V(z)$, and the current at each point is $I(z)$. The impedance of the transmission line provides for a relationship between the voltage and the current. At the reference point, the total voltage is $V(0)$ and is equal to V_1 ; the total current is $I(0)$. The power delivered to the load can be described as

$$P_L = P^F - P^R \quad (1.3)$$

where P^F is called the *forward power*, and P^R is called the *reverse power*. To put this in terms of the voltage and current of Figure 1.1, the total voltage at the port can be defined as the sum

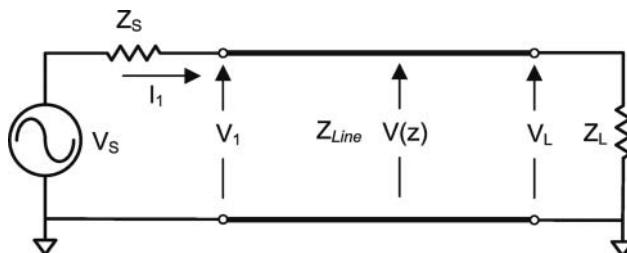


Figure 1.1 Voltage source and two-wire system.

of the forward voltage wave traveling into the port and the reverse voltage wave emerging from the port.

$$V_1 = V_F + V_R \quad (1.4)$$

The forward voltage wave represents a power traveling toward the load, or transferring from the source to the load, and the reflected voltage wave represents power traveling toward the source. To be formal, for a sinusoidal voltage source, the voltage as a function of time is

$$v_1(t) = V_1^p \cos(\omega t + \phi) = \text{Re}(V_1^p e^{j(\omega t + \phi)}) \quad (1.5)$$

From this it is clear that V_1^p is the peak voltage and the root-mean-square (rms) voltage is

$$V_1 = \frac{V_1^p}{\sqrt{2}} \quad (1.6)$$

The $\sqrt{2}$ factor shows often in the following discussion of power in a wave, and it is sometimes a point of confusion; but if one remembers that rms voltage is what is used to compute power in a sine wave, and is used to refer to the wave amplitude of a sine wave in the following equations, then it will make perfect sense.

Considering the source impedance Z_S and the line or port impedance Z_0 , and simplifying a little by making $Z_S = Z_0$ and considering the case where Z_0 is pure-real, one can relate the forward and reverse voltage to an equivalent power wave. If one looks at the reference point of Figure 1.1 and one had the possibility to insert a current probe as well as had a voltage probe, one could monitor the voltage and current.

The source voltage must equal the sum of the voltage at port 1 and the voltage drop of the current flowing through the source impedance.

$$V_S = V_1 + I_1 Z_0 \quad (1.7)$$

Defining the forward voltage as

$$V_F = \frac{1}{2}(V_1 + I_1 Z_0) \quad (1.8)$$

we see that the forward voltage represents the voltage at port 1 in the case where the termination is Z_0 . From this and Eq. (1.4), one finds that the reverse voltage must be

$$V_R = \frac{1}{2}(V_1 - I_1 Z_0) \quad (1.9)$$

If the transmission line in Figure 1.1 is long (such that the load effect is not noticeable) and the line impedance at the reference point is the same as the source, which may be called the port reference impedance, then the instantaneous current going into the transmission line is

$$I_F = V_S \left(\frac{Z_0}{Z_0 + Z_S} \right) = \frac{V_S}{2Z_0} \Big|_{Z_0=Z_S} \quad (1.10)$$

The voltage at that point is same as the forward voltage and can be found to be

$$V_F = V_S \left(\frac{Z_0}{Z_0 + Z_S} \right) = \frac{V_S}{2} \Big|_{Z_0=Z_S} \quad (1.11)$$

The power delivered to the line (or a Z_0 load) is

$$P_F = V_F I_F = \left(\frac{V_F^2}{Z_0} \right) = \frac{V_S}{4Z_0} \quad (1.12)$$

From these definitions, one can now refer to the incident and reflected power waves using the normalized incident and reflected voltage waves, a and b as (Keysight Technologies 1968).

$$a = \frac{V_F}{\sqrt{Z_0}}, \quad b = \frac{V_R}{\sqrt{Z_0}} \text{ provided } Z_0 \text{ is real} \quad (1.13)$$

Or, more formally as a power wave definition

$$a = \frac{1}{2} \left(\frac{V_1 + I_1 Z_0}{\sqrt{|ReZ_0|}} \right), \quad b = \frac{1}{2} \left(\frac{V_1 - I_1 Z_0^*}{\sqrt{|ReZ_0|}} \right) \quad (1.14)$$

where Eq. (1.14) includes the situation in which Z_0 is not pure real (Kurokawa 1965). However, it would be an unusual case to have a complex reference impedance in any practical measurement.

For real values of Z_0 , one can define the forward or incident power as $|a|^2$ and the reverse or scattered power as $|b|^2$ and see that the values a and b are related to the forward and reverse voltage waves, but with the units of square root of power. In practice, the definition of Eq. (1.13) is typically used, as the definition of Z_0 is almost always either 50 or 75 Ω . In the case of waveguide measurements, the impedance is not well defined and changes with frequency and waveguide type. It is recommended to simply use a normalized impedance of 1 for the waveguide impedance. This does not represent 1 Ω but is used to represent the fact that measurements in a waveguide are normalized to the impedance of an ideal waveguide. In Eq. (1.13) incident and reflected waves are defined, and in practice the incident waves are the independent variables, and the reflected waves are the dependent variables. Consider Figure 1.2, a 2-port network.

There are now sets of incident and reflected waves at each port i , where

$$a_i = \frac{V_{Fi}}{\sqrt{Z_{0i}}}, \quad b_i = \frac{V_{Ri}}{\sqrt{Z_{0i}}} \quad (1.15)$$

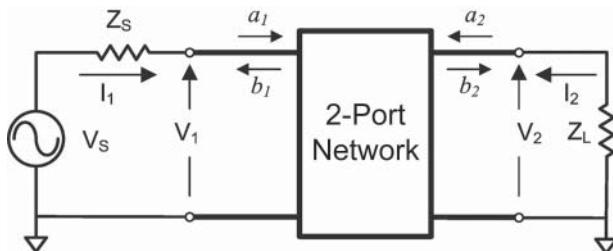


Figure 1.2 2-port network connected to a source and load.

The voltages and currents at each port can now be defined as

$$\begin{aligned} V_i &= \sqrt{Z_{0i}}(a_i + b_i) \\ I_i &= \frac{1}{\sqrt{Z_{0i}}}(a_i - b_i) \end{aligned} \quad (1.16)$$

where Z_{0i} is the reference impedance for the i th port. An important point here that is often misunderstood is that the reference impedance does not have to be the same as the port impedance or the impedance of the network. It is a “nominal” impedance; that is, it is the impedance that we “name” when we are determining the S-parameters, but it need not be associated with any impedance in the circuit. Thus, a 50Ω test system can easily measure and display S-parameters for a 75Ω device, referenced to 75Ω .

The etymology of the term *reflected* derives from optics and refers to light reflecting off a lens or other object with an index of refraction different from air, whereas it appears that the genesis for the scattering or S-matrix was derived in the study of particle physics, from the concept of wavelike particles scattering off crystals. In microwave work, scattering or S-parameters are defined to relate the independent incident waves to the dependent waves; for a 2-port network they become

$$\begin{aligned} b_1 &= S_{11}a_1 + S_{12}a_2 \\ b_2 &= S_{21}a_1 + S_{22}a_2 \end{aligned} \quad (1.17)$$

which can be placed in matrix form as

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \quad (1.18)$$

where a's represent the incident power at each port, that is, the power flowing into the port, and b's represent the scattered power, that is, the power reflected or emanating from each port. For more than two ports, the matrix can be generalized to

$$\begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} S_{11} & \cdots & S_{1n} \\ \vdots & \ddots & \vdots \\ S_{n1} & \cdots & S_{nn} \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \quad \text{or } [b_n] = [S] \cdot [a_n] \quad (1.19)$$

From Eq. (1.17) it is clear that it takes four parameters to relate the incident waves to the reflected waves, but Eq. (1.17) provides only two equations. As a consequence, solving for the S-parameters of a network requires that at least two sets of linearly independent conditions for a_1 and a_2 be applied, and the most common set is one where first a_2 is set to zero, the resulting b waves are measured, and then a_1 is set to zero, and finally a second set of b waves are measured. This yields

$$\begin{aligned} S_{11} &= \left. \frac{b_1}{a_1} \right|_{a_2=0} & S_{12} &= \left. \frac{b_1}{a_2} \right|_{a_1=0} \\ S_{21} &= \left. \frac{b_2}{a_1} \right|_{a_2=0} & S_{22} &= \left. \frac{b_2}{a_2} \right|_{a_1=0} \end{aligned} \quad (1.20)$$

which is the most common expression of S-parameter values as a function of a and b waves, and often the only one given for their definition. However, there is nothing in the definition of S-parameters that requires one or the other incident signals to be zero, and it would be just as valid to define them in terms of two sets of incident signals, a_n and a'_n , and reflected signals, b_n and b'_n .

$$\begin{aligned} S_{11} &= \left(\frac{b_1 a'_2 - a_2 b'_1}{a_1 a'_2 - a_2 a'_1} \right) & S_{12} &= \left(\frac{b_1 a'_1 - a_1 b'_1}{a_2 a'_1 - a_1 a'_2} \right) \\ S_{21} &= \left(\frac{b_2 a'_2 - a_2 b'_2}{a_1 a'_2 - a_2 a'_1} \right) & S_{22} &= \left(\frac{b_2 a'_1 - a_1 b'_2}{a_2 a'_1 - a_1 a'_2} \right) \end{aligned} \quad (1.21)$$

From Eq. (1.21) one sees that S-parameters are in general defined for a pair of stimulus drives. This will become quite important in more advanced measurements and in the actual realization of the measurement of S-parameters, because in practice it is not possible to make the incident signal go to zero because of mismatches in the measurement system.

These definitions naturally lead to the concept that S_{nn} parameters are reflection coefficients and are directly related to the DUT port input impedance and S_{mn} parameters are transmission coefficients and are directly related to the DUT gain or loss from one port to another.

Now that the S-parameters are defined, they can be related to common terms used in the industry. Consider the circuit of Figure 1.3, where the load impedance Z_L may be arbitrary and the source impedance is the reference impedance.

From inspection one can see that

$$V_1 = V_S \left(\frac{Z_L}{Z_L + Z_0} \right), \quad I_1 = V_S \left(\frac{1}{Z_L + Z_0} \right) \quad (1.22)$$

which is substituted into Eq. (1.8) and Eq. (1.9), and from (1.15) one can directly compute a_1 and b_1 as

$$a_1 = \frac{V_S}{2\sqrt{Z_0}}, \quad b_1 = \frac{V_S}{2\sqrt{Z_0}} \left(\frac{Z_L - Z_0}{Z_L + Z_0} \right) \quad (1.23)$$

From here S_{11} can be derived from inspection as

$$S_{11} = \frac{b_1}{a_1} = \frac{Z_L - Z_0}{Z_L + Z_0} \quad (1.24)$$

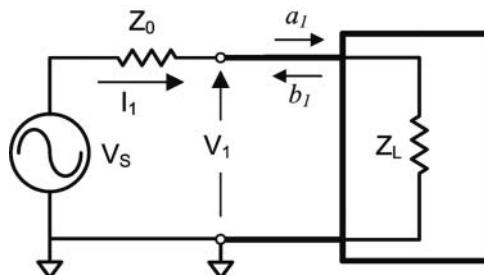


Figure 1.3 1-port network.

It is common to refer to S_{11} informally as the input impedance of the network, where

$$Z_{In} = \frac{V_1}{I_1} \quad (1.25)$$

This is clearly true for a 1-port network and can be extended to a 2-port or n-port network if all the ports of the network are terminated in the reference impedance; but in general, one cannot say that S_{11} is the input impedance of a network without knowing the termination impedance of the network. This is a common mistake that is made with respect to determining the input impedance or S-parameters of a network. S_{11} is defined for any terminations by Eq. (1.21), but it is the same as the input impedance of the network only under the condition that it is terminated in the reference impedance, thus satisfying the conditions for Eq. (1.20). Consider the network of Figure 1.2 where the load is not the reference impedance; as such, it is noted that a_1 and b_1 exist, but now Γ_1 (also called Γ_{In} for a 2-port network) is defined as

$$\Gamma_1 = \frac{b_1}{a_1} \quad (1.26)$$

with the network terminated in an arbitrary impedance. As such, Γ_1 represents the input impedance of a system comprised of the network and its terminating impedance. The important distinction is that S-parameters of a network are invariant to the input or output terminations, providing they are defined to a consistent reference impedance, whereas the input impedance of a network depends upon the termination impedance at each of the other ports. The value of Γ_1 of a 2-port network can be directly computed from the S-parameters and the terminating impedance, Z_L , as

$$\Gamma_1 = \left(S_{11} + \frac{S_{21}S_{12}\Gamma_L}{1 - S_{22}\Gamma_L} \right) \quad (1.27)$$

where Γ_L computed as in Eq. (1.24) is

$$\Gamma_L = \frac{Z_L - Z_0}{Z_L + Z_0} \quad (1.28)$$

or in the case of a 2-port network terminated by an arbitrary load then

$$\Gamma_L = \frac{a_2}{b_2} \quad (1.29)$$

Similarly, the output impedance of a network that is sourced from an arbitrary source impedance is

$$\Gamma_2 = \left(S_{22} + \frac{S_{21}S_{12}\Gamma_S}{1 - S_{11}\Gamma_S} \right) \quad (1.30)$$

Another common term for the input impedance is the voltage standing wave ratio, called VSWR (also simply called SWR), and it represents the ratio of maximum voltage to minimum voltage that one would measure along a Z_0 transmission line terminated in some arbitrary load impedance. It can be shown that this ratio can be defined in terms of the S-parameters of the network as

$$VSWR = \left(\frac{1 + |\Gamma_1|}{1 - |\Gamma_1|} \right) \quad (1.31)$$

If the network is terminated in its reference impedance, then Γ_1 becomes S_{11} . Another common term used to represent the input impedance is the reflection coefficient, ρ_{In} , where

$$\rho_{In} = |\Gamma_{In}| \quad (1.32)$$

It's also common to write

$$VSWR = \left(\frac{1 + \rho}{1 - \rho} \right) \quad (1.33)$$

Another term related to the input impedance is *return loss*, which is alternatively defined as

$$RL = 20 \cdot \log_{10}(\rho), \text{ or } RL = -20 \cdot \log_{10}(\rho) \quad (1.34)$$

with the second definition being most properly correct, as loss is defined to be positive in the case where a reflected signal is smaller than the incident signal. But, in many cases, the former definition is more commonly used; the microwave engineer must simply refer to the context of the use to determine the proper meaning of the sign. Thus, an antenna with 14 dB return loss would be understood to have a reflection coefficient of 0.2, and the value displayed on a measurement instrument might read -14 dB.

For transmission measurements, the figure of merit is often gain or insertion loss (sometimes called *isolation* when the loss is very high). Typically this is expressed in dB, and similarly to return loss, it is often referred to as a positive number. Thus

$$Gain = 20 \log_{10}(|S_{21}|) \quad (1.35)$$

Insertion loss or isolation is defined as

$$Insertion\ Loss = Isolation = -20 \log_{10}(|S_{21}|) \quad (1.36)$$

Again, the microwave engineer will need to use the context of the discussion to understand that a device with 40 dB isolation will show on an instrument display as -40 dB, due to the instrument using the evaluation of Eq. (1.35).

Notice that in the return loss, gain, and insertion loss equations, the dB value is given by the formula $20\log_{10}(|S_{nm}|)$, and this is often a source of confusion because common engineering use of dB has the computation as $X_{dB} = 10\log_{10}(X)$. This apparent inconsistency comes from the desire to have power gain when expressed in dB be equal to voltage gain, also expressed in dB. In a device sourced from a Z_0 source and terminated in a Z_0 load, the power gain is defined as the power delivered to the load relative to the power delivered from the source, and the gain is

$$Power\ gain = 10\log_{10} \left(\frac{P_{To_Load}}{P_{From_Source}} \right) \quad (1.37)$$

The power from the source is the incident power $|a_1|^2$, and the power delivered to the load is $|b_2|^2$. The S-parameter gain is S_{21} and in a matched source and load situation is simply

$$S_{21} = \frac{b_2}{a_1}, \quad |S_{21}|^2 = \left| \frac{b_2}{a_1} \right|^2 = \frac{|b_2|^2}{|a_1|^2} = Power\ Gain \quad (1.38)$$

So computing power gain as in Eq. (1.37) and converting to dB yields the familiar formula

$$\text{Power Gain}_{dB} = 10\log_{10}(|S_{21}|^2) = 20\log_{10}(|S_{21}|) \quad (1.39)$$

A few more comments on power are appropriate, as power has several common meanings that can be confused if not used carefully. For any given source, as shown in Figure 1.1, there exists a load for which the maximum power of the source may be delivered to that load. This maximum power occurs when the impedance of the load is equal to the conjugate of the impedance of the source, and the maximum power delivered is

$$P_{\max} = \frac{|V_S|^2}{4 \cdot \text{Re}(Z_S)} \quad (1.40)$$

But it is instructive to note that the maximum power as defined in Eq. (1.40) is the same as $|a_1|^2$ provided the source impedance is real and equals the reference impedance; thus, the incident power from a Z_0 source is always the maximum power that can be delivered to a load. The actual power delivered to the load can be defined in terms of a and b waves as well.

$$P_{\text{del}} = |a|^2 - |b|^2 \quad (1.41)$$

If one considers a passive two-port network and conservation of energy, power delivered to the load must be less than or equal to the power incident on the network minus the power reflected, or in terms of S-parameters

$$|S_{21}|^2 \leq 1 - |S_{11}|^2 \quad (1.42)$$

which leads the well-known formula for a lossless network

$$|S_{21}|^2 + |S_{11}|^2 = 1 \quad (1.43)$$

1.3.2 Phase Response of Networks

While most of the discussion thus far about S-parameters refers to powers, including incident, reflected, and delivered to the load, the S-parameters are truly complex numbers and contain both a magnitude and phase component. For reflection measurements, the phase component is critically important and provides insight into the input elements of the network. These will be discussed in great detail as part of Chapter 2, especially when referencing the Smith chart.

For transmission measurements, the magnitude response is often the most cited value of a system, but in many communications systems, the phase response has taken on more importance. The phase response of a network is typically given by

$$\phi_{S21} = \arctan \left[\frac{\text{Im}(S_{21})}{\text{Re}(S_{21})} \right] \quad (1.44)$$

where the region of the arctangent is usually chosen to be $\pm 180^\circ$. However, it is sometimes preferable to display the phase in absolute terms, such that there are no phase discontinuities

in the displayed value. This is sometimes called the *unwrapped* phase, in which the particular cycle of the arctangent must be determined from the previous cycle, starting from the DC value. Thus, the unwrapped phase is uniquely defined for an S_{21} response only when it includes all values down to zero frequency (DC).

The linearity of the phase response has consequences when looking at its effect on complex modulated signals. In particular, it is sometimes stated that linear networks cannot cause distortion, but this is true only of single-frequency sinusoidal inputs. Linear networks can cause distortion in the envelope of complex modulated signals, even if the frequency response (the magnitude of S_{21}) is flat. That is because the phase response of a network directly affects the relative time that various frequencies of a complex modulated signal take to pass through the network. Consider the signal in Figure 1.4.

For this network, the phase of S_{21} defines how much shift occurs for each frequency element in the modulated signal. Even though the amplitude response is the same in both Figure 1.4a,b, the phase response is different, and the envelope of the resulting output is changed. In general, there is some delay from the input to the output of a network, and the important definition

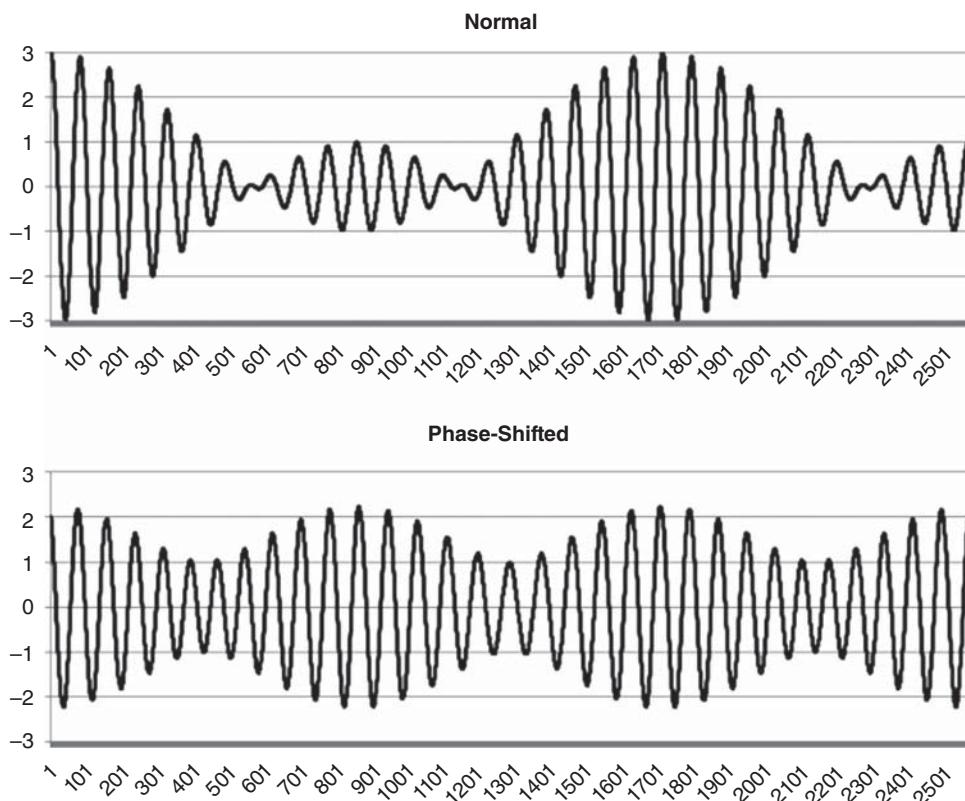


Figure 1.4 Modulated signal through a network showing distortion due to only phase shift: normal (upper), shifted (lower).

that is most commonly used is the group delay of the network, defined as

$$\tau_{GD} = -\frac{d\phi_{S21}^{rad}}{d\omega} = \frac{-d\phi_{S21}^{\circ}}{360 \cdot df} \quad (1.45)$$

While easily defined, the group delay response may be difficult to measure and/or interpret. This is because measurement instruments record discrete values for phase, and the group delay is a derivative of the phase response. Using discrete differentiation can generate numerical difficulties; Chapter 5 shows some of the difficulties encountered in practice when measuring group delay, as well as some solutions to these difficulties.

For most complex signals, the ideal goal for phase response of a network is that of a linear phase response. Deviation from linear phase is a figure of merit for the phase flatness of a network, and this is closely related another figure of merit, group delay flatness. Thus, the ideal network has a flat group delay, meaning a linear phase response. However, many complex communications systems employ equalization to remove some of the phase response effects. Often, this equalization can account for first- or second-order deviations in the phase; thus, another figure of merit is deviation from parabolic phase, which is effectively a measure of the quality of fit of the phase response to a second order polynomial. These measurements are discussed further in Chapter 5.

1.4 Power Parameters

1.4.1 Incident and Reflected Power

Just as there are a variety of S-parameters, which are derived from the fundamental parameters of incident and reflected waves a and b , so too are there many power parameters that can be identified with the same waves. As inferred earlier, the principal power parameters are incident and reflected, or forward and reverse, powers at each port, which for Z_0 real, are defined as

$$P_{Incident} = P_F = |a|^2, \quad P_{Reflected} = P_R = |b|^2 \quad (1.46)$$

The proper interpretation of these parameters is that incident and reflected power is the power that would be delivered to a nonreflecting (Z_0) load. If one were to put an ideal Z_0 directional-coupler in line with the signal, it would sample or couple the incident signal (if the coupler were set to couple the forward power) or the reflected signal (if the coupler were set to couple the reverse power). In simulations, ideal directional-couplers are often used in just such a manner.

1.4.2 Available Power

The maximum power that can delivered from a generator is called the *available power*, or $P_{Available}$, and can be defined as the power delivered from a Z_S

$$P_{Available} = P_{AS} = \frac{|a_S|^2}{(1 - |\Gamma_S|^2)} \quad (1.47)$$

where Γ_S is computed as in Eq. (1.24) as

$$\Gamma_S = \frac{Z_S - Z_0}{Z_S + Z_0} \quad (1.48)$$

This maximum power is delivered to the load when the load impedance is the conjugate of the source impedance, $Z_L = Z_S^*$.

1.4.3 Delivered Power

The power that is absorbed by an arbitrary load is called the *delivered power*, and it is computed directly from the difference between the incident and reflected power.

$$P_{del} = |a|^2 - |b|^2 \quad (1.49)$$

For most cases, this is the power parameter that is of greatest interest. In the case of a transmitter, it represents the power that is delivered to the antenna, for example, which in turn is the power radiated less the resistive loss of the antenna.

1.4.4 Power Available from a Network

A special case of available power is the power available from the output of a network, when the network is connected an arbitrary source. In this case, the available power is only a function of the network and the source impedance and is not a function of the load impedance. It represents the maximum power that could be delivered to a load under the condition that the load impedance was ideally matched and can be found by noting that the available output power is similar to Eq. (1.47) but with the source reflection coefficient replaced by the output reflection coefficient of the network Γ_2 from Eq. (1.30) such that

$$P_{Out_Available} = P_{OA} = \frac{|b_2|^2}{(1 - |\Gamma_2|^2)} \quad (1.50)$$

When a 2-port network is connected to a generator with arbitrary impedance, the output scattered wave into matched load is

$$b_2 = \frac{a_S S_{21}}{1 - \Gamma_S S_{11}} \quad (1.51)$$

Here the incident wave is represented as a_S rather than a_1 as an indication that the source is not matched, and Γ_1 is defined by Eq. (1.27). The output power incident to the load is

$$|b_2|^2 = \frac{|a_S|^2 |S_{21}|^2}{|1 - \Gamma_S S_{11}|^2} \quad (1.52)$$

Combining Eqs. (1.52) and (1.50), the available power at the output from a network that is driven from a generator with source impedance of Γ_S is

$$P_{OA} = \frac{|b_2|^2}{(1 - |\Gamma_2|^2)} = \frac{|a_S|^2 |S_{21}|^2}{|1 - \Gamma_S S_{11}|^2 (1 - |\Gamma_2|^2)} \quad (1.53)$$

With Γ_2 defined as in Eq. (1.30).

1.4.5 Available Gain

Available gain is the gain that an amplifier can provide to a conjugately matched load from a source or generator of a given impedance and is computed with the formula

$$G_A = \frac{(1 - |\Gamma_S|^2)|S_{21}|^2}{|1 - \Gamma_S S_{11}|^2(1 - |\Gamma_2|^2)} \quad (1.54)$$

$$\text{where } \Gamma_2 = \left(S_{22} + \frac{S_{21}S_{12}\Gamma_S}{1 - S_{11}\Gamma_S} \right)$$

Other derived values such as maximum available gain and maximum stable gain are discussed in detail in Chapter 6.

1.5 Noise Figure and Noise Parameters

For a receiver, the key figure of merit is its sensitivity, or the ability to detect small signals. This is limited by the intrinsic noise of the device itself, and for amplifiers and mixers, this is represented as noise figure. Noise figure is defined as a signal-to-noise ratio at the input divided by signal to noise at the output expressed in dB.

$$NF \equiv N_{Figure} = 10\log_{10} \left(\frac{\text{Signal}_{Input}/\text{Noise}_{Input}}{\text{Signal}_{Output}/\text{Noise}_{Output}} \right) = 10\log_{10} \left(\frac{(S/N)_I}{(S/N)_O} \right) \quad (1.55)$$

Its related value, noise factor, which is unitless, is

$$N_F \equiv N_{Factor} = \left(\frac{\text{Signal}_{Input}/\text{Noise}_{Input}}{\text{Signal}_{Output}/\text{Noise}_{Output}} \right) = \frac{(S/N)_I}{(S/N)_O} \quad (1.56)$$

Here the signal and noise values are represented as a power; traditionally, this is available power, but incident power can be used as well with a little care. Rearranging Eq. (1.55), one can obtain

$$N_{Factor} = \frac{N_O}{Gain \cdot N_I} = \frac{N_{O_Avail}}{G_{Avail} \cdot N_{I_Avail}} \quad (1.57)$$

In most cases, the input noise is known very well, as it consists only of thermal noise associated with the temperature of the source resistance. This is the noise available from the source and can be found from

$$N_{Avail} = N_a = kTB \quad (1.58)$$

where k is Boltzmann's constant (1.38×10^{-23} J K $^{-1}$), B is the noise bandwidth, and T is the temperature in Kelvin. Note that the available noise power does not depend upon the impedance of the source. From the definition in Eq. (1.57) it is clear that if the temperature of the source impedance changes, then the noise figure of the amplifier using this definition would change as well. Therefore, by convention, a fixed value for the temperature is presumed, and this value, known as T_0 , is 290 K.

This is the noise power that would be delivered to a conjugately matched load. Alternatively, the noise power can be represented as a noise wave, much like a signal, and one can

define an incident noise (sometimes called the *effective noise power*), which is defined as the noise delivered to a nonreflecting nonradiating load and is found as

$$N_{\text{Incident}} = N_E = N_A(1 - |\Gamma_S|^2) \quad (1.59)$$

which is consistent with the definition of Eq. (1.47). Since the available noise at the output of a network doesn't depend upon the load impedance, the available gain from a network similarly doesn't depend upon the load impedance, and the available noise at the input of the network can be computed as Eq. (1.58), the measurement of noise figure defined in this way is not dependent upon the match of the noise receiver. One way to understand this is to note that the available gain is the maximum gain that can be delivered to a load. If the load is not conjugately matched to Γ_2 , both the available gain and the available noise power at the output would be reduced by equal amounts, leaving the noise figure unchanged and independent of the noise receiver load impedance. Thus, for the case of noise measurements, the available noise power and available gain have been the important terms of use historically.

Recently more advanced techniques have been developed and made practical based on incident noise power and gain. If the impedance is known, the incident noise power can be computed as in Eq. (1.59); and if the output incident noise power N_{OE} can be measured, then one can compute the output available noise as

$$N_{OA} = \frac{N_{OE}}{(1 - |\Gamma_2|^2)} \quad (1.60)$$

Substituting into Eq. (1.57) to find

$$N_F = \frac{1}{G_A} N_{OA} \frac{1}{N_{IA}} = \frac{|1 - \Gamma_S S_{11}|^2 (1 - |\Gamma_2|^2)}{(1 - |\Gamma_S|^2) |S_{21}|^2} \frac{N_{OE}}{(1 - |\Gamma_2|^2) (kTB)} = \frac{|1 - \Gamma_S S_{11}|^2 N_{OE}}{(1 - |\Gamma_S|^2) |S_{21}|^2 (kTB)} \quad (1.61)$$

When the source is a matched source, this simplifies to

$$N_F = \frac{N_{OE}}{|S_{21}|^2 (kTB)} \quad (1.62)$$

Thus, for a simple system of an amplifier sourced with a Z_0 impedance and terminated with a Z_0 load, the noise factor can be computed simply from the noise power measured in the load and the S21 gain. However, Eq. (1.61) defines the noise figure of the amplifier in terms of the source impedance, and this is a key point. In general, although the 50Ω noise figure is the most commonly quoted, it is measured only when the source impedance provided is exactly 50Ω . In the case where the source impedance is not 50Ω , the 50Ω noise figure cannot be simply determined.

1.5.1 Noise Temperature

Because of the common factor of temperature in many noise figure computations, the noise power is sometimes redefined as available noise temperature.

$$T_A = \frac{N_A}{k_B} \quad (1.63)$$

From this definition, the noise factor becomes

$$N_F = \frac{T_A}{G_A 290} = \frac{T_{RNA}}{G_A} \quad (1.64)$$

where T_{RNA} is the relative available noise temperature, expressed in Kelvin above 290 K.

1.5.2 Effective or Excess Input Noise Temperature

For very low noise figure devices, it is often convenient to express their noise factor or noise figure in terms of the excess power that would be at the input due to a higher temperature generator termination, which would result in the same available noise temperature at the output. This can be computed as

$$T_e = 290(N_F - 1) \quad (1.65)$$

Thus, an ideal noiseless network would have a zero input noise temperature, and a 3 dB noise figure amplifier would have a 290 K excess input noise temperature, or 290 K above the reference temperature.

1.5.3 Excess Noise Power and Operating Temperature

For an amplifier under test, the noise power at the output, relative to the kTB noise power, is called the *excess noise power*, P_{NE} , and is computed as

$$P_{NE} = N_F |S_{21}|^2 \frac{(1 - |\Gamma_S|^2)}{|1 - \Gamma_S S_{11}|^2 (1 - |\Gamma_2|^2)} \quad (1.66)$$

For a matched source and load, it is the excess noise, above kTB, that is measured in the terminating resistor and can be computed as

$$P_{NE} = (|S_{21}|^2 N_F) \quad (1.67)$$

which is sometimes called the *incident relative noise* or RNP_I (as opposed to available, or RNP). Errors in noise figure measurement are often the result of not accounting properly for the fact that the source or load impedances are not exactly Z_0 . A related parameter is the operating temperature, which is analogous to the input noise temperature at the amplifier output, and is computed as

$$T_O = \frac{T_{OA}}{(1 - |\Gamma|^2)} \quad (1.68)$$

While the effect of load impedance may be overcome with the use of available gain, which is independent of load impedance, the effect of source impedance mismatch must be dealt with a much more complicated way, as shown next.

1.5.4 Noise Power Density

The excess noise is measured relative to the kTB noise floor and is expressed in dBc relative to the T_0 noise floor. However, the noise power could also be expressed in absolute terms

such as dBm. But the measured noise power depends upon the bandwidth of the detector, and so the noise power density provides a reference value with a bandwidth equivalent to 1 Hz. Thus, the noise power density is related to the excess noise by

$$P_{\text{Noise Power Density}} = \frac{P_{NE}}{B} = k(T_0 + T_e) \quad (1.69)$$

1.5.5 Noise Parameters

The formal definition of noise figure for an amplifier defines the noise figure only for the impedance or reflection coefficient of current source termination, but this noise figure is *not* the 50-Ω noise figure. Rather, it is the noise figure of the amplifier for the impedance of the source. In general, one cannot compute the 50 Ω noise figure from this value without additional information about the amplifier. If one considers the amplifier in Figure 1.5, with internal noise sources, the effect of the noise sources is to produce noise power waves that may be treated similarly to normalized power waves, a and b .

The source termination produces an incident noise wave a_{NS} and adds to the internal noise created in the amplifier, which can be represented as an input noise source a_{Namp} . There are scattered noise waves represented by the noise emitted from the input of the amplifier, b_{NI} , and the noise incident on the load is b_{N2} . From this figure, one can make a direct comparison to the S-parameters and see that reflected noise power might add or subtract to the incident noise power and affect the total noise power. However, at the input of the amplifier, the noise generated inside the amplifier is in general not correlated with the noise coming from the source termination so that they don't add together in a simple way. Because of this, the noise power at the output of the amplifier, and therefore the noise figure, depends upon the source impedance in a complex way. This complex interaction is defined by two real valued parameters and one complex parameter, known collectively as the *noise parameters*. The noise figure at any source reflection coefficient may be computed as

$$N_F = N_{Fmin} + \frac{4R_n}{Z_0} \frac{|\Gamma_{opt} - \Gamma_S|^2}{|1 + \Gamma_{opt}|^2(1 - |\Gamma_S|^2)} \quad (1.70)$$

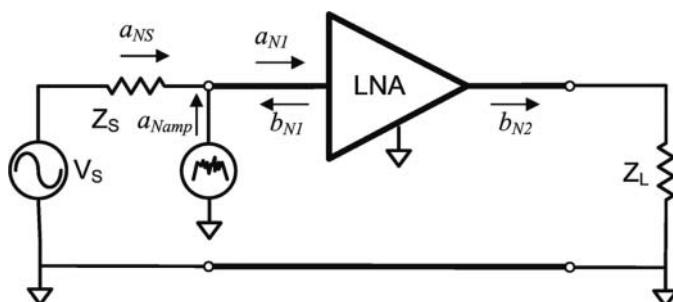


Figure 1.5 An amplifier with internal noise sources.

where $N_{F\min}$ is the minimum noise figure; Γ_{Opt} , called *gamma-opt*, is the reflection coefficient (magnitude and phase) that gives the minimum noise figure; and R_N , sometimes called the *noise resistance*, describes how the noise figure increases as the source impedances varies from the gamma-opt. The characterization required to determine these values is quite complex and is covered in Chapter 6.

1.6 Distortion Parameters

Up to now, all the parameters described have been under the consideration that the DUT is linear. However, when a DUT, particularly an amplifier, is driven with a large signal, non-linear transfer characteristics become significant, leading to an entirely new set of parameters used to describe these non-linear characteristics.

1.6.1 Harmonics

One of the first noticeable effects of large signal drive is the generation of harmonics at multiples of the input frequency. Harmonics are described by their order and either by their output power or, more commonly, by the power relative to the output power of the fundamental, and almost always in dBc (dB relative to the carrier). Second harmonic is short for second-order harmonic and refers to the harmonic found at two times the fundamental, even though it is in fact the first of the harmonic frequency above the fundamental; third harmonic is found at three times the fundamental, and so on. Surprisingly, there are not well-established symbols for harmonics; for this book, we will use $H_2, H_3 \dots H_n$ to represent the dBc values of harmonics or order 2, 3 ... respectively. In Chapter 6, the measurements of harmonics are fully developed as part of the description of X-parameters and utilize the notation $b_{2,m}$ to describe the output normalized wave power at port 2 for the m th harmonic. A similar notation is used for harmonics incident on the amplifier.

One important attribute of harmonics is that for most devices the level of the harmonics increases in dB value as the power of the input increases and to a rate directly proportional to the harmonic order, as shown in Figure 1.6. In this figure, the x-axis is the drive power, and the y-axis is the measured output power of the fundamental and the harmonics.

1.6.2 Second-Order Intercept

This pattern of increasing power as the input power is increased, but to the slope related to the order of the harmonic, cannot continue indefinitely or the harmonic power would exceed the fundamental power. While theoretically possible, in practice the harmonic power saturates just as the output power does and never crosses the level of the output power. However, if one uses the lower power regions to project a line from the fundamental and each of the harmonics, they will intersect at some power, as shown in Figure 1.6. The level that these lines converge is called the *intercept point*, and the most common value is the second-order intercept (SOI), and intercept points beyond third order are seldom used.

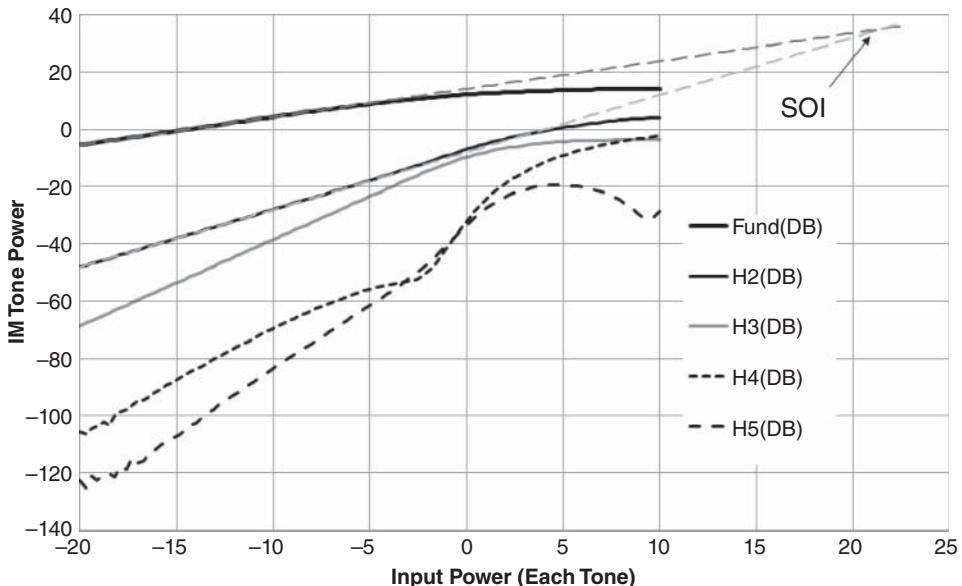


Figure 1.6 Output power of harmonics of an amplifier.

There is sometimes confusion in the use of the term SOI; while it is most commonly used to refer to the second harmonic content, in some case, it has also been used to refer to the two-tone SOI, which is a distortion product that occurs at the sum of the two tones. Most properly, one should always use the term two-tone SOI if one is to distinguish from the more common harmonic SOI.

1.6.3 Two-Tone Intermodulation Distortion

While the harmonic measurement provides a direct characterization of distortion, it suffers from the fact that the harmonic frequencies are far away from the fundamental, and in many circuits, the network response is such that the harmonic content is essentially filtered out. Thus, it is not possible to discern the non-linear response of such a network by measuring only the output signal. Of course, if the gain is measured, compression of the amplifier will show that the value of S21 changes with the input drive level. But it is convenient to have a measure or figure of merit of the distortion of an amplifier that relies only on the output signal. In such a case, two signals of different frequencies can be applied at the amplifier input, at a level sufficiently large to cause a detectable non-linear response of the amplifier. Figure 1.7 shows a measurement of a two-tone signal applied to the input of an amplifier (lower trace) and measured on the output of the amplifier (upper trace).

It is clear that several other tones are present at the output and are the result of higher-order products mixing in the amplifier due to its non-linear response and creating other signals. The principal signals of interest are the higher and lower intermodulation (IM) products, PwrN_Hi and PwrN_Lo, where N is the order of intermodulation distortion (IMD). Normally,

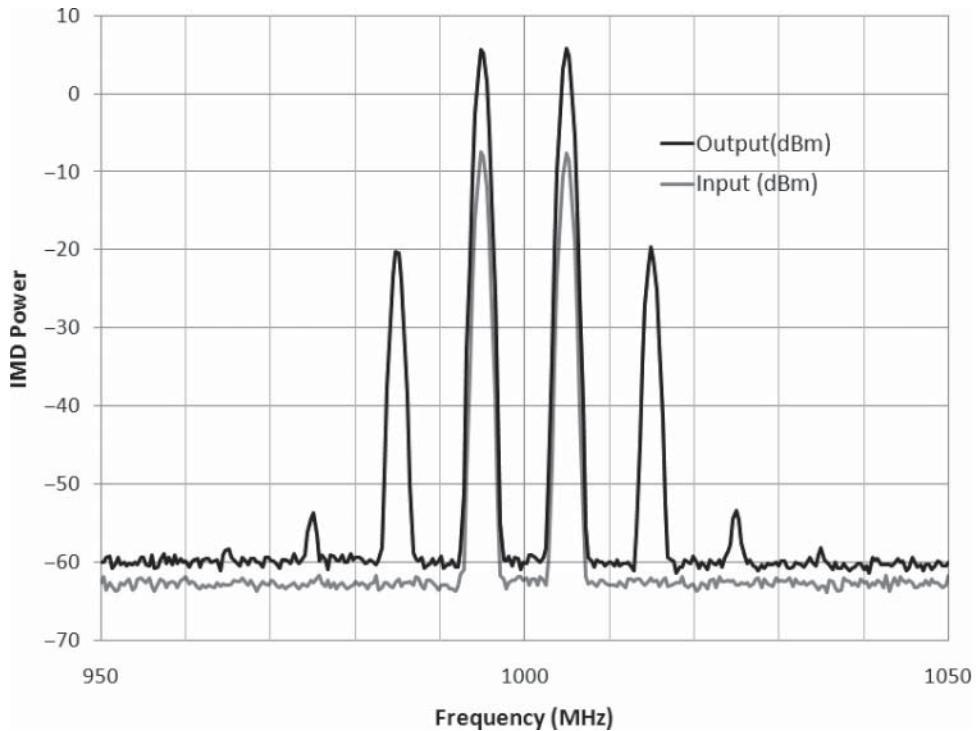


Figure 1.7 Measurement of a two-tone signal at the input and output of an amplifier.

IM products refer to the power of the IM product relative to the carrier, in dBc, and these terms are called IMN_Hi and IMN_Lo. For example, the power in the lower third-order tone is Pwr3_Lo; the level of the upper third-order tone relative to the carrier is called IM3_Hi. The frequencies of the higher and lower tones are found at

$$f_{3Hi} = 2f_{Hi} - f_{Lo}, \quad f_{3Lo} = 2f_{Lo} - f_{Hi} \quad (1.71)$$

And more generally

$$\begin{aligned} f_{mHi} &= \left(\frac{m+1}{2}\right)f_{Hi} - \left(\frac{m-1}{2}\right)f_{Lo}, \quad f_{mLo} = \left(\frac{m+1}{2}\right)f_{Lo} - \left(\frac{m-1}{2}\right)f_{Hi} \Big|_{m \text{ odd}} \\ f_{mHi} &= (m-1)f_{Hi} + (m-1)f_{Lo}, \quad f_{mLo} = (m-1)f_{Lo} - (m-1)f_{Hi} \Big|_{m \text{ even}} \end{aligned} \quad (1.72)$$

In Figure 1.7, the amplifier is driven such that the fifth-order IM product is just visible above the noise floor in the upper trace.

IM products have the same attribute as harmonics with respect to drive power, and the power in the IM product (sometimes called the *tone power*, or PWR_m for the *m*th-order IM power) increases in direct proportion to the input power and the order of the IM product. Thus, if the tone power is plotted along with the output power against an x-axis of input power, the plot will look like Figure 1.8, where the extension of the slope of the output power

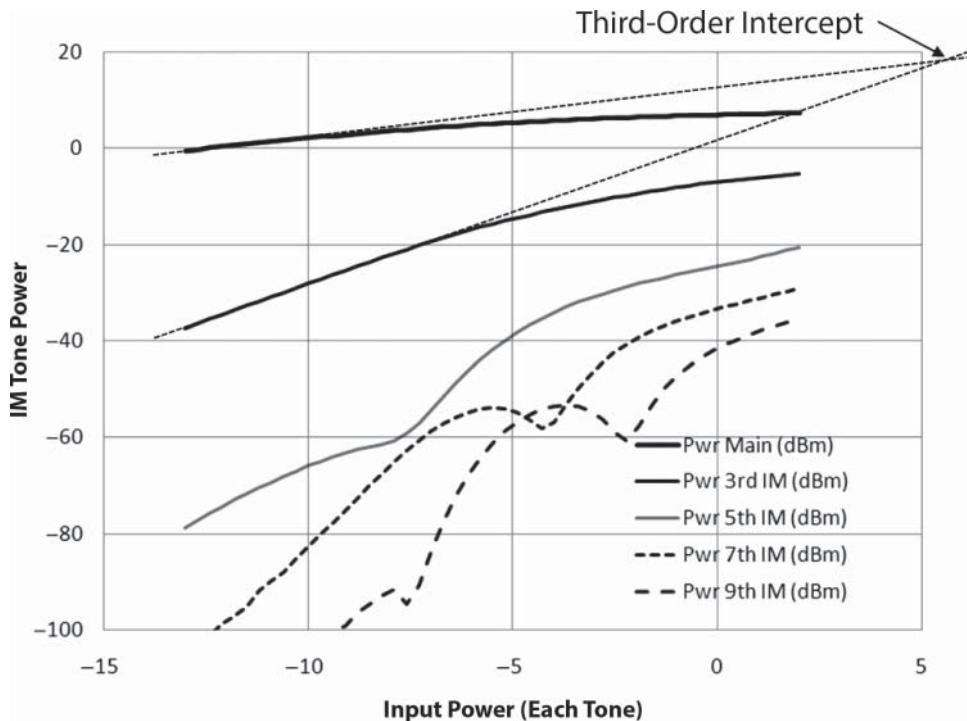


Figure 1.8 Output power and IM tone-power versus input power.

and IM tone-powers at low drives will intersect. This point of intersection for the third-order IM product is known as the third-order intercept point, or IP3. Similarly, IP5 is the fifth-order intercept point, etc.

It is also interesting to note that in general at high powers, the IM tone-powers may not increase but may decrease or have local minima. This is because of the effect of high-order IM products re-mixing and creating significant signals that lie on the lower-order products and can increase or decrease their level, depending upon the phasing of the signals.

There is often some confusion about third-order IM products (IM3) and third-order intercept point (IP3), and both are sometimes referred to as *third-order intermod*. For clarity, in this book, the intercept point will always be referred to as IP.

Finally, for amplifiers used as a low-noise amplifier (LNA) at the input of a receiver chain, it is often desired to refer the IP level to the input power, which would produce an intercept point at the output. This is distinguished as the input intercept point (IIP), and in the case of ambiguity, the normal intercept point referencing to the output power should be most properly referred to as the output-referred intercept point (OIP). The most common intercept points are the third-order ones, OIP3 and IIP3. The input and output intercept points differ by the gain of the amplifier at drive level where the measurements are made.

The details of two-tone IM measurements are discussed at length in Chapter 8.

1.6.4 Adjacent Channel Power and Adjacent Channel Level Ratio

One figure of distortion common with modulated signals is the adjacent channel power (ACP) and adjacent channel level ratio (ACLR). Sometimes a third term, adjacent channel power ratio (ACPR), is used instead of ACLR. All are measures of out-of-channel spectral regrowth caused principally by the third-order intermodulation distortion occurring because of a modulated signal. During testing, a modulated signal waveform is applied to the DUT. Figure 1.9 shows the output spectrum of a signal modulated with 16 quadrature amplitude modulation (16 QAM) over a 40 MHz BW, applied to an amplifier.

It is a repetitive periodic waveform from an arbitrary waveform generator, which must be comprised of a multiple sinewave signals, typically thousands of tones, each of which can intermodulate with each other one. In a typical modulated signal, each tone can have a nearly random amplitude and phase, so it is quite complicated to measure each distortion product directly. In general, this figure of merit measures the intermodulation products, which appear in the adjacent channel to the channel under test, as a total integrated power using band power measurements.

In the figure, the lower and upper ACP region is identified, and the signal here is caused by the third-order distortion in the amplifiers. Also identified is the outline of the distortion profile of the amplifier. ACP is used as a figure of merit as it is easy to discern the distortion level

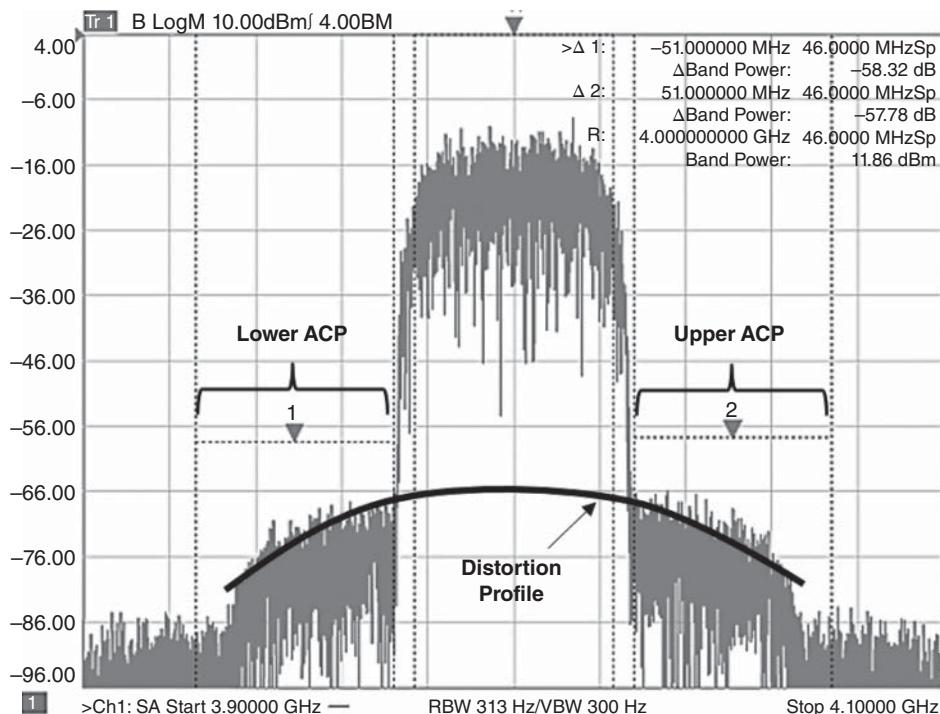


Figure 1.9 Spectral regrowth causing ACP in a 16 QAM signal.

in the adjacent channel where there is no signal. However, the distortion occurs in-channel as well as out-of-channel and is usually a bit higher in the center of the channel. The sloped response of the distortion profile is typical and can be understood by considering the density of signals that can create the intermodulation distortion. Note the outer edges of the adjacent channel where the distortion signal is lowest; only the outermost tones of the main signal can intermodulate to create a signal at these outer reaches of the adjacent channel. At the edge of the adjacent channel nearest the main signal, any two signals that are separated by one-half the main-signal bandwidth can intermodulate to create a signal here. The density of these signals is quite high, roughly half the power of the main signal. In the center of the main signal, where the distortion is not apparent because it is masked by the main-signal power, it is outlined by the distortion profile curve in the figure; any two closely spaced signals can cause distortion power here. The density of such signals is over the whole bandwidth, so the distortion level here is roughly twice that at the close-in edge of the ACP signal. Even though this distortion is masked by the main signal, it is still present and causes errors in the transmitted signal.

The total integrated power is the ACP. The ratio of the ACP to the total power in the main channel is the ACLR, shown by the Markers 1 and 2 in the figure (they are set to be a delta-marker with respect to the reference Marker R, which shows the main tone absolute power). Often, test system noise can mask the ACP or ACLR to some extent and becomes the limitation of the measurement. Details of the ACP and ACLR measurements are found in Chapter 8.

1.6.5 Noise Power Ratio (NPR)

Widely found in the satellite communications industry, noise power ratio (NPR) is a measure of distortion, and not of noise at all. In the early days of satellite development, the industry needed a measure of distortion for satellite components but could not use the more common IMD or ACP. Most satellite systems have strongly channelized amplifiers, where the communication signals fill an entire channel and are filtered at the output so adjacent channel distortion would be filtered away, and could not be used as a figure of merit for the in-channel distortion. Furthermore, the communications protocols for satellites could change over the life of the satellite, and often many different communication methods could be used in the same channel. NPR was developed to emulate a densely loaded communications channel but still provide a means to determine distortion.

In the early days, NPR signals were generated by using a noise diode followed by a filtered amplifier. This would produce a noise signal at high power, of the specified channel. This was followed by a narrow band-stop filter, which blocked the noise signal in the middle of the channel. When this signal was applied to the system component, distortion of the amplifier could be seen in the notch of the NPR signal. Figure 1.10 shows an example NPR signal, after passing through an amplifier. This is not one created by a noise diode, but rather using an arbitrary waveform generator, which is programmed to produce an additive-white-gaussian-noise (AWGN) signal with a notch at its center. In fact, the use of noise diodes to produce NPR signals has been essentially replaced throughout the industry with arbitrary waveform-generated signals. In this example, the AWGN signal is created in a low-frequency baseband generator

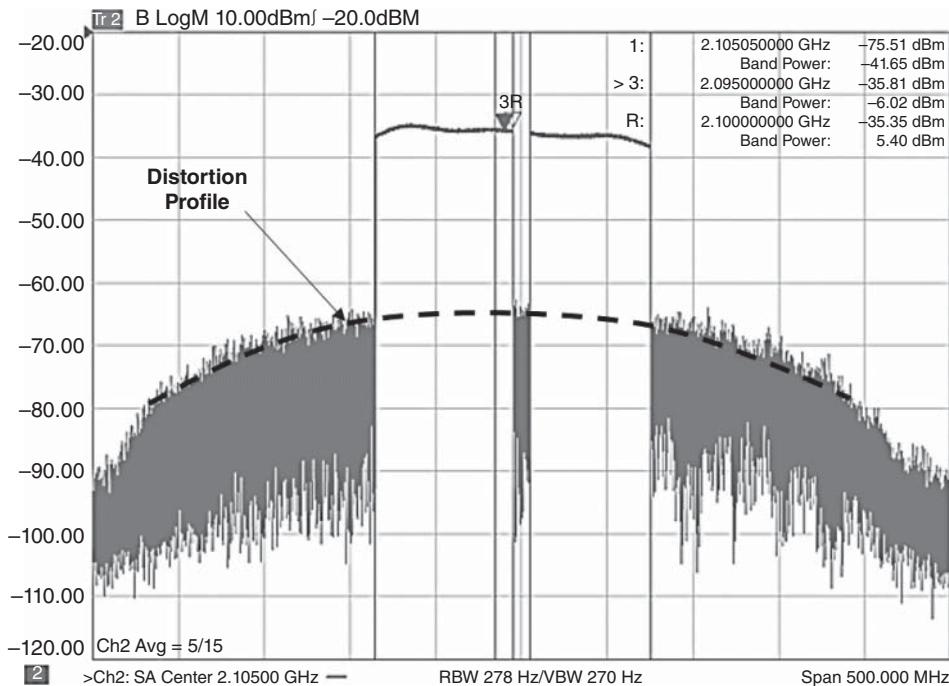


Figure 1.10 An NPR signal showing the total power and ratios of band power.

and then upconverted inside the signal source to the desired center frequency. Some unflatness is apparent in the passband of the signal due to frequency response of the signal source.

Apparent in the figure is also the ACLR level, which is nearly the same at the edge of the main signal as the NPR signal in the middle. It is clear from this figure that ACP and NPR are closely related. Imagine, though, if the DUT is followed by a sharp channelizing filter; the ACLR would be removed by the filter and could not be used to determine the distortion but the NPR signal allows one to see the in-channel distortion. NPR measurements are covered extensively in Chapter 8.

1.6.6 Error Vector Magnitude (EVM)

Error vector magnitude (EVM) is a figure of merit used in communications systems to describe the quality of a modulated signal compared to an idealized signal. In most cases, it is a measure in the so-called IQ plane of the vector difference between the measured signal and the idealized signal, which is determined by recovering the modulation pattern from the measured signal and re-creating the idealized signal. It is used when the errors are small and becomes inaccurate with large errors as the recovered signal may not be the correct signal when the EVM is quite large.

The sample point for determining the error is determined by the signaling method, and the idealized signal is time-shifted to line up with the measured signal to find the difference at the sample point. In some signaling methods, the EVM is determined by taking the fast Fourier transform (FFT) of the modulated signal and idealized signal and measuring the vector difference in the frequency domain.

EVM is affected primarily by distortion of the channel (usually in the transmitter amplifier), nonuniform frequency response (ripples or roll-off in the channel components), and noise in the system. For a transmitter component, which is the principal contributor to EVM, the noise contribution is generally not significant. In many modulation schemes, such as orthogonal frequency domain multiplexing (OFTM), the signal is broken into many narrow channels, such that the frequency response changes are small over each channel, and thus frequency flatness errors don't contribute to the EVM in these modulation schemes. In other cases, the measurement receiver has the ability to apply frequency response compensation, a kind-of inverse filtering, to remove the effects of the nonideal frequency response. This is sometimes called *equalization*, and the EVM measurement is called *equalized EVM*. After equalization, the frequency response does not contribute significant errors to the EVM signal.

This leaves only distortion as the predominant contribution to EVM, and as such EVM has become a common figure of merit for distortion in these systems. EVM measurements generally require a full demodulation to evaluate the signal quality, and at this time such capabilities are not generally available in VNAs, but this is likely to change as EVM becomes a significant figure of merit in more systems.

Recently, several papers have been presented (Sombrin 2011; Freiberger et al. 2017) that demonstrate a corresponding relationship between EVM and NPR. These works are compelling and lead one to infer that with further development, the time is near when EVM can be determined without the need for full demodulation, as illustrated in Chapter 8.

1.7 Characteristics of Microwave Components

Microwave components differ from other electrical devices in a few respects. The principal discerning attribute is the fact that the components' size cannot be ignored. In fact, the size of many components is a significant portion of a wavelength at the frequency of interest. This size causes the phase of the signals incident on the device to vary across the device, implying that microwave devices must be treated as distributed devices. A second, related attribute is that the reference ground for the device is not defined by a point but is distributed as well. Indeed, in many cases the ground is not well defined. In some situations, grounds for a device are isolated by sufficient distance that signal propagation can occur from one device ground to another. Further, even if devices are defined as series only (with no ground contact), one must realize that there is always an earth ground available so there can always be some impedance to this ground. In practice, the earth ground is actually the chassis or package of the device, or a power or other ground plane on a printed circuit board (PCB).

Finally, only in microwave components can one find the concept of wave propagation. In waveguide components, there is no "signal" and no "ground." Rather, a wave of electric-magnetic (EM) field is guided into and out of the device without regard to a specific ground plane. For these devices, even the transmission structures, waveguide for example, are a large percentage of a signal wavelength. Common concepts such as impedance become ambiguous in the realm of waveguide measurements and must be treated with special care.

1.8 Passive Microwave Components

1.8.1 Cables, Connectors, and Transmission Lines

1.8.1.1 Cables

The simplest and most ubiquitous microwave components are transmission lines. These can be found in a variety of forms and applications, and they provide the essential glue that connects the components of a microwave system. RF and microwave cables are often the first exposure an engineer has to microwave components and transmission systems, the most widespread example being a coaxial cable used for cable television (CATV, aka Community Antenna TeleVison).

The key characteristics of coaxial cables are their impedance and loss. The characteristics of coaxial cables are often defined in terms of their equivalent distributed parameters (Magnusson 2001), as shown in Figure 1.11, described by the *telegraphers' equation*

$$\frac{dv(z)}{dz} = -(r + j\omega l) \cdot i(z) \quad (1.73)$$

$$\frac{di(z)}{dz} = -(g + j\omega c) \cdot v(z) \quad (1.74)$$

where $v(z)$, $i(z)$ are the voltage and current along the transmission line, and r , l , g , c are the resistance, inductance, conductance, and capacitance per unit length.

For a lossless cable, the impedance can be computed as simply

$$Z = \sqrt{\frac{l}{c}} \quad (1.75)$$

but it becomes more complicated when loss is introduced, becoming

$$Z_{lossy} = \sqrt{\frac{r + j\omega l}{g + j\omega c}} \quad (1.76)$$

In many applications, the conductance of the cable is negligible, particularly at low frequencies, so that the only loss element is the resistance per unit length, yielding

$$Z_{lossy} = \sqrt{\frac{r + j\omega l}{j\omega c}} \quad (1.77)$$

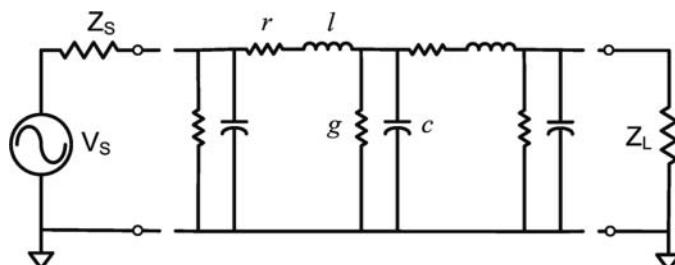


Figure 1.11 A transmission line modeled as distributed elements.

Inspection of Eq. (1.77) shows that the impedance of a cable must increase as the frequency goes down toward DC. Figure 1.12 demonstrates this with a calculation the impedance of a nominal $75\ \Omega$ cable, with a $0.0001\ \Omega\ mm^{-1}$ loss and capacitance of $0.07\ pF\ mm^{-1}$ (typical for RG 6 CATV coax). In this case, the impedance deviates from the expected value at 300 kHz by over $10\ \Omega$; and by $1\ \Omega$ at 1 MHz.

This low-frequency response of impedance for any real transmission line is often unexpected by those unfamiliar with Eq. (1.77), and it is sometimes assumed that this is a result of measurement error. However, all real transmission lines must show such a low frequency characteristic, and verification methods must take into account this effect.

An “airline” coax consists of a cable with an air dielectric, sometimes supported by dielectric beads at either end or sometimes supported only by the center conductor of the adjacent connectors, as shown in Figure 1.13. This type of cable has virtually no conductance, so series

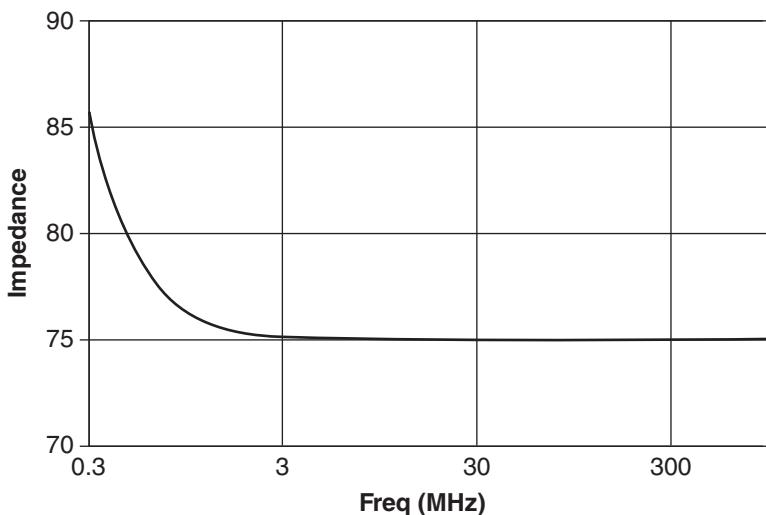


Figure 1.12 Impedance of a real transmission line at low frequency.



Figure 1.13 An airline coaxial transmission line.

resistive loss is the only loss element. The small white ring on the airline sometimes used to prevent sagging at the male end of the pin so that it may be more easily mated.

In some special applications, such as using measurements of a transmission line loaded with some material to determine the properties of the material, none of the elements of the telegraphers' equation can be ignored.

At higher frequencies, the loss of a cable is increased due to skin effect, which can be shown to increase as the square root of frequency (Collin 1966).

$$r = \sqrt{\frac{\omega\mu}{2\sigma}} \quad (1.78)$$

Thus, the insertion loss of an airline coaxial cable depends only upon the resistance per unit length of the cable, and so the insertion loss (in dB) per unit length, as a function of frequency, can be directly computed as

$$\begin{aligned} Loss(f) &= 8.68 \frac{r}{4\pi Z_0} \left(\frac{1}{R_a} + \frac{1}{R_b} \right) \\ &= A \cdot f^{1/2} \end{aligned} \quad (1.79)$$

where R_a and R_b are the inner and outer conductor radius and r contains the square root of frequency. Thus, all the attributes can be lumped into a simple single loss-term, A . Figure 1.14 shows the loss of a 10 cm airline as well as the idealized loss, as described in Eq. (1.79), where good agreement to theory is seen. However, the introduction of dielectric loading of the coaxial line will add some additional loss due to the loss tangent of the dielectric.

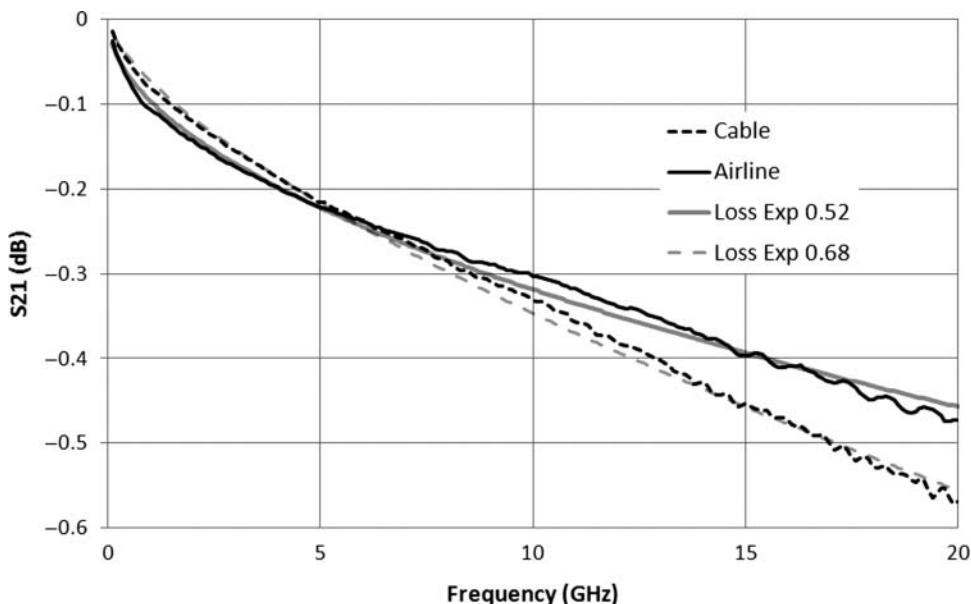


Figure 1.14 Loss of a 15 cm airline and a 15 cm semi-rigid Teflon-loaded coaxial line.

This additional loss often presents itself as an equivalent conductance per unit length, and this loss is often more significant than the skin-effect loss. Because of dielectric loss, the computed loss of (1.79) fails to fit many cables. The equation can be generalized to account for differing losses by modifying the exponent to obtain

$$\text{Loss}(f) = A \cdot f^b \quad (1.80)$$

where the loss is expressed in dB, and A and b are the loss factor and loss exponent. From the measured loss at two frequencies, it is possible to find the loss factor and loss exponent directly, although better results can be obtained by using a least-squares fit to many frequency points. Figure 1.14 shows the loss of a 15 cm section of 0.141 in. semi-rigid coaxial cable. The values for the loss at one-fourth and three-fourths of the frequency span are recorded. From these two losses, the loss factor and exponent are computed as

$$L_1 = A \cdot (f_1)^b, L_2 = A \cdot (f_2)^b$$

Taking the log of both sides, this can be turned into a linear system as

$$\begin{aligned} \log(L_1) &= \log(A) + b \cdot \log(f_1) \\ \log(L_2) &= \log(A) + b \cdot \log(f_2) \end{aligned} \quad (1.81)$$

This system of linear equations can be solved for the loss factor A and the loss exponent b .

$$A = \exp\left(\frac{\log(f_1) \cdot \log(L_2) - \log(f_2) \cdot \log(L_1)}{\log(f_1) - \log(f_2)}\right) \quad (1.82)$$

$$b = \frac{\log(L_1) - \log(L_2)}{\log(f_1) - \log(f_2)} \quad (1.83)$$

The computed loss for all frequencies from Eq. (1.80) is also shown, with remarkably good agreement to the measured values over a wide range. Ripples in the measured response are likely due to small calibration errors, as discussed in Chapter 5.

The insertion phase of a cable can likewise be computed; in practice, a linear approximation is typically sufficient, but the phase of a cable will vary with frequency beyond the linear slope due to loss as well.

The velocity of propagation for a lossless transmission line is

$$v = \frac{1}{\sqrt{l \cdot c}} \quad (1.84)$$

The impedance of a lossy cable *must* be complex from Eq. (1.77), and thus the phase response must deviate from a pure linear phase response, due to the phase velocity changing with frequency at lower frequencies. A special case for airlines, which have no dielectric loss, is

$$v_{prop} \approx \sqrt{\frac{2\omega}{rc}} \Big|_{\omega \cdot l \ll r} \quad (1.85)$$

For cables in general, the dielectric loss will cause a deviation in the velocity of propagation similar to that seen for loss. So far the discussion has focused on ideal low-loss cables,

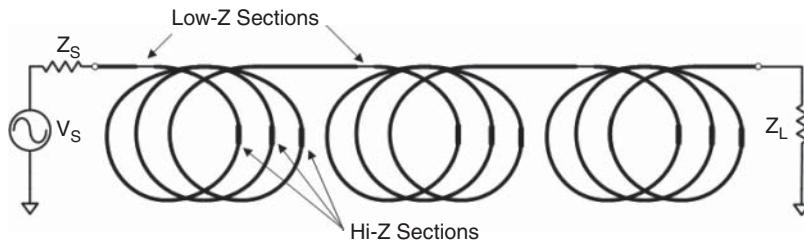


Figure 1.15 A model of a coax line with periodic impedance disturbances.

but in practice cables have defects that cause the impedance of the cable to vary along the cable. If these defects are occasional, they cause little concern and are typically overlooked unless they are so large as to cause a noticeable discrete reflection (more of that in Chapter 5). However, during cable manufacturing it is typical that the processing equipment contains elements such as spooling machines or other circular equipment (e.g. pulleys, spindles). If these have any defects in the circularity, or even a discrete flaw like a dimple, it can cause minute but periodic changes in the impedance of the cable. A flaw that causes even a one-tenth Ω deviation of impedance periodically over a long cable can cause substantial system problems called *structural return loss* (SRL), as shown in Figure 1.15. These periodic defects add up all at one frequency and can cause very narrow (as low as 100 kHz BW) very high return loss peaks, and thereby cause insertion loss dropouts at these same frequencies. In practice, the SRL test is the most difficult for low-loss, long-length cables such as those used in the CATV industry. Figure 1.16 shows a simulation of a SRL response caused by a 15 mm long, 0.1Ω impedance variation, every 30 cm, and another -0.1Ω variation every 2.7 m, each on the same 300 m coaxial cable with an insertion loss typical for main-line CATV cables. In the figure, two SRL effects are shown; a smaller effect every 50 MHz or so, due to the 2.7 m periodic variation, and a much higher effect every 500 MHz or so due to the 30 cm impedance variation. The higher impedance variation occurs more often, and so the periodic error will have a greater cumulative effect resulting in a nearly full reflection, as shown in the figure.

1.8.2 Connectors

Connectors provide the means to transition from one transmission media to another. They are often not considered as part of the device or measurement system, but their effects can dominate the results of a measurement, particularly for low-loss devices. Connectors can be distinguished by the quality and application. One remarkable aspect of connectors is the great difficulty in measuring them with any kind of accuracy. This is because most connectors provide a transition between different media, such as from a coaxial cable to a connector interface or from a PC board to a connector interface. While the connector interface is often well defined, the “back-end” of the connector is poorly defined.

Connectors that are “in series” provide transitions from male to female and provide interconnections between components. These are easiest to characterize as the ports are well defined and typically calibration kits are available and calibration methods are well understood. Connectors that are “between series” are equally well defined, but until recently they have been difficult to characterize as there were not well-defined standards for between-series

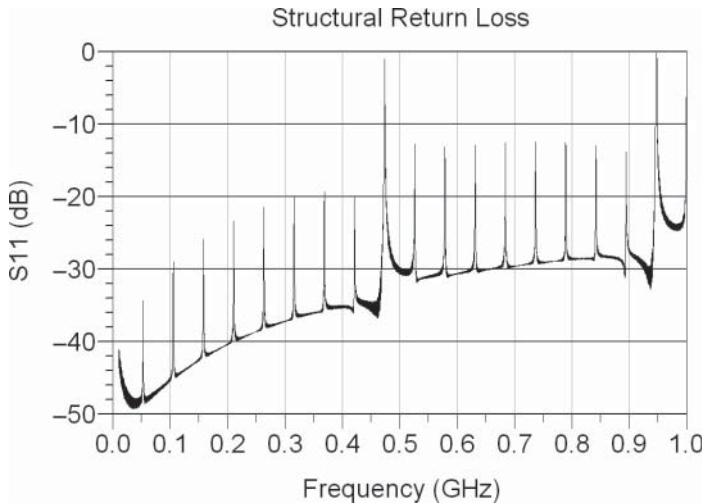


Figure 1.16 The return loss of a line with structural return loss.



Figure 1.17 In-series and between-series connectors.

adapters. Recent improvements in calibration algorithms have essentially eliminated any difficulty with characterizing these between-series adapters. Figure 1.17 shows some examples of in-series and between-series connectors.

For microwave work, there are some commonly utilized connector types that are found on the majority of components and equipment. Table 1.1 lists these common connectors along

Table 1.1 Test connectors used for RF and microwave components

Name	Outer Conductor Diameter (mm)	Rated Frequency (GHz)	First Mode (GHz)	Maximum Usable Frequency (GHz)
Type-N ($50\ \Omega$) precision	7	18	18.6	26.5 ^a
Type-N ($50\ \Omega$) commercial	7	12	12.5	15
Type-N ($75\ \Omega$) precision	7	18	18.6	18
Type-N ($75\ \Omega$) commercial	7	12	12.5	15
7 mm	7	18	18.6	18
SMA	3.5	18	19	22
3.5 mm	3.5	26.5	28	33
2.92 mm ("K")	2.92	40	44	44
2.4 mm	2.4	50	52.5	55
1.85 mm ("V")	1.85	67	68.5	70
1 mm	1	110	120	130

^aSome instrument manufacturers place this connector on 26.5 GHz instruments because it is rugged; it has the same first modes as Type-N and 7 mm.

with their normal operating frequency range. These are divided into three broad categories: precision sexless connectors, precision male/female connectors, and general-purpose or utility connectors. These connectors are typically $50\ \Omega$, but a few can be found as $75\ \Omega$ versions as well.

From Table 1.1 one can see that there are three frequencies associated with connectors: the generally understood operating frequency (often dictated by the calibration kit's maximum certified frequency), the frequency of the first mode, and the maximum frequency determined by the waveguide propagating mode of the outer conductor. The operating frequency is always below the first mode and usually by several percent. The first mode in many connectors is due to the support structure for the center pin. It is often made of some plastic material and thus has higher dielectric constant and a lower frequency to support a mode. In connectors and cables, *modes* is the term used to refer to non-transverse-electromagnetic (non-TEM) propagation that can occur in a circular waveguide mode defined by the inside dimension of the outer conductor. Adding dielectric in the bead that supports the center pin can theoretically lower the mode frequency, but if the bead is short, the mode will be evanescent (non-propagating) and may not affect the quality of the measurement. At a somewhat higher frequency, there will be a propagating mode in air for the diameter of the center conductor, but if the cable attached to the connector is sufficiently small, this mode may not propagate as well. It is the propagating modes that cause the significant dips in the response, and more importantly, these dips cannot be removed with calibration because they are not localized and because reflections in the mode of transmission far removed for the connector interface can interact with these connector modes, causing the frequency response of the mode effect to change when different devices are connected. If the response of the mode does not change when other devices are connected, it can be calibrated out.

The precision sexless connectors are now found only in metrology labs. Their chief benefit was a repeatable connector that has identical characteristics for each connector. As such, it was easy to create a system calibration, and any part with such connectors could be inserted

between two cables in either direction. This was important because in the past it was difficult to deal with “non-insertable” devices from a calibration sense (a non-insertable device is one with the same sexed connector on each port, e.g. female-female). The 7 mm connector is often found on precision attenuators and airlines used as transfer standards. The 7 mm connector is also known as the GPC-7 for general precision connector, and often as the APC-7™ for amphenol precision connector. Because these connectors are sexless, there is no need for adapters to provide interconnections between devices or between devices and cables.

1.8.2.1 7 mm Connector (APC-7, GPC-7)

The 7 mm connector has a couple of interesting attributes: the center pin has no slots but contains spring-loaded center collets that protrude slightly from the mating surface, as shown in Figure 1.18.

When mated, the collet from each connector floats against each other, providing a good center contact. There is a slight gap in the slotless outer sleeve of the center pin. As with almost all RF connectors, the outer conductor forms the physical mating plane. On most connectors, there is a slip-ring-threaded sleeve surrounded by a coupling nut. To mate, the threaded sleeve is extended on one connector and retracted on the other. On the retracted connector, the coupling nut is extended to engage the other’s sleeve and is tightened. Only one coupling nut should be tightened, although it is common but incorrect practice to tighten the other coupling nut. In fact, tightening both coupling nuts can result in the center pins pulling apart and a poorly matched contact. Occasionally, one sees parts that contain only a solid threaded outer conductor (serving the purpose of the threaded sleeve) and no coupling nut. These are more common on older test fixtures intended to mount directly the 7 mm connectors of network analyzer test sets.

1.8.2.2 Type-N 50 Ω Connector

The Type-N connector is common in lower-frequency and higher-power radio frequency (RF) and microwave work. It has the same outer diameter (7 mm) as the 7 mm connector but is sexed. In fact, this connector has the unusual attribute of having the mating surface for the



Figure 1.18 A 7 mm connector.

outer conductor (which is almost always the electrical reference plane) recessed for the female connector. Thus, the female pin protrudes (in an electrical sense) from the reference plane, and the male pin is recessed. Thus, the calibration standards associated with Type-N connectors have electrical models that are highly asymmetric for male and female standards.

The Type-N connector has precision forms, including ones with slotless connectors (metrology grade), ones with precision six-slotted collets and solid outer conductor sleeves (found on most commercial test equipment), and commercial forms with slotted outer conductor sleeves and four or even two slotted female collets. Slotless connectors have a solid hollow cylinder for the female connector with an internal four- or six-finger spring contact that takes up tolerances of the male center pin. As such, the diameter of the female center pin does not depend at all on the radius of the male pin. Typical female contacts with collets expand or contract to accept the male pin, and thus their outer dimension (and thereby their impedance) varies with the diameter tolerance of the male pin.

The commercial forms are found on a variety of devices and interconnect cables. The male version of these commercial-grade parts present two common and distinct problems: there is often a rubber “weather-seal” o-ring in the base of the connector, and the outer nut of the male connector is knurled but has no flats to allow using a torque wrench. The first problem exacerbates the second, as the mating surface of the outer conductor of the male connector is often prevented from contacting the base of the female connector because the outer (supposedly non-mating) surface of the female connector touches the rubber o-ring and prevents the male outer conductor from making full contact. If one can fully torque a Type-N connector, the rubber o-ring would compress, and the contact of the male outer conductor would occur, but as there are no flats for a torque wrench, it is difficult to sufficiently torque the Type-N connector to get good repeatable connections. This one issue is the cause of hundreds of hours of retest when components don’t pass their return-loss specs. The solution is quite simple: remove the rubber o-ring from the base of the male connector, always, before any measurement. A pair of tweezers and a needle-nose pliers are indispensable for the process of removing this annoying o-ring. One will note that none of the precision versions of Type-N connectors contains such an o-ring. Figure 1.19 shows some examples of Type-N connectors; the upper two are commercial grade, and the lower two are precision grade. Figure 1.20 shows the insertion loss measurement of a male-to-male Type-N adapter mated to a female-to-female Type-N adapter for a precision pair and a commercial-grade pair, where the loss is normalized to the length of the adapter. The commercial-grade pair is operational only to about 12 GHz, due to moding in the connector. The precision N is mode free beyond 18 GHz.

1.8.2.3 Type-N 75Ω Connector

Type-N connectors also have a 75Ω version, which has the same outer dimensions but a smaller center conductor. This is in some ways unfortunate as the smaller female collet of the 75Ω version can be damaged when inserted with a 50Ω male pin. There are a couple of versions of the 75Ω female collet, one with short slots and six fingers, and one with long slots and four fingers. A precision slotless version is also available. The short slot version has the potential for better measurements, as the slots expand less so there is less uncertainty of the open capacitance. However, on many products with 75Ω N-connectors, the long slot connector is used; the long slots were designed to accept a 50Ω male pin, at least for a few



Figure 1.19 Examples of Type-N connectors: commercial (upper) and precision (lower).

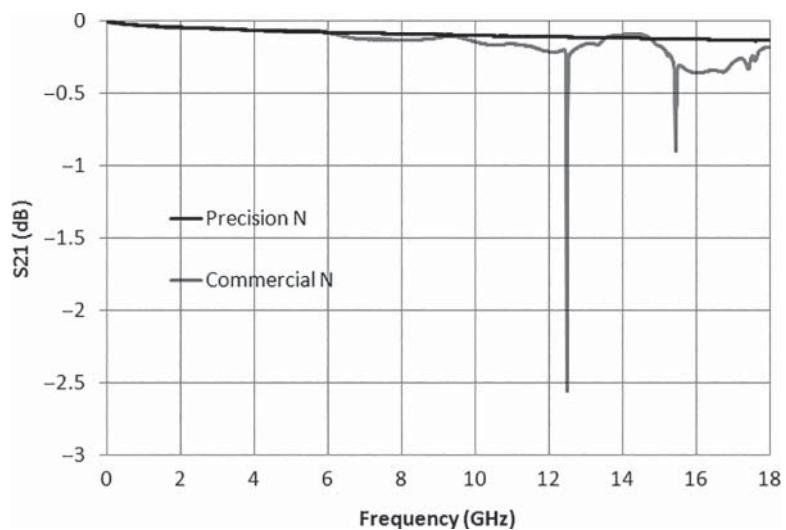


Figure 1.20 Performance of a precision and a standard Type-N connector.



Figure 1.21 $75\ \Omega$ Type-N connectors: commercial (upper) and precision (lower).

insertions, without damage. Often the $75\ \Omega$ components have an extra machined ring or line on the outer nut to help identify it. Versions of $75\ \Omega$ Type-N connectors are shown in Figure 1.21. An example of the insertion loss measurement of a mated pair of a male-to-male adapter with a female-to-female adapter is shown in Figure 1.22, where the loss is normalized for length of the adapter. The frequency limit of Type-N 75 is often stated as 2 or 3 GHz, but that is because the commercially available calibration kits were rated only to those frequencies. In practice, these connectors could be used up to 7 or 8 GHz without difficulty. The response of the commercial-grade connector is likely limited not due to moding (since the loss signature is quite low Q) but rather due to poor impedance control in the center pin support bead, causing impedance mismatch.

1.8.2.4 3.5 mm and SMA Connectors

The 3.5 mm connector is in essence half the scale of the N connector and provides higher-frequency coverage. The center pin of the 3.5 mm connector is supported by a plastic bead, rather than solid dielectric, meaning it has mode-free operation to a much higher frequency than Type-N. Traditionally, 3.5 mm connectors are specified up to 26.5 GHz,

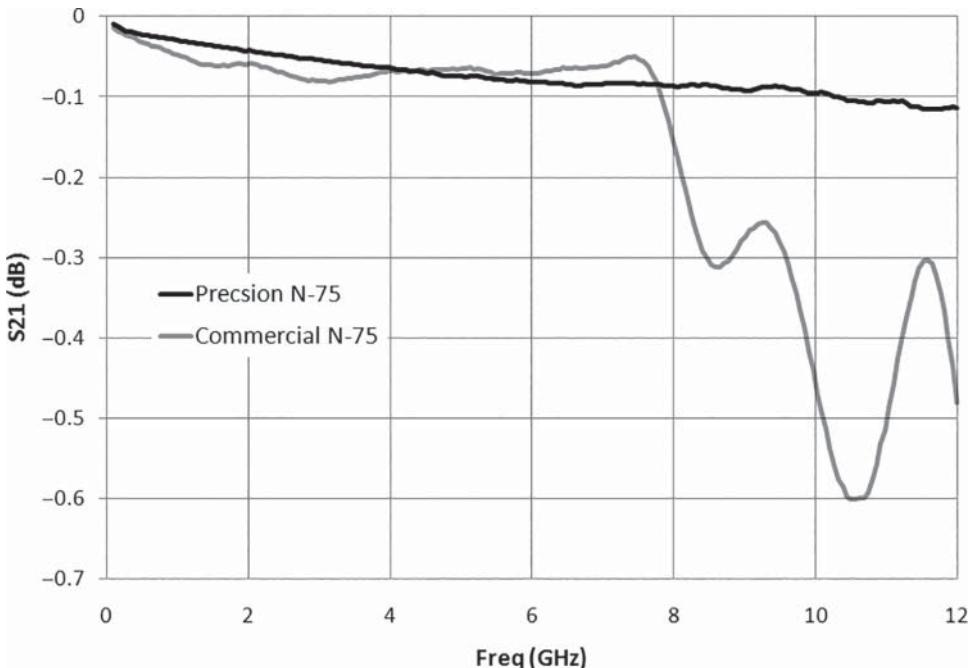


Figure 1.22 Insertion loss of 75Ω connectors.

but their first mode is nearly 30 GHz, and they are functional up to about 38 GHz. An interesting aspect of modes is that the first mode of a 3.5 mm connector is due to the bead (and its increased effective dielectric), but this mode is non-propagating, so it is reasonable to use these connectors to even higher frequencies. The 3.5 mm female connector comes with several versions of center pin, the main varieties being a four-slot collect and a slotless precision connection, found now on most calibration kits. Interestingly, even though the slotless connectors may have the center spring contact damaged due to oversized or misaligned male pins (under the microscope one or more fingers may be crushed back into the hollow of the female pin), the RF performance is almost unaffected due the robust solid outer conductor. In fact, one typically can tell if a slotless connector is damaged only by visual inspection, as the RF performance is substantially unchanged, as long as even one finger is left to make contact.

The SMA connector is mechanically compatible with the 3.5 mm connector but has a solid Teflon dielectric and thus a lower operating frequency due to moding. SMA is traditionally considered to be an 18 GHz connector, but the first propagating mode is well above 20 GHz, depending upon the type of cable that is connected to the SMA connector. The chief advantage of SMA connectors is low cost, especially when mounted to semi-rigid coaxial cables. The dimensions are such that the center wire of the coax can be used a connector pin for SMA, and only an outer conductor sleeve needs to be added to the coax outer conductor to form a male connector, shown in the lower-right picture of Figure 1.23. But these cables are notoriously bad at maintaining the proper dimensions for the center pin, and often the center pins are poorly trimmed and improperly chamfered so that they cause mating problems with their



Figure 1.23 3.5 mm (f) and (m) (upper left); SMA (f) and (m) connectors (upper right); 3.5 mm (lower left) and SMA adapters (lower right).

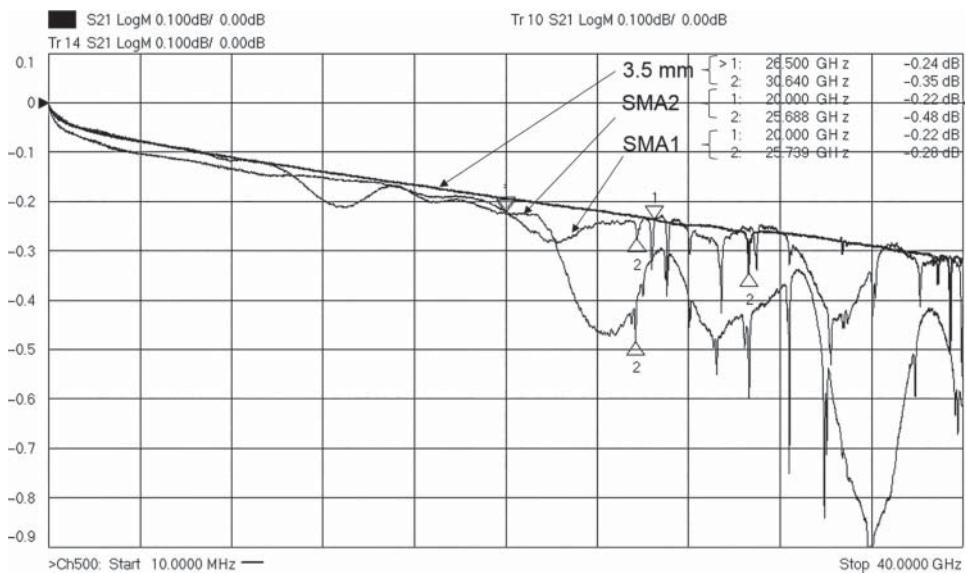


Figure 1.24 Performance of SMA and 3.5 mm mated-pair connectors.

female counterparts. This is particularly true when mating them to 3.5 mm female connectors, slotless ones in particular. Figure 1.23 shows examples of 3.5 mm and SMA connectors, with 3.5 mm on the left and SMA on the right.

Figure 1.24 shows measurement plots of a mated pair of 3.5 mm male-to-male with a 3.5 mm female-to-female, as well as two SMA examples. The moding of the SMA connector

is clearly seen above 25 GHz (Marker 2 on the SMA1 and SMA2 trace). The moding of the 3.5 mm connector is seen just above 30 GHz (Marker 2) and again at 34 and 38 GHz. There are two typical construction types for SMA, one with a press-fit of the Teflon and center conductor (SMA1 in the measured response) and one where the Teflon is held in with a small dot of epoxy through a hole in the outer conductor (SMA2 in the measured response). The second method usually gives a poorer match, and we can see that with the small dip in the S21 response of SMA2 near 12 GHz and the larger dip just above 20 GHz.

1.8.2.5 2.92 mm Connector

The 2.92 mm connector is scaled down from the 3.5 mm connector and can be mechanically mated to both the 3.5 mm and the SMA connectors. The smaller diameter outer conductor means that its mode-free operation extends proportionally higher, to 40 GHz, and is usable to perhaps 46 GHz. The female connector has a two-slot collet that provides sufficient compliance to mate with the center pin of the larger 3.5 mm and SMA connectors but that makes it less suitable for precision measurements due to increased uncertainty of the contact point on the center pin radius, which now depends upon the radius of the pin that is inserted. A further point is that the metal wall of the female collet on the 2.92 connector is quite thin and prone to damage if the mating pin is not well aligned or oversize. It's not uncommon to find 2.92 female adapters missing one of the collet fingers. The 2.92 mm connector was popularized by the Anritsu company (formally Wiltron), which introduced it as the K connector, and it is common to hear any 2.92 mm connectors referred to by that name.

Figure 1.25 shows some examples of 2.92 connectors. The key difference is in the diameter of the inside of the outer conductor. Figure 1.26 shows the insertion loss of a mated pair of 2.92 mm female-to-female adapters with a 2.92 male-to-male adapter, along with an example of a 3.5 mm mated adapter pair. The moding of the 3.5 mm pair is clearly seen above 30 GHz, but the connector is generally usable up to 38 GHz as the first small modes are bead modes and are able to be calibrated out as they generally don't propagate through the cable.

1.8.2.6 2.4 mm Connector

The 2.4 mm connector is essentially a scaled version of the 3.5 mm connector, with an associated scaling in maximum frequency. It is used extensively on 50 GHz applications, though it can be used up to 60 GHz. This connector cannot be mated to any of the SMA, 3.5 mm or 2.92, and in fact was designed to prevent damage if one tried to mate to these types. It comes with both slotted and slotless female center pins, much like the 3.5 mm connector.

1.8.2.7 1.85 mm Connectors

There are two variants of the 1.85 mm connectors, designed originally by Anritsu and Agilent. The Anritsu variety is called the V connector, and the Agilent variety is called the 1.85 mm connector. They are mechanically compatible and were originally designed for 67 GHz operation, usable to above 75 GHz. These connectors are mechanically compatible



Figure 1.25 A 3.5 mm connector compared with 2.92 mm female (upper) and male (lower).

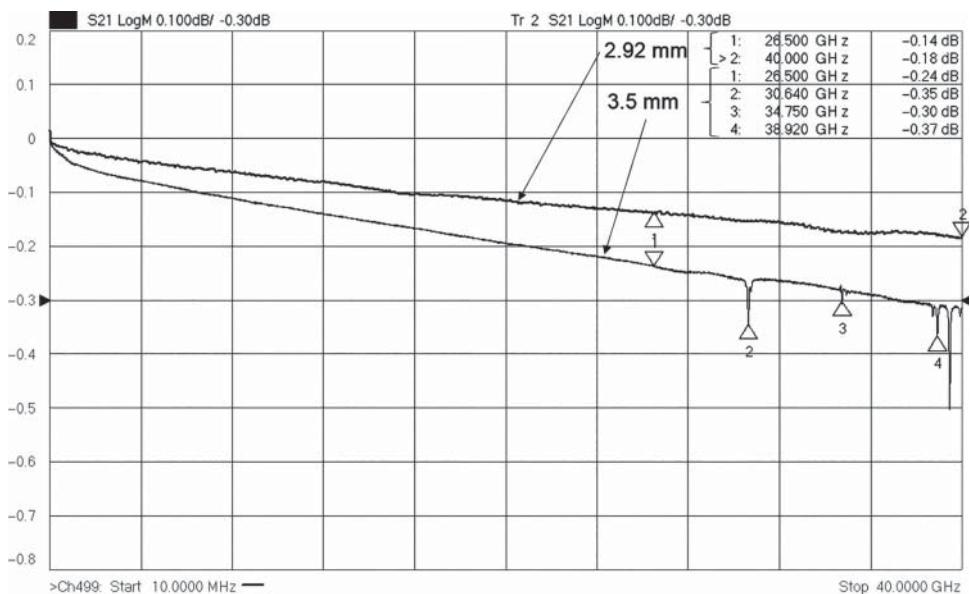


Figure 1.26 Performance of a mated pair, 2.92 compared with 3.5 mm.

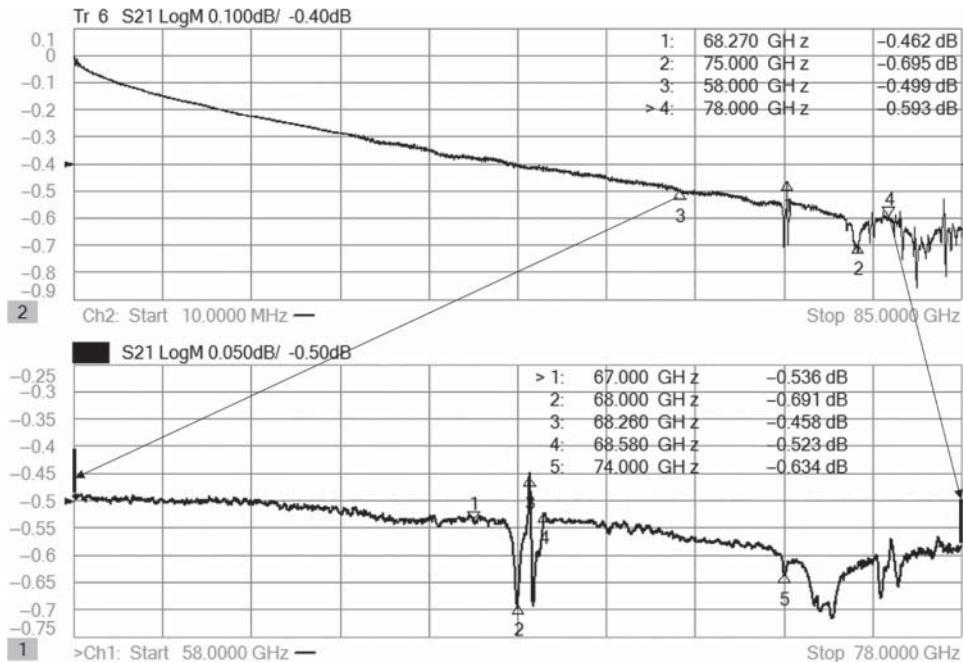


Figure 1.27 Response of mated pair of male-to-male and female-to-female 1.85 mm connectors (upper), with zoomed-in view of the first mode (lower).

with the 2.4 mm connector. Figure 1.27 shows the wideband response to 85 GHz of the 1.85 mm connector in the upper plot and a zoomed-in view of the first mode in the lower plot. This first mode is a “bead-mode” and is caused by the increased dielectric constant of the bead, which holds the center pin, lowering the first mode of the 1.85 mm coaxial line. In general, this mode is non-propagating (since it is contained in the bead) and may be calibrated out in some circumstances. For example, if this connector is used with an on-wafer probe and the coax from the connector to the probe-tip is mode free, then the bead mode will act like a small, stationary resonance that can be removed with a calibration. If the mode is propagating, then changes in the termination impedance change the effects of the mode, and it cannot be calibrated out (it is not stationary with respect to an external impedance); but if it is non-propagating and there is a sufficient length of mode-free line (such a cable) between this bead mode and the reference plane, the evanescent fields associated with the mode will die off before arriving at the reference plane and thus will not couple to the terminating impedance. As this first mode is less than 0.2 dB, in many cases it is not significant. The 1.85 mm connector has been used out past 75 GHz.

1.8.2.8 1 mm Connector

The 1 mm connector is essentially a scaled version of the 1.85 mm connector but cannot be mated to it. It is typically specified to 110 GHz performance but is usable to above 120 GHz,

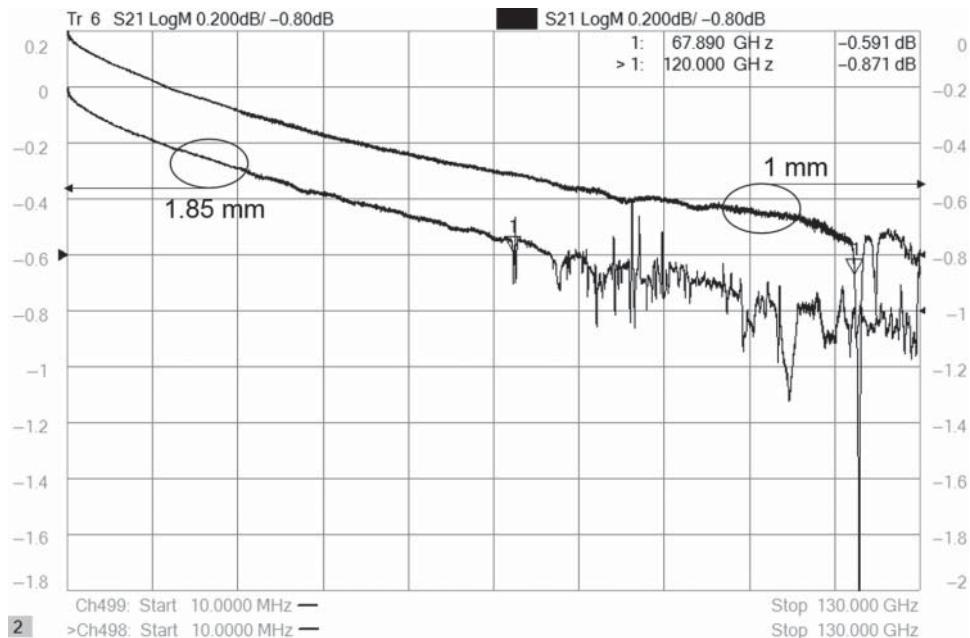


Figure 1.28 Response of a 1 mm mated pair and a 1.85 mm mated pair.

with some versions being specified up to 120 GHz and used up to 140 GHz. Figure 1.28 shows the response of a mated pair of 1 mm male-to-male and female-to-female adapters (right scale) along with the same 1.85 mm mated pair of Figure 1.27 (left scale). The reference line is offset by 1 division to make it easier to see the traces; without the offset, the traces would lie nearly on top of each other. The mode for 1 mm is a bit of a deeper mode but is now out past 120 GHz. There is a second mode at 124 GHz, but both are also non-propagating, so it may be possible to remove them with calibration. The depth of the mode does imply that it may be from the multiple beads in the mated pair and not be very stable with changes in temperature.

1.8.2.9 PC Board Launches and Cable Connectors

For many design and measurement applications, the circuit of interest is embedded in a PC board. There are many types and styles of PC board launches, which typically have an SMA connector on one end and PC board contacts at the other, as well as miniature versions such as the QMA connector. These can come in edge launch as well as right angle, and their performance depends greatly upon the mounting pattern on the PC board trace. These can be difficult to characterize because only one end is available in a standard connector. An example of a common PC board launch is shown in Figure 1.29. Measurement techniques for these devices, as well as methods to remove their effects from the measurement of on-board PC components, are discussed in Chapter 11.

Connectors designed for coaxial-cables provide similar challenges, as the cable to which they are attached affects the quality of the connection, and the common practice of attaching

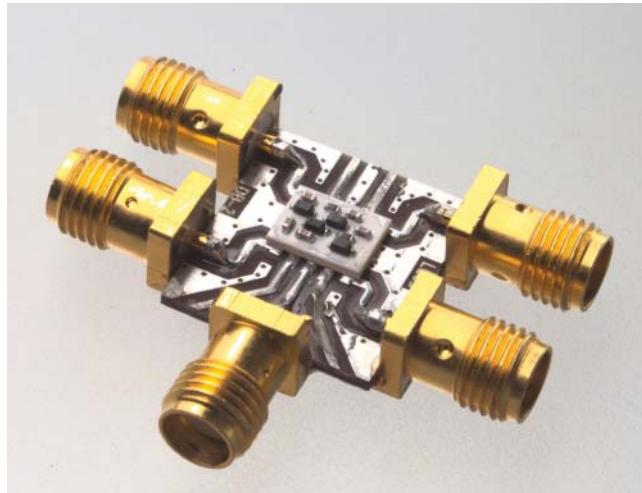


Figure 1.29 PC board SMC launches.

two connectors to each end of cable makes it difficult to separate the effects of one from the other. Time-domain techniques can be applied to remove these unwanted effects, as described in Chapter 5.

1.8.3 Non-coaxial Transmission Lines

Transmission lines provide the interconnection between components, typically in a microcircuit or a PC board. These are distinguished from a measurement perspective because they are typically much shorter, often not shielded, and the interface to them is not easy to make and sometimes not well defined. While there have been whole books written on the subject, a short review of some common transmission line structures and their attributes are described next, with a focus on attributes important to measurement. Transmission lines are characterized by the same three parameters: impedance, effective dielectric constant, and loss.

1.8.3.1 Microstrip

Certainly the most widespread transmission line must be the microstrip line, shown in Figure 1.30. This is found in planar structures such as PC boards and micro-circuits. Consisting of a thin strip of metal on a dielectric substrate, over a ground plane, it is used for connection between components as well as creating transmission line components such as couplers and filters (Hong and Lancaster 2001).

The computation of the transmission parameters has been fully documented in many forms, but for measurement purposes these lines are typically 50Ω (or the equivalent system impedance) even though as a design element they can take on any value. For most applications, the dielectric constant is 10 or less, so the w/h ratio is greater than 1 for 50Ω .

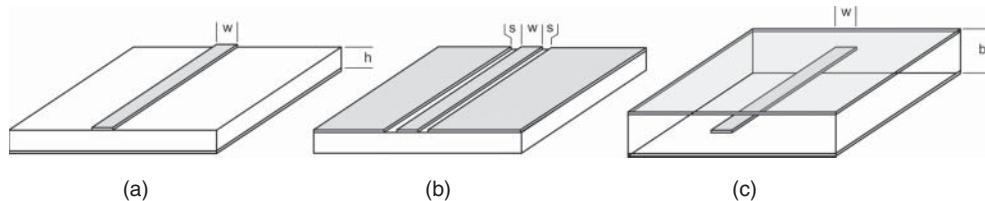


Figure 1.30 Planer transmission lines: microstrip (a), coplanar waveguide (b), strip line (c).

The approximate impedance can be computed as (Pozar 1990)

$$Z_{\mu\text{strip}} = \begin{cases} \frac{60}{\sqrt{\epsilon_{re}}} \ln \left(\frac{8h}{w} + \frac{w}{4h} \right) & \text{for } \frac{w}{d} \leq 1 \\ \frac{377}{\sqrt{\epsilon_{re}} \left[\frac{w}{h} + 1.393 + 0.677 \ln \left(\frac{w}{h} + 1.444 \right) \right]} & \text{for } \frac{w}{d} \geq 1 \end{cases} \quad (1.86)$$

where ϵ_{re} is the effective relative-dielectric-constant, found from

$$\epsilon_{re} = \left(\frac{\epsilon_r + 1}{2} \right) + \left(\frac{\epsilon_r - 1}{2} \right) \cdot \left(1 + 12 \frac{h}{w} \right)^{-1/2} \quad (1.87)$$

The effective relative-dielectric constant sets the velocity factor of the transmission line, but in microstrip, some of the fields travel in the substrate and some in air. Therefore, the transmission is not purely transverse-electromagnetic (TEM), and some structures become more difficult to design, particularly coupled lines, the even and odd mode velocity factors of which are not the equal. Since the line is not pure TEM, at high frequency, dispersion effects will become apparent where the effective delay of the line is not constant with frequency.

The loss of microstrip lines is difficult to compute accurately because it depends upon many factors including the conductivity of the microstrip line and the ground plane, the dielectric loss of the substrate, radiated loss to the housing or shield, and losses related to both surface roughness and edge roughness. These roughness losses can be significant in PC board and low-temperature cofired-ceramic (LTCC) applications and are dependent upon the particular processes used. While there are high-quality PC board materials (Duriod™ or GTEK™ are common trade names), the material known as FR4 is most common, and the dielectric constant and loss of this PC board material can be uncertain. The finished substrate can be comprised of layers of board material sandwiched together with glue, and the final thickness can depend upon processing steps, so it is best when evaluating microstrip transmission lines to produce sample structures that can help determine the exact nature of the material.

One high-performance material used is single-crystal sapphire, and it has the unusual property of having a dielectric constant that has a directionality, with a higher constant of 10.4 in one of the three dimensions, and a lower constant of 9.8 in the other two. A second, common high-performance dielectric is ceramic found in thin-film, thick-film, and LTCC applications. It has a uniform dielectric constant typically between 9.6 and 9.8 depending upon the purity and grain structure of the ceramic.

1.8.3.2 Other Quasi-Microstrip Structures

For many applications, the size of 50Ω microstrip line is not suitable for connections to very large devices. Some common modifications are *suspended substrate* microstrip line, where the ground plan has been removed some distance from the dielectric. This has the effect of lowering the effective dielectric constant and raising the impedance of the line. In this way, a wider line can be used to connect to a wide component and still maintain a matched impedance. A *shielded* microstrip line is entirely enclosed (the theoretical models of microstrip lines assume no top shield), and the top metal tends to lower the impedance of the line. This is particularly true for suspended microstrip lines.

1.8.3.3 Coplanar Waveguide

One difficulty with microstrip transmission lines is that the ground and signal conductors are on different physical planes. Coplanar waveguide (CPW), as the name implies, provides a coplanar structure of ground-signal-ground, as shown in Figure 1.30b. An alternative is grounded coplanar where the backside is a conductor as well, and in practice, all coplanar lines have associated package ground, but the ground may be ignored if there is a substantial air-gap between the substrate and the package ground. The references provide some computations of coplanar waveguide impedance for various configurations (Wen 1969; Simons 2001). In microwave measurements, CPW is used extensively as a contacting means for on-wafer measurements and is used to provide extremely low ground inductance for measuring microwave transistors and circuits, as shown in Figure 1.31, with either topside grounds (left) or backside grounds (right). Note that since the impedance depends only upon the scale of width to space, this allows contacts of large scale (such as probes) to be transitioned to small scale such as IC devices.

CPW has some inherent problems due to the ground being on a surface plane or sheet. In many instances the CPW line is mounted in a metal package, and the ground plane is grounded at the package wall. If the distance from the package wall to the ground plane edge approaches a quarter-wavelength at the frequency of interest, or multiples thereof, then a transmission line mode can form such that the ground of the CPW appears as an open relative to the package ground. This concept of “hot grounds” for CPW has been observed in many situations and

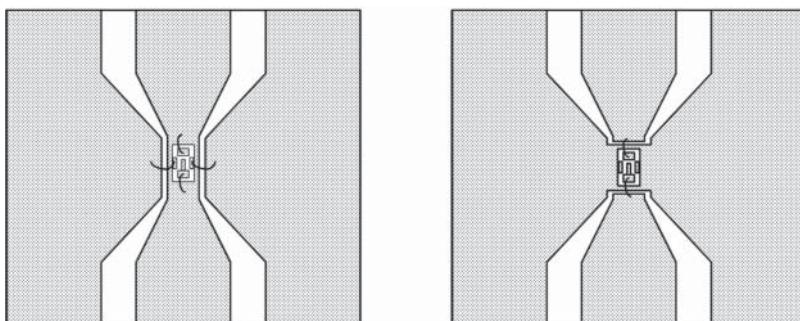


Figure 1.31 CPW-mounted IC.

is sometimes avoided by periodically grounding one side to the other through a small cross connection on the backside of the CPW. Another method is to provide a lossy connection to the sidewall ground through absorptive material or thin film resist material to suppress energy in the unwanted mode. Another alternative is treating the CPW as a suspended substrate only under the gap between ground and conductor and “stitching” the CPW ground to the backside ground through a series of conductive vias. The impedance of these structures is lowered by the added ground, so an adjustment of the center line width is usually made to accommodate the additional ground paths.

1.8.3.4 stripline

More common as a transmission line on an inner layer of a PC board, strip line consists of a thin strip or rectangle of metal sandwiched between two ground planes embedded in a uniform dielectric constant, as shown in Figure 1.30c. The impedance of these lines is much lower than the equivalent-width microstrip line, but they have an advantage of being fully TEM in nature and so often the design of components such as coupled lines is easier as the even- and odd-mode velocity factors are the same. An approximate formula for computing the value of a stripline impedance with a zero thickness strip is (Pozar 1990)

$$Z_{\text{stripline}} = \frac{30\pi}{\sqrt{\epsilon_r}} \frac{b}{W_e + 0.441b} \quad (1.88)$$

$$W_e = \begin{cases} w & \text{for } w/b > 0.35 \\ (w - (0.35 - w/b)^2 \cdot b) & \text{for } w/b < 0.35 \end{cases} \quad (1.89)$$

More complex formulas that include a broad range of applicability and include effects for finite strip thickness and asymmetric placement of the strip can be found in many references (IPC 2004; Cohn 1954).

1.9 Filters

Filters come in a variety of types including low pass, band pass, high pass, and band stop. Multiport filters form diplexers or multiplexers, which are used to separate or combine signals of different frequency from a common port to a port associated with the different frequencies of interest. Diplexers are sometimes called *duplicators*, but duplexing is a function of the operation of a communication system. That is, a system that can transmit and receive at the same time is said to be operating in a duplex mode. A diplexer is used to support the duplex operation by keeping the transmit signal from saturating the receiver.

The structure and variety of filters are almost endless, but they all share these common attributes: low loss in the pass band, low reflection in the pass band, high reflection, and high loss in the stop band. In nearly every case, the goal of the design is to minimize unwanted loss, and this quality of a filter is often referred to as the *Q* of the filter. In microwave cases, filters are designed to operate into a matched impedance, so there is always loss associated with power from the source being absorbed by the load. The Q of a filter in operation is fixed by the loading of the ports and can never be infinite. The quality of a filter is usually defined by its unloaded Q, which accounts for the (desired) power loss from the source to the load.

For many filters, the desired qualities are a trade-off between creating a maximally flat passband and creating a maximally sharp cutoff. Thus, the measurement of the transmission response of the filter is critical in evaluating the quality of a filter design. For most filters used in communications, the transmission responses is desired to be equally flat (rather than maximally flat) across the passband, resulting in filters that have Chevyshev-type response (equal ripple) in the passband (Zverev 1967). The desire for sharp cutoffs has led to many filters employing an elliptic response, which provides for finite zeros in the transmission response. Stopband performance of high-performance filters can also require careful consideration in measuring, with some requirements going beyond 130 dB of isolation over selected regions of the stop band. These extreme isolation requirements put tremendous burdens on the design of the filter, as well as the design and use of the measurement systems.

In modern communications systems using complex modulation, the phase response of the filters is also critical, and a significant design parameter is controlling the phase of the filter to follow a linear response, with a key measurement parameter being deviation from linear phase. Closely aligned to that is maintaining a constant group delay through the passband. Equalization techniques are utilized that can remove higher-order phase responses, such that another measure of filter phase response is deviation from parabolic phase, where the phase is fitted to a second-order response, and the deviation of the phase from this second-order response is the measurement criteria. Some filters are used as part of a feed-forward or matched system network where their phase response as well as absolute phase and delay must be carefully controlled.

The reflection response of filters is also a key measurement parameter. To the first order, any signal that is reflected is not transmitted so that high reflections lead to high transmission loss. However, the loss due to reflection for most well-matched filters is much less than the dissipation loss. Still, low reflections at the test ports are required to avoid excess transmission ripple from concatenated components, and even moderate reflections from filters in a high-power transmission path can cause damage to the preceding power amplifier. Thus, very low return loss is often a critical parameter of filters and also a difficult parameter to measure well. This becomes especially true in the case of diplex and multiplex filters, where the loading of any port affects the return loss of the common port.

For high-power applications, the filter itself can become a source of IM distortion, and the attribute passive inter-modulation (PIM) has become common in the measurement of these high-power filters. Poor mechanical contacts between components in a filter, poor plating on a filter, or the use of magnetic materials in the plating or construction of the filter can lead to hysteresis effects that cause IMD to be created in an otherwise passive structure. The level of IMD typically found in these filters is less than -155 dBc, but this can be a difficult spec to meet without careful design and assembly.

Most of these high-performance communication filters are designed using coupled-resonator designs (Cameron et al. 2007; Hunter 2001). Because of manufacturing tolerances, these filters cannot be manufactured to specification from the start; they require tuning of the resonators as well as the inter-resonator couplings. Techniques to optimize the response of these filters are highly sought and a key aspect of the filter measurement task, requiring fast precise response of the transmission and reflection response in real time.

Another type of filter commonly found in the intermediate frequency (IF) paths of receivers is a surface acoustic wave (SAW) filter. The frequency of these SAW filters has been steadily increasing, and they are sometimes found in the front end of a receiver. SAW filters can be

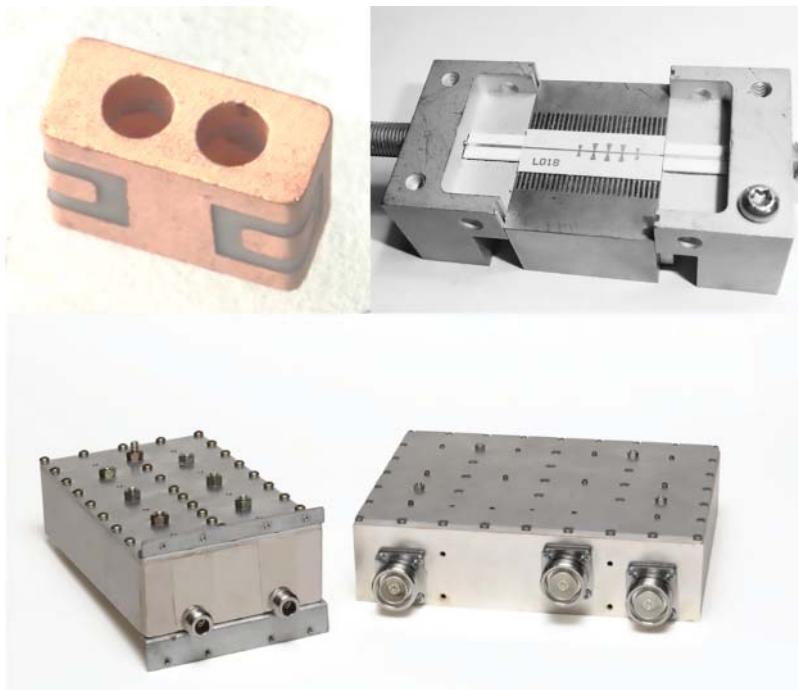


Figure 1.32 Examples of microwave filters: cellular phone handset filter (upper left), thin film filter (upper right), and cellular phone base station filters (bottom).

made to high orders and can have large delays (in the order of microseconds). Because of these long delays, special measurement techniques are required when attempting high-speed measurements. Another type of acoustic wave filters are the film bulk acoustic resonator (FBAR) filters, which are small in size and have been used as RF/TX duplexers in handset cell phones.

Ceramic coupled resonator filters are also used extensively in cell phone and radio applications. Because of manufacturing tolerances, the filters are often required to be tuned as part of the manufacturing process, and tuning consists of grinding or laser-cutting electrodes until the proper filter shape is obtained. This presents some difficulty in coupled resonator filters as the tuning is often “one way,” and once the resonator frequency has been increased, it cannot be reduced again. This has led to the need for high-speed measurements to ensure that the latency between measurement and tuning is as small as possible.

Some examples of filters are shown in Figure 1.32.

1.10 Directional Couplers

Directional couplers separate the forward and reverse waves in a transmission system (see Section 1.3). A directional-coupler is classically defined as a 4-port device, often with a good load on the fourth port, as shown in Figure 1.33; but in practice a load element is almost always permanently attached. The directional-coupler has four key characteristics: insertion

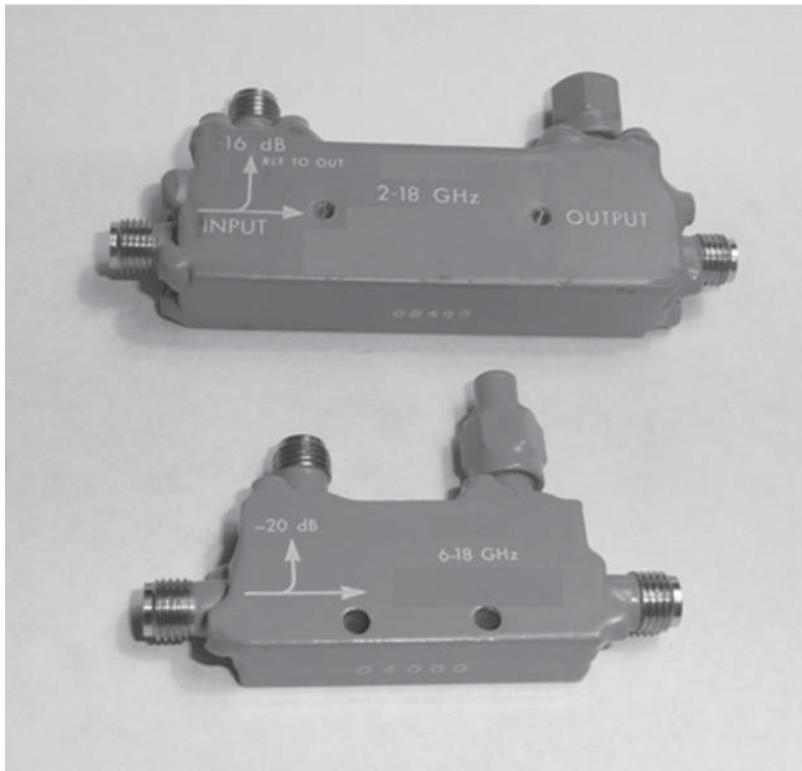


Figure 1.33 Directional couplers.

loss, coupling factor, isolation, and directivity. In fact, directivity is related to the other three factors in a specific way.

$$\text{Directivity} = \frac{\text{Isolation}}{\text{Coupling} \cdot \text{Loss}} \quad (1.90)$$

Most couplers have a nearly lossless structure so that the directivity is nearly equal to the isolation/coupling, but for lossy structures, such as directional bridges, the earlier definition provides the proper description. In fact, consider the case of a directional-coupler with 20 dB of coupling, 50 dB of isolation, and 0.05 dB of insertion loss, setting the directivity at nearly 30 dB. If a 10 dB pad is added to the input, as shown in Figure 1.34, the isolation is increased by 10 dB, the loss is increased by 10 dB, and the coupling stays the same. Thus, the simple but incorrect definition of directivity as isolation/coupling would yield an increase of 10 dB.

In fact, a better way of looking at directivity is the ability of the power at the coupled port to represent a change in reflection at the test port. Again considering Figure 1.34, if a signal of 0 dBm is injected into the input port and a full reflection (an open or short) is applied to the test port, the coupled port will show a power of about -30 dB (10 dB loss, plus a full reflection, plus 20 dB coupling; here the isolation term is ignored for the moment). If a load is applied to the test port, the signal at the input sees a 10 dB loss and 50 dB isolation for a

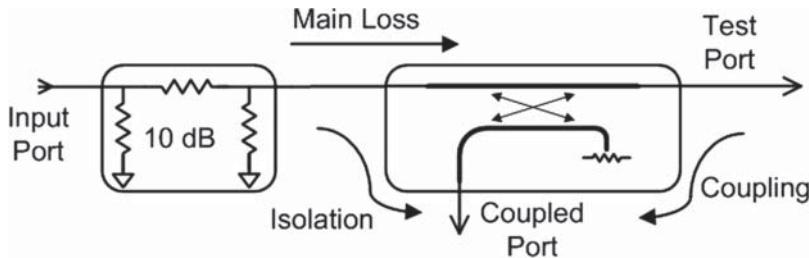


Figure 1.34 The effect of attenuation at the input of a coupler.

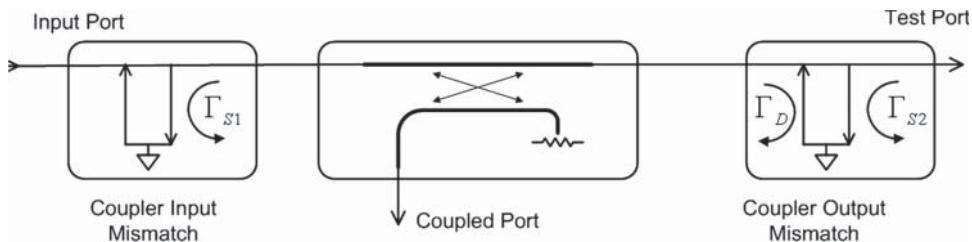


Figure 1.35 Coupler with mismatch after the test port flow graph.

value at the coupled port of -60 dBm . The difference between the open and the load is 30 dB ; hence, the directivity is 30 dB , and adding the pad at the input has no effect.

In practice, the output match of a directional-coupler is critical, and the test port mismatch can dominate the directivity. This signal flow is demonstrated in Figure 1.35. Mismatch at the output of the directional-coupler affects directivity on a one-for-one basis. This mismatch is combined with the coupler input mismatch to create the overall source-match. The source-match affects the power measured at the coupled port when measuring large reflections at the test port. This “source mismatch” causes some reflected signal from the test port to re-reflect from the input port, reflect a second time off the test port termination, and add or subtract to the main reflection, causing error in the coupled port power.

However, the output mismatch is a direct error and causes reflection back into the coupler, thereby adding directly to the coupler directivity.

1.11 Circulators and Isolators

While most passive components are linear and bilateral (that is, the forward loss equals the reverse loss), a particular class of devices based on the ferromagnetic effect doesn't follow this rule. These devices comprise circulators and isolators. A circulator is a 3-port device, with low loss in one direction between ports, say from port 1 to 2, port 2 to 3, and port 3 to 1. But it has high loss (called the *isolation*) in the reverse direction, from port 2 to 1, 3 to 2, or 3 to 1. An isolator is a special case of a circulator with a good load applied to port 3, such that it becomes a 2-port device. Circulators pose a particular measurement difficulty as the isolation

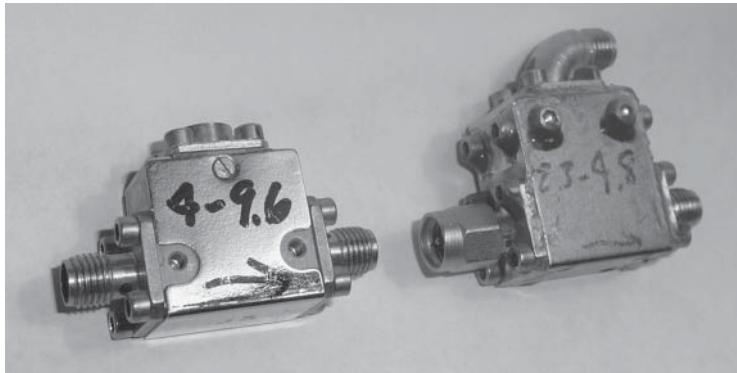


Figure 1.36 Isolator (left) and circulator (right).

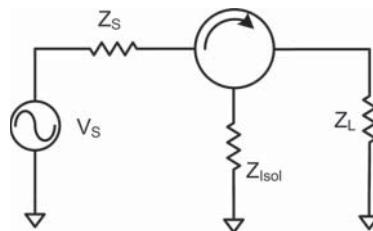


Figure 1.37 Schematic representation of a circulator.

between ports depends upon the match applied to a third port. Thus, for good measurement quality, the isolation measurement requires a good effective-match on the ports.

Further, circulators are often tuned by magnetizing a permanent magnet attached to the circulator, and it's desired that the measurement system can determine the isolation of all three ports in a single connection step, and with good speed. Thus, multiport (more than two ports) systems were developed to simplify the connections, and multiport calibration techniques were developed to satisfy the need for high-quality correction.

Even though they are passive devices, circulators and isolators are sometimes tested for their high-power response, such as compression and IMD. The ferromagnetic effect has hysteresis properties that can produce IMD and compression when driven with sufficiently high powers. Figure 1.36 shows an isolator (left) and a circulator (right). Figure 1.37 shows the signal flow for a circulator. In the figure, the isolator on the left has an internal load mounted at the top, and the circulator on the right has three ports, with an SMA connector in place of the isolator load.

1.12 Antennas

As the air interface for all communications systems, antenna performance is the first (in a receiver) and the last (in a transmitter) characteristic that affects the overall system performance. An antenna can be small and simple, such as a whip antenna found on a

handset, or quite complicated such as those found in phased-array radar systems. Antennas have two key attributes: reflection and gain pattern.

Antenna reflection is essentially a measure of the power transfer efficiency from the transmitter to the over-the-air signal. Ideally, the antenna should be impedance matched to the transmitter's output impedance. In fact, it is typically the case that the antenna is matched to some reference impedance, typically $50\ \Omega$, while the transmitter is likewise matched to the same reference impedance. This implies that while the two may be matched, in many cases they can be exactly mismatched if the phase of the antenna mismatch is not the conjugate of the phase of the transmitter's mismatch. The tighter the mismatch specification is for each, the less variation in transmitter power one sees when phasing causes the two mismatches to be on opposite sides of the reference impedance.

Further, simple antennas are matched to a rather narrow range of frequencies, and it is a significant aspect of antenna design to extend the impedance match across a broad range of frequencies. One common form is a bi-conical antenna, often found for use in testing the radiated emissions from electrical components. On the other end of the spectrum is the desire for a narrowband antenna to have a low return loss over a small frequency range to minimize reflected power back to the high-power transmitter.

Antenna gain, or antenna gain pattern, describes the efficiency of an antenna in radiating into the desired direction (or beam) relative to a theoretical omni-directional antenna, often referred to as an *isotropic radiator*. This figure of merit is known as dBi, or decibels relative to an isotropic antenna.

Antenna pattern measurements are the measurement of the antenna radiation pattern, typically plotted as a contour of constant dBi on a polar plot, where the polar angle is relative to the main beam or “bore-sight” of the antenna. Antenna pattern measurements can range from simple gain measurements on an antenna on a turntable to near-field probing of complex multi-element phased array structure. While these complex measurements are beyond the scope of the book, many aspects of antenna return loss measurement, including techniques to improve these measurements, will be covered.

1.13 PC Board Components

While a wide-ranging topic, the measurement of passive PC board components is focused on the measurement of surface mount technology (SMT) resistors, SMT capacitors, and SMT inductors. These components comprise the majority of passive elements used in radio circuits and also create some of the most undesirable side effects in circuits because of the nature of their parasitic elements. Here is a review of the models of these elements; during measurement, the difficulty is in understanding the relative importance of aspects of these models and extracting the values of the model elements.

1.13.1 SMT Resistors

Resistors are perhaps the simplest of electronic elements to consider, and Ohms law is often the first lesson of an electronic text.

$$R = \frac{V}{I} \quad (1.91)$$

However, the model of an RF resistor becomes much more complex as frequencies rise and distributed effects and parasitic elements become dominant. In this discussion the focus will be on surface-mount PC board components, as they are used almost exclusively today in modern circuits. Thin film or thick film hybrid resistors have similar effects, and although the parasitic and distributed effects tend to hold off until higher frequencies, much of this discussion applies to them as well.

A good model for a resistor consists of a resistive value in series with an inductance, both shunted by a capacitance. This is a reasonable model for an SMT resistor in isolation, but the values and effects of the model are modified greatly by the mounting scheme of the component. For example, if it is mounted in series with a microstrip transmission line and the impedance is such that the resistor is much narrower than the transmission line, then this model works well for predicting circuit behavior; on the other hand, if it is mounted on a narrow line, then the contact pads will provide additional shunt capacitance to ground, and the model must include some element to account for this effect. At lower frequencies, some shunt capacitance will do well, but at higher frequencies, a length of low impedance transmission line might be a better choice.

A resistor used in shunt mode to ground can have an entirely different model when it comes to parasitic effects from that of a series resistance. While the RF value of the resistive element may stay almost the same as the series value (close to the DC value), the effective inductance can be substantially higher as the inductance of the ground via adds to that of the resistor in a microstrip configuration. A larger pad on the ground via, surprisingly, can add even more effective inductance as it resonates with the inductance of the via to increase the apparent inductance of the pair. Meanwhile, the shunt capacitance of the resistor may be absorbed in the transmission line width. Figure 1.38 shows a model for a resistor mounted in the series and shunt configurations. Measurement examples to illustrate extracting these values will be shown in Chapter 11.

In many instances, one of the two parasitic elements will dominate the model for first-order high-frequency effects. In fact, one can use some simple calculations to estimate a rough order of magnitude for these parasitic elements. Take, for example, an 0603 resistor, which has dimensions of approximately 0.6 mm width, 0.4 mm height (considering some excess plating, and some edge effect), and 0.76 mm length. If one considers the contact of the resistor wrapped around the body, one might reasonably divide the effective length by 3, to about 0.25 mm. Remembering that SMT resistors are often constructed on ceramic substrates, with a relative dielectric constant of about 10, then the capacitance can be computed as

$$C = \epsilon_r \epsilon_0 \frac{W \cdot H}{L} = 10 \cdot 8.85 \times 10^{-15} (F/mm) \cdot \frac{0.6 \cdot 0.4}{.25} = 0.085 pF \quad (1.92)$$

The actual value may be substantially greater or less depending upon the exact attributes of the electrodes, but this gives a starting estimate. For inductance one can look to the formula

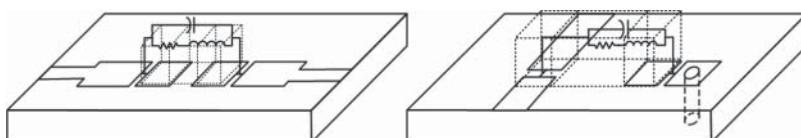


Figure 1.38 Models for a series resistor (left) and shunt resistor (right).

of a transmission line, and assuming that the resistor is mounted on a narrow line, such that its impedance is high, the inductance is computed as

$$l = \mu_0 \cdot L = 4\pi 10^{-10} (H/mm) \cdot 0.8mm = 0.8nH \quad (1.93)$$

Thus, from the model one can compute the values of resistance for which the inductance or capacitive term dominates, at some frequency. For example, at 3 GHz, the inductance has a value of about 15Ω reactance in series with the resistive element; the capacitance has a value of about 1250Ω reactive in shunt. At 50Ω , the inductance value dominates, at 300Ω , the capacitive value is the dominant parasitic effect. For low values of resistance and high frequency, the inductance becomes dominant, and the series impedance is larger than expected, causing the loss through the resistance to be larger than expected because of this effect. At high values of resistance, the parasitic capacitance reduces the series impedance, and the expected loss through the resistor is less than expected. The values change with the physical size of the component; thus, cross-over points differ in resistance and frequency, but with similar effects. This can be used to advantage as there exists a crossover point where the inductive and capacitive effects cancel somewhat, and the resistor behaves in a more ideal way, to higher frequencies, than for values above or below the crossover value. Using this value of resistance can, in series or parallel arrangements, provide a range of resistances that avoid parasitic effects until higher frequencies. For these values, a 50Ω resistor terminated to ground will have about -18 dB return loss at 3 GHz; however, two 100Ω resistors terminated to ground will have about -36 dB return loss, thus providing a better RF resistance of 50Ω than a single resistor. Thus, characterization of the parasitic effects, and proper compensation, can allow use of SMT parts to much higher than expected frequencies (Dunsmore 1988). Figure 1.39 shows the effective impedance of a single 50Ω SMT resistor and two 100Ω SMT resistors in parallel, when used as a 50Ω load. This effect also occurs for SMT inductors and capacitors.

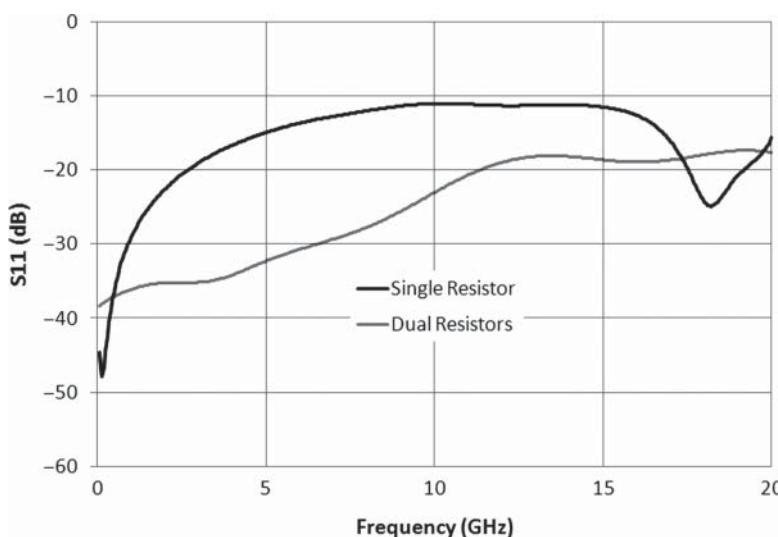


Figure 1.39 Input match of a single SMT resistor and two in parallel.

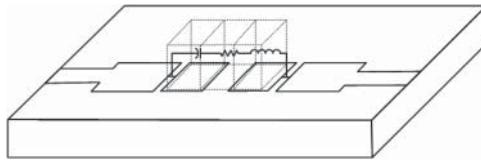


Figure 1.40 Model of an SMT capacitor.

1.13.2 SMT Capacitors

SMT capacitors have a different model from resistors. To a first order, their parasitic effects tend to be all in series, as shown in the model of Figure 1.40. The series inductance is due primarily to the package size and is similar to that of a resistor. The series resistance is due to the manufacturing characteristics of the capacitor and thus cannot be easily estimated. If an SMT capacitor is used in a resonant structure, this resistance will have the principal effect on the Q of the resonator. However, its effect is typically small in most wideband applications, where the capacitor is used as a series DC blocking capacitor or a shunt RF bypass capacitor. This is because the series inductance will dominate the series resistance in these use cases and cause the impedance of the capacitor to rise with frequency (rather than go to zero). At high frequencies, there may also exist a parasitic shunt capacitance across the entire package, which may cause the impedance to fall again.

The case where the series resistance is of consequence is when the capacitor is used in a tuned circuit, where the package inductance may be subsumed in the resonating inductor and thus at resonance the series resistance adds to a degradation of the Q of the capacitor. With careful design, the capacitance value may be compensated for by the including the effects of the series inductance; this effect is to make the capacitor look larger than its prescribed value. In fact, where the reactance of the parasitic inductance equals the reactance of the capacitance, the effective value of capacitance goes to infinite and the series impedance becomes just the parasitic resistance. So, for characterizing capacitors for use in tuned circuits, one must really assess their value near the frequency on which they will be having the most effect on a circuit. Consider a one-pole filter, where the cutoff of the filter starts to occur when the reactance of the capacitor reaches 50Ω . In many cases, the inductance is quite significant and already altering the effective value of the capacitor. Thus, it is important to evaluate the effective capacitance near this point. A good rule of thumb is to evaluate a capacitor where the reactance is $j50 \Omega$.

A further characteristic of capacitors that is significant is the internal assembly structure. Capacitors are typically formed by a set of interleaved parallel plates with alternate plates connected at each end to the terminals. The plates can be parallel to or vertical to the PC board. For some cases, the capacitor body itself can form a dielectric resonator at high frequency, but below that the capacitor can act as a single, large conductive block on a PC board trace, typically resulting in a model that might best be considered a transmission line of somewhat lower impedance than the mounting line.

Capacitors used as bypass capacitors have an additional parasitic effect from the series inductance of the ground via, and from the pad above the ground via.

1.13.3 SMT Inductors

Inductors are perhaps the most complicated of the simple passive components. Because they are constructed of coils of very fine wire, sometimes multiple layers of coils, their parasitic elements are greatly affected by the details of their construction. Some inductors have the axis of the coil parallel to the PC board, and some are wound with the axis perpendicular. In both cases, the model for the inductor is essentially the same as the resistor, as shown Figure 1.39, but with the value of the series inductance equal to the DC value of the inductor, and the series resistance equal to the DC resistance. Inductors, because of the nature of their construction, have very large relative parasitic capacitances. In cases where an inductor is used for a bias element (relying on its impedance to be high at high frequencies) one often finds that the parasitic capacitance will become the main effect over the band of interest. Thus, in many cases the value of inductance used is carefully selected based on the overall effective inductance and sometimes utilizes the shunt capacitance to provide a high impedance at a particular frequency of interest. It may quite difficult to make a single inductor provide good RF performance over a wide band.

When inductors are used as elements in filters, the parasitic capacitance can often have significant effects for use in band-pass filters, and the inductance must be evaluated for each use to find the effective value considering the parasitic capacitance.

A common figure of merit for inductors is the self-resonant frequency (SRF), above which they act more like a capacitor (impedance goes lower with increasing frequency) than an inductor. The value of the SRF can be estimated in one way by looking at the length of the wire used in making the inductor. The SRF will be less than the frequency for which the wire is one-quarter wavelength.

1.13.4 PC Board Vias

The PC board via is perhaps the most common PC board component, and often the most overlooked. The effect of a via depends greatly upon how it is structured in the circuit. A single via to ground in the center of a transmission line appears as almost a pure inductance. However, a via between RF traces can have aspects of inductance and some parasitic capacitance (due to pads around the via) that can cancel, in part or all, the inductive effect. When a via is used in a mounting pad for a shunt element, such as a resistor used as a load, or a bypass capacitor, the mounting pad and via form a resonant structure such that the size of the mounting pad can increase the effective impedance of the via. Further, several vias are often used in parallel to ground devices, sometimes to lower their effective inductance and sometimes to provide greater heat sinking of an active device. Putting vias in parallel does lower their effective inductance, but not in a simple way. Rather than halving the inductance, mutual inductance between vias means that the value of effective inductance doesn't reduce as expected. For example, putting two 100Ω resistors at the end of a line to ground, placed in parallel, may show much larger inductive effect the same two 100Ω resistors place in a T pattern, where the ground vias are separated and the mutual inductance is less.

1.14 Active Microwave Components

With a few exceptions, passive components follow some fundamental rules that greatly simplify their characterization; principally, they are linear, so their characterization doesn't depend on the power of the signal used to characterize them, but only on the frequency. Active components, on the other hand, are sensitive to power, and their responses to both frequency stimulus and power stimulus are important. Often, passive components are operated well below any power level that causes a change in their response, but more and more active components are being driven into higher-power operation to optimize their efficiency.

1.14.1 *Linear and Non-linear*

In the measurement sense, one definition of linear devices is that they are devices in which the output power is a linear function of the input power. If the input power is doubled, the output power is doubled. Almost all passive devices follow this rule, and many active devices as well. An alternative definition of linear is one for which only frequencies that are available at the input appear at the output. In practice, the first definition is more useful for system response. Some important characteristics of active components are discussed next.

1.14.2 *Amplifiers: System, Low-Noise, High Power*

1.14.2.1 **System Amplifiers**

System amplifiers are simply gain blocks used to boost signal levels in a system, while providing reverse isolation. They can have higher noise figures than LNAs as they are used in signal paths where the signal is well above the noise floor. They often follow an LNA stage, frequently after some pre-filtering. They are also often used in the frequency converters as a local oscillator (LO) amplifier to isolate the RF signal from leaking out the LO port, or as an isolation amplifier to prevent LO leakage out the RF port. These tend to be broad band amplifiers, with good input and output match, emulating an idealized gain block. The important figures of merit for such amplifiers are gain (S_{21}), input and output match (S_{11} , S_{22}), and isolation (S_{12}). Occasionally, directivity of an amplifier is defined as isolation (a positive number in dB) minus the gain (in dB), or S_{12}/S_{21} . It is a measure of the effects of a load apparent at the input of the amplifier, or how the output impedance is affected by the source impedance (Mini-circuits n.d.) and is important in cases where other system components have a poor or unstable match. Since these amplifiers have wide bandwidths, it is important that they have good stability as they can have a variety of load impedances applied. Other figures of merit for system amplifiers can include gain flatness (deviation of the gain from nominal value), 1 dB compression point (the power at which the gain drops by 1 dB), harmonic distortion, and two-tone third-order IM, sometimes expressed as third-order intercept point (see Section 1.3).

1.14.2.2 **Low-Noise Amplifiers**

Low-noise amplifiers are found in the front end of communications systems and are particularly designed to provide signal gain without a lot of added noise power. The key figure

of merit is noise figure, along with gain. But, in a system sense, the noise parameters (see Section 1.3) are quite important as they describe the way in which the noise figure changes with source impedance. LNAs are used in low-power applications, and their 1 dB compression point is not a key spec; however, distortion can still be a limiting factor in their use, so a common specification is the input referred intercept point. A key trade-off made in LNAs is between lowest noise figure and good input match. The source impedance where the LNA provides the lowest noise may not be the same as the system impedance, so a key design task for LNAs is to optimize this trade-off.

1.14.2.3 Power Amplifiers

Many of the figures of merit for power amplifiers are the same as system and LNAs, but with an emphasis on power handling. In addition, the efficiency of the amplifier is one of the key specifications that one finds primarily with power amplifiers, implying that the DC drive voltage and current must also be characterized. Because power amplifiers are often used with pulsed RF stimulus, the pulse characteristics, such as pulse profile including pulse amplitude and phase droop, are key parameters.

Power amplifiers are often driven into a non-linear region, so the common linear S-parameters may not apply well to predict matching. Therefore, load-pull characterization is often performed on power amplifiers. Gain compression and output-referred intercept point are common for power amplifiers. Some amplifier designs such as traveling-wave-tube amplifiers (TWT) have a characteristic that causes the output power to reach a maximum and then decrease with increasing drive power, and the point of maximum power is called *saturation*. Gain at rated output power is another form of a compression measurement where rather than specifying a power for which the gain is reduced by 1 dB, it specifies a fixed output power at which the gain is measured.

Power amplifiers are often specified for their distortion characteristics including IMD and harmonic content. In the case of modulated drive signals, other related figures of merit are ACPR and adjacent channel power level (ACPL). A figure of merit that combines many others is EVM, which is influenced by a combination of compression, flatness, and inter-modulation distortion among other effects.

1.14.3 Mixers and Frequency Converters

Another major class of components is mixers and frequency converters. Mixers convert an RF signal to an IF frequency (aka down-converters) or IF frequencies to RF (up-converters) through the use of a third signal, known as the *local oscillator* (LO). Typically, the output or input that is lower in frequency is called the IF (for *intermediate frequency*). The LO provided to the mixer drives some non-linear aspect of the circuit, typically diodes or transistors that are switched on and off at the LO rate. Using the second definition, frequency converters and mixers would not be considered linear. In fact, in their normal operation, they are linear (under the first definition) with respect the desired information signal, and ideally the frequency conversion does not change the linearity of the input/output transfer function.

The input is transferred to the output through this time-varying conduction, and the output signal includes the sum and difference of the input and the LO. Mixers tend to be fundamental

building blocks that are combined with filters and amplifiers and sometimes other mixers to form frequency converters. In practice, frequency converters are typically filtered at the input to prevent unwanted signals from mixing with the LO and creating a signal in the output band, and they are typically filtered at the output to eliminate one of the plus or minus products of the mixing process. Some converters, known as *image reject mixers*, have circuitry that suppresses the unwanted product, sometimes called the *sideband*, without filtering. These are typically created by two mixers driven with RF and LO signals phased such that the output signals of the desired sideband are in phase and added together to produce a higher output, and the undesired sideband is out of phase; thus they cancel and produce a smaller output. Monolithic microwave integrated circuits (MMICs) sometimes blur the line between converters and mixers as they may contain several amplifier and mixing stages, but they typically don't contain filtering.

Mixers have fundamental parameters that include conversion loss (or gain in the case of MMICs with amplifiers), isolation (of which there are 12 varieties, to be discussed in Chapter 7), compression level, noise figure (which for passive mixers is typically just the conversion loss), input and output match, and most of the other parameters found for amplifiers.

Mixers are sometimes referred to as passive mixers or active mixers. Passive mixers don't contain amplification circuits, are typically constructed of diodes (with a ring or star configuration being most common), and have baluns in some paths to provide improved isolation and reduced higher-order products. Higher-order mixing products, also called *spurious* mixing products or *spurs*, refer to signals other than the simple sum and difference of the input and LO frequencies. They are often referred to by the order of harmonic related to creating them; for example, a 2:1 spur is created at the frequency of two times the LO plus and minus one times the input. The level of the mixer spurious products, which are sometimes called *mixer intermod products* (even though the two tones are not applied to the input), change with respect to the RF drive signal at the rate of the order of the RF portion of the spur. For example, a 2:1 spur will increase 1 dB in power for each dB in power the RF signal increases. However, since the LO is creating the non-linearity in the mixer, as well as the spurious signal, it is difficult to predict how the spurious power will change with respect to LO power. In many cases, driving the LO power higher produces a higher spurious signal, as the relative magnitude of the RF signal to the LO power is reduced; in other cases, the transfer impedance of the non-linear element becomes more consistent across the RF drive level. The spurious higher-order products of a mixer are sometimes defined as a spur table, which shows the dBc values of higher-order products relative to the desired output. Chapter 7 has more details on measuring these behaviors.

Mixers with baluns can suppress some of the higher-order products, with baluns on the RF port suppressing products that have even-order LO spurious, and baluns on the LO port suppressing products that have even-order RF spurious. These are called *single-balance mixers*, and it is typical that the LO port is balanced. Double-balanced mixers have baluns on the RF or IF ports. Refer to Chapter 7 for more details on mixer configurations. Triple-balanced mixers are usually comprised of a pair of double-balanced mixers, adding a balun to the IF port. Their main advantage is to divide the RF signal power between the two diode quads, lowering the RF relative to the LO, whereby the spurious signals created will be lower; then the outputs are combined to recover the power. This provides a mixer with the same conversion loss and lower spurious products at the same output power level. The disadvantage is that since the LO drive is also divided, higher LO drive power is needed to achieve the same linearity for each diode quad.

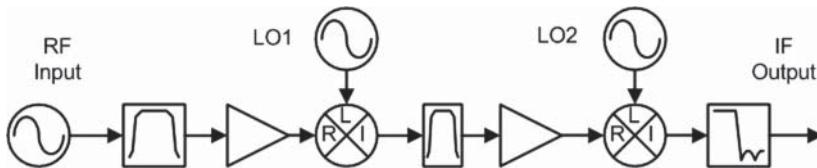


Figure 1.41 Dual-LO frequency converter.

The creation of spurious mixer products is a key aspect of system design, with the goal of eliminating spurious signals from the IF output. Unfortunately, some frequency plans are such that the spurious products must fall into the band of interest. In such cases, system designers move to multiple conversion stages to create a first stage, which produces an output free of spurious signals over the range of input signals of interest, and then has a second conversion stage that produces the frequency of interest at the output. This multiple conversion or “dual-LO” system is typically called a *frequency converter* and often contains additional filtering and amplification, as shown in Figure 1.41.

Because of the multiple components used, the frequency response of converters often has gain ripple and phase ripple, which can distort the information signal. Key figures of merit for converters are gain flatness, group delay flatness, and the related phase flatness, which is also known as deviation from linear phase, and represents residual ripple after fitting the phase data to a straight line. Modern systems employ equalization techniques that can remove some of the flatness effects provided they follow simple curvatures; as such, another specification found on converters is deviation from parabolic phase, which is the residual ripple in the phase data when it is fitted to a second order curve.

Mixers often have quite poor input or output match because of the switching nature of their operation, so their effect on system flatness when assembled into a converter can be quite dramatic. Until recently, it was difficult to predict the effects of output load of a mixer on its input match, but the mathematical tools to model mixers as system elements have been developed (Williams et al. 2005) that describe these relationships. Mixers that produce an output that is the sum of the input and output signals are relatively simple to describe, but mixers that produce an output that is the difference between input and LO have a more complicated behavior in the case where the input frequency is less than the LO frequency. These are sometimes called *image mixers*, and their unusual characteristic is that as the input frequency goes up, the output frequency goes down; this also applies to phase: a negative phase shift of the input signal results in a positive phase shift of the output signal. How these special cases affect the system performance will be described in more precise mathematical terms in Chapter 7.

1.14.4 Frequency Multiplier and Limiters and Dividers

Mixers are not the only way to create new frequencies at the output; frequency multipliers are also used to generate high-frequency signals, particularly when creating mm-wave sources. Frequency multipliers produce harmonics by changing a sine-wave input signal into non-linear wave. The basic doubler is a half-wave or full-wave rectifier, such as a diode bridge. A pair of back-to-back diodes turns a sine wave into a square wave, which is rich in odd harmonic content. This is essentially the same as a limiter.

The key figure of merits of a multiplier is the conversion loss from fundamental drive to the desired harmonic. Other important characteristics are fundamental feed-through and higher-order harmonics.

Limiters have the key characteristic of maximum output power; that is the power at which they limit. Also important is the onset of limiting and the compression point. Ideal limiters are linear until the onset of limiting, and then they effectively clip the output voltage above that level.

Other multiplier types are step-recovery diodes, and non-linear transmission lines that, when driven with a sine wave, effectively “snap” on to produce a sharp edge. Depending upon the design, the on-time can be short, which produces an output rich in harmonics. Some digital circuits can also be used to create narrow pulses from a sine-wave input as a pulse generator. Such a pulse will also be rich in harmonics.

One aspect of a multiplier that is not easily discerned is the group delay through it. That is because for some change in input frequency one will see a multiplied change in output frequency. FM that passes through a doubler will have the same rate as the original FM, but twice the deviation. For this reason, doublers or multipliers are seldom used in the signal or communication path of RF or microwave systems but can be used in the base carrier paths and in the LOs of many systems.

1.14.4.1 Frequency Dividers

Frequency dividers provide for a lower value of frequency than the input frequency. Like multipliers they are highly non-linear and often produce square-wave outputs. Some key specifications of dividers are minimum and maximum input power to ensure proper operation of the divider, output power and harmonics, and additive phase noise. Typically, for each divide by two stage, the phase noise is reduced by 6 dB. But noise or jitter in the divide circuitry can add noise to the signal; the added phase noise at the output relative to the phase noise at the input is called *additive phase noise*. This is also a concern with mixers, where the LO phase noise can be added to the output signal, and to a lesser extent amplifiers.

1.14.5 Oscillators

Oscillators in some ways represent the most non-linear of electrical circuits with frequencies created at the output with no input (other than noise). Oscillators have a wide variety of characteristics that are important to characterize, including output frequency, output power, harmonics, phase noise, frequency pushing (change in frequency with change in DC power), frequency pulling (change in frequency due to change in load impedance), and output match.

Voltage-controlled oscillators (VCOs) have the additional ability to control the frequency of the output due to a voltage change at the input. The voltage-to-frequency control factor is a key attribute of a VCO. A related microwave component is a yttrium-iron-garnet (YIG) oscillator, which uses a spherical YIG resonator as the frequency control element of the YIG-tuned-oscillator (YTO). The YIG resonator has the characteristic that the resonant frequency changes with magnetic field. YTOs have wide tuning bandwidth (up to 10:1) and low phase noise. Tuning is performed by changing the current in an electromagnet but can be very low bandwidth due to the large inductance of the magnet. YTOs often have a second,

lower inductance coil (called the *FM coil*), which provides small change to frequency but with high bandwidth.

As the focus of this book is stimulus/response measurements, the measurement of oscillators will not be covered.

1.15 Measurement Instrumentation

1.15.1 Power Meters

Perhaps the simplest and most common of microwave instruments is the power meter. It consists of a sensor, which absorbs or detects RF power and converts it to a DC signal, and measurement circuitry, which accurately measures this DC signal and applies correction and calibration factors to it, to produce a reading of the RF power level. Power meters come in a variety of forms and complexities, some of which are noted next.

1.15.1.1 Calorimeters

Often considered the most accurate and traceable of power measurement systems, the calorimeter consists of a thermally isolated RF load that absorbs the RF energy. This load is kept in a heat exchanger, and a thermopile is used to sense the change in temperature. Since the fundamental measurement is temperature, the traceability of these systems to fundamental SI units is good. These systems can handle large power but are slow to respond, are heavy, and are typically not used by practicing RF engineers, except in special cases.

1.15.1.2 RF Bolometers and Thermistor

An RF bolometer or thermistor is a system where the RF measuring element is a thermally sensitive resistor used as part of a DC bridge system. The DC bridge is electrically balanced, and when an RF signal is applied to the bolometer element, the element heats and its DC resistance changes. The DC bridge is nulled using an offset voltage, and the measure of the offset voltage can be related directly to the power absorbed by the bolometer. The key aspect of the bolometer is that it is equally sensitive to RF or DC power; thus, a precision DC source can be used to produce a known power at the bolometer, and the balancing circuit is thus calibrated relative to the DC power absorbed. The heating effect of the RF power produces the same offsets as the DC power and thus is easily calibrated. Bolometers have a relatively small dynamic range (the range of input powers over which they operate) but have linearity (the ability to correctly measure differences in input power) derived from a bridge circuit using DC substitution. Typically, bolometers are found only in precision metrology laboratories and are not in common use among RF engineers.

1.15.1.3 RF Thermocouples

Until recently, RF thermocouples were the most common type of power sensor used. These thermocouples convert heating directly into a DC voltage and, because their small size and

thus small thermal mass, are much faster responding and have a larger dynamic range than either thermistors or calorimeters. As with other sensors, these require calibration with a precision source but are typically DC blocked, so the source must be a low-frequency AC source. These sensors are commonly used throughout the RF industry but have the detriment of being somewhat slow-responding (with response times in the several to tens of millisecond range), but they are extremely linear and relatively non-responsive to harmonics. That is, harmonic power will be detected as an RMS error of the power of desired signal. Since harmonics 20 dBc or lower represent less than 1% of the power of the main signal, the error due to harmonics is quite low.

1.15.1.4 Diode Detectors

For modern power meter applications, the diode or multi-diode power sensor is often the preferred choice. These sensors employ one or more diodes that rectify the RF signal and produce an equivalent DC signal. Occasionally, the DC signal is “chopped” or modified in such a way as to produce a square wave to the measurement portion of the power meter, typically a precision analog-to-digital converter (ADC). Chopping the signal helps compensate for DC offsets in the ADC input.

Older diode detectors used only a single diode, and the top 20 dB of the detector range was often described as the “linear” range; below that range, the diode would operate in “square-law” mode where the output voltage would be a function of the square of the input RF signal. In the low-power range, the output voltage would be linearly related to the square of the input voltage of the RF signal, thus be linearly proportional to the detected power. In such a region, they operated almost as well as the thermistor sensors but with much faster speeds and much wider dynamic range. At the top of their measurement range, in the linear region, the output circuitry and measurement algorithms are adjusted to compensate for the change to the linear mode of operation. However, in the linear mode, the power in the harmonics has a much greater effect, and a 20 dBc harmonic signal can have up to a 10% change in the measured power of the fundamental, even though it contains only 1% of the power. This is due to the peaking effect that the harmonic can have on the RF voltage. Out of the square-law region (also known as the *linear region*, which is in fact where the power meter is not as linear in the usual sense of the word), the power meter may not give accurate readings for complex modulated signal or signal with high harmonic content or high peak-to-average envelope power.

More modern diode sensors use a multitude (two or more) of embedded diode elements, some of which are padded with larger attenuation to allow them to operate at higher powers and still be in the square-law region. Complex algorithms in the power meter instrumentation detect when the power from one sensor exceeds the square-law region and change to take their readings for power from one of the attenuated diodes. This extends the useful range of the power sensor over more common older diode sensors.

1.15.2 Signal Sources

1.15.2.1 Analog Sources

While not a measurement instrument in their own right, signal sources or signal synthesizers, or simply sources, are used as accessory equipment in a variety of measurement tasks. They

can provide CW signals in place of a mixer LO or provide an input signal to an amplifier or filter. These are typically called *analog sources*, and their key attributes are frequency range, output power range (minimum and maximum), phase noise and spectral purity, and frequency switching speed.

While the first two attributes are obvious, the phase noise and spectral purity are key attributes when making measurements close to the carrier such as IMD measurements or when making other distortion measurements such as harmonics.

Switching speed becomes important in automated test systems (ATSs) when using the source as a swept frequency stimulus. Commonly, stand-alone signal sources make a trade-off between lower-phase noise and slower switching speeds.

1.15.2.2 Vector Sources

Another class of signal sources are vector signal generators, which have an internal I/Q modulator that allows an almost infinite variety of signals to be created. Some of these vector sources (also called *digital sources* because they can create signals using digital modulation techniques) have built-in arbitrary-waveform generators (AWGs, or *arbs* in the vernacular), while others have broadband I/Q inputs to allow external AWGs to drive their vector modulators directly.

With vector sources, the AWGs can be used to create a wide variety of signals including extremely fast switching CW sources (within the bandwidth of the I/Q modulator), two-tone or multitone signals, pseudo-random noise waveforms, and complex modulated signals following the formats used in digital communications and cellular phones.

Some key attributes of vector sources are the modulation bandwidth of the I/Q inputs, the modulation bandwidth or speed of the arbitrary waveform generator (if it is built-in), the memory of the AWG (which affects the length of signals that can be created), and the I/Q fidelity or linearity of the modulator. This linearity limits the ability of the vector source to produce clean signals. For example, a two-tone signal can be created by doing a double sideband suppressed carrier modulation, but if there is imbalance or non-linearity in the modulator, there will be carrier leakage between the two tones.

The output power amplifiers of vector sources are important as their distortion will directly affect the modulated signal, causing TOI, and the spectral spreading of modulated signals.

1.15.3 Spectrum Analyzers

A spectrum analyzer (SA) is a specialized type of receiver, which displays the power of a signal on the y-axis versus the frequency of the signal on the x-axis. As such, it could be considered a frequency-sensitive power meter.

The key attributes of a spectrum analyzer are its displayed average noise level (DANL) and its maximum input power. The maximum input is set by the compression of the input mixer in the SA and can be increased by adding input attenuation. However, adding attenuation degrades the DANL by an equal amount. A further limitation in measuring signals, for example, TOI, is the self-generated distortion of the input mixer, which will generate TOI signals at the same frequencies as that as the TOI from the signal under test. The data sheet for an SA will typically specify the distortion in dBc relative to some input level at the mixer.

This, coupled with the noise floor, will set the measurement range of the SA. Lower-resolution bandwidth will lower the noise floor at the cost of speed of measurement. Similar effects are present for the measurement of harmonics.

Another key attribute of a spectrum analyzer is its frequency flatness and power linearity specifications. Flatness specifications of a spectrum analyzer are usually quite large, as much as ± 2.0 dB for 26 GHz microwave version, although typical performance is much better, and this flatness can be compensated for with an amplitude calibration. Microwave and mm-wave systems can be even worse. The large value for frequency response comes from the interactions of the pre-selector (which is usually a swept YIG filter) and the first converter in the SA. To a first order, these are stable and can be corrected for, but there will still be a residual flatness error even after calibration related to the post-tuning drift of the pre-selector; that is, it does not always tune its peak value to the same frequency for the same settings. Another source of uncorrected error is the mismatch between the SA input and the output match of the signal source being measured. In some cases, this can be a quite high number, up to ± 1 dB or more.

As the name implies, the key role of spectrum analyzers is in determining the quality of unknown spectrums. The use of spectrum analyzers in microwave component test applications is primarily as a means to the frequency response or distortion response of a system as a known stimulus is applied.

In recent times, these applications of spectrum analyzers are being augmented by advanced VNAs, which have higher-speed receivers, built-in sources, and advanced calibration capabilities, as well as PXI-based spectrum analyzers. Spectrum analyzers are available in modular (PXI) formats and can be configured as multi-channel spectrum analyzers. Further, some spectrum analysis has moved into the mm-wave range, typically with external down-converters, but some manufacturers are producing spectrum analyzers with internal mm-wave converters to provide broadband (up to 110 GHz) capabilities. These newer implementations may not have hardware preselection for image rejection, but rather use advanced forms of digital image-rejection based on high-speed FFT processing or other methods. More about this appears in Chapter 8.

1.15.4 Vector Signal Analyzers

With the advent of digitally modulated signals for RF and microwave communications, spectrum analyzers have evolved into much more complex systems that include the ability to do wideband de-modulation of these signals. These specialized spectrum analyzers are often called *vector signal analyzers* (VSAs) and play an important role in component tests.

For many active components, a key figure of merit is the distortion that they apply to the vector modulated signal in the form of amplitude or phase error relative to an ideal signal. The composite of all errors over a set of digital symbols is called the *error vector*, and the average magnitude of this error is the error-vector-magnitude (EVM). While EVM is a signal figure of merit, as it is compared to an ideal waveform, the EVM from an amplifier is a combination of the EVM of the input signal and the errors added by the amplifier. From this it is clear that EVM is not a microwave-component parameter; but a related value, residual or added EVM, is and is described as the EVM at the output signal relative to the input signal. In practice, high-quality sources are used to produce the digitally modulated signals, so the input effect on EVM is small, but with higher data rates and wider bandwidths of modulation, these input

effects are becoming more important. Thus, there is a need for a multi-channel VSA that can compare input to output signals. A normal SA with a VSA capability does not provide such dual-channel capability, but some manufacturers supply a specialized dual-channel receiver for the VSA, while other implementations of a VSA use a wideband digitizer, or even a digital oscilloscope, to do direct digitization of the modulated signal. To date, up to four simultaneous channels have been reported in such a VSA.

1.15.5 Noise Figure Analyzers

An offshoot of a spectrum analyzer, a noise figure analyzer (NFA) is a specialized test instrument that is designed particularly for measuring noise figure. NFAs started out as specialized spectrum analyzers, with improvements in the quality of the receiver and with electronically switched gains to allow the noise figure of the test equipment to be minimized relative to the signal being measured. Some of the things needed to accomplish this, such as adding high-gain LNAs in front of the first converter, reduced the maximum input power of the instrument so that it was no longer suitable for general-purpose SA applications.

On the other hand, several spectrum analyzer manufacturers have added a noise figure personality to their SA offerings so that there is quite a lot of overlap between the capabilities of the two systems. However, most SA implementations require the use of an added LNA, at least over some of the band. Newer SAs have an IF structure almost as flexible as an NFA to optimize the performance of the system.

All of these systems of NFA utilize the “hot/cold” or “Y-factor” method of measuring noise figure (more about this in Chapter 9) using as an input to the DUT a noise source that can be turned on and off. From careful measurement of the output noise, the gain and noise figure of the DUT can be discerned.

More recently, VNAs have been modified to operate as NFAs, utilizing an entirely different technique called the *cold-noise* method. In this method, the output noise power is measured, along with the gain using the normal VNA measurement of gain, and the noise figure is computed from these values. No noise source is used in the measurement. This has an advantage of being faster (only one noise measurement is needed) but does have the disadvantage of being sensitive to drift in the gain of the VNA noise receiver. The Y-factor method does not depend upon the gain of the NFA receiver, but this advantage is often offset by the fact that the gain measurements of the NFA are sensitive to match errors as are the noise measurements, and these are not compensated for.

The ultimate in noise figure analysis is a noise parameter test system. This system properly accounts for all mismatch effects. Some systems use both a VNA and an NFA to measure the gain and noise power, respectively. All noise parameter systems include an input impedance tuner to characterize the change in noise power versus impedance value. Recently, tuners have been combined with VNA-based NFAs to produce compact, high-speed noise parameter test systems. These newer systems provide the ultimate in speed and accuracy available today.

1.15.6 Network Analyzers

Network analyzers combine the attributes of a source, and a tracking spectrum analyzer, to produce a stimulus/response test system ideally suited to component tests. These systems have been commercialized for more than 40 years and provide some of the highest-quality

measurements available today. While there are many distinct manufacturers and architectures, network analyzers broadly fall into two categories: scalar network analyzers (SNAs) or VNAs.

1.15.6.1 Scalar Network Analyzers

These instruments were some of the earliest implementations of stimulus/response testing and often consisted only of a sweeping signal source (sometimes called a *sweeper*) and a diode detector, the output of which was passed through a “log-amplifier” that produced an output proportional the power (in dBm) at the input. This was sent to the y-axis of a display, with the sweep tune-voltage of the sweeper sent to the x-axis, thus producing frequency response trace. Later, the signal from the detector and the sweeper were digitized and displayed on more modern displays with marker readouts and numerical scaling.

Other SNA systems were developed by putting a tracking generator into a spectrum analyzer so that the source signal followed the tuned filter of the SA. This produced a frequency response trace on the SA screen.

SNAs had the attribute of being simple to use, with almost no setup or calibration required. The scalar detectors were designed to be quite flat in frequency response, and a system typically consisted of one at the input and one at the output of a DUT. However, for measurements of input and output match, or impedance, the SNA relied on a high-quality coupler or directional bridge. If there was any cabling, switching, or other test system fixturing between the bridge and the DUT, the composite match of all were measured. There was no additional calibration possible to remove the effects of mismatch. As test systems became more complex and integrated, scalar analyzers started to fall from favor, and there are virtually none sold today by commercial instrument manufacturers.

1.15.6.2 Vector Network Analyzers

For microwave component test, the quintessential instrument is a VNA. These products have been around in a modern form since the mid-1980s, and there are many units from that time still in use today. The modern VNA consists of several key components, all of which contribute to making it the most versatile, as well as most complicated, of test instruments; these are as follows:

RF or Microwave Source: This provides the stimulus signal to the DUT. RF sources in a VNA have several important attributes including frequency range, power range (absolute maximum and minimum powers), automatic-level-control (ALC) range (the range over which power can be changed without changing the internal step attenuators), harmonic and spurious content, and sweep speed. In the most modern analyzers, there may be more than one source, up to one source per port of the VNA. Older VNAs required that the source be connected to the reference channel in some way, as either the receiver was locked to the source (e.g. the HP 8510) or the source was locked to the receiver (e.g. the HP 8753). Modern VNAs, for the most part, have multiple synthesizers so that the source and receiver can be tuned completely independently.

RF test set: In older-model VNAs, the test set was a separate instrument with a port switch (for switching the source from port 1 to port 2), a reference channel splitter, and directional-couplers. The test set provided the signal switching and signal separation to find the incident and reflected waves at each port. Most modern VNAs have the test set integrated with the rest of the components in a single frame, but for some high-power cases, it is still necessary to use external components for the test set.

Receivers: A key attribute of VNAs is the ability to measure the magnitude and phase of the incident and reflected waves at the same instant. This requires sets of phase synchronous receivers, which implies that all the receivers must have a common LO. In older, RF VNAs, the reference channel was common to ports 1 and 2, and the port switch occurred after the reference channel tap. Most modern analyzers have a receiver per port, which is required for some of the more sophisticated calibration algorithms. More about that appears in Chapter 3.

Digitizer: After the receiver converts the RF signals into an IF baseband signal, they pass to a multi-channel phase-synchronous digitizer that provides the detection method. Very old VNAs used analog amplitude and phase detectors, but since at least 1985, all VNAs utilize a fully digital IF. In modern VNAs, the digital IF allows complete flexibility to change IF detection bandwidths, modify gains based on signal conditions, and detect overload conditions. Deep memory on the IF allows complicated signal processing, and sophisticated triggering allows synchronization with pulsed RF and DC measurements.

CPU: The main processor of a VNA used to be custom-built micro-controllers, but most modern VNAs take advantage of Windows™-based processors and provide rich programming environments. These newer instruments essentially contain a PC inside, with custom programming, known as *firmware*, which is designed to maximize the capability of the instruments' intrinsic hardware.

Front Panel: The front panel provides the digital display as well as the normal user interface to the measurement functions. Only the spectrum analyzer comes close to the sophistication of the VNA, and in more modern systems, the VNA essentially contains all the functions of each of the instruments mentioned so far. Thus, its user interface is understandably more complex. Significant research and design effort goes into streamlining the interface, but as the complexity of test functions increases, with more difficult and divergent requirements, it is natural that the user interface of these modern systems can be quite complex.

Rear Panel: Often overlooked, much of the triggering, synchronization, and programming interface is accomplished through rear-panel interface functions. These can include built-in voltage sources, voltmeters, general-purpose input/output (GPIO) busses, pulse generators and pulse gating, as well as LAN interfaces, USB interfaces, and video display outputs.

The detailed operation of a VNA is described in Chapter 2.

Extensions to traditional VNAs allow them to create multiple signals for two-tone measurements and to have very low noise figures for noise figure measurements. But the main attraction of VNAs is calibration. A key attribute is that since they measure the magnitude and phase of waves applied to their ports, they can use mathematical correction to remove the effects of their own impedance mismatch and frequency response in a manner that makes their measurements nearly ideal. The details of VNA calibration are covered in depth in Chapter 3.

Thus, even though there is a wide variety of test equipment available for microwave component measurement, by far the most widely used is the VNA, and while many of the topics of component measurements in this book are extensible to any of the previous instruments, the specific implementation and examples will be illustrated primarily using the VNA, as that has become the predominant component test analyzer in use today.

References

- Cameron, R.J., Kudsia, C.M., and Mansour, R.R. *Microwave Filters for Communication Systems: Fundamentals, Design, and Applications*. Hoboken, NJ: Wiley-Interscience. Print.
- Cohn, S.B. (1954). Characteristic impedance of the shielded-strip transmission line. *Transactions of the IRE Professional Group on Microwave Theory and Techniques* **2** (2): 52–57.
- Collier, R.J. and Skinner, A.D. (2007). *Microwave Measurements*. London: Institution of Engineering and Technology. Print.
- Collin, R. (1966). *Foundations for Microwave Engineering*. New York: McGraw-Hill. Print.
- Dunsmore, J. (1988). Utilize an ANA to model lumped circuit elements. *Microwaves and RF*: 11.
- Freiberger, K., Enzinger, H., and Vogel, C. (2017). A noise power ratio measurement method for accurate estimation of the error vector magnitude. *IEEE Transactions on Microwave Theory and Techniques* **65** (5): 1632–1645.
- Hong, J.-S. and Lancaster, M.J. (2001). *Microstrip Filters for RF/Microwave Applications*. New York: Wiley. Print.
- Hunter, I.C. (2001). *Theory and Design of Microwave Filters*. London: Institution of Electrical Engineers. Print.
- IPC (2004). *IPC-2141A: Design Guide for High-Speed Controlled Impedance Circuit Boards*. Northbrook, IL: IPC Print.
- Keysight Technologies. (1968). HP Application Note AN-95-1 [Online]. Available at: <http://literature.cdn.keysight.com/litweb/pdf/5952-0918.pdf> (Accessed 20 January 2020).
- Kurokawa, K. (1965). Power waves and the scattering matrix. *IEEE Transactions on Microwave Theory and Techniques* **13** (2): 194–202.
- Magnusson, P. (2001). *Transmission Lines and Wave Propagation*. Boca Raton, FL: CRC Press. Print.
- Marks, R.B. and Williams, D.F. (1992). A general waveguide circuit theory. *Journal of Research of the National Institute of Standards and Technology* **97** (5): 535–562.
- Mini-Circuits. (n.d.) Amplifier Terms Defined AN-60-038 [Online]. Available at: <http://www.minicircuits.com/app/AN60-038.pdf>. (Accessed 11 Feb. 2012).
- Pozar, D.M. (1990). *Microwave Engineering*. Reading, MA: Addison-Wesley. Print.
- Simons, R.N. (2001). *Coplanar Waveguide Circuits Components and Systems*. Wiley. Print.
- Sombrin, Jacques, 2011. On the formal identity of EVM and NPR measurement methods: Conditions for identity of error vector magnitude and noise power ratio, Microwave Conference (EuMC), 2011 41st European, IEEE, 337–340.
- Wen, C.P. (1969). Coplanar waveguide: a surface strip transmission line suitable for nonreciprocal gyromagnetic device applications. *IEEE Transactions on Microwave Theory and Techniques* **17** (12): 1087–1090.
- Williams, D.F., Ndagijimana, F., Remley, K.A. et al. (2005). Scattering-parameter models and representations for microwave mixers. *IEEE Transactions on Microwave Theory and Techniques* **53** (1): 314–321.
- Zverev, A.I. (1967). *Handbook of Filter Synthesis*. New York: Wiley. Print.

2

VNA Measurement Systems

2.1 Introduction

S-parameter measurements of devices provide the common reference for RF and microwave circuit and system analysis. While the basic methods of S-parameter measurements were developed decades ago, many advances have occurred in just the last five years that makes obsolete the common understanding of the capabilities and limitations of these measurements. Vastly improved hardware and software capabilities provide for control of stimulus signals and analysis of response signals that allow measurement systems to extend the basic linear S-parameters to multiport, differential, and non-linear characteristics. In the past, S-parameters were limited to two ports; now up to 32-port systems are readily available. In the past, measurements were limited to linear responses; now non-linear, distortion, noise, and even load pull characterizations are possible. In the past, calibration techniques were restricted to a few limited sets of standards and algorithms and limited to devices with the same input and output frequencies; now a wide range of calibration algorithms and applications can be applied to a variety of components, with few restrictions.

A clear understanding of the underlying architecture of a vector network analyzer (VNA) is necessary to understand the full capabilities and limitations of the modern VNA. The first part of this chapter deconstructs the VNA to discuss the individual block diagram elements, their attributes and deficiencies, and how they operate together to provide the capability and applications described in later chapters. In the history of VNAs, the HP 8753 and the HP 8510 were the industry-leading RF and microwave VNAs of the 1980s and 1990s, from which many of the principal understandings of capabilities and limitations were formed. For that reason, many of the characteristics of these analyzers are discussed in the following sections to provide a context for the discussion of the modern VNA attributes. In almost all cases, many well-known limitations of these products no longer apply, and a key goal of the first section of this chapter is to illuminate to the reader these improvements.

By around the year 2000, an arms race of sorts emerged in the world of VNAs with the nearly simultaneous introduction of the PNA and ENA families from Agilent, the Ballmann S100, the ZVR and ZVK from Rohde-Schwartz, the Lightning™ and Scorpion™ from Anritsu, and the 3765 from Advantest. By 2010, Agilent and Rohde-Schwartz advanced to the modern generation of multi-function component test platforms, the PNA-X and the ZVA,

while Anristu's products remained mostly in the area of linear S-parameter test in the form of the Vectorstar™. As the author is a principle designer and architect of the Keysight products, the details on the VNA architecture, structure, and capabilities described here are derived from this knowledge. But many of the factors discussed apply equally well to all measurement systems, regardless of manufacturer, including custom-built systems sometimes seen in university research labs or national standards laboratories. Because of these advancements, many rules of thumb and common understandings based on the first generation of commercial VNAs are no longer relevant.

The second portion of this chapter describes the wide range of measurements and characteristics that can be derived from the basic measurements. In Section 2.3, the basic functionality for making measurements is described along with real-world issues and errors that affect these measurements. Particularly in VNA-based measurements, many of these errors can be characterized during a calibration process, and error correction can be applied to the results to remove, to a great extent, the effects of these errors. The calibration and error correction process will be described in detail in Chapter 3. Detailed descriptions of measurements of particular devices are covered in subsequent chapters: linear devices (Chapter 5), amplifiers (Chapter 6), mixers (Chapter 7), and balanced devices (Chapter 10).

Author's notes on the second edition: The evolution of VNA architecture has dramatically increased since the first edition was written in 2011. Almost all VNAs on the market today have a full reflectometer on each port, so the discussion the three-receiver architecture presented in the first edition of this book (principally the architecture of the HP 8753) has been greatly reduced to make room for more interesting VNA enhancements such as true multiport.

2.2 VNA Block Diagrams

The basic block diagram for a component test system is a stimulus source, which is applied to the input of the device-under-test (DUT), and a response receiver at the output of the DUT. For S-parameter measurements, the inputs consist of incident waves at all the ports, and the outputs consist of scattered waves at all the ports, so in general one would require a stimulus and two receivers at each port. In addition, there must be signal separation devices at each port to isolate the incident and scattered waves.

Early systems measured only transmission and/or reflection response, in only one direction, and thus consisted of at most a directional device (bridge or coupler) at the input and a receiver at the output. These systems were classified as transmission/reflection (TR) systems and were most commonly found as scalar network analyzers, although lower-cost VNAs sometimes had TR test sets as well. The advantage of a vector TR analyzer is that the errors in the directional device could be removed with calibration and error correction.

Figure 2.1 shows the block diagram of a TR system. For simplicity sake the reference receiver will be normally at port 1, measuring the $a1$ wave, and the test receivers were limited to the two ports as well; normally the test receiver at port 1 is the reflection receiver ($b1$), and the test receiver at port 2 is the transmission receiver ($b2$). The source is typically split using a two-resistor power splitter or a coupler to create a reference signal that is proportional to the incident signal on the DUT, followed by a directional-coupler or directional bridge, whose coupled arm goes to the reflection test receiver, measuring $b1$. After the DUT, the transmission test receiver measures $b2$. Because of advances in VNAs and integration of dual reflectometers and receivers, this architecture is seldom seen in modern VNAs.

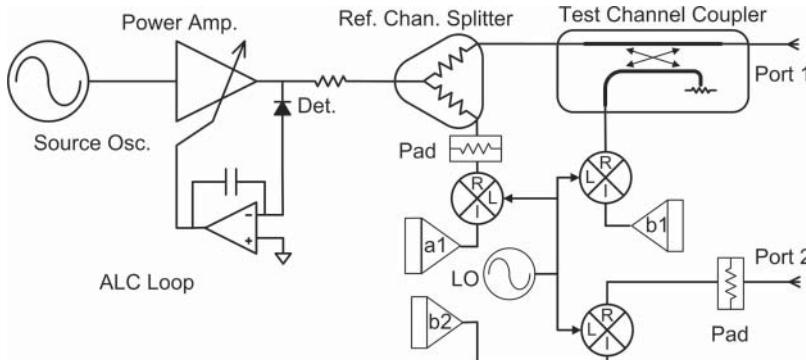


Figure 2.1 A TR network analyzer block diagram.

A full S-parameter system extends the block diagram of the TR system by adding a reflectometer at each port and provides a source at each port. Older systems would switch the source between the ports using a test port switch. Two such block diagrams are shown in Figure 2.2. There are two distinct versions that use either one or two reference receivers.

The three-receiver version (upper diagram) was common in lower-cost or RF network analyzers in the past but has largely been replaced with four receiver versions. Having individual receivers for all the reference and test port channels provides for more and better calibration choices, as will be discussed in Chapter 3.

Older analyzers such as the HP-8510 used separate external sources and switched the source between the ports; others had internal sources, but the cost of the source was a major portion of the instrument cost, so a single switched-source was used in these integrated analyzers. Often, the reference channel splitter was integrated into this switch as well. This provided a compact switch-splitter assembly and allowed a lower-cost alternative to individual splitters or directional-couplers.

Modern network analyzers make use of a hybrid approach with two or more internal sources, such that more than one port at a time can have an output signal, as shown in Figure 2.3.

While there is no requirement for having more than one port active in traditional S-parameter measurements, advanced measurements such as two-tone intermodulation distortion (IMD), active-load, or differential-device test can make good use of these extra sources. In these systems it is common to use a directional-coupler in both the reference and test arms for lower loss from the source to the test port, allowing higher maximum test port power. One synthesizer (that is, the frequency generation unit) may be shared between two ports (of a 4-port VNA) with an individual output amplifier and leveling circuits available at each of the ports. This system can provide outputs at any of the ports at the same time or provide two different frequencies out of pairs of ports, which is useful for mixer test applications.

2.2.1 VNA Source

The VNA source provides the stimulus for the S-parameter measurement. In the original VNAs these were open-loop sweepers, but since about 1985 the use of frequency synthesizers

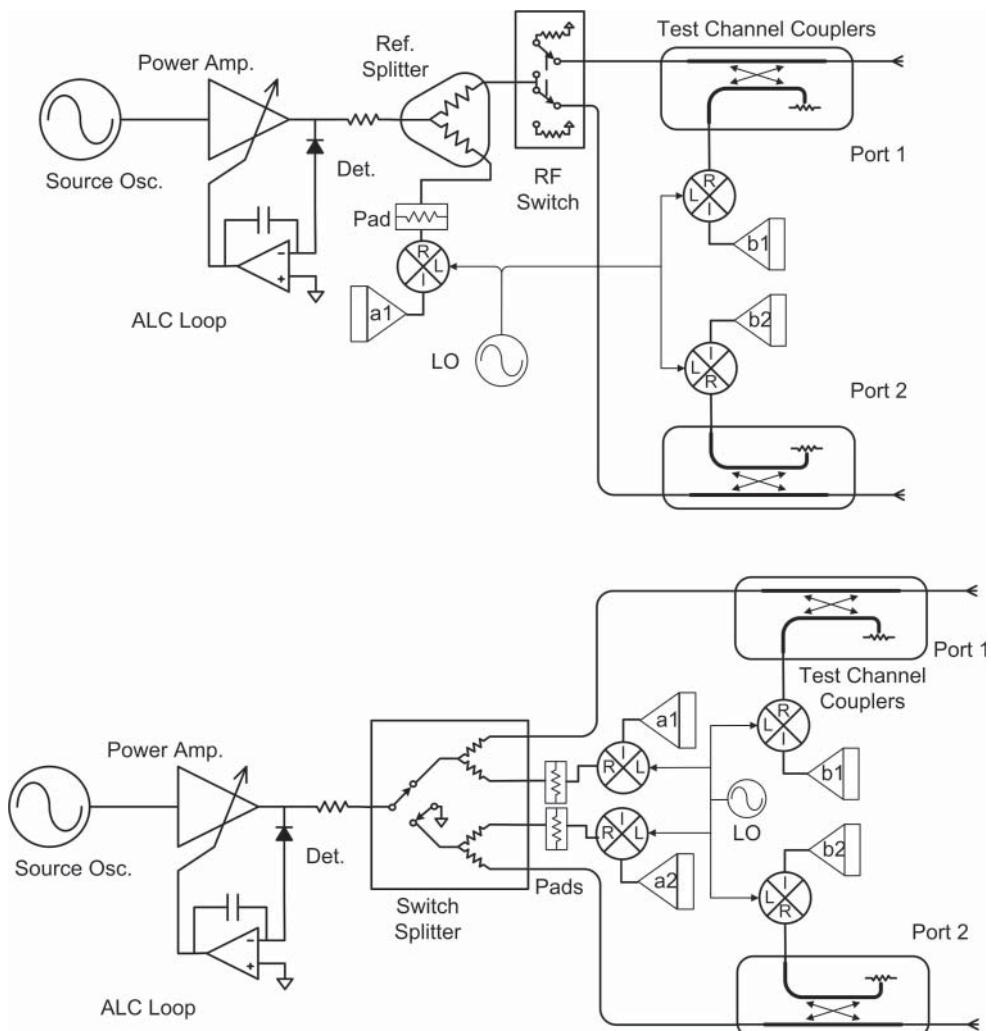


Figure 2.2 S-parameter block diagrams for a three-receiver and four-receiver VNA.

has become the norm. Sweepers used open-loop swept-frequency oscillators to produce the stimulus signals; synthesizers replaced the open-loop control with fractional-N or multi-loop signal generation where the output signal is digitally derived from a 10 MHz reference oscillator to a resolution of less than 1 Hz. Early sources were routed directly to the test sets of VNAs, which were also external, stand-alone instruments, often with the first converter assembly inside.

More recently the source is provided internally to the VNA. In the first integrated VNAs, the signal quality of the internal source was less than that of external instrumentation sources but had the advantage of being much faster in sweeping frequency. For applications such as filter tuning and test, fast sweep times across wide frequency ranges were required. Common

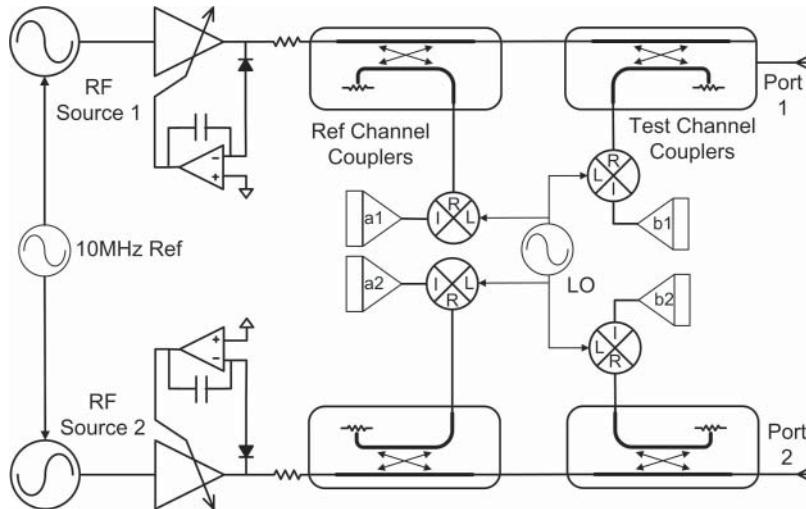


Figure 2.3 Multiple sources in a single VNA.

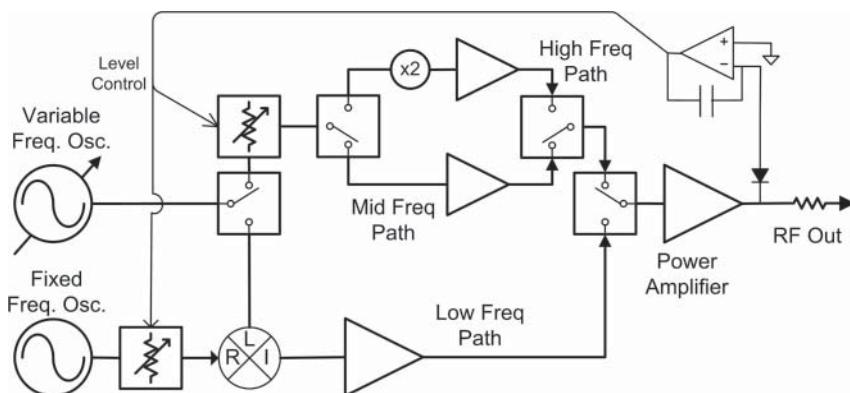


Figure 2.4 Example of a VNA source block diagram.

to VNA sources are the ability to vary power level over a prescribed range, usually called the automatic-level-control (ALC) range, which has values from 20 to over 40 dB. These ranges are often extended with integrated discrete step-attenuators after the source. With new open-loop control, the power range can be extended even farther, typically greater than 70 dB.

Most VNA sources also have a power flatness specification to provide some constant power to the DUT. In many cases, this level is digitally corrected by a factory calibration to be quite accurate at the test port. In addition, a user-performed source-power-level correction is usually available to provide additional accuracy enhancement, as described later in Chapter 3.

A typical VNA source block diagram is shown in Figure 2.4.

Generally, there is a fundamental oscillator that provides a swept frequency response over one or more octaves. Previously, the output was often switched or split to use as an local

oscillator (LO) to a low-frequency heterodyne source stage; many modern VNAs use direct digital synthesis to get the low-frequency response and divide the fundamental oscillator to get the higher range of the low-frequency path.

The swept frequency oscillator is typically phase-locked to a lower-frequency fractional-N (F.N) circuit or direct-digital synthesizers (DDSS). In older VNAs, the phase locking was accomplished through the reference receiver, which reduced cost but required that the reference path signal remain present for all measurements. Modern analyzers separate the frequency synthesis for sources from receivers for greater flexibility and no longer require the any signal be present in the reference path. Because of this, the RF signal may be pulsed modulated without losing phase lock on the synthesizer.

The output may also go to multiple stages of division or multiplication, followed by amplification and filtering. The final output signal is combined from each of the input signals, resulting in broad frequency coverage. Typically, this common output has some RF detector on it to provide for ALC loop operation, maintaining constant source power over the different frequency bands and compensating for amplifier flatness. But often the ALC attenuator drive can be operated in an open-loop mode to extend the power range beyond that of the detector diode.

The level control circuits often use an amplitude modulator before the amplification chain to complete the ALC loops. In some modern VNAs, a pulse modulator is added as well to provide for high-speed pulsed RF measurements. When a pulse-modulator is used inside the ALC loop, the ALC function must be disabled as it will attempt to respond to the pulsed signals. In this open-loop mode, more sophisticated calibration or digital control must be used to control the source output power. Recently the use of a reference or test receiver as the ALC loop power control, rather than the internal diode detector, has become more common. This form of receiver leveling provides accurate results if the receivers are calibrated, a much wider range of leveling than the diode detector can support, and the ability to be programmatically controlled to correct for any external path loss or to provide a prescribed power profile.

2.2.2 Understanding Source-Match

One of the most confusing issues with respect to VNA measurements is the idea of source-match. Source-match also affects non-VNA measurements when signal sources are connected directly to a DUT, but in such cases it is almost always ignored as its affect cannot be determined; but in VNA measurements, the port 1 reflectometer can help to determine the exact effect of mismatch between the source and the DUT and correct for these errors. In fact, there are three different and distinct source attributes that are often confused as the source-match of a VNA, and each affects different measurements in different ways.

2.2.2.1 Ratio Source-Match

The ratio source-match is that match that will affect the results of a ratio measurement, given a DUT is not perfectly matched to the reference impedance. The value of the ratio source-match, most commonly called the *raw source-match* or *uncorrected source-match*, is derived from a combination of the quality of the reference channel signal-separation device and any mismatch between this device and the input port of the DUT. This is always the value used to compute the uncertainty or accuracy of a gain or return loss measurement. But this match

applies only to parameters that have a ratio of some receiver to the reference receiver. In this case the reference receiver is used to measure the change in source drive power incident to the DUT, and the error-correction math removes the ripple at the output of the DUT associated with the measured incident ripple in power. Thus, for linear measurements such as S21, where the output signal is linearly related to the input signal, the ratio is not affected by changes in incident power to the extent that the reference receiver properly measures the actual incident wave at the DUT.

Errors in the measurement of the actual incident wave are mostly attributed to the ratio source-match of the VNA. The ratio source-match can be determined during the calibration process and is shown in Figure 2.5 for two cases of reference channel signal separation: the upper trace is using a two-resistor power splitter, and the lower trace is using a directional-coupler. While the detailed response is different, the overall quality is quite similar between the two cases.

When a splitter is used, since the splitter uses equal $50\ \Omega$ resistors in most cases, the input match to the splitter (as it appears from the source) is nominally $50\ \Omega$, and the loss through the splitter is about 6 dB. The mathematical process of taking the ratio has the effect of creating a virtual ground at the common node of the splitter, so the ratio source-match is a measure of the quality of the internal $50\ \Omega$ resistor.

Interestingly, for the case where the reference comes from a coupler, in the absence of other sources of mismatch after the coupler, the ratio source-match will be identical to the

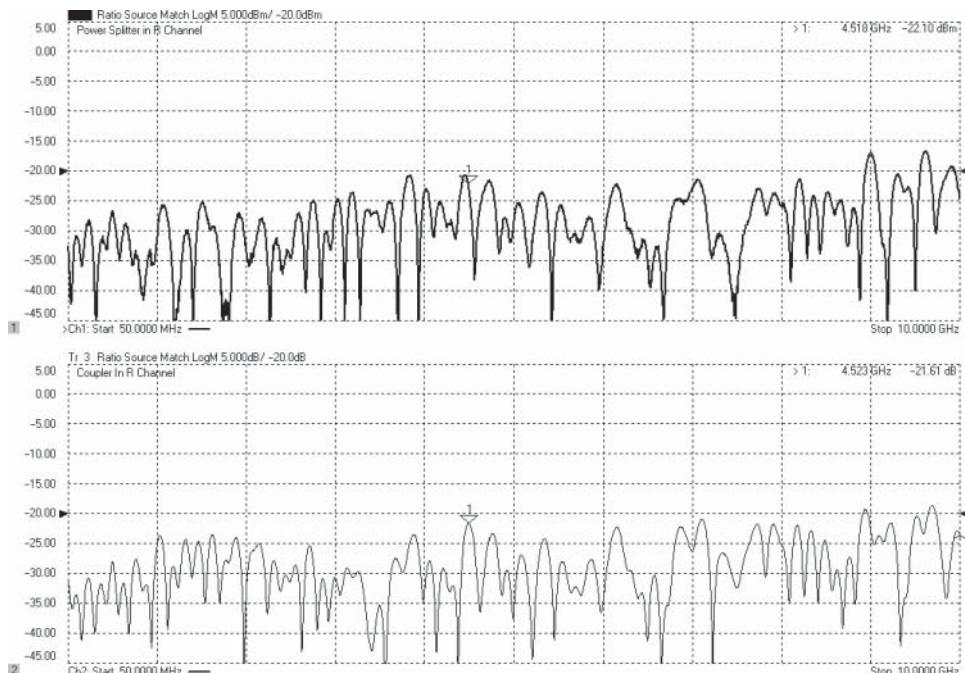


Figure 2.5 Ratio source match: trace when using a power splitter (upper) and trace when using a directional-coupler (lower).

directivity of the reference coupler. This makes sense as directivity is a measure of the reverse signal leaking into the coupled port, and this signal adds to the reference receiver reading even though it is not part of the incident signal, thus causing an error in ratio measurements.

2.2.2.2 Power Source-Match

The power source-match is the value that describes how the output power of the source varies with the applied load. If the power match is zero (perfectly matched), then the output forward wave, $a1$, would not be affected at all by the load. If the reference coupler is not ideal (that is, it has some leakage), it is possible have an ideal power source-match and a non-ideal ratio source-match. Conversely, if the reference coupler is perfect but the source has a mismatch, the ratio source-match may be ideal, but there is a larger power source-match error. In some ideally constructed S-parameter measurement architectures, where all components are ideal, the power source-match is not zero. Consider the block diagram of Figure 2.1 simplified in Figure 2.6, where a two-resistor power splitter is used, and the source impedance is also 50Ω .

From the test port one sees a series of 50Ω resistance (of the splitter), behind which is the 50Ω source impedance in parallel with 100Ω (50Ω from the splitter, 50Ω from the reference receiver, in series), to generate a power match of

$$Z_{PwrSrcMatch} = Z_{\text{splitter_main}} + \frac{1}{\frac{1}{Z_S} + \frac{1}{Z_R + Z_{\text{splitter_R}}}} = 50 + \frac{1}{\frac{1}{50} + \frac{1}{50 + 50}} = 83.3 \text{ ohms}$$
(2.1)

as the Thevenin equivalent impedance. From this it is clear that the for the two-resistor splitter case, even in an ideal case the power source-match cannot be Z_0 .

In the case where a directional-coupler is used in the reference channel, the nominal match may be much closer to Z_0 . The result of this non- 50Ω equivalent power source-match is that when the DUT is not matched, there will be a reflected signal that, while it will be detected by the reference channel and thus compensated for in gain measurements, will cause the $a1$ wave to vary from the value one sees when the port is terminated in 50Ω and thus cause error or ripple in the drive power to the DUT, as illustrated in Figure 2.7. The figure shows, in the dark trace, the incident power at $a1$ for a load at the test port. It is not perfectly flat because of other mismatches in the system after the reference coupler. The light trace shows the result of an open at the test port. This large ripple is because of the poor power source-match and generates an error in the incident signal of nearly 1 dB. Since this is power measured at the $a1$ receiver, it

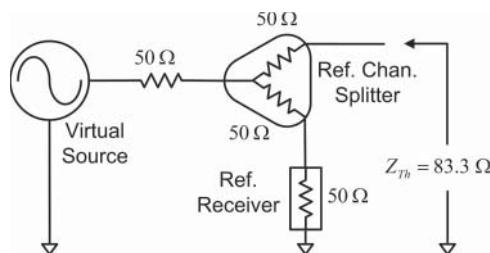


Figure 2.6 Simplified diagram of source power match.

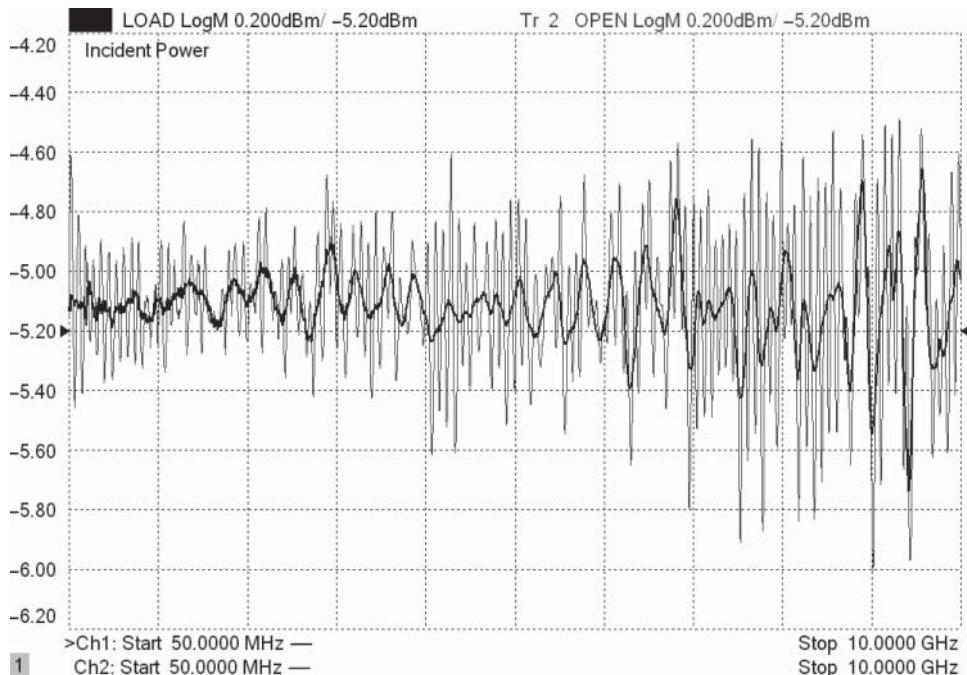


Figure 2.7 Measured incident power into a load termination and an open termination for a VNA with a coupler in the reference channel.

is related to the ripple in the output power, but other loss or reflections between the reference splitter and the test port can affect the output power. Other reflections past the reference splitter will add to the power source-match but are not represented in the $a1$ measurement. In the case of a linear device, which S-parameters presume, this is of no consequence; but in the case of non-linear devices, such as amplifiers in compression, this will directly affect the reported output power. In such a case, the drive power from the VNA can be higher or lower than the displayed power setting, so the amplifier will be further, or lesser, in compression and the power reported will be in error if based on the input drive level.

The power source-match is quite difficult to determine as it is apparent only when there is a mismatch applied at the test port. In essence, one must vary the impedance applied to the test port and measure the change in power coming out of the drive port to infer the power source-match; it cannot be directly measured. By using a “long-line” technique, a line-stretcher, a sliding mismatch, or an impedance tuner as a termination of the test port, and adding a coupler to sample the incident wave (shown in Figure 2.8), one can determine the power source-match from a series of measurements, where the line stretcher or tuner is changed.

A *line-stretcher* is a transmission line structure, usually stripline, that allows the length of the transmission line to be varied. These are sometimes called *trombone-lines* because the center conductor is constructed as a trombone-like slider. An example of one is shown in Figure 2.9.

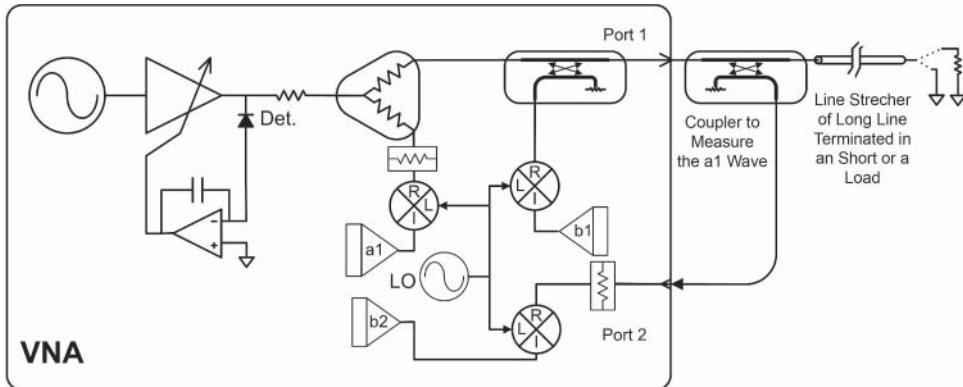


Figure 2.8 Block diagram for measuring power source-match.

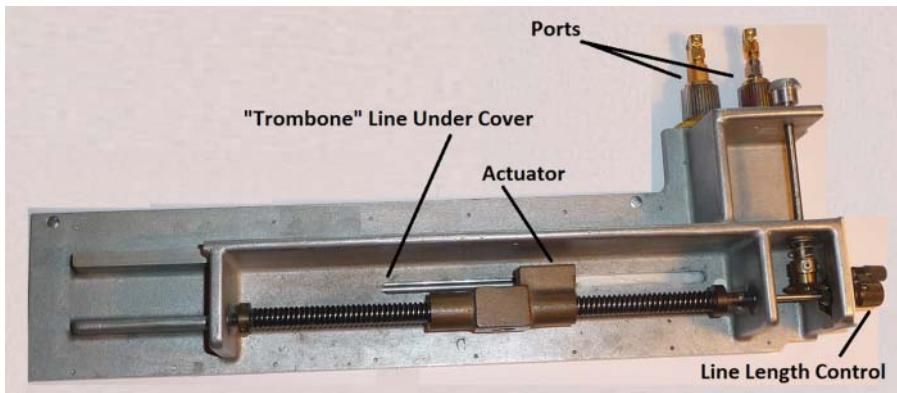


Figure 2.9 A line stretcher used for match measurements.

The traces shown in Figure 2.10 measure the $a1$ wave from the test port output by adding an external coupler and routing the coupled arm to the $b2$ receiver. The main arm of the coupler is connected to a power meter, and the power is set to obtain -10 dBm; then the $b2$ receiver is calibrated to this output power. Next the power meter is removed, and a long line terminated in full reflection (short or open) is put in its place. The ripple on the trace is an indication of the power mismatch. Though not shown, in this case the effective source-match (or ratio source-match) is good, but the output power has ripple related to the power mismatch. When the line is terminated with a short or open, the peak-to-peak ripple is exactly the voltage standing-wave-ratio (VSWR) of the power source-match. The upper trace is a measurement of a system with a power splitter used for the reference signal separator; the lower trace shows the same measurement, but this time the power splitter is replaced with a directional-coupler for the reference-channel signal separation. Clearly, there is an improvement in the power source-match using the coupler.

At any frequency, the VSWR of the power source-match can be determined from this response; the envelope of the response can be used directly, or the line stretcher may be

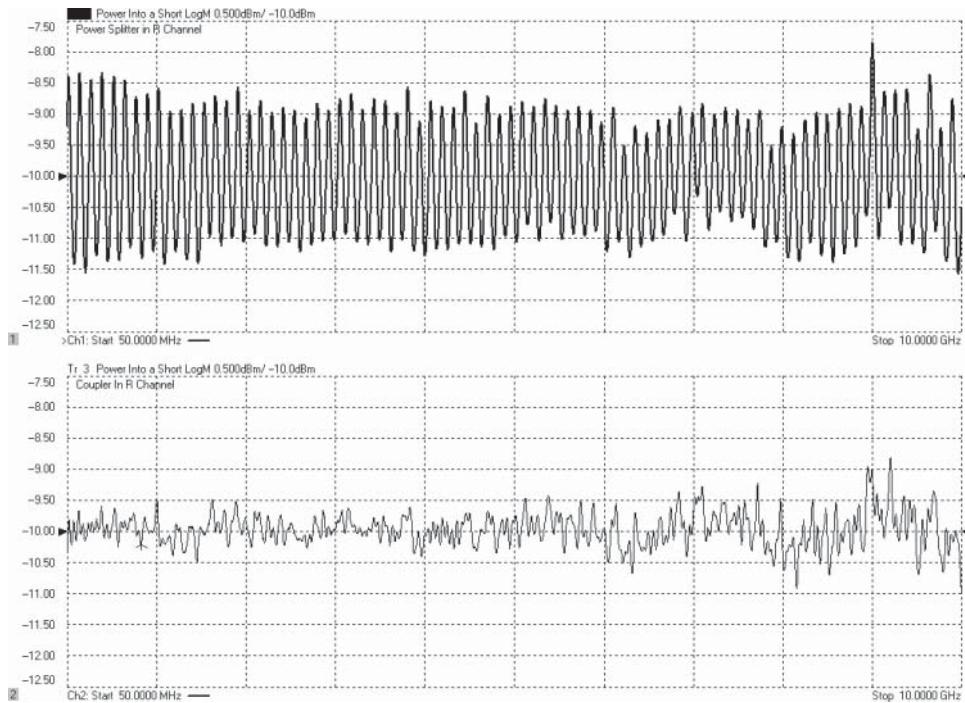


Figure 2.10 Measurement of long line indicating power source-match using an external coupler terminated in a short: a two-resistor power splitter (upper) and a trace for a coupler in the reference path (lower).

adjusted to obtain a peak and valley at any frequency of interest, from which the power source-match is computed as

$$SM_{PWR} = 20 \log \left(\frac{1 - 10^{\frac{VSWR}{20}}}{1 + 10^{\frac{VSWR}{20}}} \right) + 2 \cdot L_{CM} \quad (2.2)$$

where $VSWR$ is the peak-to-peak ripple in dB found at the output of the monitoring coupler and L_{CM} is the loss in the main arm of the monitoring coupler. In the upper trace of the previous example, the p-p ripple at low frequency is about 3.0 dB, and the mainline loss of the external coupler is about 1.6 dB, so the power source-match is

$$SM_{PWR} = 20 \log \left(\frac{1 - 10^{\frac{3}{20}}}{1 + 10^{\frac{3}{20}}} \right) + 2 \cdot 1.6 = -15.3 + 3.2 = -12.1 \text{ dB} \quad (2.3)$$

This is almost exactly the power match expected from a 50Ω splitter (83.3Ω or -12.05 dB). The lower trace shows a power source-match for a directional-coupler of around -21.6 dB at low frequencies, and -18 dB at higher frequencies.

Extracting the effective match from a mismatch ripple is a technique that will be useful for many other analyses in component measurement. An alternative test method uses a mismatch

pad connected to another of the VNA ports, and the ripple is measured at the second port. In such a case, the VSWR is computed into the value of the mismatch, and the return loss value of the mismatch pad is subtracted to the measured VSWR of the source to find the true value. The ripple here is different from the reference channel ripple of the measured wave at the *a1* receiver shown in Figure 2.7, as the effects of mismatch after the reference splitter are not apparent at the *a1* receiver.

The power source-match is normally set by the ALC loop in the source, and as such the reflections from the DUT are sensed by the detector diode in the ALC loop such that the source power is adjusted to maintain a constant voltage at the detector.

2.2.2.3 Source Output Impedance

The effective output match of the VNA source is the same as the power source-match over the region within the ALC loop bandwidth of the source signal. In that region, the ALC loop responds to reflection signals as described earlier. Outside the ALC loop, the source presents a different output impedance. This is the reflection that would be measured in response to a signal not related to the source output signal. This source reflection is important in cases where the DUT presents other signals reflected into the source, such as from a mixer, or inter-modulation products from the input of an amplifier. This value represents the manner in which these other signals would be reflected back out of the source. It can be measured directly as any other reflection coefficient, using a separate VNA reflectometer; an example measurement is shown in Figure 2.11, where the internal source is set to 1 GHz. Here, the internal VNA source frequency is shown as a large spike at 1 GHz on the source output-impedance measurement. This is again a case where the coupler in the R-channel used to sample the incident wave provides better match (lower trace) than a two-resistor splitter (upper trace), at least at low frequencies where mismatch after the reference splitter is minimal.

2.2.3 VNA Test Set

2.2.3.1 Test Set Switch

In some VNAs the source is switched between ports using a test set switch, which can come before or after the reference channel splitter. The termination of this switch provides the load match of the port when the source is not active on that port. This load match is not the same as the source-match (ratio or power) and so some advanced calibration techniques that rely on the port match being consistent whether the port is a source or a load must be modified, as discussed in the next chapter. If the switch comes before the reference channel splitter, there will be a reference channel receiver for each port (four-receiver VNA, Figure 2.2 lower). If the switch comes after the reference channel splitter (three-receiver VNA, Figure 2.2 upper), the reference channel is shared between ports. It samples the source signal only when the source is active. This three-receiver architecture does not support some calibration methods, such as thru-reflect-line (TRL) and so modifications and compromises to the calibration methods must be made.

The short explanation for the difficulty is that TRL calibration methods require measuring the load match of port 2, when port 1 is active. To do this measurement, the ratio of $a2/b2$

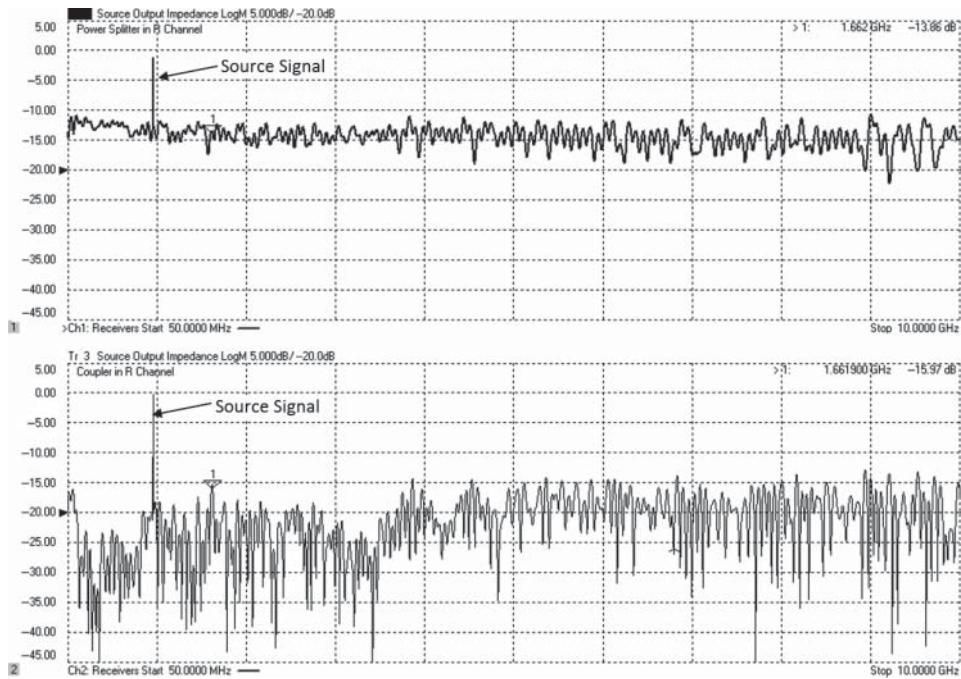


Figure 2.11 Measured source output impedance away from the source frequency: a trace using a power splitter in the R channel (upper) and a trace using a coupler (lower).

is acquired during the thru step. But there is no a_2 receiver available in the three-receiver architecture. Modifications can be made that assume the source-match and load match of the port are identical, but this case is not common unless attenuation is added after the reference channel split. Attenuation added reduces the difference between source and load match at a port by twice the attenuation value. Pre-characterization of the difference in source and load match, called the *delta-match*, can be performed and removes the need for characterization at the time of calibration. This allows three-receiver architectures to support the same calibration as four-receiver architectures and is found in some of the more modern low-cost analyzers.

2.2.3.2 Step Attenuator Effects

In some VNA designs, a step attenuator is added between the reference coupler and the test coupler to allow a greater change in the source power setting beyond that which the source ALC circuit can produce. This step attenuator has the additional benefit of providing a good match to the test port. In cases where the power source-match and ratio source-match are not the same, the step attenuator reduces the difference between these values based on twice the attenuation value. Reducing the difference between power source-match and ratio source-match allows one to compute the error in source power from the ratio source-match, which is determined as part of the normal calibration process. In general, the power source-match is not characterized during any normal calibration.

Another issue to be concerned about with step attenuations is their effect on the quality of the measurement when the attenuation value is changed. In most newer VNAs, the nominal attenuator value is known, and the effective value of the reference receiver is compensated for when the attenuator value is changed. The source ALC power is also changed, so that changing the attenuator value causes only a slight change in the value of the power coming from the test port; the nominal attenuator value is usually within 0.25–0.5 dB of the actual attenuator value. Since the port power stays the same, the internal source power must be raised by the amount of increase in the step attenuator. Since the reference receiver comes before the step attenuator, it will see a larger signal value; as it is desired to have the reference receiver power display the same value as the port power, its reading is also decreased by the value of the step attenuator.

Placing the step attenuator at this point in the block diagram has a distinct advantage in that it allows a large signal in the reference channel even when a small signal is needed at the test port, providing a low noise signal.

The loss of the step attenuator is well compensated for, but its effect on the match is not. If the preset condition has the step attenuator set to 0 dB, the match of the port is terminated back in the test port switch. When even a single stage of test port attenuation is used, the predominate source and load match characteristic is set by the match of the attenuator, which is typically quite good. Thus, it is good practice to use some source step attenuation if the maximum test port power is not required for the measurement. And since the raw match is better for any attenuator setting other than the 0 dB step, the effect on a calibrated measurement if the attenuator is switched to a different, non-zero, value is smaller if the calibration is performed with some attenuation applied.

Some older VNAs did not allow changing the attenuation value after calibration and did not compensate for the nominal value of an attenuator change; error correction in these VNAs is often turned off for a step attenuator change.

In general, changing a step attenuator will change all the raw error terms on the port that has the step attenuator. Techniques are discussed in Chapter 3 that can compensate for much of this change.

2.2.3.3 Test Set Reflections

In addition to the source-match effects produced by the source impedance, power source-match, and ratio source-match, reflections from within the test set of the VNA will also exist, as well as from the test port cables and from any fixtures that provide an interface from the VNA to the DUT. These sources of mismatch are common to all of the previously mentioned source-match effects and will add to them in a similar way. However, since they are common, their effects on port power and gain are also the same.

The reflection and mismatch between the reference channel split and the test port coupler affect the incident signal, $a1$, but are not monitored by the reference channel receiver. Reflections after the test port coupler also affect the $a1$ signal but will be apparent in changes measured on the reflected signal, $b1$. However, their composite effect will add to the overall source-match, and their effects on measurements can be compensated provided they remain stable. In addition, mismatch and loss after the test port coupler can be characterized in such a way that changes to these values, such as due to drift in a test port cable, also can be compensated in some cases. Mismatch correction in power measurements is discussed in detail in Chapter 3.

2.2.4 Directional Devices

One vital VNA component is the directional device used at the test port to separate the reflected wave from the incident wave. This is most often a directional-coupler or directional bridge, although simpler structures have been proposed as well. These devices are characterized by their main-line loss (the attenuation of the $a1$ signal), the coupled-arm loss (the attenuation of the $b1$ signal), and their directivity (the ability to separate the $b1$ signal from the $a1$ signal). In addition, any mismatch in the directional device will contribute to the port match and source-match. If there is mismatch before the directional device, it will have no effect on its directional characteristics (directivity or isolation). However, any mismatch after the directional device, such as in a test port cable or fixture, will contribute equally to mismatch and degradation of directivity, as described in Section 1.10.

2.2.4.1 RF Directional Bridges

Most RF VNAs make use of a directional bridge, which has the important characteristic of maintaining good coupling and isolation over very wide frequency ranges and at very low frequencies. While the most common implementation of a bridge is a balanced Wheatstone bridge, this simple implementation can be modified to create a component that has characteristics similar to a directional-coupler, but with much wider frequency range and low frequency of operation. A bridge is often used in metrology applications where balance in a DC resistive path provides a measure of some quantity such as the power absorbed by a load (see Section 1.15). To understand how these bridges can be configured as directional-couplers, with low loss and high isolation, consider the diagram in Figure 2.12, which is a common representation of a Wheatstone bridge.

In this configuration, the signal from the source is applied across the top and bottom of the bridge, and if the ratio of $R1/R2$ is equal to $R4/R3$, the net voltage across R_{det} (which in a common bridge represents the meter movement) will be zero, and no current will flow through the detector.

In a thermistor, all the resistors are $50\ \Omega$, and one of them represents the RF input of the power sensor, typically $R3$. A DC signal is applied from the source across the bridge, and the imbalance is measured as the voltage difference across the R_{det} resistor. In an RF bridge it is desired to isolate the bottom node of the bridge from ground, and so a transformer is added for this purpose, as shown in Figure 2.13. This 1:1 transformer performs the function

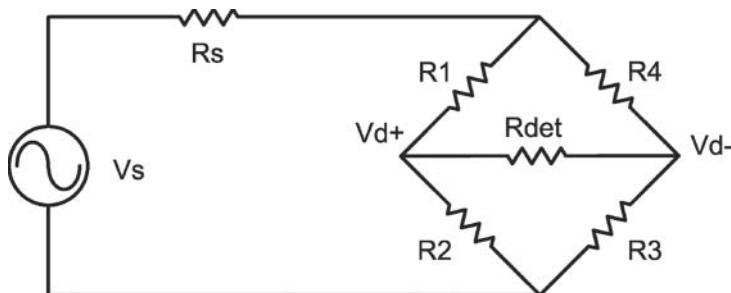


Figure 2.12 Schematic of a directional bridge.

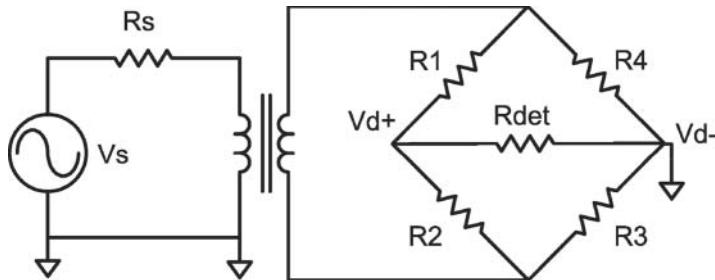


Figure 2.13 Adding a transformer between the source and the bridge.

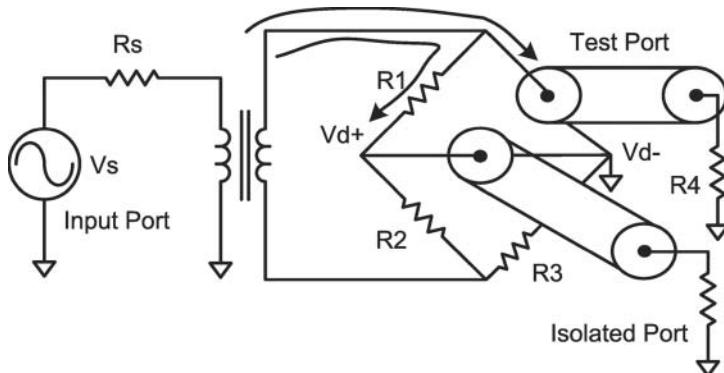


Figure 2.14 Replacing bridge elements with RF ports.

of a BALanced-UNbalanced transformer (or balun), changing the unbalanced (or grounded) source into a balanced signal across the bridge. Doing this allows grounding a different leg of the bridge, which as will be seen is key to making a bridge act as a directional-coupler.

From this modification, the RF implementation of the bridge can be better understood. Since the low side of the detector is now ground, the resistor represented by Rdet and R4 can be replaced with transmission line structures of similar impedances, representing the RF ports of the directional bridge, as shown in Figure 2.14. In this figure, the Rdet resistor is replaced with the coupled port of the bridge, and one can see that the RF energy flowing from the source appears equally at both the center conductor and the ground of isolated port.

However, since the RF current appears at the test port, relative to ground, a portion of the RF signal will appear across R4; the relative value of the voltage on R4 to Vs/2 is the insertion loss of the directional bridge. If the bridge uses equal resistors, then R1, R2, R3, and R4 as well as Rs are all $50\ \Omega$. With these values, it is easy to see that Vs is applied equally to R1 and R2, as well as R3 and R4, so that the voltage across R4 is one-fourth the source voltage. Therefore, the loss of an equal resistor balanced bridge is one-half voltage applied at the bridge input, or -6 dB . In general, the insertion loss of a bridge, where $R_S = R_4 = Z_0$ is

$$L_{Bridge} = 20 \log \left(\frac{Z_0}{Z_0 + R_3} \right) \quad (2.4)$$

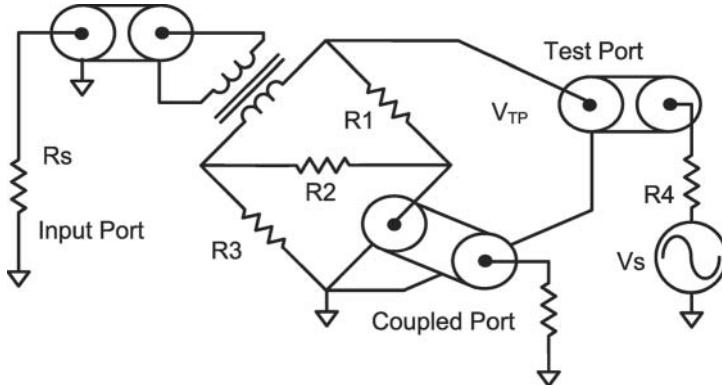


Figure 2.15 A bridge redrawn to show the coupling factor.

From this description we can see that in the case where the bridge is terminated in Z_0 , there is no signal in the isolated port, demonstrating that this bridge isolates the incident signal. The first criteria of a directional device is satisfied. The second criteria is that the bridge does respond to the reflection signal from the test port. To understand how that occurs, it is useful to redraw the bridge, bringing the ground point of the test port down to the bottom of a redrawn circuit, as shown in Figure 2.15.

In this drawing, the source has been moved from the input to the output, but the bridge circuit is topologically identical to the previous figure. When driven from the test port (or when measuring a reflected signal), the isolated arm becomes the coupled arm, and the coupling factor of the coupled arm can be computed as

$$C_{Bridge} = 20 \log \left(\frac{Z_0}{Z_0 + R_1} \right) \quad (2.5)$$

For the case of an equal resistor bridge, the coupling factor is equal to the loss, -6 dB. If R_1 is not equal to Z_0 , R_3 can be computed as

$$R_3 = \frac{(Z_0)^2}{R_1} \quad (2.6)$$

Note that the loss is directly proportional to the coupling as

$$L_{Bridge} = 20 \log(1 - C_{Bridge}) \quad (2.7)$$

For RF VNAs, it is common to use a directional bridge in the test set. Directional bridges of this type have been used since the 1970s, and an example of such a bridge used in the HP 8753B is shown in Figure 2.16. This bridge has been modified to have an unequal coupling and loss, so the insertion loss is lower than normal (around -1.5 dB), and the coupling is higher than normal (around -16 dB) for a Wheatstone bridge.

The RF performance of such a microwave bridge is shown in Figure 2.17. The insertion loss increases with frequency due to the loss of the coax balun and increased coupling due to parasitic series inductance in R_3 . This same inductance causes a degradation of directivity in

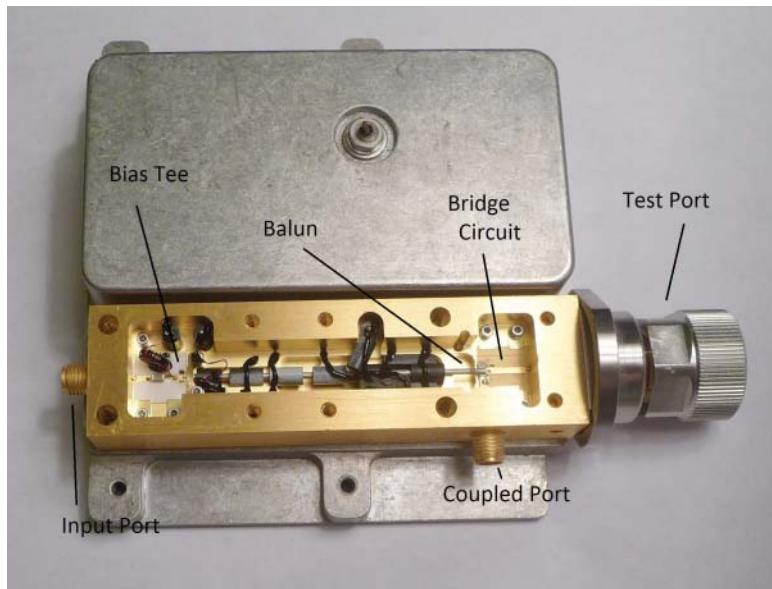


Figure 2.16 An example of a directional bridge from the HP 8753B.

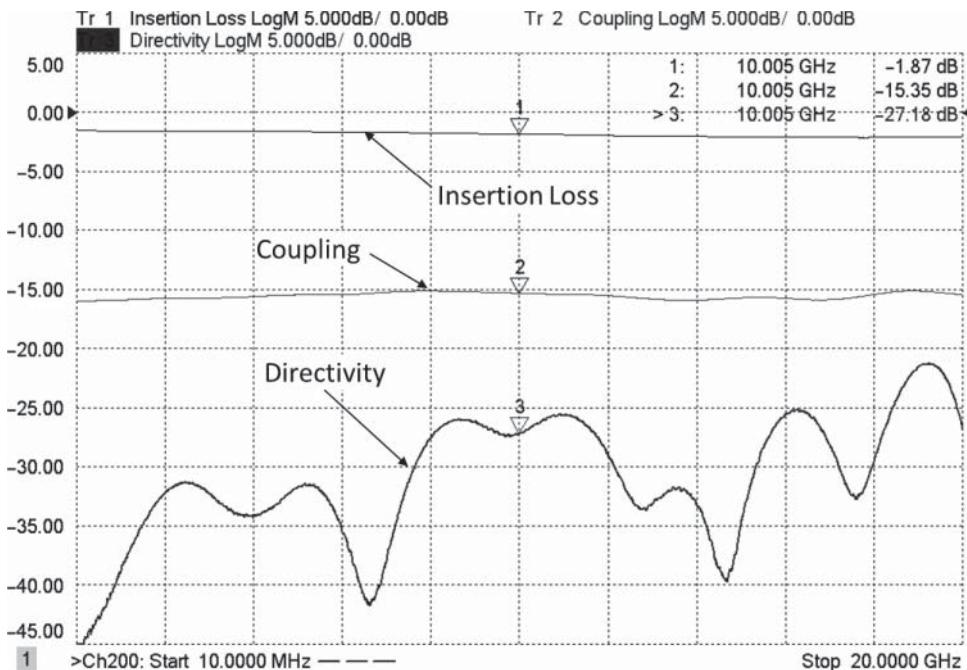


Figure 2.17 RF performance of a directional bridge.

the bridge as frequency increases. Bridges are inherently lossy structures, where some of the power is absorbed by the resistive elements in the bridge. The power absorbed by the bridge is equal to the insertion loss of the bridge minus the power coupled to the coupled port.

Bridges of this type have been used successfully up to 27 GHz.

2.2.4.2 Directional Couplers

Directional couplers are more often used in higher-microwave frequency ranges because of the difficulty of maintaining good bridge performance at high frequencies. Directional coupler design is a broad topic, and much literature has been devoted to structures that can be used as couplers. However, for use in VNAs, there are some particular characteristics that are critical. In general, commercial directional-couplers are designed to maintain a flat coupling factor over their bandwidth, and the bandwidth is limited by this coupling factor. Couplers used for VNA reflectometers require wide bandwidths, so rather than a flat response, they are often designed with an equal-ripple or Chebyshev response. Ripple in the loss or coupling factor is not much concern in a modern VNA, where calibration techniques can remove almost any frequency response error. Isolation is an important criteria in VNA couplers. One attribute about directional-couplers that distinguish them from bridges is that they are ideally lossless devices such that all the power applied is either coupled (to the coupled port or the internal load) or transmitted through the coupler. The relationship between insertion loss and coupling factor is

$$L_{Coupler} = 20 \log [1 - (C_{Coupler})^2] \quad (2.8)$$

Directional couplers typically come in one of three forms: waveguide couplers, microstrip couplers, and stripline couplers.

Waveguide couplers are most common at mm-wave frequencies but have the inherent limitation of narrowband operation due to the narrowband nature of waveguides. The structure of waveguide couplers is a 4-port device with the main arm connected in such a way as to have irises (or holes) to a second waveguide. The second waveguide can have two ports or one port internally terminated. The nature of the coupler is symmetrical. In theory either port can be the coupled port; in practice a load is often embedded in the coupled arm. Because of the fundamental function of a waveguide coupler, the forward coupled wave comes out of the waveguide port nearest the test port. This often causes confusion in the symbols used.

A microstrip or stripline coupler uses a different electric-magnetic (EM) configuration to perform coupling, and the coupled arm of these couplers is the one farthest from the test port. Microstrip couplers often suffer from the fact that there is some dispersion in microstrip lines, and since the even- and odd-mode waves in the coupled lines experience different effective dielectric constants, they will have different velocities of propagation. This makes it more difficult to create microstrip couplers with good isolation. For this reason, many VNA couplers are in the form of stripline (or slabline, which is similar to stripline but with a rectangular center conductor thickness), suspended in air. These couplers are designed to have very stable coupling and isolation factors. For a VNA, it is not so important what the exact directivity is, as long as it is completely stable. Figure 2.18 shows an example of a directional-coupler used in VNAs. The test port connector is one attribute that differentiates this from a commercially available directional-coupler that might be used as component in



Figure 2.18 A directional coupler used in VNAs.

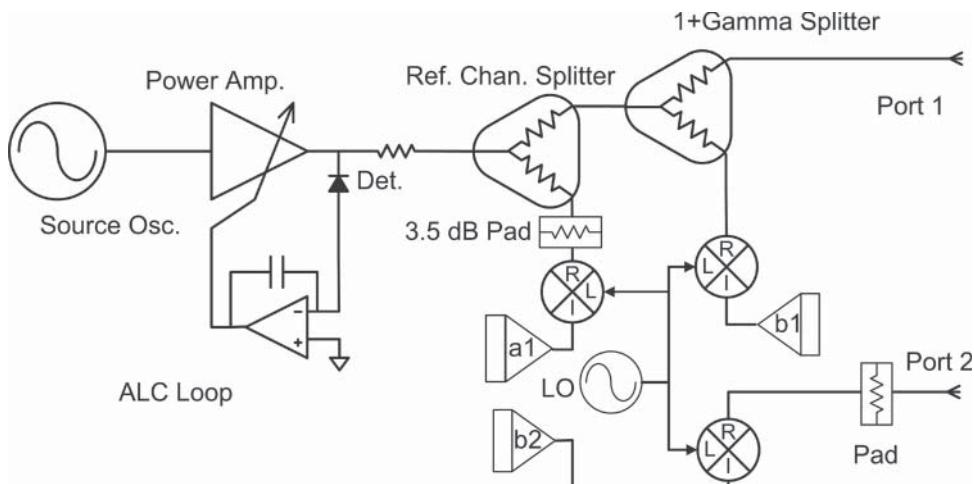


Figure 2.19 Block diagram of a 1+gamma reflectometer.

a different system. This connector is designed to be firmly mounted to the VNA front panel and withstand numerous connections and reconnections. This coupler has an integrated load and so exposes only three ports.

2.2.4.3 1+Gamma

Another proposed reflectometer structure is a 1+gamma structure, whose name comes from the block-diagram architecture, shown in Figure 2.19. As the name implies, the signal at the b_1 receiver is a combination of the incident (a_1) and reflected (gamma) signal.

In this configuration, the signal in the test or b_1 receiver never goes to zero; rather, it is minimum with a short, maximum with an open, and nominal 1 when there is a load attached. Also, the signal variation between an open and short is about 14 dB less than that for a bridge

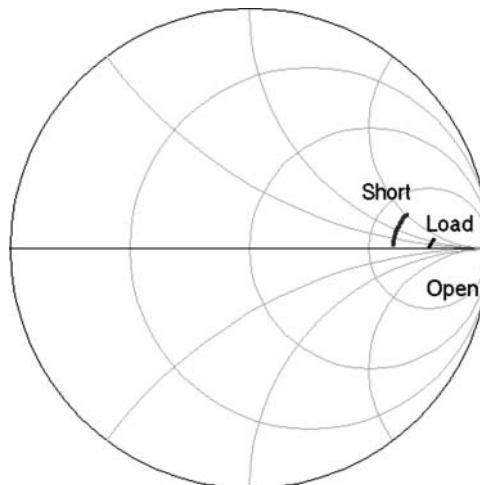


Figure 2.20 Smith chart showing reflections of a 1+gamma bridge with an open, short, and load.

or directional-coupler. Put another way, the reflection gain of the 1+gamma bridge is lower than for a directional-coupler or bridge. Consider the Smith chart in Figure 2.20; an open, short, and load (all non-ideal with fringing capacitance and series inductance) are shown for each on a 1+gamma reflectometer.

The value of attenuation in the reference channel is adjusted to set the value of the open circuit reflection to 1. For a directional-coupler, the load gives a zero reflection (ideally), and the short gives a -1 reflection. For the 1+gamma bridge, the open is also 1, but the short is $+0.6$, and the load is $+0.75$; thus, the difference between the open and the short moves from 2 to only 0.4. These reflections are mapped to the full Smith chart through the error correction math, in such a way that the values from the reflections, and any instability, are multiplied by 5. Also, since the load condition has a large signal in the $b1$ receiver, any instability in that signal is apparent as a directivity error, which is also multiplied by 5. In theory, if directivity is defined as the average of the open/short response relative to the load response, then the directivity of a 1+gamma reflectometer is about 0 dB (remember that directivity for a coupler or bridge is always positive, often 20 dB or more).

Theoretically, any directivity error can be corrected for by a calibration, but in practice, certain unstable errors can cause uncorrectable errors when the directivity is poor. Thus, 1+gamma structures have largely disappeared from use. Also, this same multiplying effect causes any slight drift in the test port cable to cause a considerable change in the measured reflection coefficient, after calibration.

2.2.5 VNA Receivers

The final RF components in a VNA block diagram are the test and reference receivers. Dynamic range is a key specification of a VNA and is sometimes referred to as the difference between the maximum signal level that the receiver can accept while still operating and the noise floor of the receiver. In most cases, the maximum damage level is

significantly above the maximum operating level of the receiver, which is usually limited by the input compression level of the receiver. The maximum operating level is set by the structure of the components, but for most modern VNAs, it is around the -5 dBm at the receiver mixer input, or about $+10$ dBm at the test port, after considering the coupling factor of the test port coupler. The noise floor of the receiver is set primarily by the type of mixing down-converter used, of which the two principal types are sampling down-converters (or samplers) and mixers.

2.2.5.1 Samplers

The sampling down-converters are circuits that are driven by a low frequency pulse, which has very high harmonic content. The example circuit shown in Figure 2.21 is typical of older VNA sampling receivers such as found in the HP8753 or HP8510 VNAs. In the circuit, the diode pair acts like a switch, which is driven with a very short pulse from a pulse generator driven by a voltage-controlled oscillator (VCO), operating at relatively low frequencies. The short conduction angle (the amount of a cycle that the diodes conduct) of the pulse means that the frequency content is very high, sometimes referred to as a *harmonic-comb*, and the sampler can convert frequencies much higher than the VCO drive.

It is not unusual to use harmonics of up to 200 times the VCO frequency. Because the conduction angle is so short, the effective input impedance is high, essentially multiplied by the maximum harmonic number, and this means the effective noise figure of the sampler is high as well. Since the conduction angle does not depend upon the frequency being measured, the noise figure does not depend upon the frequency being measured. The conduction angle and conversion efficiency can be adjusted by adjusting the diode bias so that they are not quite turned on, and the pulse provided by the VCO gives the extra current to turn on the diodes fully.

Some advantages of samplers are that they do not require a high frequency local oscillator, as a mixer would, and they can simultaneously down-convert a signal and all its harmonics to the intermediate frequency (IF). This capability is exploited in a sampling oscilloscope, and now in some of the modern versions of sampling VNAs used for non-linear measurements.

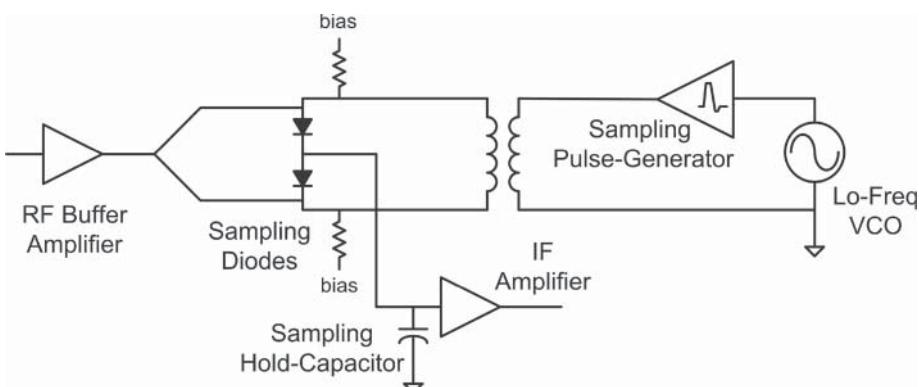


Figure 2.21 Schematic of a sampler.

However, for the most part samplers have been largely abandoned in VNAs due to a number of difficult problems that they present.

The foremost problem is the degraded noise floor in the sampler-based VNA. The effective noise floor is further reduced as the conversion efficiency of higher-order harmonics typically degrades near the top of the sampler frequency range. Almost all modern VNAs use some form of mathematical response correction on the sampler response so that frequency response of the VNA receiver to a constant input power appears flat over its entire frequency range. This response correction, which removes the effect of roll-off in real conversion loss, has the consequence of increasing the apparent noise floor of the sampler at higher frequencies.

A secondary problem with samplers is that the rich harmonic-content of the VCO dictates that the sampling receiver has many regions where it is sensitive to other signals, such as harmonics or spurious signals of the VNA source or DUT, and spurious signals present at the DUT output. This makes sampling receivers particularly poor at measuring mixers or frequency converters, where the sampling comb-tooth can cross mixer output signals at many different frequencies. Figure 2.22 shows an example of spurs from a source generating responses in a sampling receiver. In this case, the source signal is generated by mixing a 3.8 GHz fixed RF signal with a 3.8–6.8 GHz swept yttrium-iron-garnet (YIG) oscillator LO. The mixed product provides the desired 0–3 GHz source output, but spurious signals at $2 \cdot RF - LO$ and $3 \cdot RF - 2 \cdot LO$ do show up in the 0–3 GHz VNA measurement receiver band. While small, these spurious responses do degrade the S21 accuracy.

Another difficulty with sampling mixers becomes apparent when measuring filters in the stopband and is caused by remixing signals reflected off the DUT back into the input reflection receiver such as the *b1* receiver. Because this effect has the appearance of a signal bouncing off the input reflection of the filter stopband and then bouncing off the *b1* mixer (at a different frequency), it is sometimes called *sampler bounce* (or *mixer bounce* in the case of mixers). Designs of these components must be carefully considered to avoid these bounce signals, and the basic design of samplers make them especially susceptible to this particular type of crosstalk.

For these reasons, and the fact that creating wideband mixers with full-band RF-frequency LOs has become much more cost effective, the use of samplers in VNAs has been phased out.

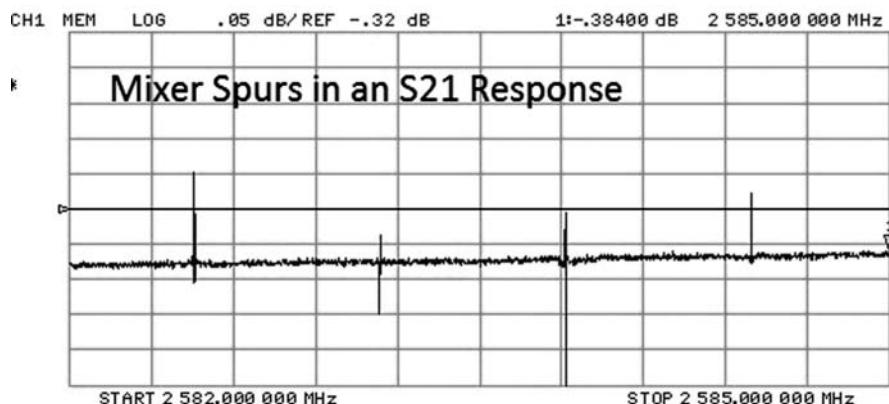


Figure 2.22 Spurs from a source crossing a harmonic of the VCO.

2.2.5.2 Mixers

Almost exclusively, modern VNAs use mixers as VNA receivers. The mixers are driven by a fundamental or low-harmonic-order LO. This provides a much larger conduction angle for frequency conversion and as such provides a lower noise floor. The trade-off, of course, is the cost of the LO, which may be due to its increased complexity.

2.2.5.3 Noise Floor

With mixers used as VNA receivers, the critical performance attributes are noise floor and input compression point. A fundamental LO provides best performance; using a third-order harmonic of the LO to obtain higher RF frequency response theoretically degrades the conversion loss by 9 dB, and fifth harmonic degrades it by 14 dB, based the idea that the LO drives the mixer into square-wave conversion. This degradation in conversion loss is represented as an increase in noise floor for most VNAs, due to the internal factory-based response correction.

2.2.5.4 Spurious Responses

Mixers have much lower unwanted spurious responses compared to sampler-based receivers. The primary spurious (or unintended) conversion occurs at the IF image of the desired RF signal. If the RF is above the LO (low-side mixing), the image will be one IF below the LO (hi-side mixing). Because mixers used as VNA receivers have lower noise floor and lower spurious responses, a much wider class of measurements that can be performed including noise figure, two-tone IMD, and even some modulated measurements.

2.2.5.5 Phase Noise

The LO distribution attributes are critical in the performance of mixer-based systems. The phase noise of the LO contributes directly to errors (primarily trace noise) in the phase of the measured signal. Even the trace noise of the amplitude response can be degraded by phase noise as the signal moves across the IF bandwidth filter. Providing a common and coherent LO to all the mixers improves the trace noise of the measurement by allowing the effects of LO phase noise to be reduced when the measured parameter is a ratio, such as gain or return loss.

2.2.5.6 Isolation and Crosstalk

Isolation between VNA receivers is important when measuring high-dynamic-range devices such as filters. In almost all measurements, the reference path has a large signal (as it measures the incident wave) and so is a constant source of leakage signals. Partly for this reason, it is common to provide additional loss in front of the reference channel mixer (5 or 10 dB) to lower the incident signal level and provide greater reverse isolation for mixer bounce. This also helps to keep the reference channel mixer operating in its linear region for higher source power signals, avoiding compression in the reference mixer.

In fact, there are four primary crosstalk paths in VNAs, listed here:

1. RF signal from the internal source or reference receiver to the transmission test port. This crosstalk signal is independent of the DUT characteristics, and its level remains constant regardless of the DUT properties but will change with test frequency. Leakage of the reference mixer IF to the test mixer IF path has similar characteristics, but its value doesn't change with frequency.
2. RF signal from the reflection test receiver (*b1*) that leaks to the transmission test receiver. This signal depends upon the input reflection of the DUT; if the DUT is well matched, there will be no signal at the reflection receiver. Since this signal depends upon the DUT, correcting for it is more complicated.
3. RF signal that leaks from the test set switch, to port 2, and reflects off the DUT output match into the port 2 transmission test mixer. As this signal depends upon the DUT characteristics, correcting for it can also be more complicated. Modern VNAs that use separate sources instead of test set switches eliminate this source of crosstalk.
4. The final source of signal leakage is related to any test fixture or probing done to connect to the DUT. Leakage from port 1 to port 2 of the probes or fixtures is usually electric field radiation or magnetic field coupling between the ports. Since these fields are non-TEM, they do not remain constant with changes in the DUT characteristics, and their effect may not be well understood. Probe-to-probe isolation is a key problem in measurements, but one that is not well accounted for. Careful fixture or probe design that includes shielding is perhaps the best solution to this final leakage effect.

In most modern VNAs, the design of the mixers and LO isolation networks are such that the level of the first three sources of crosstalk are at or below the noise floor of the receiver. As such they can be ignored except in special cases where extended dynamic range is desired, as discussed in Chapter 6. The fourth cause of crosstalk is inherent in the fixtures or probes, and it can sometimes be removed with calibration. But since the source is often due to radiation from one port to the other, this radiation pattern depends in a complex way on the actual loading of the port and the structure of the DUT. For example, in a probed situation, leaving the probes up as an “open” calibration standard can cause the probes to act as E-field antennas and can produce crosstalk between the probes. Grounding the probes, to produce a short, can cause magnetic field coupling between the probes, again producing crosstalk. Both of these crosstalk terms are non-TEM, meaning they have E and H fields that propagate in the direction from port 1 to port 2. Normal calibration methodologies do not correct for non-TEM crosstalk as their values do not remain constant if the DUT configuration changes.

2.2.6 *IF and Data Processing*

The final hardware portion of the VNA block diagram is the IF processing chain. The VNA receiver converts the RF signal to a first IF frequency, which is further converted and detected in the IF processing path. In older analyzers, such as the HP-8510, this consisted of a synchronous analog second-converter that produced two DC outputs proportional the real and imaginary portions of the RF voltage at the receiver input. These DC voltages were measured with DC analog-to-digital converters (ADCs) that produced a digital representation of the real

and imaginary values. More modern IF structures such as in the HP-8753 or HP-8720 used a second stage of IF down-conversion to bring the IF signal down to a frequency where an AC ADC could directly sample the waveform. The final IF frequency was set by the sampling rate of the ADC.

2.2.6.1 ADC Design

Now, most modern VNAs incorporate a high-speed ADC and perform direct sampling of the first IF signal. An example of a VNA digital-IF block diagram is shown in Figure 2.23. The IF signal is preconditioned with adjustable gain to optimize the signal-to-noise ratio in the ADC. For some applications, it is useful to have a narrowband pre-filter before the ADC so that the IF can be switched between a wideband IF and a narrowband response. An anti-alias filter is used just before the ADC, with a bandwidth of about one-third to one-fourth that of the ADC clock rate.

The FPGA that processes the ADC readings can be configured as a digital second converter of flexible IF frequency, so the final digital IF frequency can be quite arbitrary. There are several modes of operation for the digital IF. For these high-speed ADCs, the raw ADC readings have very high bit rates. Some of the latest designs for VNAs have four channels of data, at 16 bits and 100 mega-samples per second to produce a data rate of 6.4 Gbps. Specialized conditioning of the signal and advanced digital signal processing (much of which is proprietary) can improve the performance of the IF ADCs to many more effective bits.

At these high data rates, the main CPU cannot process the data fast enough to keep up, so an FPGA is used to decimate and filter the signals before the processed data is sent to the main processor using shared DMA memory. The function of decimation and filtering is the basic data processing step of any digital IF; in this function, a measurement is performed by setting the source and receiver frequencies so that the first IF contains the signal of interest. The ADC samples the IF signal, typically with two to four times over sampling, although it can be as much as 60 or 100 times over sampled. A finite set of samples is processed by the FPGA to produce a final result that represents the real and imaginary parts of the signal being measured. For example, if the digital IF is operating at 100 Msps, the IF frequency is 10 MHz, and the IF filter is set to 100 kHz IF BW, then approximately 10 μ s of data are captured, or approximately 1000 data samples. These 1000 samples are processed by a multiply-add chain in the FPGA to both filter the response and extract the real and imaginary values. In this way, the 1000 samples are reduced to two samples.

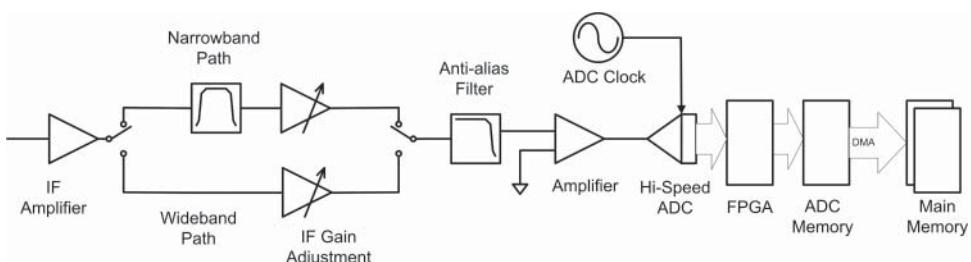


Figure 2.23 Digital IF block diagram.

A second mode of operation for the digital IF is “ADC capture” mode. In this mode, the FPGA does not process the data; rather, the data samples are simply captured into the local memory in the digital IF for a finite period of time. The entire ADC data stream is available for any further processing that might be beyond the algorithms available in the FPGA. Some modern VNAs have memory depths up to 4 Gb allowing deep memory captures. This mode of operation, while not typical, is useful for capturing anomalous effects such as transient or pulsed responses, as well as more complicated functions such as de-modulation of IF signals.

2.2.7 Multiport VNAs

For a class of RF and microwave devices, the normal port count of two to four ports is not sufficient, and multiport measurements are required. There are two distinct classes of multiport test requirements that generate the need for two distinct RF architectures to support the measurements. In the past, RF switching test sets provide the basis for these multiport extensions, but recently true-multiport VNAs have been introduced.

The first class of devices requires multiple sets of 2-, 3-, and 4-port measurements. As such, the native mode measurements of the VNA are sufficient, and all that is required is RF switching to route the VNA ports to the various port pairs of the DUT. One example of such a DUT is a satellite multichannel-diplexer (or multiplexer), which filters and separates signals from a common antenna path to each of several output channels, as shown in Figure 2.24. This unit has waveguide filters and interconnections to provide for the lowest possible loss.

This device requires two-port measurements for each path from the common port so a 2-port VNA with one common port and one switch port can make all the required measurements. These are sometimes known as *switching test sets* or *simple switch trees*.

The second class of devices requires a measurement from each port to every other port, and in general the response of any path depends upon the loading or match applied to every other port. A “Butler matrix” is a kind of signal dividing network used in phased-array radar systems, which has this attribute. An eight-port Butler has four inputs and four outputs, and the proper description is an 8×8 S-parameter matrix. To measure such a device, a switch matrix must be able to allow measuring every path of the device. Informally, these types of switch matrixes are called *full cross-bar switches*, which implies that from the two ports of the VNA, any path of the DUT can be measured.

There is a further requirement on the Butler matrix; a full N-by-N port calibration measurement must be able to be performed to correct for the imperfect match of at each port. This requires not only a full cross-bar matrix but one that supports N-by-N calibration as well. A third style of test set allows such N-by-N S-parameters called an *extension test set*, which extends or adds to the number of test ports from a VNA, but these have largely been replaced with true-multiport VNAs.

More recently, several vendors have developed high-port-count VNAs with up to 24 internal ports (R&S ZNBT) or configurable module multiport VNAs (Keysight M9875), in which PXI-based VNA modules can be flexibly configured to large numbers of ports (more than 66 ports).

The various forms of multiport configurations are described next.

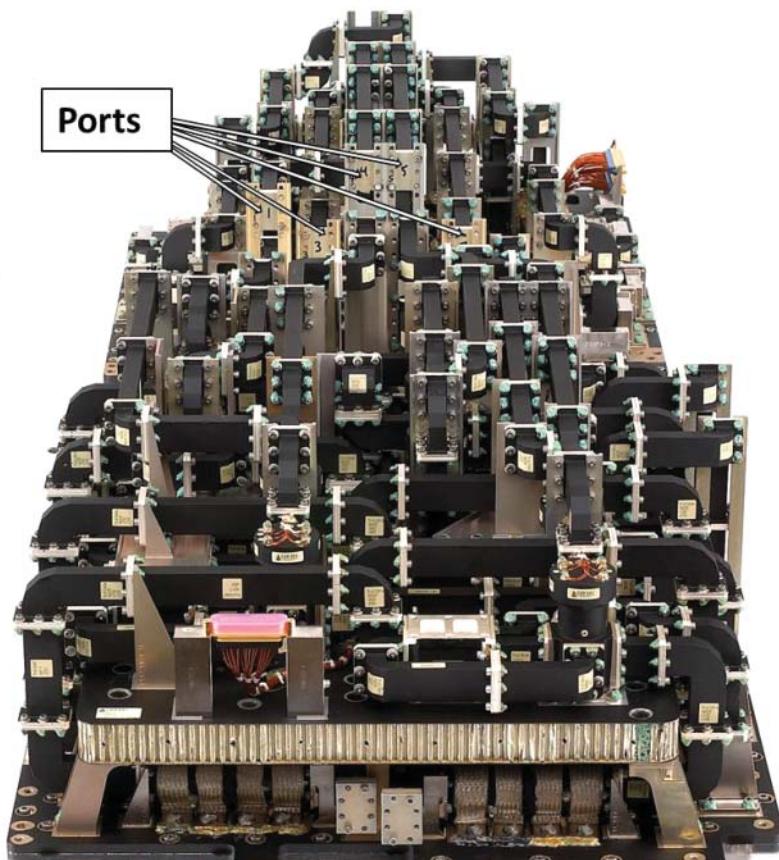


Figure 2.24 A satellite multiplexer with many outputs. Source: Courtesy ComDev Ltd., with permission.

2.2.7.1 Switching Test Sets

Switching test sets contain only RF switches formed in a matrix to provide the needed measurement paths. Figure 2.25 shows the block diagram of a simple switch tree test set. These test sets are typically constructed from either 1×2 RF switches or 1×4 to 1×6 RF switches. The 1×2 RF switches are sometimes used as some versions provide for an RF load on the unused ports. The 1×4 or 1×6 are typically mechanical switches and may not load the unused ports. If a multiport device has a path response between two ports that depends on the load match of a third port, the switch matrix must provide a load on the unused port. Larger switch configurations that have loads are often not available above 40 GHz, so 1×2 matrix arrays are used. Electronic 1×2 switches are available over a wide range of frequencies, but there are few electronic switches with higher port counts, so electronically switched test sets are typically configured from 1×2 RF switches.

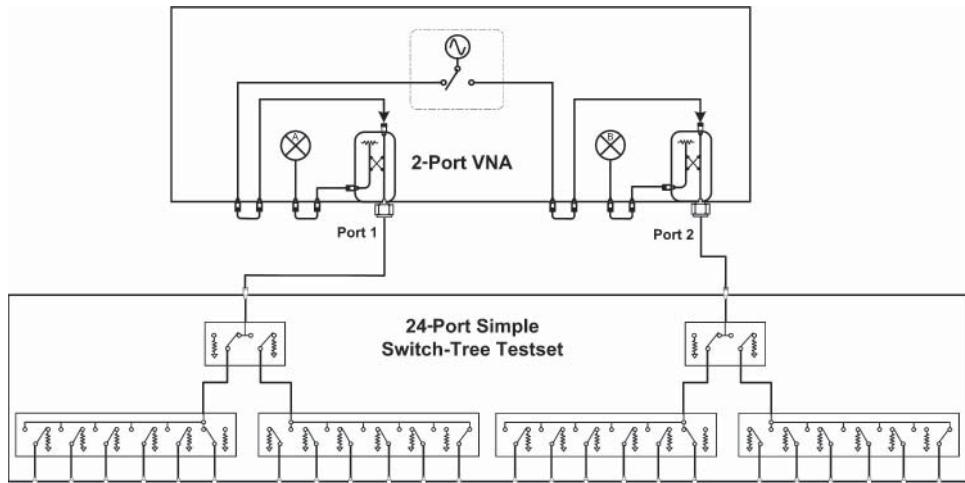


Figure 2.25 Simple switch tree test set.

The simple switch matrixes of Figure 2.25 can be viewed as having a port 1 switch set and port 2 switch set, and any path from the port 1 side to the port 2 side can be measured, but no measurements are available between ports on the port 1 of the switch set, nor between ports on the port 2 side. While there are 24 ports available in the test set, only 12 paths can be measured from any one of the 12 input ports. Thus, this simple switch tree test set can support 144 paths, but a full 24 port device actually has 276 paths. There are 66 paths on the VNA port 1 side that cannot be measured, and there are 66 paths on the VNA port 2 side that cannot be measured. To obtain a full matrix of paths, a so-called full cross-bar switch matrix is required.

To accomplish full cross-bar testing, the configuration of the test set shown in Figure 2.26 is used. In the general configuration, sets of $1 \times n$ switch trees are cross connected to 1×2 switches at each port. This configuration provides for any path to be measured, but the unused ports are terminated back in the $1 \times n$ switches, which are terminated internally in a load. If the $1 \times N$ switches are not internally terminated (rather, they are left open), then the 1×2 switch must provide a termination for an unused port. Figure 2.26 shows a full cross-bar switch constructed of a 1×2 port switch connecting to a pair of $1 \times n$ switches. With this configuration, every port that is not connected to the VNA is terminated in a switch load. However, it is difficult to use this type of switch matrix to perform full N-by-N calibrations as the exact value of the load termination of any port changes depending upon the switch settings of other ports.

For example, if test set ports 1 and 6 are the active ports, ports 2–5 are terminated in the 1×6 switch on the left. If test set port 5 is made active, then port 6 may be terminated in the 1×6 switch on the right. The fact that the termination of the port depends on the path selected makes calibration beyond the two ports selected more difficult.

Custom switching test sets might have a reduced number of paths, forming a combination of full cross-bar on some ports and simple switch trees on other ports. For high speed and

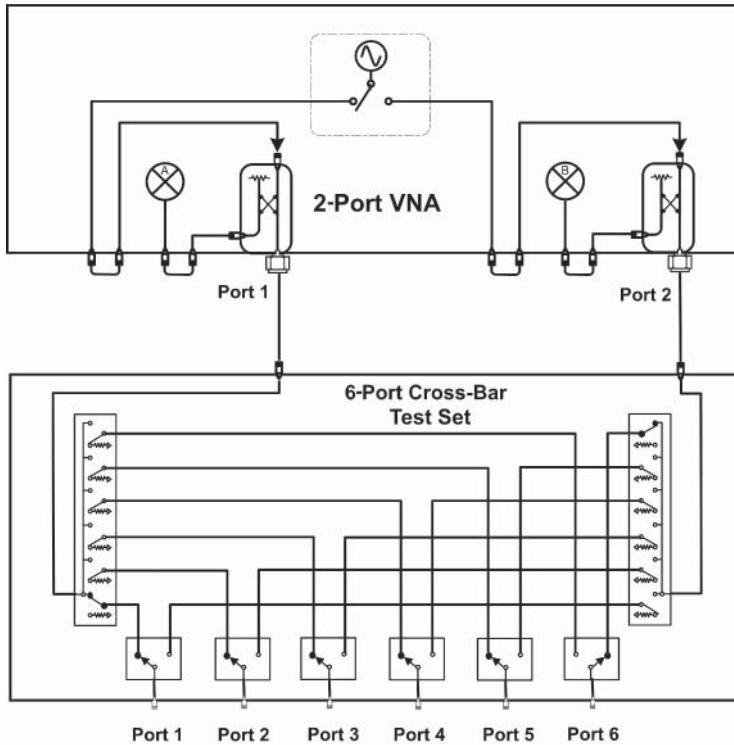


Figure 2.26 Full cross-bar switching test set.

reliability, solid-state switching is preferred. Mechanical switches have almost no loss, but solid-state switches can have considerable loss at microwave frequencies. This loss is after the directional-coupler and dramatically degrades the RF performance of the system. On the other hand, mechanical switches can have slight changes in return loss for each switch cycle, also leading to instabilities. Thus, this architecture of switches after the directional-couplers of the VNA is a simple one, but at a cost of substantially reduced stability and performance.

2.2.7.2 Extension Test Sets

To satisfy the requirement for making full N -by- N calibrated measurements, often referred to as *full N-port cal measurements*, a test set design has been developed that includes both directional-couplers and switches. The original implementation of this style of extension test sets was configured to supply two additional ports to a two-port VNA to create a 4-port VNA for making the first balanced and differential measurements. The general idea of an extension test set is to essentially extend the source switch matrix of the VNA to more outputs through a source switch and also extend the internal receivers to more ports through a receiver switch. This requires that an additional test port coupler be provided for each additional port. Because the switching occurs behind the VNA directional-couplers, they are still available as test ports:

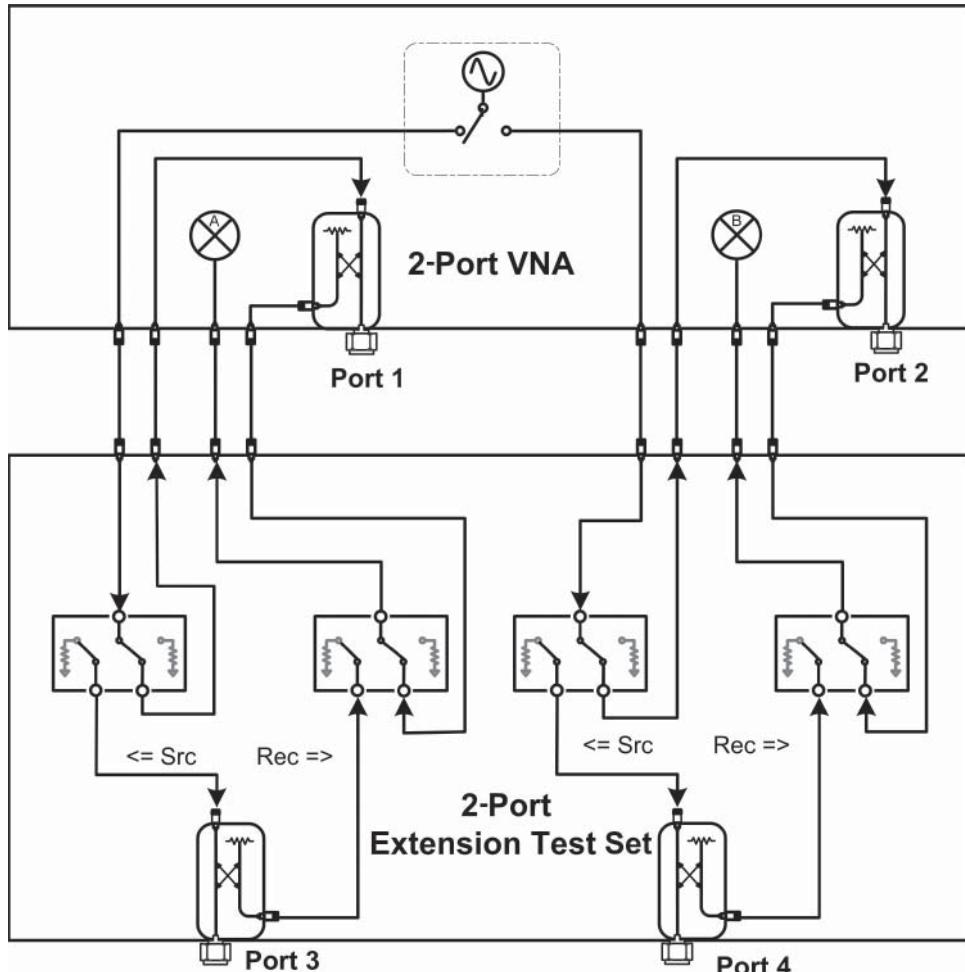


Figure 2.27 Extension test set block diagram.

the ports on the test set extend the total number of ports available, which is why it's called an *extension* test set. Figure 2.27 shows block diagrams for a simple two-port extension test set.

One key point of the block diagram is that the test set breaks into the source and receiver loops behind the test port coupler. Since any number of switch paths can be supplied behind the test couplers, there is in theory no limit to the number of ports that can be used. Further, this block diagram allows additional test sets to be added so that any number of test ports can be created by stacking extension test sets. Common configurations are 4-port extension test sets for a 4-port VNA to extend to a total of 8 ports, 10-port extension test sets for a 2-port VNA to achieve a total 12 ports, and 12-port extension test sets for a 4-port VNA to achieve a total of 16 ports. Figure 2.28 shows a 4-port VNA with two 4-port extension test sets to create a 12-port system.



Figure 2.28 12-port system using a 4-port VNA and two extension test sets. Source: Photo courtesy of Keysight Technologies.

The switches may be either mechanical switches or solid-state switches. Because all the switching occurs behind the test port couplers, the stability and performance of the measurements are much better than that of switching test sets, and loss in the switch, while it reduces the dynamic range, has no effect on stability of the measurements.

In some cases, an option may be provided to add a low-noise amplifier (LNA) between the coupled port of the test coupler and the switch input. This improves performance as the gain of the LNA improves the dynamic range. Adding amplifiers in between the coupled arm and the switch also removes another source of error. In some cases, the source-match of a port changes when the source and test port share the same VNA receiver, for example ports 1 and 3 in Figure 2.27. This error is typically small as the difference between the match of the VNA receiver and the match of the switch is small (on the order of -10 dB) and is further reduced by twice the coupling loss (32 dB) resulting in a typical source-match error smaller than -40 dB . In most cases it has a negligible effect, but in some measurements, particularly

circulators or couplers, it can become significant and is not removed in calibration, so adding an amplifier ensures that the match presented to the coupled arm is constant. The test ports also change load characteristics depending upon if they are terminated in a switch or the VNA internal load; however, the N-port calibration methods characterize both these states and fully correct for the difference.

2.2.7.3 True-Multiport VNAs

While the extension test set provides a directional-coupler on each port of the test system, the reference coupler and the measurement receivers are shared, so the number of ports that can be measured simultaneously is limited to the number of receivers in the base instrument. Recently, improved integration has made it possible to include a full VNA test receiver on each port, so true-multiport VNAs are now available. These come in a variety of form factors, but for the most part they are intended for manufacturing operations, where size and footprint are important.

One of the first offerings for a large-port-count true-multiport VNA was the ZNBT from Rohde & Schwarz. It provides options from 8 to 24 ports, with a faceless instrument. In this configuration, it had six independent sources (one for each four ports) as well as receivers on each port.

A modular form of multiport VNAs has been introduced in a PXI format, which allows for configuring from 2 to more than 68 ports, potentially up to 100 ports, depending upon the number and model of VNA modules used. Figure 2.29 shows a modular system with eight 6-port modules (Keysight model M9804-006) and one 2-port module (Keysight model M9804-002), configured as a 50-port VNA system. There is one source per module, but a full dual-reflectometer and dual RF receiver for each port. Thus, the 2-port modules have 1 source and 4 receivers; the 4-port modules have 1 source and 8 receivers, and the 6-port modules have 1 source and 12 receivers.

Multiport VNAs in a modular format require the local oscillator to be shared across all modules to get the best trace noise performance. These systems provide a daisy chain approach

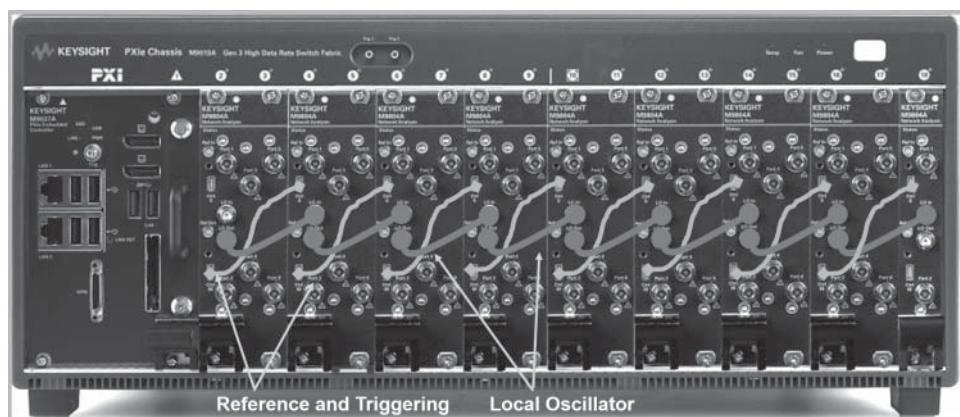


Figure 2.29 A 50-port VNA system comprised of 6-port and 2-port modules.

Table 2.1 Sweeps needed for N-port calibration

Total Ports	Total Paths	Switched 2-Port	Switched 4-Port	True Multiport
8	28	56 sweeps	24 sweeps	8 sweeps
16	120	240 sweeps	64 sweeps	16 sweeps
24	276	552 sweeps	144 sweeps	24 sweeps

to the connect the LO and the 10 MHz reference to each of the modules. The big advantage of a modular approach is the test system is easily reconfigured to support different test needs. For example, a 16-port system, comprised of eight 2-port modules, can be reconfigured into four sets of 4-port VNAs.

While more expensive than a switched version of a VNA, the economics of a true-multiport system readily become apparent when one considers the overall measurement time and number of sweeps needed to complete an N-port calibrated measurement. Table 2.1 shows the number of sweeps needed to complete an N-port calibrated measurement. From a strict sweep time point of view, a true multiport VNA greatly reduces the overall test time requirement.

2.2.7.4 Calibration of Multiport VNAs

Calibration is often a concern with multiport test systems. Traditional S-parameter calibrations require measurements between every path of a test system. However, new techniques have greatly reduced the total number of calibration steps to the point where a full N-by-N port S-parameter calibration can be achieved with a single one-port return loss calibration and N-1 thru measurements using the quick-short-open-load-thru (QSOLT) calibration. More details of these new calibration methods will be discussed in Chapter 3. Some manufacturers provide high-port-count electronic calibration modules that can simplify the calibration process.

2.2.8 High-Power Test Systems

Most VNAs have a maximum test port operating level on the order of 10–15 dBm, with a damage level on the order of +30 dBm. Beyond the operating level, the receiver will be in substantial compression, so the data is not valid. Many VNAs provide internal receiver attenuators that allow reducing the power to the receiver, providing operation to much higher levels. The maximum input power to test port couplers are often rated higher than the maximum level of other components behind the directional-coupler so that with proper padding and isolation, the VNA can be operated to levels as high as +43 dBm, depending on the model. Operation above these levels is possible but requires substantial external components including external couplers to ensure the power level at the VNA components is below the power damage level. Details of high power test configuration are shown in Chapter 6.

Another common practice is to add fixed attenuators between the DUT output and the VNA test port. This works well so long as the total attenuation between the test port and the DUT is less than 10 dB. Adding attenuation after the test port coupler degrades the directivity by

two times the attenuation (in dB), as will be shown at the end of Section 2.3.2. In practice, up to 10 dB of external attenuation can be added and compensated for with normal calibration techniques. If between 10 and 20 dB are added, the system becomes somewhat unstable, and for more than 20 dB of added attenuation, different techniques for calibration must be used, and the S22 measurements become unreliable.

For testing devices that require high-power drives, it is common to add an amplifier to increase the power level normally available from the VNA. One method is to simply add a booster amplifier to the port 1 output and drive the DUT directly from the booster amplifier. This generally results in poor measurements of DUT due to mismatch and gain errors in the booster amplifier. In this approach, it is common to add a booster amplifier, normalize the S21 trace, and then add the DUT and measure the resulting gain relative to the normalized booster amplifier response. However, the normalization has errors due to mismatch between the booster amplifier and the load port. And the measurement has errors due to mismatch between the booster amplifier and the high-power DUT. Further, the input match or S11 of the DUT cannot be reliably measured because booster amplifier isolation eliminates the ability to measure a signal reflected from the DUT. A second error in gain measurements often occurs with this direct approach due to gain drift or gain compression of the booster amplifier.

A better systematic approach is to add the booster amplifier behind the test port coupler, and use a second coupler as the reference-channel tap to generate a signal proportional to the booster amplifier output signal that can be routed to the reference channel. In this scenario, a directional-coupler rather than a power splitter is typically used to provide lower loss after the booster amplifier. The output of the reference channel is directed through the test port coupler of port 1 so that the S11 of the DUT can be accurately measured. In almost all cases, an accurate measurement of high-drive-power devices requires a booster amplifier followed by a reference coupler. Chapter 6 provides a detailed discussion of high-power amplifier measurements, including several alternative block diagram configurations to support various power levels.

2.2.9 VNA with mm-Wave Extenders

Frequencies above 30 GHz are technically millimeter wave (as in they literally have a wavelength less than a cm), and it is common to consider mm-wave above 50 GHz to over 120 GHz. At these high frequencies the losses in the test leads become so great that it often makes testing quite difficult because of the power loss from the VNA port to the DUT reference plan. This is particularly true in on-wafer applications where the large size of the VNA means it must be placed some distance from the probes, and that means long test cables.

Thus, at these high frequencies many equipment manufacturers make mm-wave extenders or remote heads. These heads typically have built-in frequency multipliers for the source and LO, built-in mixers, and built-in directional couplers. Figure 2.30 shows a representative block diagram of a mm-wave extender. The actual implementation can vary a lot, but modern mm-wave heads all have this similar style.

The mm-head shown represents a *broadband head* and will have a coaxial test port (typically 1 mm or smaller) that can pass DC. The low-band path takes the RF input signal directly and passes it through to the test port. There are a couple of basic methods to provide the

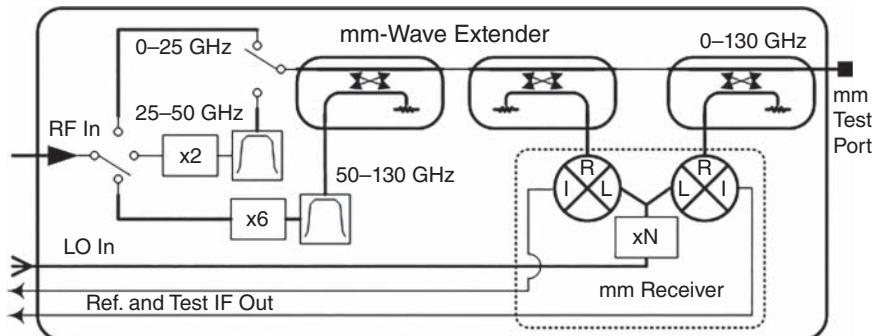


Figure 2.30 mm-Wave Head block diagram with broadband capability.

low-band measurement capability: some manufacturers use the directional-couplers in the base box (port 1, for example) to drive the RF input and will measure S-parameters from low frequency to the first multiplier band using the normal VNA receivers in the base box. This reduces the need for broadband couplers in the head but may add instability due to the long test port cables from the base box to the mm-input. One manufacturer uses broadband receivers in the mm-receiver portion of the head, as well as broadband coupler. The RF signal is only a source, and even at low frequency the mm-head couplers are used. This provides a quite stable result, but it is a difficult design to achieve such a broadband coupler. The LO multiplier either can be a discrete multiplier or can rely on higher-order harmonic conversion in the mixer. The reference and test IF signals are returned to the base box where they are processed to produce the desired detection.

Older mm-wave broadband systems (not shown) placed the combiner in front of the mm-port. This allowed the use of a full waveguide coupler in the head (which of course cannot pass low frequencies), but the loss of the combiner at the output leads to system instabilities, and this style of broadband head has largely been obsoleted.

One version of mm-heads, so-called banded-heads, typically use waveguide-based couplers and cover only the waveguide band. The test port is a waveguide as well. These heads have only a single source and LO multiplier, and of course they have no low frequency path at all. Banded heads operate over nearly all the waveguide bands from V band (50–75) up to THz frequencies of 1.5 THz.

The reference and test IF paths, just like in the broadband head, are returned to the base box for signal conditioning and detection in the digital IF.

Figure 2.31 shows a picture of a broadband, 4-port mm-wave system. This particular unit allows measurements from 900 Hz to 130 GHz, with full S-parameter calibration as well as other features of modern VNA such as broadband spectrum analysis (see Chapter 8) and noise figure capability. The long test umbilical contains a composite cable that has all the RF and IF cables plus DC bias, DC power, and switch control. The switch control is needed to switch between filter paths in the head to maintain a good level of spurious performance. The length is needed as these systems are often used in on-wafer applications. Fortunately, because the test port couplers are inside the head, there is good stability even with the long interface cables.

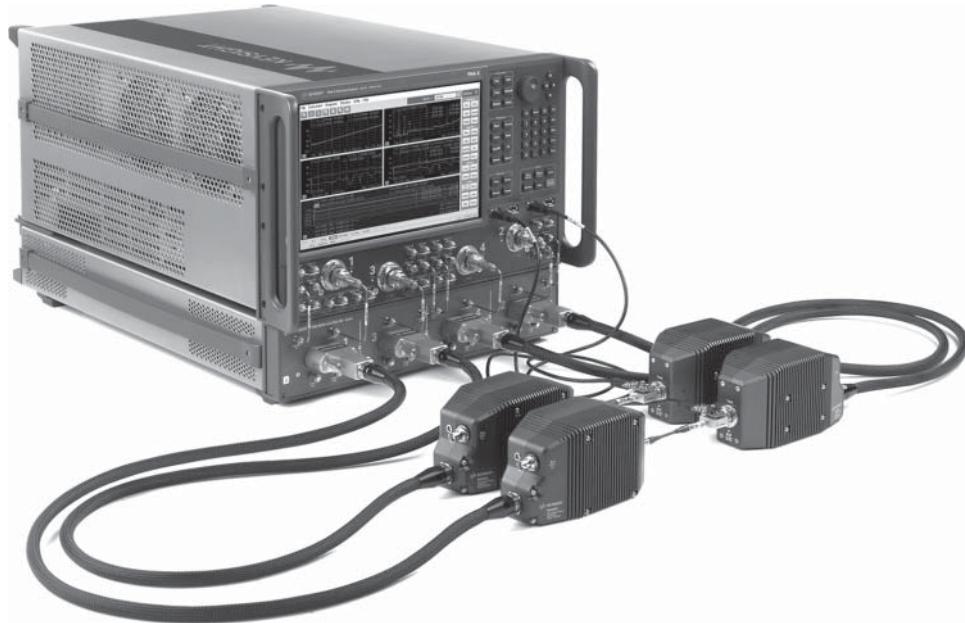


Figure 2.31 A 4-port, 900 Hz to 130 GHz VNA system using mm-wave extenders. Source: Photo courtesy of Keysight Technologies.

2.3 VNA Measurement of Linear Microwave Parameters

In this section the fundamentals of making microwave measurements are discussed for a variety of parameters, along with the consequences of the practical limitations of the RF hardware detailed in the previous section. This section discusses measurement methods and the sources of errors and other complications to making microwave measurements related to test equipment limitations.

2.3.1 Measurement Limitations of the VNA

The systematic error terms (described in Chapter 3) are well known, and many methods are used to essentially eliminate their effects on the S-parameter measurement results. However, other hardware limitations in a VNA cannot be so easily removed, and special care must be taken to diminish the effects of these limitations.

2.3.1.1 Noise Floor

The consideration of system noise floor is often not included in the specifications of measurement accuracy, under the assumption that noise effects can be eliminated with sufficient reduction in IF bandwidth or increase in averaging factor. While this is theoretically true,

in many circumstances the increase in measurement time makes this impractical. In some situations, such as real-time tuning of microwave filters, the IF bandwidth must be increased to achieve a real-time update rate. The noise effects increase 10 dB for each 10 \times increase in IF BW, so it is a simple matter to compute the effective noise floor at any IF BW given a noise floor at some IF BW, typically 10 Hz.

There exist two distinct noise effects in S-parameter measurements: noise floor and high-level trace noise. The noise floor is easily understood as the effect of added noise at the input of the receiver, due to the noise figure of the VNA receiver. The coupling factor of the test port coupler reduces the measured signal further so that the effect of noise floor is more dominant. The effect of noise floor on a measurement can be determined by taking the RMS noise floor, converting it to an equivalent linear amplitude wave, and then adding it to the amplitude of the signal at the measured receiver.

The conversion to the linear b_2 noise is

$$b_{2_Noise} = 10^{\frac{NoiseFloor_{dBm}}{20}} \quad (2.9)$$

Note that the raw measured noise floor on a VNA receiver will be the square root of the noise power, as the a and b waves are in units of square root of power.

Often, the noise floor of a VNA is expressed as a dBc value relative to a 0 dB insertion loss measurement. Of course, for a constant noise power in the receiver, the relative noise floor will depend upon the source drive power.

The RMS trace noise apparent on an S-parameter trace can be computed by adding the RMS noise floor to the amplitude of the signal at the b receiver.

$$TraceNoise_{dB} = 20 \log_{10} \left[\frac{b_{2_noise} + b_{2_signal}}{b_{2_signal}} \right] \quad (2.10)$$

when the noise floor is sufficiently below the measurement of interest. Of course, when the noise floor is above the measured value, the measurement becomes meaningless.

Take, for example, a filter with 80 dB of insertion loss ($S21 = -80$ dB), with a drive power from the source of 0 dBm, a VNA with an RMS noise floor of -127 dBm in a 10 Hz bandwidth. If it is measured using a 10 kHz IF bandwidth, as shown in Figure 2.32, the trace noise due to noise floor at any insertion loss can be computed.

The effective noise floor is 30 dB greater than the 10 Hz spec, for a level of -97 dBm. The measured b_2 noise would be

$$10^{\frac{-97}{20}} = 1.41 \cdot 10^{-5} \quad (2.11)$$

The output signal is

$$10^{\frac{[-80]}{20}} = 1 \cdot 10^{-4} \quad (2.12)$$

The RMS trace noise level would then be

$$TraceNoise = 20 \log \left[\frac{(1.41 \cdot 10^{-5} + 1 \cdot 10^{-4})}{1 \cdot 10^{-4}} \right] = 1.15 dB_{RMS} \quad (2.13)$$

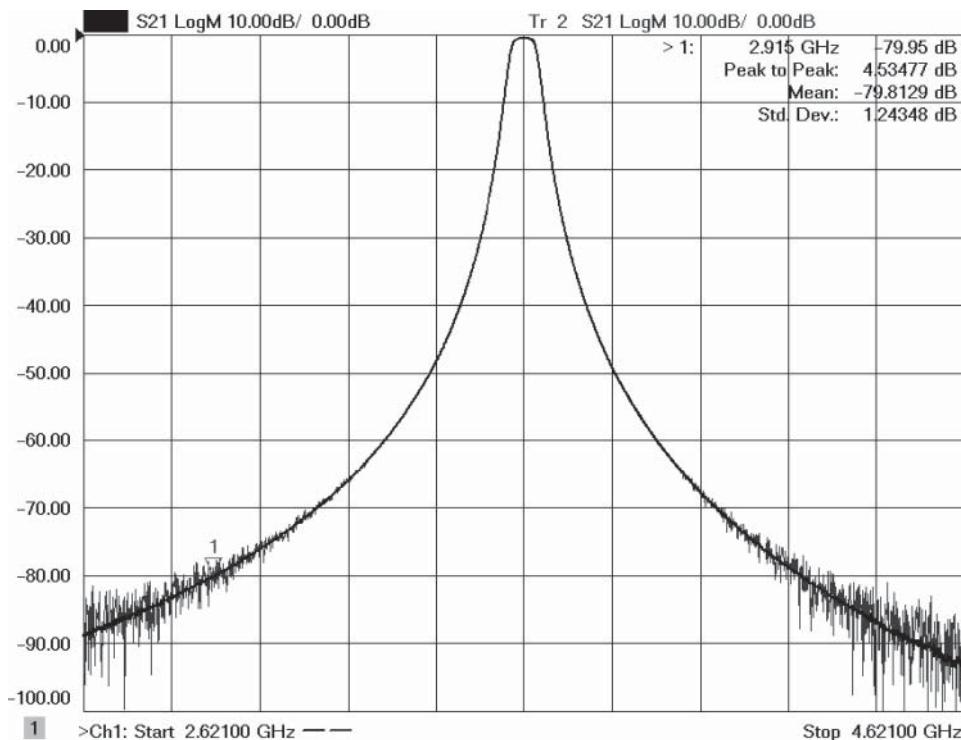


Figure 2.32 Effects of noise floor on an S21 measurement.

This value is close to the measured trace noise, shown as trace statistics computed near Marker 1 on Figure 2.32 and displayed as $SDEV = 1.24 \text{ dB}$ (trace statistics measures the variation of signal of a trace, and in this case the computation is restricted to be a 5% region near the marker position). Thus, one would see substantial noise on the filter stopband measurement. The RMS trace noise represents one standard deviation of noise. For this example, about 21 points are used to compute the trace noise near the marker. One would expect a peak-to-peak trace noise of about four standard deviations in the worst case or approximately 4.6 dB of peak-to-peak noise on a typical measurement. However, since noise can take on any value for any single instance, the RMS value is almost always used when describing noise-related values. As the S-parameter signal rises above the noise floor of the VNA, trace noise diminishes at a rate of about three times (in dB) for each 10 dB increase in signal level. But this 3-for-10 reduction doesn't continue at high signal levels.

A second cause of trace noise in a measurement is called *high-level* trace noise. At high signal levels, the noise from the source signal, typically due to the phase noise of the source, can rise above the VNA noise floor and dominate the trace noise in the measurement. Further, if the source in the VNA has substantial internal amplification, the broadband noise floor from the source can dominate the phase noise far from the carrier. In this region, the trace noise stays approximately the same as the S-parameter signal increases. Consider the trace

noise on the skirt of a filter: when the signal through the filter is sufficiently high, the trace noise on the measurement decreases as the signal level rises above the noise floor, until the source phase noise or pedestal noise, as it is sometimes called, becomes dominant. Above this level the trace noise stays constant as a dBc level even as the S-parameter loss diminishes. The problem of high-level trace noise is more commonly found on older VNAs where phase noise was typically worse than that of stand-alone signal sources due to the difficulty of integrating sources internally. The problem is also seen on more modern VNA systems at mm-wave frequency, where multipliers are used to increase the source frequency. With each 2x multiplication of frequency, the phase noise increases by 6 dB. These problems are typically seen only at high power levels because the use of attenuators for power-level control reduces the source signal and the phase noise in the same manner.

Figure 2.33 shows the phase noise of a VNA source as it is increased to where the phase noise is higher than the noise floor. The memory trace, shown in light gray, is for a power level of -10 dBm . At this level phase noise is below the receiver noise floor. The dark trace shows the phase noise when the source power is increased to $+10 \text{ dBm}$. Here the phase noise is about 15 dB above the noise floor and will limit the high-level trace noise. This data was measured with a 10 kHz resolution bandwidth (RBW), so the actual maximum phase noise is about -110 dBc Hz^{-1} at offsets below about 50 kHz.

Figure 2.34 shows a plot of trace noise as a function of received power. In this normalized response, the trace noise limit is apparent in the high-level region starting at about -10 dBm , where the trace noise no longer decreases directly as a function of increased signal level, indicated on the figure as the “Hi-Level Trace Noise” region.

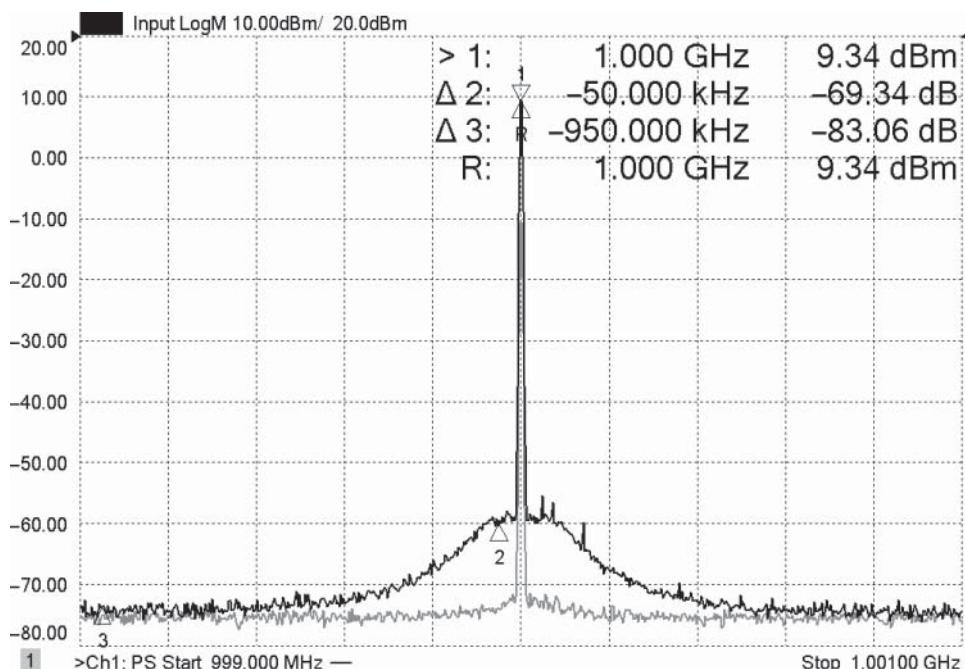


Figure 2.33 VNA source signal where phase noise rises above noise floor.

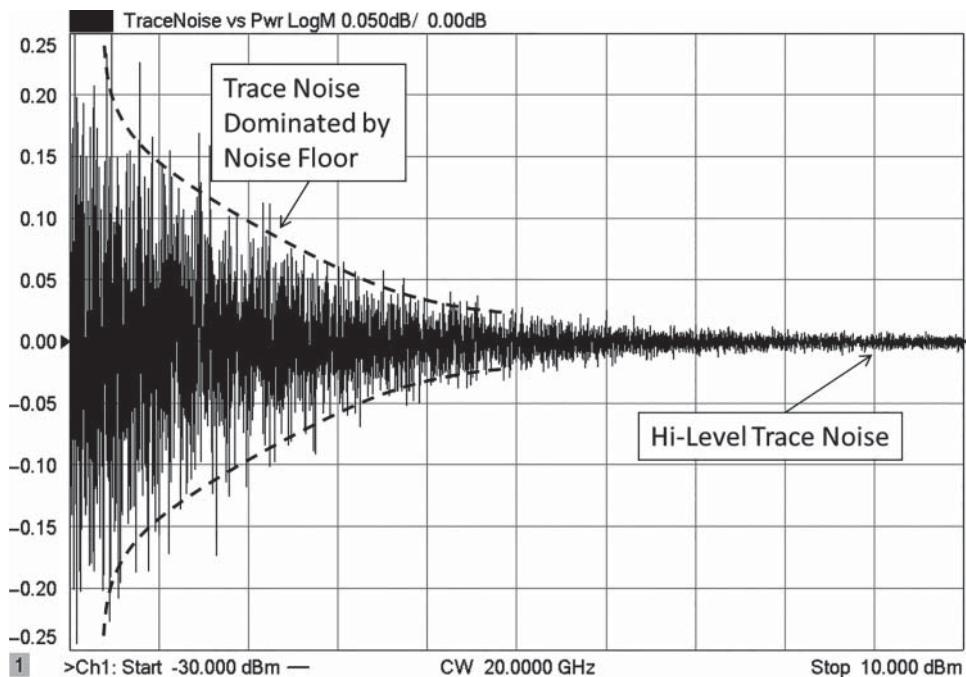


Figure 2.34 Example of trace noise decreasing with increased signal level, until high level noise limit is reached.

2.3.2 Limitations Due to External Components

Often, the performance of external components used to connect from the VNA to the DUT presents the largest contribution of errors to the measurement. These errors can come in a variety of configurations, and each has its own peculiarities that can affect measurements in different ways. The most common causes of external errors are cables and connectors.

Cables, connectors, and adapters are ubiquitous when using VNAs to measure most devices. The quality and particularly the stability of the cable and connector can dramatically affect the quality of the measurement.

The first-order effect of cables is added loss and mismatch in a measurement. For short cables, the loss is not significant, but the mismatch can add directly to the source-match and directivity of the VNA to degrade performance. With error correction, the effect of mismatch can be substantially reduced (to the level of the calibration standards quality) if it is stable, but cable instability limits the repeatability of the cable mismatch and often is the dominant error in a return-loss measurement.

For transmission measurements, the effects of mismatch do directly affect calibration, though it is reduced to a small level by the quality of the calibration standards. Often, the major instability in a cable is the phase response versus frequency. Even if the amplitude of the cable is stable, if the phase response changes, the VNA error correction will become corrupted because of phase shift of the cable mismatch error. Methods for determining the quality of cable and the effects of flexure will be described in Chapter 9.

2.4 Measurements Derived from S-Parameters

S-parameter measurements provide substantial information about the qualities of a DUT. In many cases, the transformation and formatting of these parameters is necessary to more readily understand the intrinsic attributes of the DUT. Some of these transformations are graphical in nature, such as plotting on a Smith chart; some are formatting such a group delay and SWR; and some are functional transformation such as time-domain transforms. Some of the more important transformations are discussed next, with emphasis on some particularly interesting results.

2.4.1 The Smith Chart

The Smith chart is a visualization tool that every RF engineer should strive to master. It provides a compact form for describing the match characteristics of a DUT, as well as being a useful tool for moving the match point of a device to a more desired value. Invented by Philip Smith (1944), it maps the normalized complex value of a termination impedance onto a circular-based chart, from which the impedance effects of adding lengths of transmission line onto the termination impedance are easily computed. The original intention for the use of a Smith chart was for the computing of impedances presented to a generator as lengths of transmission line were added to a load and was intended particularly for the use of telephone line impedance matching. Adding a length of transmission lines changes the apparent termination impedance, Z_T , according to

$$Z_S = Z_0 \frac{(Z_0 + Z_T \tanh[(\alpha + j\beta)z])}{(Z_T + Z_0 \tanh[(\alpha + j\beta)z])} \quad (2.14)$$

where α and β are the real and imaginary propagation constants, and z is the distance from the load. This computation was tedious, in part because the argument of the hyperbolic tangent is complex, so a nomographic approach was desirable. A Smith chart solves by this mapping impedance to reflection coefficient (Γ), and plotting the return loss on a polar plot, as

$$\Gamma = \frac{(Z - Z_0)}{(Z + Z_0)} \quad (2.15)$$

The genius of the Smith chart is recognizing that rotating an impedance value through a length of transmission line is the same as rotating the phase of the reflection coefficient value on the chart. The Smith chart maps the impedance onto the polar reflection coefficient plot, but with the graticule lines marked with circles of constant resistance and circles of constant reactance. As such, any return loss value can be plotted, and the equivalent resistance and reactance can be determined immediately. To see the effect of adding some Z_0 transmission line, the impedance is simply rotated on the polar plot by the phase shift of the transmission line. If the line is lossy, the return loss is modified by the line loss (two times the one-way loss of the line), and from this new position, the resistance and the reactance are directly read.

2.4.1.1 Series and Shunt Elements

The original intent of the Smith chart was to show S_{11} at a fixed frequency and use the chart to derive the change in impedance due to a change in distance from the generator. But the use

of the Smith chart in VNAs differs from the original intent in that the display shows return loss or S11 as a function of frequency, and the phase rotation displayed is due to a phase shift in a transmission line or device caused by the increase in frequency. Various characteristics, such as capacitance, inductance, loss, and delay, can be directly inferred from the Smith chart trajectory displayed on a VNA, and it is often more informative than just the LogMag plot or the Phase plot individually. In many instances, the Smith chart is useful for determining the principal component characteristics of the DUT. Since by most designs, the DUT should ideally be matched, the deviation from matched conditions is due to some parasitic series or shunt element. Series elements show up in a Smith chart trajectory as following a contour of constant resistances. Shunt elements are not intuitively deduced from a Smith chart but can be deduced from an admittance chart (also called an *inverse* Smith chart), which follows the same conformal mapping of a Smith chart (impedance chart) but with the inverse of impedance (admittance) displayed as lines of constant conductance or susceptance.

For high-frequency measurements, shunt capacitance or series inductance is almost always the parasitic values that must be dealt with. Note that the parasitic effect of a series capacitance or a shunt inductance actually diminishes with frequency, with the capacitor becoming a short, and the inductor an open, and these elements typically cause only low-frequency degradation.

Figure 2.35 shows both an impedance plot (left) and admittance plot (right), each with two circuit elements: a $40\ \Omega$ load with a shunt capacitance (dark trace, left), and a $60\ \Omega$ load with a series inductance. (right).

Note that on the impedance chart, the highlighted marker values for the series inductive circuit display constant resistance and inductance, even as the frequency varies. The impedance plot of the shunt RC circuit (dark trace) does not show either constant resistance or constant capacitance. However, the same trace on the admittance chart (right) does show constant

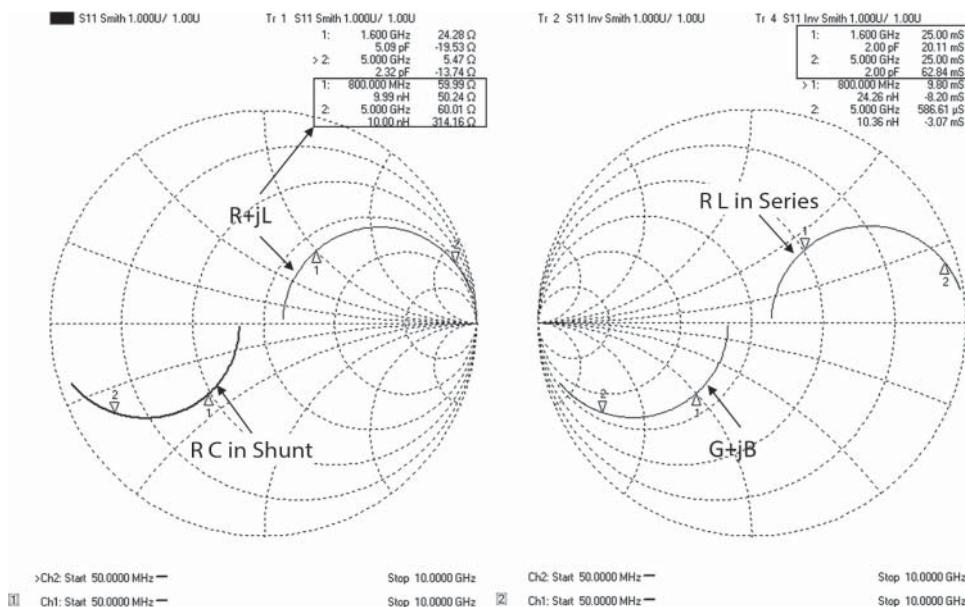


Figure 2.35 Impedance and admittance Smith charts.

conductance and constant capacitance, where the series inductive LR circuit does not show constant value. From these charts, it is clear that the impedance chart allows determination of series elements easily as their trajectory follows constant resistance circles, and the admittance chart allows determination of shunt elements as their trajectory follows constant conductance circles.

In evaluating practical responses for input impedance, it is often the case the series or shunt element is at the end of some short, or long, transmission line that wraps the response around the Smith chart. In this case, for useful information to be discerned, it is necessary to remove excess delay from the measurement. However, it is sometimes difficult to know the exact delay that must be removed. In such a case, it is possible to use two marker readouts of the VNA to attempt to determine the value of a parasitic element by means of unwrapping the phase of the response until the underlying value is determined. Figure 2.36 shows a set of responses where the DUT characteristic is delayed by a short length of transmission line, as might be found in a printed circuit board (PCB) fixture or on-wafer probe, followed by a series resistance and inductance (left, upper) or a shunt capacitance and conductance (right, upper). These are the two most common cases of parasitic characteristics.

The trajectories are distorted by a frequency dependent phase shift due to the delay of a portion of transmission line between the VNA reference port and the parasitic. For the inductive case, the elements are series elements, and the normal Smith chart, or impedance chart, is shown. The marker readout gives the resistance, in Ω ; the reactance, in Ω ; and the equivalent inductance. It is clear that the apparent value of the reactive element is not constant. Most VNAs include the equivalent inductance or capacitance value associated with the reactance for the marker position and frequency.

For the capacitive example, since it is a shunt impedance, an inverse Smith chart or admittance chart is used. The value displayed for the real part of the admittance is the conductance, in milli-Siemans (or mS, inverse Ω), and for the imaginary part is the susceptance, also in milli-Siemans. The reactive part is converted to an equivalent shunt capacitance or inductance, determined by the sign of the imaginary part of the admittance. Again, it is clear that the apparent value of the shunt reactive element is not constant. In fact, both trajectories show an attribute of having a resonance, since they cross the real axis. However the fact that the magnitude of reflection is not a minimum at the crossing indicates that this is not a true resonant structure, but rather a device whose phase response is distorted by a length or delay of a transmission line between the measurement plane and the discrete impedance or admittance.

It is reasonably simple to investigate the effect of removing the delay, by using two markers, spaced in frequency. By reading the value of the imaginary element of each marker while adding in electrical delay, the phase shift of the delay line can be removed, and the resulting underlying element characteristics are revealed. When both marker readings show the same value for the reactive element, then the proper delay has been removed, as shown in the middle portion of Figure 2.36. In this case, the left plots give a capacitance of 1 pF in shunt with 100 Ω , and the right plots show an inductance of 3 nH in series with a resistance of 25 Ω .

The lower traces show the same measurement, but with even more electrical delay removed from the response. Electrical delay is a common scaling function in VNAs that provides a linear phase shift versus frequency for any particular trace. A related function is port extension, which also provides a phase shift but that shift is associated with the port of the analyzer, rather than with just the particular trace. With electrical delay scaling, only the trace that is active has the delay applied, and different traces of the same parameter can have different delays. With port extension, all traces that are associated with a particular port, for example, S11

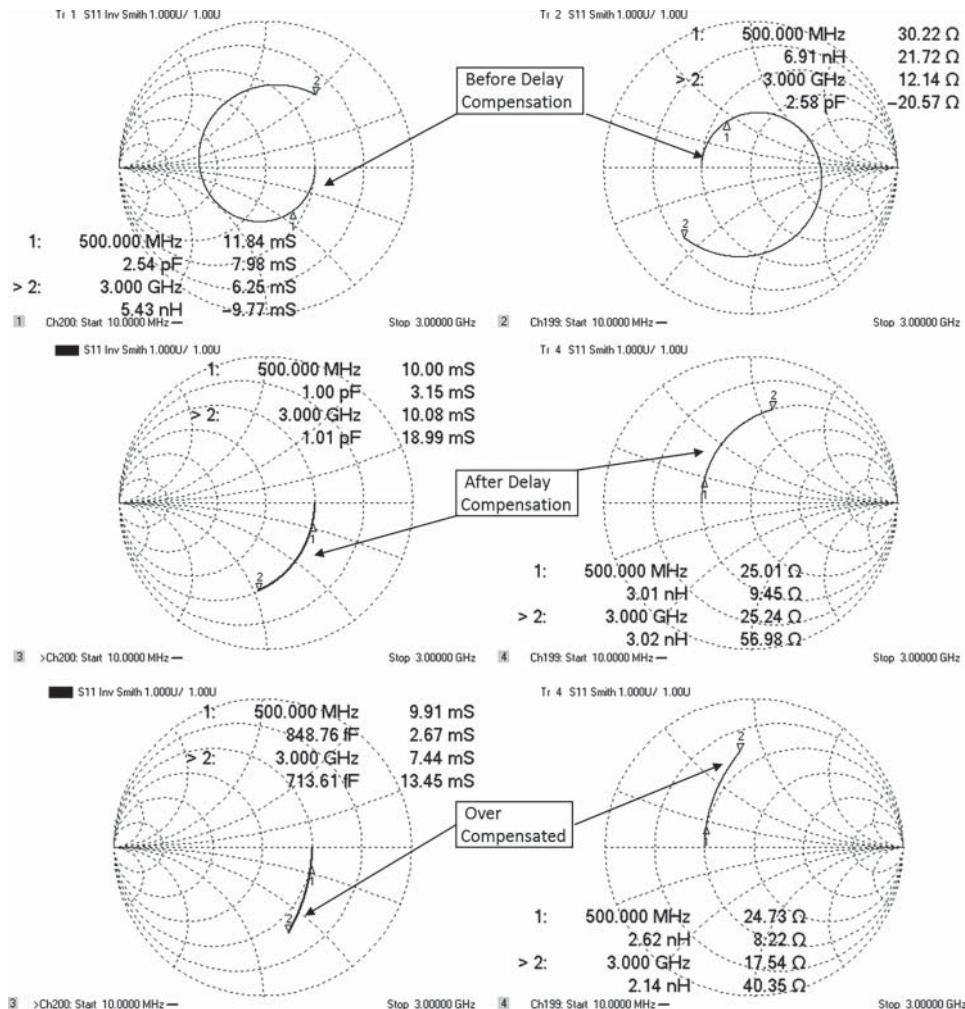


Figure 2.36 Smith chart (right) and admittance chart (left) with wrapped phase (upper), unwrapped phase (middle), and overcompensated (lower); for an inductor (right) and a capacitor (left).

and S21 with port 1, will have their phase response modified by the port extension. Electrical delay applies the same phase shift regardless of the parameter type, but port extensions properly account for a two-times phase shift for reflection parameters in contrast with a one-times phase shift for transmission parameters. Therefore, it is perhaps better to use port extension to accommodate changes in reference plane and reserve electrical delay for when one wants to remove the linear phase shift of a particular parameter.

The delay or port extension is adjusted until the trace rotation is minimized, while still maintaining a trajectory that follows a clockwise rotation. Foster demonstrated that all real devices should have phase that increases with frequency causing the clockwise rotation, so the proper amount of delay to be removed can often be determined by looking at the rotation direction of the trace trajectory. This is demonstrated in the lower traces of Figure 2.36, where

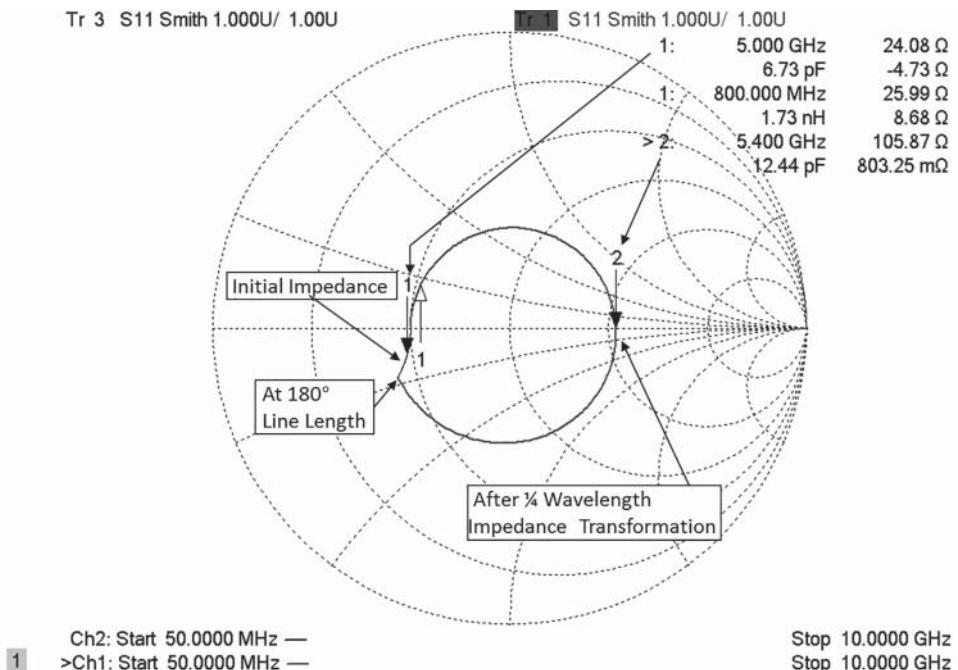


Figure 2.37 An impedance value rotated by 180° 50 Ω line.

an additional 10% of the delay from the middle traces has been removed, making the response overcompensated and causing a reactive element value at the two markers to be different.

2.4.1.2 Impedance Transformation

One aspect of rotation on the Smith chart that is often misunderstood is that the rotation around the *center* of the chart for a transmission line delay occurs only if the transmission line impedance matches the reference impedance of the Smith chart. Consider a case of a termination consisting of a 25 Ω resistor to ground, shunted by a 3 pF capacitor, and evaluated from DC to 10 GHz. The impedance trajectory is shown in the light trace in Figure 2.37, which shows a small deviation from 25 Ω due to the shunt capacitance. The darker trace shows the same impedance, but at the end of a transmission line that has 180° of phase shift at 10 GHz. The value of the impedance trajectory centers on 50 Ω, and the value of the trace at 180° phase shift matches that exactly of zero phase shift. At the frequency where the phase shifts 90° due to the transmission line (5.4 GHz) plus the slight phase shift of the DUT, the impedance is nearly 100 Ω. This is a well-known aspect of $\frac{1}{4}$ wave (or 90°, or $\lambda/4$) transmission line transformers. If impedance of the line is Z_0 , then the impedance at the end of a $\frac{1}{4}$ wave section is

$$Z_{\lambda/4} = \frac{Z_0^2}{Z_T} \quad (2.16)$$

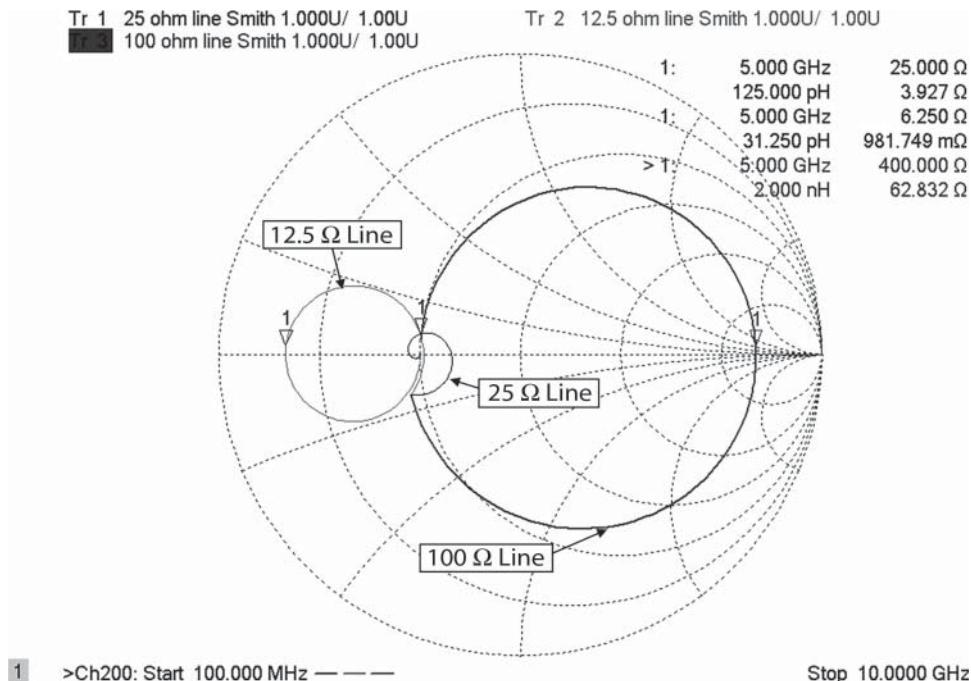


Figure 2.38 25 Ω termination proceeded by half-wavelength segments of 12.5, 25, and 100 Ω lines.

One consequence of this is that the maximum deviation of impedance due to a transmission line depends completely on the impedance of the transmission line. Figure 2.38 shows the Smith chart trajectories for the same termination, but this time with a 12.5 Ω line, a 25 Ω line, and a 100 Ω line before the termination. Of course, at 180°, no transformation of impedance takes place, and the impedance value at the end of the line is identical to that at the 0° phase shift. It is interesting to note that the smallest deviation of impedance is for the case where the line matches the impedance of the termination, rather than matching the system impedance, as in Figure 2.37.

The other important aspect to note is that when the transmission line is of greater impedance than Z_L , the resulting impedance will transform to a higher value, while when the transmission line is of lower impedance, the resulting impedance will be lower than Z_L .

2.4.2 Transforming S-Parameters to Other Impedances

While it is most common to define S-parameters in a 50 Ω impedance, or 75 Ω for cable-television applications, situations arise where it is necessary to define an S-parameter matrix in other than 50 Ω, or to have it defined with 50 Ω on one port and with a different impedance on another port. This requirement occurs for matching circuits, or impedance transformers, as well as the use of waveguide adapters where it is common practice to define the terminal impedance as 1 Ω. Unfortunately, while S-parameter definitions don't prohibit different impedances on different ports, the most common data files for S-parameters, the

so-called Touchstone™ or S2P files, provide for only a single impedance in their definition. (Recently a second revision of the S2P file format has been defined that allows different impedances on different ports, but it has not yet been widely implemented.) Thus, it is often necessary to transform S-parameters from one reference impedance to another. If the complete S-parameter matrix is available, then a matrix transformation (Tippet and Speciale 1982) can be used to convert the impedance of by applying

$$[S'] = [X]^{-1}([S] - [\Gamma])([I] - [\Gamma][S])^{-1}[X] \quad (2.17)$$

$$\text{where } [X] = \begin{bmatrix} x_1 & 0 & 0 & 0 \\ 0 & x_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & x_n \end{bmatrix}, \quad x_n = 1 - \Gamma_n, \quad [\Gamma] = \begin{bmatrix} \Gamma_1 & 0 & \cdots & 0 \\ 0 & \Gamma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \Gamma_n \end{bmatrix}, \quad \Gamma_n = \frac{Z'_n - Z_n}{Z'_n + Z_n}$$

This is a generalized formula so that an impedance, $[Z_n]$, may be defined for any port of the original $[S]$ matrix and any other impedance $[Z'_n]$ may be defined for any other port for the new $[S']$ matrix. However, the two most common cases are where the transformation occurs for all impedances at the ports being equal, so that each element in the X matrix and Γ matrix are identical, and where in the 2-port case only one impedance is transformed, as when the S-parameters of a network are defined in two different impedances.

If the measurement system impedance is pure-real, an alternative method for obtaining S-parameters at a different real impedance than the measurement system is to de-embed an ideal transformer at each port, with the turns ratio set to the square root of the impedance change. De-embedding methods are discussed in Chapter 9.

2.4.3 Concatenating Circuits and T-Parameters

In many instances, it is convenient to concatenate devices, and signal-flow charts provide a useful tool for understanding the interactions and determining the resulting S-parameter matrix. With appropriate transformations, the concatenation of S-parameter devices can be greatly simplified. One such transformation is from S-parameters to T-parameters, which also depend upon the wave functions but in a different relationship.

Figure 2.39 shows a concatenation of two devices, with a and b waves for each independently identified. Using normal signal-flow properties, the combined S-parameters of two devices is

$$\begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} = \begin{bmatrix} S_{11A} + \frac{S_{11B} \cdot S_{21A} \cdot S_{12A}}{(1 - S_{22A} \cdot S_{11B})} & \frac{S_{12A} \cdot S_{12B}}{(1 - S_{22A} \cdot S_{11B})} \\ \frac{S_{21A} \cdot S_{21B}}{(1 - S_{22A} \cdot S_{11B})} & S_{22B} + \frac{S_{22A} \cdot S_{21B} \cdot S_{12B}}{(1 - S_{22A} \cdot S_{11B})} \end{bmatrix} \quad (2.18)$$

However, signal-flow-graph techniques get tedious for concatenating a long series of devices, and other transformations make this work easier and more programmatic.

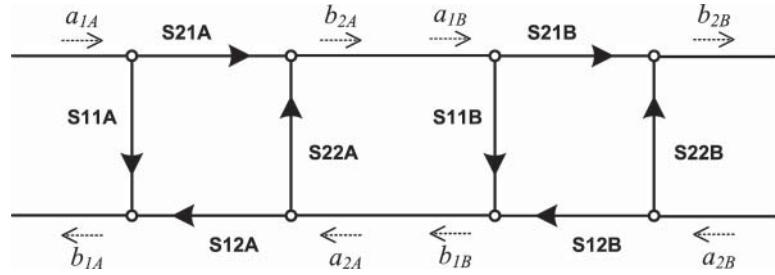


Figure 2.39 Concatenation of two devices.

The T-parameters (Keysight Application Note 154 n.d.-b) create a new functional relationship between input and output waves, with the independent variables being waves on the right and the dependent variables being waves on the left.

$$\begin{aligned} b_1 &= T_{11}a_2 + T_{12}b_2 \\ a_1 &= T_{21}a_2 + T_{22}b_2 \end{aligned} \quad (2.19)$$

Or in the matrix form

$$\begin{bmatrix} b_1 \\ a_1 \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \begin{bmatrix} a_2 \\ b_2 \end{bmatrix} \quad (2.20)$$

From this, the T-matrix describing the first and second devices are

$$\begin{bmatrix} b_{1A} \\ a_{1A} \end{bmatrix} = \begin{bmatrix} T_{11A} & T_{12A} \\ T_{21A} & T_{22A} \end{bmatrix} \begin{bmatrix} a_{2A} \\ b_{2A} \end{bmatrix}, \quad \begin{bmatrix} b_{1B} \\ a_{1B} \end{bmatrix} = \begin{bmatrix} T_{11B} & T_{12B} \\ T_{21B} & T_{22B} \end{bmatrix} \begin{bmatrix} a_{2B} \\ b_{2B} \end{bmatrix} \quad (2.21)$$

From inspection, one can recognize that the waves $a_{2A} = b_{1B}$ and $b_{2A} = a_{1B}$ so that the concatenation becomes this simple result

$$\begin{bmatrix} b_{1A} \\ a_{1A} \end{bmatrix} = \begin{bmatrix} T_{11A} & T_{12A} \\ T_{21A} & T_{22A} \end{bmatrix} \begin{bmatrix} T_{11B} & T_{12B} \\ T_{21B} & T_{22B} \end{bmatrix} \begin{bmatrix} a_{2B} \\ b_{2B} \end{bmatrix} \quad (2.22)$$

Or

$$\begin{bmatrix} b_{1A} \\ a_{1A} \end{bmatrix} = [T_A] [T_B] \begin{bmatrix} a_{2B} \\ b_{2B} \end{bmatrix} \quad (2.23)$$

Using this definition of T-parameters, the following conversions can be defined

$$\begin{aligned} \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} &= \frac{1}{S_{21}} \begin{bmatrix} -(S_{11}S_{22} - S_{21}S_{12}) & S_{11} \\ -S_{22} & 1 \end{bmatrix}, \\ \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} &= \frac{1}{T_{22}} \begin{bmatrix} T_{12} & (T_{11}T_{22} - T_{21}T_{12}) \\ 1 & -T_{21} \end{bmatrix} \end{aligned} \quad (2.24)$$

Note that in this conversion, S21 always appears in the denominator. This can cause numerical difficulties in devices with transmission zeros and can sometimes cause de-embedding functions to fail. More robust de-embedding algorithms check for this condition and modify the method of concatenation in such a case.

Other definitions of T-parameter type relationships have been described, which exchange the position variables a_1 and b_1 on the dependent variable side and exchange the position of a_2 and b_2 on the independent variable side (Mavaddat 1996). This version has similar properties, but care must be taken not to confuse the two methods as, of course, the resulting T-parameters are different. Another definition, which might seem more intuitive, would set the input terms a_1 and b_1 as the independent variable. Unfortunately, this has the undesirable effect of setting S12 in the denominator of the transformation parameter and thus gives difficulties when applied to unilateral gain devices such as amplifiers.

2.5 Modeling Circuits Using Y and Z Conversion

One common desire in evaluating the performance of a component is to model that component as an impedance comprised of a resistive element with a single series or shunt reactive element, as demonstrated in Section 2.4.1.1. This desire was furthered by some built-in transformation functions on VNAs, first introduced with the HP8753A but common now on many models. The goal was to model a device in such a way that the S-parameters mapped to a single resistive and reactive element in the so-called Z-transform case (not to be confused with the discrete time z-transform) or a single conductance and susceptance in the Y-transform case. These are quite simple models and represented in Figure 2.40.

2.5.1 Reflection Conversion

Reflection conversions are computed from the S11 trace and are essentially the same values as presented by impedance or admittance readouts of the Smith chart markers. Thus, Z-reflection conversion would be used with the circuit description from Figure 2.40a and display the impedance in the real part of the result and the reactance in the imaginary part of the result. Y-reflection would be used with the circuit of Figure 2.40b and display the conductance in the real part and the susceptance in the imaginary part of the result. The computations for these conversions are

$$Z_{\text{Refl}} = Z_0 \frac{(1 + S_{11})}{(1 - S_{11})} , \quad Y_{\text{Refl}} = \frac{(1 - S_{11})}{Z_0(1 - S_{11})} \quad (2.25)$$

Typically, these conversions would be used on one-port devices and measurements. If it is used on a 2-port device, one must remember that the load impedance will affect the measured value of the Z- or Y-reflected conversion.

2.5.2 Transmission Conversion

These reflection conversions are already well known as the models represented by the Smith chart, but a similar conversion can be performed for a simple transmission measurement. In

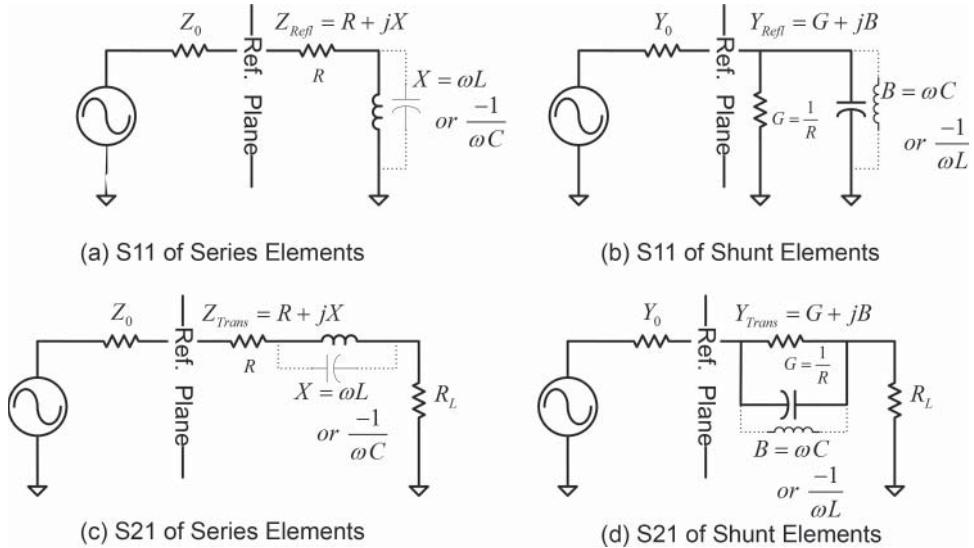


Figure 2.40 Y and Z conversion circuits.

this case, the circuits of Figure 2.40c and Figure 2.40d are the reference circuits for these conversions. They are useful when analyzing the series element models, such as coupling capacitors, and the models for series resistors and inductors. The underlying computation for the transmission conversions is

$$Z_{Trans} = Z_0 \frac{2(1 - S_{21})}{S_{21}} , \quad Y_{Trans} = \frac{S_{21}}{2Z_0(1 - S_{21})} \quad (2.26)$$

The Z-transmission conversion would be well suited to view the series resistance of a coupling capacitor. The Y-transmission would show the resistive value of a series-mounted surface-mount technology (SMT) resistor with a shunt capacitance as a constant conductance with a reactance increasing as $2\pi f$, forming a straight reactance line.

These conversions are often confused with conversion to Y- or Z-parameters, but they are not, in general, related. These provide simple modeling functions based on a single S-parameter, whereas the Y-, Z-, and related parameters provide a matrix result and require knowledge of all four S-parameters as well as the reference impedance. These other matrix parameters are described in the next section.

2.6 Other Linear Parameters

Even though a VNA measures S-parameters as its fundamental information, many other figures of merit may be computed directly from these measurements, through the use of transformations, found in several references (Hong and Lancaster 2001; Keysight Technologies n.d.-a). Most of these common parameters relate the voltage and current at the ports, rather than the *a* and *b* waves. Many of these transformations arise out of different definitions of terminal conditions as applied to Figure 1.2. These definitions arise out of DC or low-frequency

measurements, where it is an easy matter to short a terminal, meaning $Z_L = 0$, or open a terminal, meaning $Z_L \rightarrow \infty$. An often confusing point is that it is not necessary that the terminals actually be opened or shorted, but most commonly the parameter is described in those terms. Just as it is most common to terminate the a 2-port network in Z_0 to define S21, making $a_2 = 0$, it is not necessary to do so, and S21 can be determined with any terminal impedance as long as sufficient changes in a_1 and a_2 are made to solve Eq. (1.17), as shown in Eq. (1.21). Since the voltage and current relationships on the terminals of a DUT are easily determined from the S-parameters, many other linear parameters can be determined as well. Unless otherwise noted, these transformations apply to the simple case where the S-parameters are defined with a single, real-valued reference impedance.

2.6.1 Z-Parameters, or Open-Circuit Impedance Parameters

Z-parameters are one of the more commonly defined parameters and often the first characterization parameter introduced in engineering courses on electrical circuit fundamentals.

The Z-parameters are defined in terms of voltages and currents on the terminals as

$$\begin{aligned} V_1 &= Z_{11} \cdot I_1 + Z_{12} \cdot I_2 \\ V_2 &= Z_{21} \cdot I_1 + Z_{22} \cdot I_2 \end{aligned} \quad (2.27)$$

where the V's and I's are defined in Figure 1.2. If we apply the condition of driving a voltage source into the first input terminal and opening the first output terminal, which forces I_2 to zero, and measure the input and output voltages, we can determine two of the parameters; similarly, the other two parameters are determined by driving the output terminal and opening the input terminal. Mathematically this can be stated as

$$\begin{aligned} Z_{11} &= \left. \frac{V_1}{I_1} \right|_{I_2=0} & Z_{12} &= \left. \frac{V_1}{I_2} \right|_{I_1=0} \\ Z_{21} &= \left. \frac{V_2}{I_1} \right|_{I_2=0} & Z_{22} &= \left. \frac{V_2}{I_2} \right|_{I_1=0} \end{aligned} \quad (2.28)$$

But, the conditions for measurement of these parameters directly cannot be realized in RF and microwave systems for several key reasons, listed here:

1. When the ports of an RF circuit are left open, fringing capacitance from the center pin of the RF terminal to ground reduces the impedance at high frequency, and phase shift from the DUT reference plane makes the practical value of an open circuit deviate from the ideal.
2. The measuring equipment to sense V_1 , I_1 , V_2 , and I_2 has a parasitic impedance to ground, which also shunts some of the terminal current. At the driving port, this means that the measured current does not match the actual current into the DUT, and at the open port it means that while being measured, the output impedance does not match that of an open circuit
3. For many active devices, the DUT is only conditionally stable and may oscillate if larger reflections are presented at the ports. Phase shift of the open circuit from the terminal

port to the DUT active device can cause the reflection to take on almost any phase, and it ensures that at some frequency the reflection at the port will be such that the device will oscillate. This is perhaps the primary reason for S-parameters being used on active devices; they provide a consistent low-reflection load, which in general prevents oscillations of the DUT.

Of course, Z-parameters are not restricted to just two ports, and the Z-parameters can be put in a matrix form of

$$\begin{bmatrix} V_1 \\ \vdots \\ V_n \end{bmatrix} = \begin{bmatrix} Z_{11} & \cdots & Z_{1n} \\ \vdots & \ddots & \vdots \\ Z_{n1} & \cdots & Z_{nn} \end{bmatrix} \cdot \begin{bmatrix} I_1 \\ \vdots \\ I_n \end{bmatrix} \quad \text{or } [V_n] = [Z] \cdot [I_n] \quad (2.29)$$

where $[Z]$ is called the Z-matrix.

The Z-matrix and S-matrix can be computed from each other, given the reference impedance is known for the S-parameters, and the same on each port; then

$$\begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} = \frac{Z_0}{\Delta S} \begin{bmatrix} (1 + S_{11})(1 - S_{22}) + S_{21}S_{12} & 2S_{12} \\ 2S_{21} & (1 - S_{11})(1 + S_{22}) + S_{21}S_{12} \end{bmatrix} \quad (2.30)$$

where $\Delta S = (1 - S_{11})(1 - S_{22}) - S_{21}S_{12}$

$$\begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} = \frac{1}{\Delta Z} \begin{bmatrix} (Z_{11} - Z_0)(Z_{22} + Z_0) - Z_{21}Z_{12} & 2Z_{12}Z_0 \\ 2Z_{21}Z_0 & (Z_{11} + Z_0)(Z_{22} - Z_0) - Z_{21}Z_{12} \end{bmatrix} \quad (2.31)$$

where $\Delta Z = (Z_{11} + Z_0)(Z_{22} + Z_0) - Z_{21}Z_{12}$

An attribute of the Z-matrix is that if a DUT is lossless, the Z-matrix will contain only pure imaginary numbers; this is commonly found in filter design applications. If $Z_{21} = Z_{12}$, then the DUT is reciprocal, and if also $Z_{11} = Z_{22}$, the network is symmetrical. Note that in general $Z_{In} \neq Z_{11}$ except for a 1-port network. Z_{In} represents the ratio of V_1 and I_1 for the DUT as it is terminated, normally in the system reference impedance Z_0 , where Z_{11} is the ratio of V_1 and I_1 when all the other ports are open circuited, a not very useful case in practice. Another important attribute of the Z-matrix is that its values do not depend upon the measurement system, unlike S-parameters whose values depend upon the reference impedance for each port and whose values can change for the same network, if these reference impedances change. Put another way, the S11 of a 50 Ω load will be quite different when measured in a 75 Ω reference impedance, but the Z-parameters will not change.

2.6.2 Y-Parameters, or Short-Circuit Admittance Parameters

Y-parameters are essentially an inverse, of Z-parameters, and in fact the Y-matrix is the inverse of the Z-matrix. The definition of Y-parameters is derived from

$$\begin{aligned} I_1 &= Y_{11} \cdot V_1 + Y_{12} \cdot V_2 \\ I_2 &= Y_{21} \cdot V_1 + Y_{22} \cdot V_2 \end{aligned} \quad (2.32)$$

From this, the common description of Y-parameters are

$$\begin{aligned} Y_{11} &= \left. \frac{I_1}{V_1} \right|_{V_2=0} & Y_{12} &= \left. \frac{I_1}{V_2} \right|_{V_1=0} \\ Y_{21} &= \left. \frac{I_2}{V_1} \right|_{V_2=0} & Y_{22} &= \left. \frac{I_2}{V_2} \right|_{V_1=0} \end{aligned} \quad (2.33)$$

The Y-parameters can also be defined for more than 2-port devices, and the matrix form is

$$\begin{bmatrix} I_1 \\ \vdots \\ I_n \end{bmatrix} = \begin{bmatrix} Y_{11} & \cdots & Y_{1n} \\ \vdots & \ddots & \vdots \\ Y_{n1} & \cdots & Y_{nn} \end{bmatrix} \cdot \begin{bmatrix} V_1 \\ \vdots \\ V_n \end{bmatrix} \text{ or } [I_n] = [Y] \cdot [V_n] \quad (2.34)$$

where the Y-matrix refers to $[Y]$. The Y-matrix is related to the Z-matrix through its inverse, as

$$[Y] = [Z]^{-1} \quad (2.35)$$

The conversions between the S-matrix and the Y-matrix are

$$\begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix} = \frac{Y_0}{\Delta_Y S} \begin{bmatrix} (1 - S_{11})(1 + S_{22}) + S_{21}S_{12} & -2S_{12} \\ -2S_{21} & (1 + S_{11})(1 - S_{22}) + S_{21}S_{12} \end{bmatrix} \quad (2.36)$$

where $\Delta_Y S = (1 + S_{11})(1 + S_{22}) - S_{21}S_{12}$

$$\begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} = \frac{1}{\Delta Y} \begin{bmatrix} (Y_0 - Y_{11})(Y_0 + Y_{22}) + Y_{21}Y_{12} & -2Y_{12}Y_0 \\ -2Y_{21}Y_0 & (Y_0 + Y_{11})(Y_0 - Y_{22}) + Y_{21}Y_{12} \end{bmatrix} \quad (2.37)$$

where $\Delta Y = (Y_0 + Y_{11})(Y_0 + Y_{22}) - Y_{21}Y_{12}$

2.6.3 ABCD Parameters

Just as the T-parameters provide for easy concatenation of devices using *a* and *b* waves, a similar matrix representation can be used when the terminal characteristics are defined in terms of voltage and current. These are sometimes called *transfer parameters* (reminiscent of the T-parameters) or *chain parameters* as networks can be chained together in a matrix multiplication manner.

The functional definition of ABCD parameters is found in at least two different forms, one of which is

$$\begin{aligned} V_1 &= A \cdot V_2 - B \cdot I_2 \\ I_1 &= C \cdot V_2 - D \cdot I_2 \end{aligned} \quad (2.38)$$

The second form replaces the minus sign with a plus sign with resulting changes in the derived values.

From Eq. (2.38) the values for ABCD parameters can be defined as

$$\begin{aligned} A &= \frac{V_1}{V_2} \Big|_{I_2=0} & B &= \frac{V_1}{-I_2} \Big|_{V_2=0} \\ C &= \frac{I_1}{V_2} \Big|_{I_2=0} & D &= \frac{I_1}{-I_2} \Big|_{V_2=0} \end{aligned} \quad (2.39)$$

The transformations between the ABCD-matrix and the S-matrix are

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \frac{1}{2S_{21}} \begin{bmatrix} (1 + S_{11})(1 - S_{22}) + S_{21}S_{12} & Z_0[(1 + S_{11})(1 + S_{22}) - S_{21}S_{12}] \\ \frac{1}{Z_0}[(1 - S_{11})(1 - S_{22}) - S_{21}S_{12}] & (1 - S_{11})(1 + S_{22}) + S_{21}S_{12} \end{bmatrix} \quad (2.40)$$

$$\begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} = \frac{1}{\Delta} \begin{bmatrix} A + \frac{B}{Z_0} - CZ_0 - D & 2(AD - BC) \\ 2 & -A + \frac{B}{Z_0} - CZ_0 + D \end{bmatrix} \quad (2.41)$$

where $\Delta = A + \frac{B}{Z_0} + CZ_0 + D$

2.6.4 H-Parameters or Hybrid Parameters

Because of its intrinsic transfer function, as a voltage-controlled current source, transistor performance has often been described using hybrid parameters. Their functional definition is

$$\begin{aligned} V_1 &= H_{11} \cdot I_1 + H_{12} \cdot V_2 \\ I_2 &= H_{21} \cdot I_1 + H_{22} \cdot V_2 \end{aligned} \quad (2.42)$$

From this, the definition of individual H-parameters is

$$\begin{aligned} H_{11} &= \frac{V_1}{I_1} \Big|_{V_2=0} & H_{12} &= \frac{V_1}{V_2} \Big|_{I_1=0} \\ H_{21} &= \frac{I_2}{I_1} \Big|_{V_2=0} & Y_{22} &= \frac{I_2}{V_2} \Big|_{I_1=0} \end{aligned} \quad (2.43)$$

The H-matrix is most simply defined in terms of the other impedance matrixes as

$$[H] = \begin{bmatrix} \frac{1}{Y_{11}} & \frac{Z_{12}}{Z_{22}} \\ -\frac{1}{D} & \frac{1}{Z_{11}} \end{bmatrix} = \begin{bmatrix} \frac{(1 + S_{11})(1 + S_{22}) - S_{21}S_{12}}{Z_0(1 - S_{11})(1 + S_{22}) + S_{21}S_{12}} & \frac{2 \cdot S_{12}}{(1 - S_{11})(1 + S_{22}) + S_{21}S_{12}} \\ \frac{-2 \cdot S_{21}}{(1 - S_{11})(1 + S_{22}) + S_{21}S_{12}} & \frac{1}{Z_0} \cdot \frac{(1 - S_{11})(1 - S_{22}) - S_{21}S_{12}}{(1 - S_{11})(1 + S_{22}) + S_{21}S_{12}} \end{bmatrix} \quad (2.44)$$

2.6.5 Complex Conversions and Non-equal Reference Impedances

It is important to note that all the conversions described in the previous sections are valid only for the case of $Z_{01} = Z_{02} = Z_0$ and that Z_0 be pure-real. The transformations for cases where the port impedances are not equal, and not real, have been computed and are available in papers by Marks and Williams (1992) and with using somewhat different wave definitions, by D.A. Frickey (1994). While the case of complex termination impedances is unusual, the case for differing reference impedances on ports is more common. Since the network elements do not change when reference impedance of the system is changed, Y-, Z-, and H-related parameters also do not change with reference impedance. S- and T-parameter values do change, and thus it is critical to know the reference impedance for each case.

References

- Frickey, D.A. (1994). Conversions between S, Z, Y, H, ABCD, and T parameters which are valid for complex source and load impedances. *IEEE Transactions on Microwave Theory and Techniques* **42** (2): 205–211.
- Hong, J.-S. and Lancaster, M.J. (2001). *Microstrip Filters for RF/Microwave Applications*. New York: Wiley. Print.
- Keysight Technologies (n.d.-a) S-Parameters... Circuit Analysis and Design [Online]. Available at: <http://literature.cdn.keysight.com/litweb/pdf/5952-0918.pdf> (Accessed 20 January 2020).
- Keysight Technologies (n.d.-b) S-Parameter Design [Online], Available at: <https://literature.cdn.keysight.com/litweb/pdf/5952-1087.pdf?id=1000001947:epsg:apn> (Accessed 28 December 2018).
- Marks, R.B. and Williams, D.F. (1992). A general waveguide circuit theory. *Journal of Research of the National Institute of Standards and Technology* **97**: 533–561.
- Mavaddat, R. (1996). *Network Scattering Parameters*. Singapore: World Scientific. Print.
- Smith, P.H. (1944). An improved transmission line calculator. *Electronics* **17**: 130–133.
- Tippet, J.C. and Speciale, R.A. (1982). A rigorous technique for measuring the scattering matrix of a multiport device with a 2-port network analyzer. *IEEE Transactions on Microwave Theory and Techniques* **30** (5): 661–666.

3

Calibration and Vector Error Correction

3.1 Introduction

The vector network analyzer (VNA) is perhaps the most precise electronic instrument used in RF and microwave measurements. Modern VNAs can measure signals of high and low power with better precision than any other power sensor, and they can measure the gain across frequency of an electronic device with a performance traceable to measurements of physical dimensions. Of all the electronic measurement systems, the VNA derives the largest share of its performance and quality of measurement from error correction. However, when one speaks about calibration in reference to a VNA, it is in a different context than the term *calibration* for other instruments, and this is a common point of confusion. Also, VNAs provide correction for both magnitude and phase responses, which is commonly called *vector error correction*. Here, the term *error correction* should be understood to refer in general to vector correction methods.

Most other electronic test instruments are calibrated to their performance through a process of careful measurement with other electronic equipment of a higher quality, typically once a year. This process ensures that the equipment is meeting its specified performance, and in some instances, the performance is measured and adjustments are made to optimize the performance. In the specific case of noise sources that have excess noise ratio (ENR) data table, power sensors that have a flatness-versus-frequency coefficient table, and other equipment that maintains correction arrays associated with their performance, the values for these corrections are updated when the equipment is sent in for its periodic calibration. In the case where an instrument cannot meet its performance specifications even after adjustment, repairs are made typically in the form of replacing defective modules in the instrument. For most electronic instruments, their performance pertains only to measurements made directly at the connector of the test equipment. This “calibration” process concludes with a reference sticker that states the equipment is in calibration, with a recommended date for recertification.

Most VNAs go through a similar process, where the raw hardware performance characteristics are similarly characterized. But for VNAs the actual quality of the hardware, such

as frequency flatness or coupler directivity, can be easily characterized in the field, and their effects as well as the effects of connectors, cables, fixtures, and probes can be removed to yield a resulting measurement of the device-under-test (DUT) that far surpasses the capability of the raw hardware. This process is commonly *calibration*, but a more proper term is *error correction*. Calibration implies that measurements are made to characterize a performance parameter, and then adjustments are made to improve the actual performance. In contrast, traditional error correction in a VNA is strictly a post-processing function, where error correction algorithms are applied post-measurement to raw measured data to produce the corrected result.

The VNA error correction process consists of two steps: the first step, typically called *calibration* or *VNA-cal*, is to measure known standards, such as an open/short/load, to determine the systematic VNA error terms. This might be formally called *error correction acquisition*. The second step is measuring the DUT and applying the error correction algorithms to obtain a corrected result. This might be formally called *error correction application* or more informally just *correction*. These processes are unrelated to the yearly calibration process and are sometimes called *user calibration* to indicate that it is done at the time of use, rather than at the instrument manufacturers' factory or service center.

Finally, some modern VNAs have had the error correction acquisition process performed in the factory before being shipped so that they have a built-in “factory-cal” that can be applied to raw measurements even if a user does not perform the VNA-cal acquisition. In at least one example, the HP 8752A, a test port cable connected to port 2, was characterized at the factory and included in the factory calibration error terms. In most cases, a user VNA-cal replaces the factory-cal, so the factory-cal does not have any effect on the user calibration. Some exceptions to this are discussed in Section 3.16.7.

There have been many papers and publications discussing VNA calibration, and while not completely consistent in their terminology, common terms and symbols for some cases are recognized. Many of the terms come from the original implementations in the early HP VNAs of the HP 8510 and HP 8753, which interestingly form both general types of measurement systems, the four-receiver VNA and the three-receiver VNA, respectively. Some advanced techniques utilize differing terminology in the original papers, but an attempt has been made here to maintain common terminology whenever possible. As with any subject of sufficient longevity, there are far more published papers on variations of VNA calibration than can be included even in a book of this nature. Interested readers are encouraged to review the references and bibliography for a more in-depth study of this area. Here, the theoretical treatment of VNA correction will be limited to the important results that are found in common practice today and in the practical application of those methods to real-world problems faced by the practicing engineer.

3.1.1 Error Correction and Linear Measurement Methods for S-Parameters

Linear measurements imply that the parameter being measured does not depend upon the level of the signal applied. For RF and microwave measurements, the principle linear parameters are S-parameters from which gain, match, impedance, and isolation, among others, can be derived.

3.1.1.1 Signal Flow Graphs for VNA Hardware Configurations

A VNA is used to measure the S-parameters of a device by applying the source signal to the input of the DUT and measuring the response on the VNA receivers. With the source applied to port 1 of the DUT the incident wave, a_{1M} , is measured from the reference channel, and the test signals are measured at the input and output as b_{1M} and b_{2M} , but these do not represent the actual values of incident and scattered waves from the DUT. Errors in the VNA alter the source and receiver signals such that the measured values can vary significantly from the values at the DUT reference plane. There are a variety of errors associated with S-parameter measurements, but the principle ones can be defined in a signal flow diagram (Fitzpatrick 1978; Rytting 1996a), as shown in Figure 3.1. In conventional VNA measurements, each port of the DUT is stimulated by the source signal, and the other ports provide a nominally matched termination. For 2-port devices, it's common to refer to stimulus at port 1 as the forward direction and port 2 as the reverse direction. In multiport systems, the F and R may be replaced with port numbers for the reflection port terms, such as ED1, ES1, and ER1, but the transmission terms must contain a port pair, such as ET21 or EL12. While it might seem odd that the load match is considered a transmission term, some VNA and multiport test systems present a different load impedance to the DUT depending upon the source port, so the load match must be explicitly designated in terms of both the load port and the source port. It is standard practice to name the load port first and the source port second so that ET21 is the tracking term that primarily affects the S21 parameter.

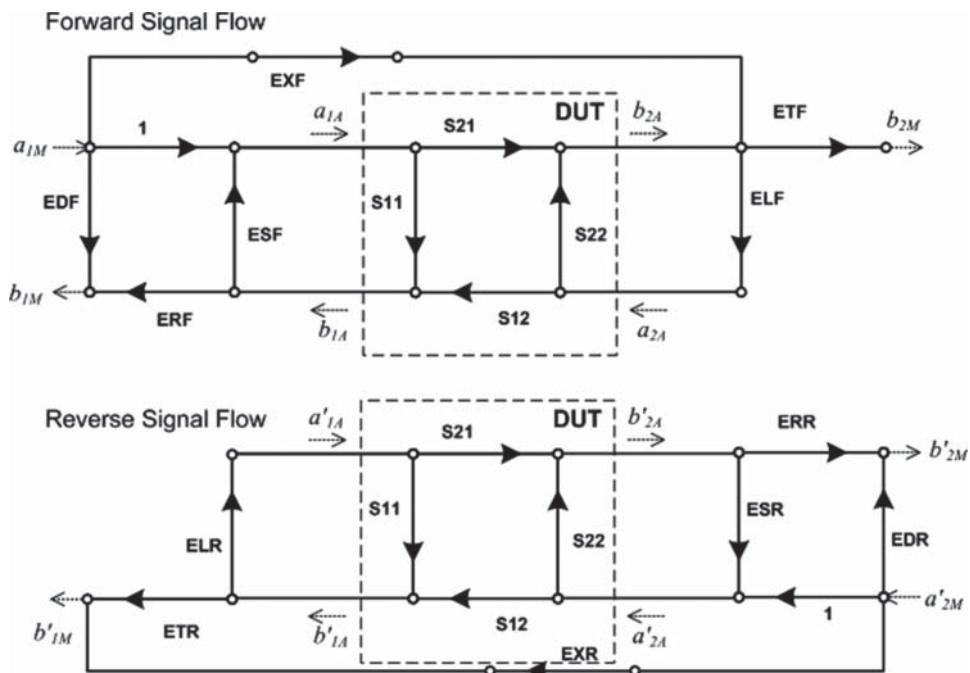


Figure 3.1 Signal flow diagram for a forward and reverse measurements of a DUT and VNA.

Table 3.1 Systematic error terms

Systematic error terms in a VNA			
EDF	Forward directivity	EDR	Reverse directivity
ESF	Forward source match	ESR	Reverse source match
ERF	Forward reflection tracking	ERR	Reverse Reflection tracking
ELF	Forward load match	ELR	Reverse load match
ETF	Forward transmission tracking	ETR	Reverse transmission tracking
EXF	Forward crosstalk	EXR	Reverse crosstalk

Table 3.2 Measurements and associated error terms

Measurement\Error	Tracking	Mismatch	Leakage
Input reflection	ERF	ESF	EDF
Forward transmission	ETF	ELF	EXF
Reverse transmission	ETR	ELR	EXR
Output reflection	ERR	ESR	EDF

The traditional systematic errors in the measurement are shown, which can be characterized and removed from the measurements of a linear device. These error terms are listed in Table 3.1.

Another way to view this is that the errors come in three types: frequency response (usually called *tracking*), mismatch, and leakage. And the errors occur for each of the four S-parameters so that they may be expressed in a 3×4 table, as shown in Table 3.2.

Later, these definitions will be used to simplify the formulas for correction. From this, it is implied that directivity has a similar effect on reflection measurements that crosstalk does on transmission measurements; a similar implication holds true for frequency response. Mismatch has a more complicated effect, as will become clear in later sections.

This signal flow graph configuration represents what is commonly known as the *12-term* error model for S-parameter measurements. Most VNAs use this as the model for which the error correction algorithms are applied. In this signal flow model, the measurements are presumed to occur with a forward sweep of the VNA where the source is applied to port 1, followed by a reverse sweep where the source is applied to port 2. Tracking terms represent relative losses (or how the receivers track each other), and match terms represent errors relative to the system Z_0 . As described in Chapter 2, the source-match error term refers to the ratio source match. Implicit in this model is that the port match at each port changes depending upon whether the source is active, as the internal switching of the VNA changes its impedance when the source is switched to a terminated state. This is represented in the naming where the port match is called the *source-match* for the active port and the *load match* for the inactive port.

The crosstalk error term is largely ignored in modern VNAs, where the isolation between ports is greater than the noise floor of the system. In such a case, the crosstalk cannot be

adequately characterized and thus is typically set to zero, reducing the 12-term error model to a 10-term model.

One subtle detail of the error terms is the unfortunate naming of the directivity error term, EDF, which in fact does not exactly represent the directivity of the test coupler but rather the ratio of the leakage signal (from the isolation of the coupler) to the loss to the $a1$ receiver. This is essentially $b1/a1$ when measuring a perfect Z_0 load. In the case where the reflection tracking is 1, the directivity error term is identical to the directivity of the test system. But if the VNA setup causes the ERF term to differ from 1 (that is, not exactly 0 dB, for example due to different losses to the $a1$ and $b1$ receiver), then the $b1/a1$ leakage response to the load will change by that offset value, even though the coupler directivity is not changed (see Figure 1.34). For example, if an attenuator is added between the coupled arm of the port 1 test port coupler and the test receiver, or between the $a1$ reference coupler and the main arm of the test port coupler, the measured load response and the EDF term will change, as will the reflection tracking response ERF, but the system directivity is not affected. Thus, the EDF error term does not represent the coupler or system directivity; rather, that can be computed as

$$\text{SystemDirectivity} = \frac{\text{EDF}}{\text{ERF}} \quad (3.1)$$

If an external attenuator is added after the test port coupler, between the coupler and the DUT, then the EDF term will not change, but the refection tracking will (by twice the attenuators' value), so the effective directivity of the test port coupler and the system will also be reduced by an amount equal to twice the added attenuation.

In Section 3.3 methods for characterizing these systematic errors and solving the flow graph for the actual or corrected S-parameters will be described. It should be remembered that while the VNA itself may have well-matched ports, with nearly ideal error terms, most of the errors in the flow diagram of Figure 3.1 occur as a result of cabling, fixturing, or wafer-probing interfaces to the DUT.

3.1.2 Power Measurements with a VNA

The signal flow diagram of Figure 3.1 is a simplified representation of the signal path that is suitable for S-parameters or other parameters that represent of a ratio of a and b waves, where it is presumed that the VNA receivers measure the a_{1M} and b_{1M} directly. In the past, VNAs evaluated only ratio types of parameters. Modern VNAs can be calibrated to report power levels directly, and to control the source and receive powers in sophisticated ways, so that a more detailed signal flow diagram is required to understand the errors associated with parameters that are a direct measurement of the a and b waves at the DUT.

Power measurements of a DUT, strictly speaking, are not linear measurements of a component, but a combination of the DUT and the source signal. For example, the output power of an amplifier depends upon the input power, unless the amplifier is completely saturated. And if the DUT is linear, the power at any port can be computed by the S-parameters and knowledge of the input power. The measurement methods for making power measurements are very similar to making standard S-parameter measurements, and the errors in power measurements are quite closely related to the errors in S-parameter measurements. In most cases, power measurements are required to quantify the non-linear behavior of a DUT, such as the

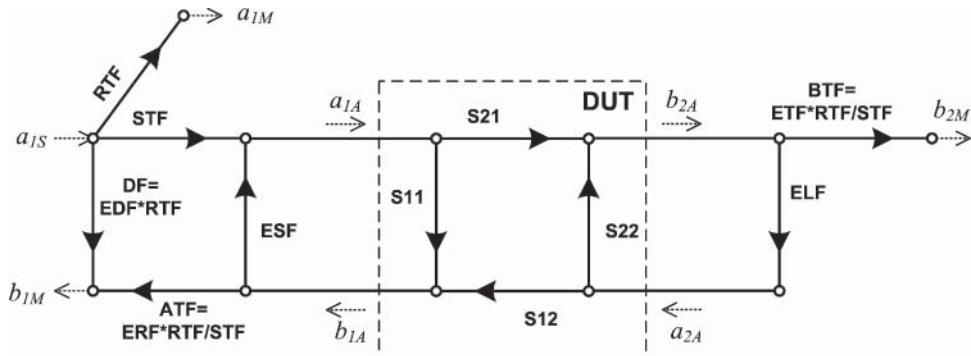


Figure 3.2 Signal flow diagram with source and receiver errors included.

power level where the device changes from linear to non-linear response. Other power-related measurements include power-added efficiency, which measures both direct current (DC) and radio frequency (RF) power, and can compute the efficiency of converting the DC power to RF output signal.

Figure 3.2 shows a signal flow diagram that includes source and receiver error terms for a source that is active in the forward direction. Unlike the S-parameter error terms, error terms associated with the source loss and receiver loss do not have standardized names, so some are defined here and used consistently throughout this chapter.

The applied source signal (a_{1S}) at the input is measured by the reference receiver, which has its own loss and frequency response described by the reference transmission forward (RTF) error and represents the loss to the reference receiver. The loss of the source path to the test port of the DUT is represented by the source transmission forward (STF) error and represents the loss in the test set. The reflection test receiver (sometimes labeled the A receiver) measures the reflected power with an A transmission in the forward flow graph (ATF); the transmission test receiver (sometimes labeled the B receiver) measures the transmitted power with a B transmission error in the forward flow graph (BTF). Many manufacturers provide a factory correction of the receivers so that the loss and frequency response of the test and reference coupler are substantially removed. Note that the ATF may be equal to the AT, but also might be different if there is some switching that occurs in the test set between the forward and reverse directions that changes the loss to the A coupler.

This particular representation of a signal flow graph will be useful in describing the source and receiver power corrections, as discussed in Sections 3.6 and 3.7. These error terms are seldom mentioned in the current literature on VNAs and error correction, but aspects of them have been described in several papers. For example, the loss of the reference and test receiver is not explicitly articulated; in some manufacturers' error correction algorithms, the error terms are identified by type and port number and sometimes called *R receiver response tracking* (R_{Tr}) or *A or B receiver response tracking* (A_{Tr}, B_{Tr}). Other nomenclature acknowledges the use of the more formally correct $a1, a2, b1, b2$ response tracking, and some allow the use of either naming convention.

The signal flow diagram of Figure 3.2 separates the individual contributors to the S-parameter tracking terms, ERF and ETF, which can be computed from the receiver

tracking terms, and the loss of the source path, as

$$ERF = \frac{STF \cdot ATF}{RTF}, \quad ETF = \frac{STF \cdot BTF}{RTF} \quad (3.2)$$

With the configuration of the signal flow diagram separating the receiver response, the directivity error term is modified such that

$$EDF \equiv \left. \frac{b_{1M}}{a_{1M}} \right|_{\Gamma_{Load}=0} = \frac{DF}{RTF} \quad (3.3)$$

EDF should be recognized as the raw measurement of b_{1M}/a_{1M} with a perfect non-reflecting load at the test port. If the test port coupler has coupling of ATF and directivity of DIR, then the isolation of the coupler is

$$CouplerIsolation = DIR \cdot ATF \quad (3.4)$$

EDF will change if the value of STF changes.

In Figure 3.2 the loss of the reference channel splitter or coupler is shown as RTF. In the past, the VNA was designed so that this loss was equal to the combination of STF and ATF for the reflection port or STF and BTF for the transmission port; as such, losses of ATF and BTF were designed to have the same values; ATF and BTF would inherently take on the same values if identical couplers are used at ports 1 and 2. In this way, the ERF and ETF error terms would approach one, and the raw measured values of the receivers would more nearly match the actual values of the DUT. Modern VNAs sometimes added additional attenuation into the reference path to ensure that even at the highest source powers, the reference channel receiver would not be in compression. This allows extending the dynamic range of the analyzer when the DUT has high loss, described in Chapter 5.

Typically, the values of RTF and BTF are measured at the manufacturers' factory, and the raw values of a_{1M} , b_{1M} , b_{2M} , and a_{2M} are mathematically adjusted to roughly compensate for this loss. This factory calibration allows raw measurements of a DUT to be close to the actual DUT values and is helpful in setting up and optimizing a measurement before starting the calibration process. Some instruments take this a step further and provide a factory 1-port calibration to the raw measured results so that the apparent source match and directivity might seem very good. This is particularly true for systems that use lower-cost or more compact structures for couplers and bridges that may have quite poor, though stable, performance. Since traditional VNAs provide reasonable measurements even without error correction, providing a hidden 1-port factory calibration can maintain the illusion of good raw hardware performance even in cases where the raw performance is not very good.

Similarly, the loss of the source tracking term, STF, is compensated in the source settings so that a_{IS} is adjusted such that the power delivered to the test port is approximately correct. Some high-performance VNAs place a step attenuator in the STF path, allowing the power at the test port to be substantially lowered while maintaining high power at the reference receiver, a_{1M} . Here, changing the step attenuator would drop the power into the test receivers while leaving the power in the reference receiver high, causing the apparent value of the S-parameters to change. Modern analyzers compensate the factory calibration by the nominal change in the attenuator so that the displayed receiver power at the reference receiver is lowered by the attenuator value. Still, the attenuator loss is not identical to its

nominal value; additionally, the mismatch of the attenuator changes with the attenuator state, so that substantial ripple is often seen after calibrating in one state and changing to another state. Even these errors can be removed by some clever steps to characterize the attenuator response, as discussed in Chapter 11. Mid-performance analyzers often combine the reference and test coupling into a single, lower-cost structure, and place the step attenuator between the source and the reference coupler. This has the advantage of keeping the test set response constant but has a disadvantage in that for low-power situations the reference signal is very low and often dominates the trace noise.

At the test port of a VNA with a factory correction, the source power, receiver power readings, and S-parameters typically have less than 1 dB of error. Of course, if a cable is used, the loss of the cable is usually not compensated for in the VNA factory calibrations.

In some cases, additional internal step attenuators are added to the VNA block diagram between the coupled arm of the test port coupler and the test receivers, which measure b_{1M} and b_{2M} . In these cases, the values of ATF and BTF may also be compensated in the VNA software by the value of the receiver attenuation, allowing the receiver attenuation state to be changed while the raw values of the DUT S-parameters remain relatively the same. In other cases, external attenuators are sometimes placed between the coupled arm of the test port couplers and the receivers especially in high-power measurement situations, where the power absorbed by the external attenuator exceeds the internal attenuators' power handling. In these cases, the ATF or BTF error terms will incorporate the value of the attenuator offsets, as the factory calibration does not compensate for external components.

3.2 Basic Error Correction for S-Parameters: Cal-Application

Correction of systematic errors of S-parameter measurements in VNAs has been available for decades; the details of how the systematic errors affect the measurement results of a 2-port S-parameter measurement are presented in this section.

There are two basic error models employed in VNAs, which rely on simultaneous measurements on either three VNA receivers, which is known as the *12-term model* (Fitzpatrick 1978; Rytting 1980), or four receivers, known as the *8-term model* (Marks 1997). In modern VNAs, both are used, and it is a simple process to move between either model. In fact, most VNAs represent the error terms strictly as the 12-term model, but 8-term methods are often more convenient to use to determine the values of the error terms. Other models include many more effects, but these are not commonly used in practice (Rytting 1996b).

3.2.1 12-Term Error Model

The 12-term error model actually consists of two separate 6-term models, one in the forward direction and one in the reverse direction. This model was introduced in Figure 3.1. In each case, three simultaneous or phase-coherent receiver measurements are required, including one incident wave and two scattered waves. Here the incident wave associated with the load port is assumed to be zero. Consider the forward error-model presented in the upper portion of the figure.

The measured values of a and b waves can be related to the actual values by computing the scattered waves from the incident waves, the error terms, and the actual S-parameters, where

here $a_2 = 0$, as ratios expressed as measured S-parameters. For the forward stimulation, this is

$$S_{11M} = \frac{b_{1M}}{a_{1M}} = EDF + \frac{ERF \left(S_{11A} + \frac{S_{21}ELF \cdot S_{12A}}{(1 - S_{22A} \cdot ELF)} \right)}{\left[1 - ESF \cdot \left(S_{11A} + \frac{S_{21}ELF \cdot S_{12A}}{(1 - S_{22A} \cdot ELF)} \right) \right]} \\ S_{21M} = \frac{b_{2M}}{a_{1M}} = \frac{(S_{21A} \cdot ETF)}{(1 - S_{11A} \cdot ESF) \cdot (1 - S_{22A} \cdot ELF) - ESF \cdot S_{21A} \cdot S_{12A} \cdot ELF} + EXF \quad (3.5)$$

An important point here is that the measured S-parameters depend upon all four actual DUT S-parameters. For the reverse stimulation, a similar analysis can be done for the lower portion of Figure 3.1, where we use primed marks for the reverse wave terms.

This results in expressions for the measured reverse S-parameters S_{12} and S_{22} .

$$S_{12M} = \frac{b'_{1M}}{a'_{2M}} = \frac{(S_{12A} \cdot ETR)}{(1 - S_{11A} \cdot ELR) \cdot (1 - S_{22A} \cdot ESR) - ESR \cdot S_{21A} \cdot S_{12A} \cdot ELR} + EXR \\ S_{22M} = \frac{b'_{2M}}{a'_{2M}} = EDR + \frac{ERR \left(S_{22A} + \frac{S_{21}ELR \cdot S_{12A}}{(1 - S_{11A} \cdot ELR)} \right)}{\left[1 - ESR \cdot \left(S_{22A} + \frac{S_{21}ELR \cdot S_{12A}}{(1 - S_{11A} \cdot ELR)} \right) \right]} \quad (3.6)$$

Again, these reverse parameters depend upon the reverse error terms and all four actual S-parameters. Between Eqs. (3.5) and (3.6), there are four equations, sufficient to solve for the four unknown actual S-parameters, provided that the error terms are all known.

The solution to the actual S-parameters is

$$S_{11A} = \frac{S_{11N} \cdot (1 + S_{22N} \cdot ESR) - ELF \cdot S_{21N} \cdot S_{12N}}{(1 + S_{11N} \cdot ESF)(1 + S_{22N} \cdot ESR) - ELF \cdot ELR \cdot S_{21N} \cdot S_{12N}} \\ S_{21A} = \frac{S_{21N} \cdot (1 + S_{22N} \cdot [ESR - ELF])}{(1 + S_{11N} \cdot ESF)(1 + S_{22N} \cdot ESR) - ELF \cdot ELR \cdot S_{21N} \cdot S_{12N}} \\ S_{12A} = \frac{S_{12N} \cdot (1 + S_{11N} \cdot [ESF - ELR])}{(1 + S_{11N} \cdot ESF)(1 + S_{22N} \cdot ESR) - ELF \cdot ELR \cdot S_{21N} \cdot S_{12N}} \\ S_{22A} = \frac{S_{22N} \cdot (1 + S_{11N} \cdot ESF) - ELR \cdot S_{21N} \cdot S_{12N}}{(1 + S_{11N} \cdot ESF)(1 + S_{22N} \cdot ESR) - ELF \cdot ELR \cdot S_{21N} \cdot S_{12N}} \quad (3.7)$$

where a normalized S-parameter is defined as $S_{11N} = \frac{S_{11M} - EDF}{ERF}$, $S_{21N} = \frac{S_{21M} - EXF}{ETF}$, $S_{12N} = \frac{S_{12M} - EXR}{ETR}$, $S_{22N} = \frac{S_{22M} - EDR}{ERR}$

This is a different form than given in other publications (Keysight Application Note 154 2014), in that it includes a normalized S-parameter that represents the measured S-parameter after subtracting the leakage terms and normalizing to the response tracking terms associated with the S-parameter. This form was anticipated in the description of error terms in Table 3.2. Here, directivity is recognized as the leakage term associated with reflection parameters, and crosstalk is the leakage associated with transmission terms.

This particular formatting of the solution gives good insight into some attributes of the error correction process. For example, if the VNA architecture is such that the forward source-match and reverse load match at any test port is the same, then the equations for S21A and S12A are simplified to the normalized S-parameter divided by a common term in all equations that represents the mismatch loop equations.

The application of error correction to raw measurements is generally well known and well understood. The difficulty ensues with determining the values of the error correction terms; the process that here is called *error correction acquisition*, or *cal-acquisition*, is discussed in the next section.

3.2.2 1-Port Error Model

For 1-port devices, it is not necessary to complete a 2-port calibration to obtain 1-port reflection measurements. The 1-port calibration application can be found from simplifying Eq. (3.7) to eliminate the two port terms, such that the 1-port correction utilizes only a single measured response and three error terms.

$$S_{11A} = \frac{\left(\frac{S_{11M} - EDF}{ERF} \right)}{\left[1 + \left(\frac{S_{11M} - EDF}{ERF} \right) \cdot ESF \right]} = \frac{(S_{11M} - EDF)}{[ERF + (S_{11M} - EDF) \cdot ESF]} \quad (3.8)$$

This simplified version of the 2-port response is sometimes called the Γ_1 correction as it provides the corrected reflection response of a 2-port network as it is currently terminated at port 2. This is useful in computing some aspects of other corrections, such as the applied power to the port, which depends only on the match apparent at the port and not on the S11 of the DUT. As a reminder, the S11 of the DUT is the reflection when all the other ports are terminated in a matched load; the Γ_1 of the port represents the reflection or impedance of the DUT port with whatever termination is applied at the other ports.

3.2.3 8-Term Error Model

The 8-term error model is distinguished from the 12-term model in that it requires measurements at all four test receivers: two incident waves and two scattered waves. The error model is shown in Figure 3.3.

The 8-term error model has the advantage that even though in practical implementation, the load match at each port changes depending upon whether the port is a source or a load, the 8-term model is not changed, as the change in load impedance is captured in changes in the incident waves at each port.

The 8-term error correction can be derived from the recognition that the measured S-parameters are a cascade of the input error box, the actual S-parameters, and the output error box. The cascaded T-parameters (see Section 2.4.3) give the measured T-parameters of the DUT in terms of the actual T-parameters of the DUT and the input and output error box T-parameters, as

$$\mathbf{T}_M = \mathbf{T}_X \cdot \mathbf{T}_{Act} \cdot \mathbf{T}_Y \quad (3.9)$$

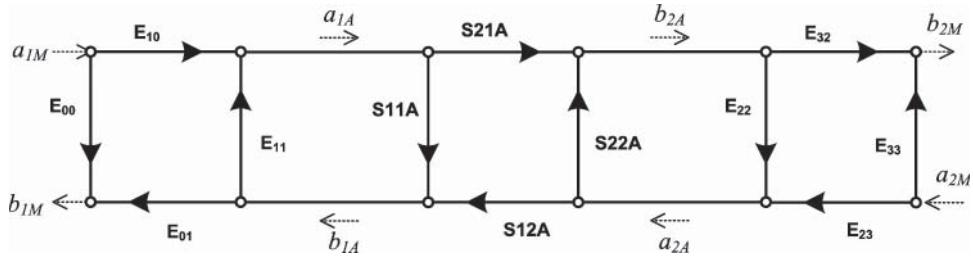


Figure 3.3 8-term error model, with four measured waves.

where X and Y represent the port 1 and port 2 error boxes. From the definition of T-parameters, the relationship between measured values and actual values can be written as

$$\begin{aligned} & \left[\begin{array}{cc} \frac{(S_{21M}S_{12M} - S_{11M}S_{22M})}{S_{21M}} & \frac{S_{11M}}{S_{21M}} \\ -\frac{S_{22M}}{S_{21M}} & \frac{1}{S_{21M}} \end{array} \right] \\ & = \frac{1}{E_{10}E_{32}} \begin{bmatrix} (E_{10}E_{01} - E_{00}E_{11}) & E_{00} \\ -E_{11} & 1 \end{bmatrix} \cdot \left[\begin{array}{cc} \frac{(S_{21A}S_{12A} - S_{11A}S_{22A})}{S_{21A}} & \frac{S_{11A}}{S_{21A}} \\ -\frac{S_{22A}}{S_{21A}} & \frac{1}{S_{21A}} \end{array} \right] \\ & \quad \cdot \begin{bmatrix} (E_{32}E_{23} - E_{33}E_{22}) & E_{22} \\ -E_{33} & 1 \end{bmatrix} \end{aligned} \quad (3.10)$$

The actual values can be computed by inverting the input and output T-matrixes as

$$\mathbf{T}_{Act} = \mathbf{T}_X^{-1} \cdot \mathbf{T}_M \cdot \mathbf{T}_Y^{-1} \quad (3.11)$$

with S-parameters determined by transforming the T-matrix back to the S-matrix. From the form of Eq. (3.10), it is clear that the measured value of S_{21} must be non-zero; similarly, the transmission tracking term, represented by $E_{10}E_{32}$, must be non-zero. Another aspect of the 8-term error model is recognition that only seven independent terms are needed to represent the model. The values E_{00} , E_{11} , E_{22} , and E_{33} represent four of the independent values. The values of E_{10} , E_{01} , E_{23} , and E_{32} always appear in the error correction in Eq. (3.11) as products, $E_{10}E_{01}$, $E_{10}E_{32}$, and $E_{32}E_{23}$; thus, the four terms represent only three independent values. A key aspect of the 8-term error model is that all four wave-terms are utilized in its determination; the 12-term error model requires measurements of only three of the four waves (one incident and two scattered).

The 8-term error model is normally transformed into the 12-term error model before it is applied, by breaking it into two 6-term models, with appropriate handling of change in load at port 2 during the forward sweep and at port 1 during the reverse sweep. From Figure 3.1 the 1-port terms can be equated as

$$\begin{aligned} EDF &= E_{00}, & ERF &= E_{10}E_{01}, & ESF &= E_{11} \\ EDR &= E_{33}, & ERR &= E_{32}E_{23}, & ESR &= E_{22} \end{aligned} \quad (3.12)$$

The other terms require a more complex evaluation. A key recognition here is that the effect caused by the port terminated as a load can be modeled as an additional error term at each port, which represents the ratio of incident to scattered waves at the measurement port. The terms in the forward and reverse direction can be defined as Γ_F and Γ_R , which are called *switch terms* and represent the change in match due to the source switch. These terms do not change with changes of external components, cables, or connections, and thus are entirely internal to the VNA. In most cases, these terms are very stable, and once determined in some manner, they can be utilized in other computations that rely on the 8-term model, replacing the measurement of the incident wave at the terminated port with a computation. A typical application is to use a 12-term model in a well-known connector type to find the switch terms and then utilize the switch terms to determine the 8-term models in different connector types such as on-wafer probes. The switch terms are defined as

$$\Gamma_F = \left. \frac{a_{2M}}{b_{2M}} \right|_{\text{Source at Port1}} \quad (3.13)$$

$$\Gamma_R = \left. \frac{a_{1M}}{b_{1M}} \right|_{\text{Source at Port2}} \quad (3.14)$$

from which the forward transmission tracking and load match terms can be determined as

$$ETF = \frac{E_{10}E_{32}}{1 - E_{33}\Gamma_F}, \quad ELF = E_{22} + \frac{E_{32}E_{23}\Gamma_F}{1 - E_{33}\Gamma_F} \quad (3.15)$$

The reverse transmission tracking and load match terms can be determined as

$$ETR = \frac{E_{10}E_{32}}{1 - E_{00}\Gamma_R}, \quad ELR = E_{11} + \frac{E_{10}E_{01}\Gamma_R}{1 - E_{00}\Gamma_R} \quad (3.16)$$

Similarly, the 8-term error model, plus the switch terms, can be computed from the 12-term model as

$$\begin{aligned} E_{00} &= EDF, & E_{11} &= ESF \\ E_{33} &= EDR, & E_{22} &= ESR \end{aligned} \quad (3.17)$$

$$\Gamma_F = \frac{E_{LF} - E_{SR}}{E_{RR} + E_{DR}(E_{LF} - E_{SR})}, \quad \Gamma_R = \frac{E_{LR} - E_{SF}}{E_{RF} + E_{DF}(E_{LR} - E_{SF})} \quad (3.18)$$

$$\frac{E_{23}}{E_{10}} = \frac{E_{TR}}{E_{RF} + E_{DF}(E_{LR} - E_{SF})} \quad (3.19)$$

$$E_{01}E_{10} = E_{RF}, \quad E_{32}E_{10} = \frac{E_{RR}E_{TF}}{E_{RR} + E_{DR}(E_{LF} - E_{SR})} \quad (3.20)$$

In essence, the 12-term error model has an explicit description for both the source-match at a port (when that port is the active source) and the load match at that same port (when it is not the active source). These matches, the switch terms of the VNA, are not the same. The 8-term error model presumes that the source-match and load match are the same and employs an additional measurement to characterize the difference.

3.3 Determining Error Terms: Cal-Acquisition for 12-Term Models

While the correction math is relatively straightforward, determining the error terms for corrections can be anything from trivially easy, for example using an electronic calibration module on a coaxial connection, to extremely difficult, as in the case of measurements at cryogenic temperatures, at power level extremes, or in cases with unusual DUT connections. In general, there are two distinct calibration methodologies based on either the 12-term error model or the 8-term error model.

The measurement requirements for cal-acquisition depend upon the number of required error terms. Essentially, one independent measurement is required for each error term. A device that has exactly known properties is called a *calibration standard*, or cal-std, and is used to generate one or more independent measurements. Depending upon the VNA architecture, more than one independent measurement may be acquired for each sweep, using a variety of measurement receivers to acquire the independent measurements on the same cal-std. For example, a 1-port calibration standard (such as an open or a load) creates one independent reflection measurement. It may also be used to create independent transmission measurements, such as a measurement of cross talk.

Calibration-acquisition for the 12-term model requires six independent measurements for each direction, as there are no common error terms between the forward and reverse directions. The crosstalk measurements are most commonly omitted, as the crosstalk of modern VNAs is below the noise floor of the measurement system, except in special cases, and so any measurements would just be a measurement of noise and impair the resulting correction rather than enhance it. Thus, the 12-term model requires just 10 terms and 10 independent measurements.

The standards for calibration come in two distinct types known as mechanical calibration standards and electronic calibration, or Ecal. Mechanical standards are physical representations of opens, shorts, loads, and in some cases thru (note, for purposes of clarity, the word *thru* is used to describe a calibration kit through-standard; this is historically how the standard appears in the VNA menu). These are usually sold together as a set and form a mechanical calibration kit or, *cal-kit* as it is commonly called. Electronic calibration kits have built-in switchable standards that provide a similar function to the open/short/load standards, as well as a through state. More details on electronic calibration are described in Section 3.4.6.

The most common form of cal-acquisition makes three measurements of 1-port standards at each port, and two measurements of a known thru standard, for 10 total measurements. Here we differentiate between a “known thru” standard, for which all four S-parameters are well known, and an “unknown thru,” which has unknown characteristics but follows the rule that $S_{21} = S_{12}$. It is sometimes called a *reciprocal Thru*. Detailed information is given in Section 3.3.3.

3.3.1 1-Port Error Terms

The 1-port error terms are most commonly found using an open/short/load calibration. For ease of comprehension, consider first the case where the reflection standards do provide the reflection of an ideal open, ($\Gamma_{Open} = 1$); an ideal short, ($\Gamma_{Short} = -1$); and an ideal

load, ($\Gamma_{Load} = 0$). The measured S11 of each can be written in terms of the actual reflection coefficient and the error terms.

$$\begin{aligned} S_{11M}^{Ideal_Open} &= EDF + \frac{ERF(1)}{[1 - ESF(1)]} \\ S_{11M}^{Ideal_Short} &= EDF + \frac{ERF(-1)}{[1 - ESF(-1)]} \\ S_{11M}^{Ideal_Load} &= EDF + \frac{ERF(0)}{[1 - ESF(0)]} \end{aligned} \quad (3.21)$$

From this, the 1-port error terms are easily computed as

$$EDF = S_{11M}^{Ideal_Load} \quad (3.22)$$

$$ESF = \frac{\left(S_{11M}^{Ideal_Open} + S_{11M}^{Ideal_Short} - 2EDF \right)}{\left(S_{11M}^{Ideal_Open} - S_{11M}^{Ideal_Short} \right)} \quad (3.23)$$

$$ERF = \frac{-2 \left(S_{11M}^{Ideal_Open} - EDF \right) \left(S_{11M}^{Ideal_Short} - EDF \right)}{\left(S_{11M}^{Ideal_Open} - S_{11M}^{Ideal_Short} \right)} \quad (3.24)$$

Of course, these equations apply only to ideal opens, shorts, and loads, but are useful in understanding the structure of the error terms. In particular, if one considers a system where the raw reflection tracking is 1 and evaluates the EDF term and ESF term graphically, one can show the relationships directly. In essence, the EDF term is just the load response. Figure 3.4 (left, upper) shows the load response as a vector diagram, where the measured load is equal to the EDF term, and with the ideal load ($\Gamma_L = 0$) at the center of the chart, one finds $\Gamma_{LM} = EDF$, thus $\Gamma_{LM} - EDF = \Gamma_L = 0$. Figure 3.4 (left, lower) shows the construction of the measured short, and Figure 3.4 (right, lower) the measured open. The upper right shows the construction of source match determination as $Open + Short - 2EDF = 2 ESF$. In practice, the ripple seen on a log magnitude (dB) trace of an open or a short represents the voltage standing-wave-ratio (VSWR) of the source-match and the directivity of the port, and the mean value of the trace represents the reflection tracking.

In general, three known reflections are used, but they are not assumed ideal. There are models for each reflection standard built into the VNA firmware, and these models provide a way to modify the ideal standards to account for real-world errors such as loss or open-end fringing capacitance.

The generalized 3-term solution (Rytting 1996a) for the error terms is

$$\begin{bmatrix} EDF \\ ERF - EDF \cdot ESF \\ ESF \end{bmatrix} = \begin{bmatrix} 1 & \Gamma_{AO} & \Gamma_{AO} \cdot \Gamma_{MO} \\ 1 & \Gamma_{AS} & \Gamma_{AS} \cdot \Gamma_{MS} \\ 1 & \Gamma_{AL} & \Gamma_{AL} \cdot \Gamma_{ML} \end{bmatrix}^{-1} \begin{bmatrix} \Gamma_{MO} \\ \Gamma_{MS} \\ \Gamma_{ML} \end{bmatrix} \quad (3.25)$$

where Γ_{AO} , Γ_{AS} , Γ_{AL} are the presumed actual reflection coefficients of the three standards (essentially, the models of the standards), and Γ_{MO} , Γ_{MS} , Γ_{ML} are the measured values of the standards. While the labels imply open/short/load, it is understood that the standards can take on any values.

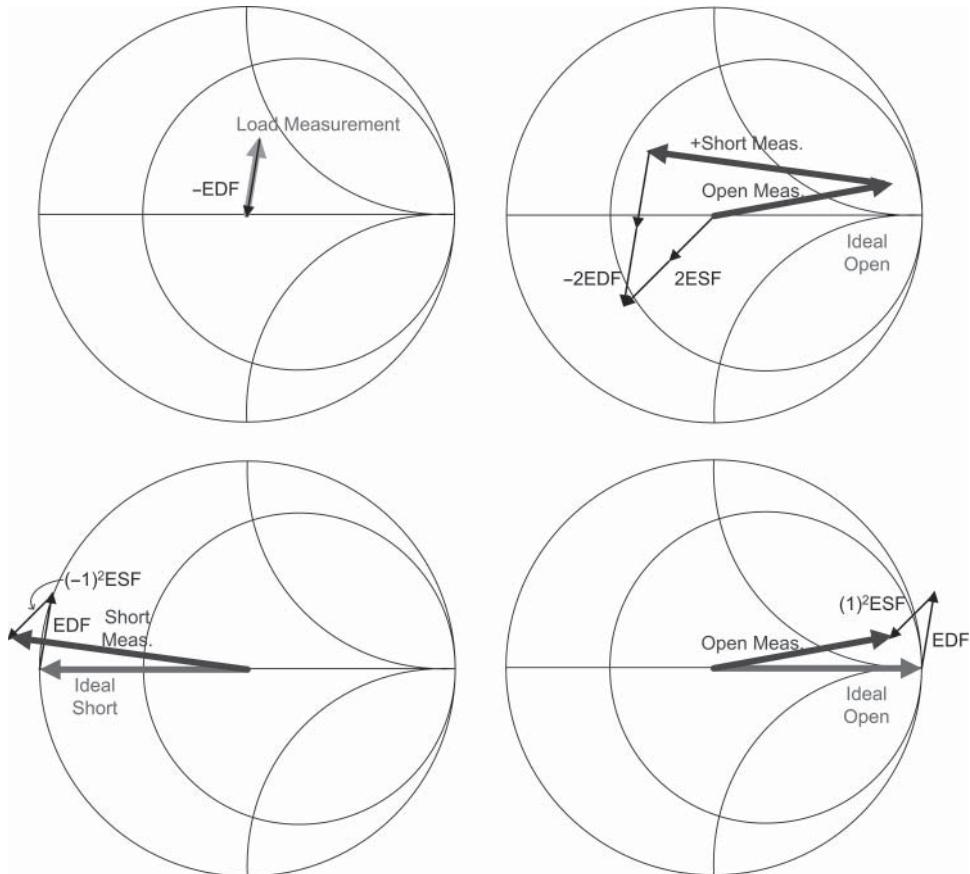


Figure 3.4 Determining the error terms graphically for open/short/load responses.

3.3.2 1-Port Standards

From Eq. (3.25), it is clear that any three known-reflections can be used for 1-port calibrations. Opens, shorts, and loads are most commonly used as they provide good separation of variables. While it's not possible to create standards that match the ideal assumptions of Eq. (3.21), at RF frequencies, the standards are quite close to ideal. Most 1-port standards are in the form of a simple circuit model for the standard plus an input transmission line in the form of an offset delay and loss.

3.3.2.1 Open Standards

First things first: there is *no such thing* as an ideal open standard. All opens have fringing capacitance, and virtually all opens have some offset length (certainly all commercial cal-kits have only offset-opens, even if they are not called that). Therefore, you will *never* see a dot

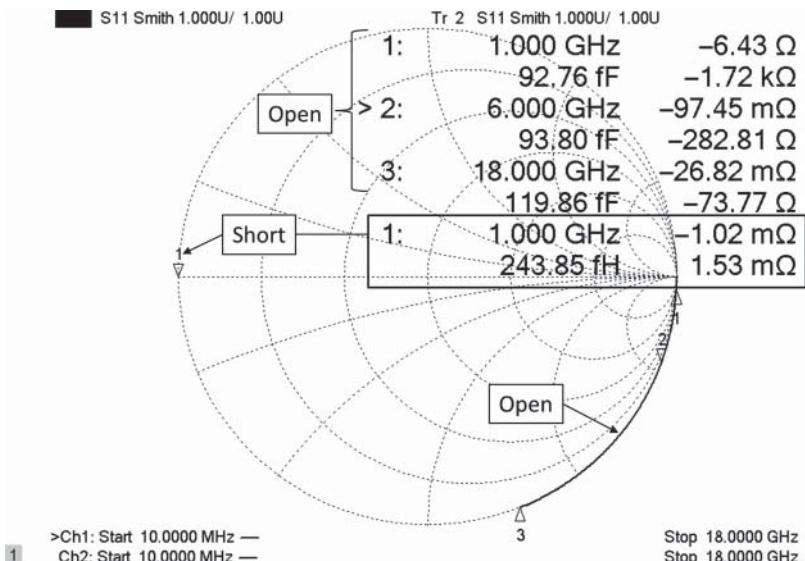


Figure 3.5 The correct re-measurement of an open and short after calibration; the arc is caused by the excess fringing capacitance on the open, while a nearly flush short may be possible at lower frequencies.

at the edge of the Smith chart at $Z = \infty$; it will *always* be an arc due to the offset open, as demonstrated in Figure 3.5.

This measurement is for a 7 mm connector, which is one of the only flush connectors commonly available. As a flush connector, it has no offset line before the standard. Every other coaxial connector type has an offset delay in both the open and the short. The offset delay produces greater than 360° of phase shift for connectors such as 3.5 mm. In the figure, the open fringing capacitance is not constant, as the model has increasing capacitance as frequency increases. The short has a tiny amount of series inductance; from the model, this amount is less than 1 pH.

The classic model for the open circuit is shown in Figure 3.6. The principal deviations from ideal is due to a change in position of the open, rotated from the reference plane by a short length of transmission line, and fringing capacitance from the end of the open.

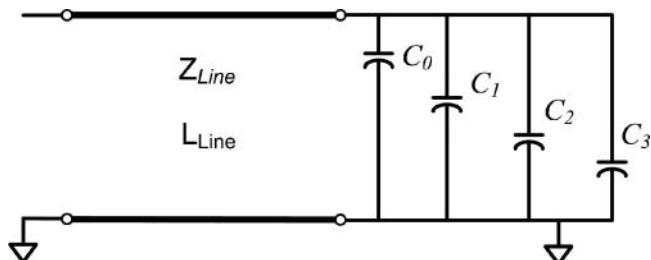


Figure 3.6 Model for an open circuit.

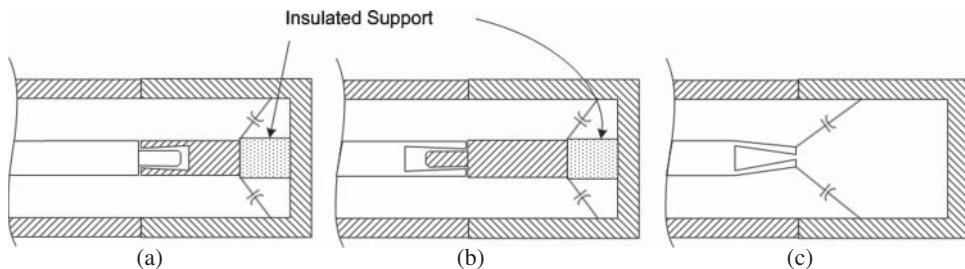


Figure 3.7 Physical construction of (a) female open, (b) male open with extender pin, (c) male open without extension.

The fringing capacitance of the open end is illustrated by the small capacitors in Figure 3.7. In most instances, the male test port (female standard) uses a center-pin extender that produces a constant diameter to match the system impedance (Figure 3.7a).

Throughout this book, the definition of the sex of a cal-std is determined by the standard and *not* the test port. In some older analyzers, the connector sex referred to the test port, but it is now common practice to refer to the connector of the DUT or the cal-std being used.

The center-pin extender produces phase shift associated with the length of the line. Fringing capacitance adds additional phase shift and is typically modeled as a capacitive versus frequency polynomial in the form of

$$C(f) = C_0 + C_1 f + C_2 f^2 + C_3 f^3 \quad (3.26)$$

The values for the capacitive coefficients are fitted to the performance of the open, and the actual extracted values may be different for RF versus microwave versions of the same standard. The polynomial fit can be adjusted to get better performance at low frequencies for the RF kit or fitted to a broader overall response for a microwave kit. Newer techniques allow for essentially any arbitrary model of any standard, using so-called data-based methods.

Figure 3.7b shows a female test port with a male center-pin extender. The extenders are typically attached to the back of the standard with an insulated support from Lexan or other plastic material. Figure 3.7c shows a female test port with a shielded open, but no center-pin extender. This is common on RF connectors for low-cost kits; for example, non-metrology Type-N calibration kits typically don't have a center-pin extender for the female test port (male standard). It is still necessary to provide a shield or outer conductor to avoid radiation from the center pin. The fringing capacitance in this case is less certain than that of the solid cylinder that is formed by the center-pin extension. One reason for the uncertainty is that the test port center-pin female fingers may have a differing amount of gap depending upon construction and the size of male pins previously used. For the male test port (female standard), almost all calibration kits include a test port extender to keep the outer diameter of the male pin constant. On older cal-kits, the male extension pin was a separate piece and often lost from the cal-kit; unfortunately, its use is critical for good calibrations.

A common problem seen with open standards is radiation of the open if the center pin is not shielded. For this reason, all precision calibration kits use a shield on the open circuit, but open circuits for fixtures, probes, and adapters are sometimes used when a no-precision kit

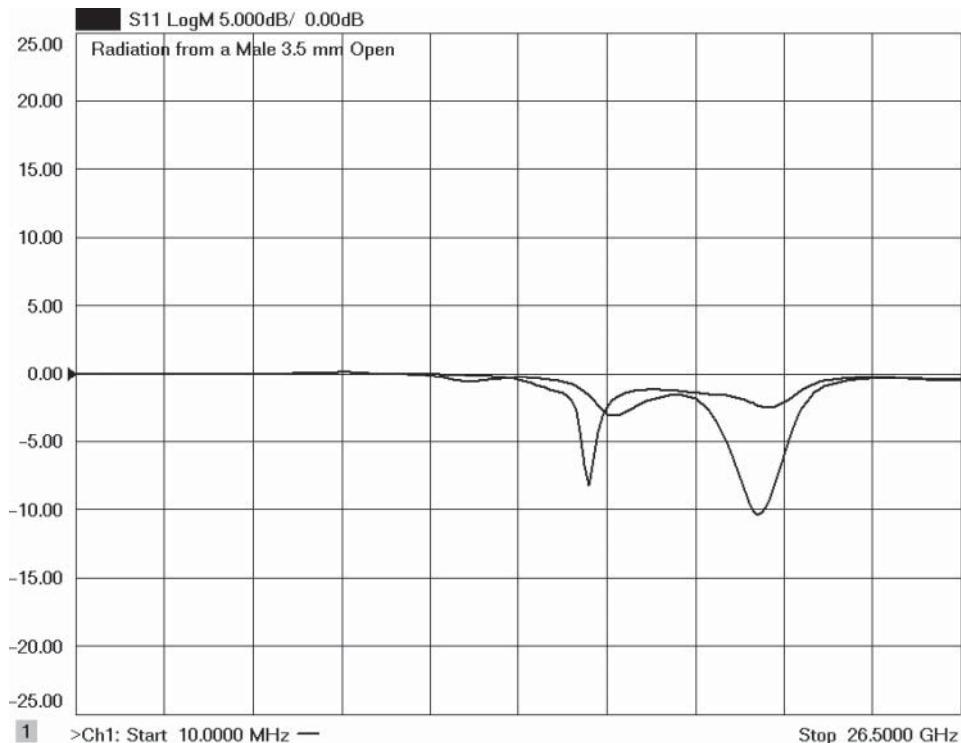


Figure 3.8 Variations in the open reflection coefficient due to radiation for an unshielded male SMA.

is available. Figure 3.8 shows two traces of a measurement of an SMA-male-open test port. The change in response is due to touching the outer coupling nut. The dip in the S11 response shows the frequency where the unshielded SMA-open is radiating.

3.3.2.2 Short Standards

Short standards are almost always more ideal than open standards, because they form nearly ideal reflections at the shorting plane. The typical model for a short standard is shown in Figure 3.9. The inductance model for a short circuit is in the form of an inductance versus frequency polynomial as

$$L(f) = L_0 + L_1 f + L_2 f^2 + L_3 f^3 \quad (3.27)$$

For older RF models, the values of all the inductance terms are zero, and the only non-ideal aspect is the transmission line model that provides the offset delay, but newer models have standardized on Eq. (3.27). Microwave models include inductance values, but they are very tiny indeed, and the deviation from ideal short is much smaller than that of an open circuit, as illustrated in Figure 3.5.

The short circuit standards are the simplest of mechanical standards and typically consist of just the center pin connected to ground, as shown in Figure 3.10. While there is no requirement

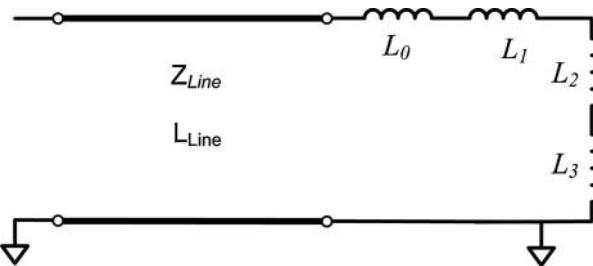


Figure 3.9 Model for a short standard.

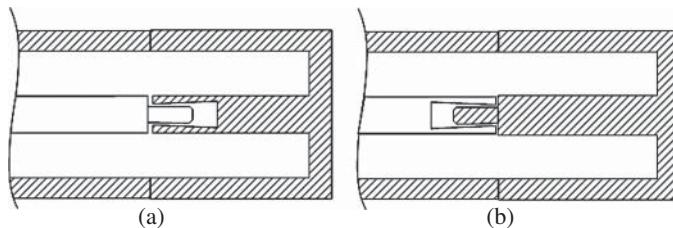


Figure 3.10 Short circuit standards (a) male test port, (b) female test port.

for the length of the short to match the length of the open, it is best to make the length just slightly longer so that phase shift versus frequency of the short matches that of the open (which has some excess phase due to the fringing capacitance). Ideally, the opens and shorts should maintain 180° of phase separate throughout the frequency range of their operation. The short standard typically has the smallest residual errors. In a band-limited calibration, shorts of different delays, called offset-shorts, may be used to provide all of the three calibration standards. Calibrations using offset-shorts are often used in metrology applications where the diameter and length of the short circuit standard can be characterized with very fine precision and the impedance and delay computed with minimal errors. These are particularly useful in high-frequency mm-wave applications as it is difficult to create a well-matched fixed load.

3.3.2.3 Load Standards: Fixed Loads

The load standard is usually the most difficult to produce, and errors associated with it may increase dramatically with frequency. The load element is formed by terminating a coaxial standard into a resistive element, often a thin-film circuit with a patch of tantalum-nitride designed to provide a constant impedance with frequency (Figure 3.11).

The typical model for a load element contains only a resistance and a delay line, as shown in Figure 3.12a. The value of the resistance can be set independently from the system Z_0 , but it is typically set to be Z_0 . The value of the transmission line impedance is also typically set to Z_0 . An alternative model for a load element is a series R-L circuit, as shown in Figure 3.12b. This is a common situation found in on-wafer load standards, and the determination of the value of inductance requires additional information during the calibration. A method called LRRM (Davidson et al. 1990) calibration provides this additional information by correlating

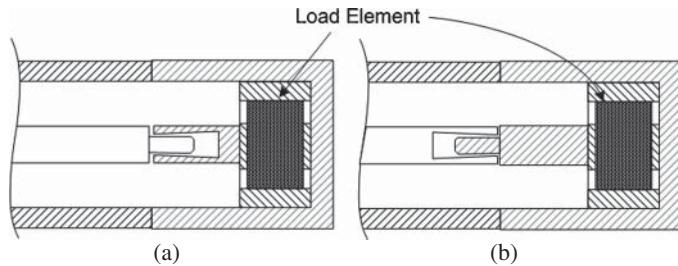


Figure 3.11 Load elements (a) male test port, (b) female test port.

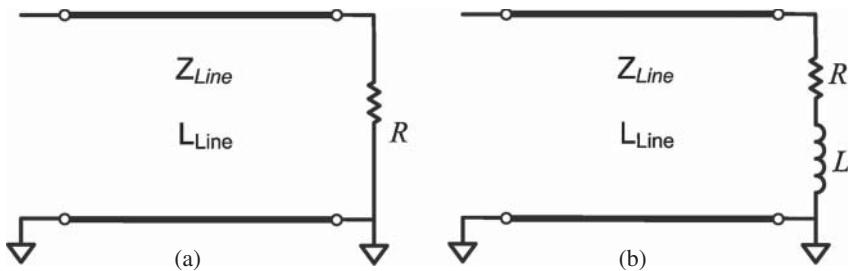


Figure 3.12 (a) Typical model for a load standard, (b) model for a load showing series inductance.

the load response information across several frequencies to determine the values for the fixed R and L values. Since most VNAs do not provide model entries for the R-L combination, the effect of the inductance is generated by setting the impedance of the load delay line to a very high value (typically 500Ω is the maximum allowed in many VNAs), and the length of the line is adjusted to give the equivalent phase shift of the effective inductance.

More details on modeling standards and determining parasitic values are discussed in Chapter 11.

3.3.2.4 Load Standards: Sliding Loads

The sliding load, which should more properly be called a *sliding mismatch*, is constructed from a length of precision airline followed by a moderately good termination, as shown diagrammatically in Figure 3.13. The center conductor of the airline portion is typically created in such a way that it can slide into place while the outer conductor is not yet mated to allow a bead-less connection. The load element is typically not a resistive element but is more commonly a tapered bead of lossy material that essentially makes the airline look like a lossy element. It is designed to have an impedance that is not quite 50Ω , normally in the range of 26–40 dB return loss. The far end of the airline is then clamped to the outer conductor at a precise connector gauge length so that any gap is only on the test port side. It is important to have this slight gap in the sliding load as all test ports must have a gap as well to avoid any interference issues with the center pin connection; if the sliding load is pressed tightly against the test port center pin, it will provide a better, less reflective contact. But when used for measurement, the port will have a gap to avoid interfering with the DUT. Thus, to properly

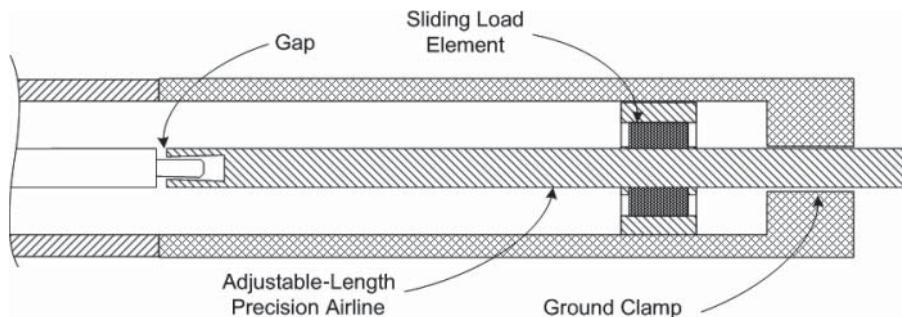


Figure 3.13 Representation of a sliding load.

characterize the test port, the sliding load must also provide a gap. The sliding load design allows the position of the center-conductor to be essential exactly at the reference plane. One might note that any DUT is likely to have a slightly greater gap depending upon its specific design to avoid any interference with other connections. That is, both the test port and any DUT normally have slightly recessed pins.

As the sliding load is moved, its apparent impedance rotates around the Smith chart, an example of which is shown at one frequency over the range of the slide in Figure 3.14. The upper plot has the Smith chart at full scale, and the lower plot shows the locus of points for the sliding load at a smaller scale. In fact, the sliding load actually presents a spiral due to the finite loss of the transmission line. For the sliding load standard, several data acquisitions (typically 5) are required, each at a different offset length for the slide. This forms a locus of points, the center of which is the impedance of the airline. The difference between this computed center and the actual center of the Smith chart determines the directivity error term, EDF.

The load element is deliberately set to be not perfectly matched so that the diameter of the circle created by the locus of load points is sufficiently large such that trace noise of the slide set position does not degrade from the computation of the center of the circle.

Some special care is needed with slide load calibrations. The quality of the airline portion of the sliding load sets the quality of the calibration. Since the length of the line is limited and the lossy element works well only above RF frequencies, the sliding load standard is used only above RF frequencies. While the actual operating frequency is determined in the cal-kit definition, the typical start point for a sliding load is between 2 and 3 GHz. Below these frequencies, a fixed load must be used.

When using a sliding load, it is recommended to change the slide positions in non-even steps, to avoid creating a periodic situation where at some frequencies the load response falls on the same phase as an earlier step. Many sliding loads come with logarithmically spaced marks along the outer conductor to provide a reference place for setting the slides. Usually, five or more slides are required to compute the center of the circle, and often the VNA code will allow more slides if desired. It is best practice to move the slide in only one direction, which minimizes stability errors. Because the quality of the airline impedance can be maintained at high frequencies, the sliding load calibration provides much better extraction of the directivity term, EDF, and therefore, also provides better source-match determination, ESF, as implied by Eq. (3.24). To the first order, the reflection tracking term is not affected much by the EDF term.

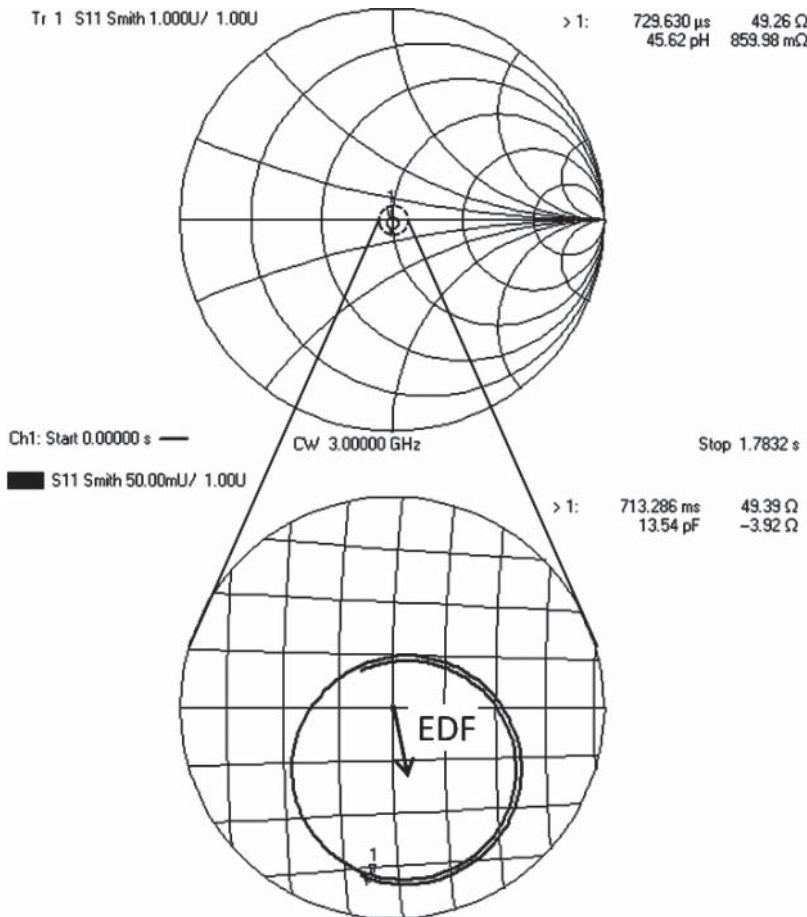


Figure 3.14 Smith chart measurement of a sliding load at a single frequency, while changing the slide position.

3.3.3 2-Port Error Terms

The 2-port, or multiport, error terms describe terms that come into play on transmission measurements, and for 2-port calibrations these terms are the transmission-tracking terms ETF and ETR, the load-match terms ELF and ELR, and the isolation terms EXF and EXR. The transmission terms are computed by measuring thru standards, and the isolation terms are computed from measurements of S21 and S12 with reflection standards on the test ports.

3.3.3.1 Isolation Standards

The isolation standard for the 12-term model is typically measured using load standards. From Eq. (3.5), the S21 and S12 measurements with loads at port 1 and port 2 have actual values

of $S_{21}^{LoadsA} = 0$, $S_{12}^{LoadsA} = 0$, resulting in

$$S_{21}^{LoadsM} = EXF, S_{12}^{LoadsM} = EXR \quad (3.28)$$

That is, the isolation or crosstalk error term is simply the S21 or S12 measurement with the load attached. Usually, substantial averaging must be used or the measurement is simply the noise floor of the VNA. In some cases, it may be advantageous to use the DUT (terminated by a load) to terminate the test port during an isolation calibration. In this way, any leakage from the reflection port (e.g. b1) due to mismatch of the DUT (as found in the stop band of a filter) may be characterized and removed.

3.3.3.2 Thru Standards: Flush Thru

The thru standard should most properly be called a “defined thru” standard and represents a 2-port standard for which all the S-parameters are known. There are two main forms of defined thru standards, sometimes called *flush thru* and *defined thru*.

The flush thru standard consists simply of the process of mating a pair of connectors together. For sexless connectors, such as the precision 7 mm, the flush thru connection is always available if the physical test ports can be moved in such a way as to be directly connected. For sexed connectors, such as the 3.5 mm connector, the flush thru is available only for a port pair that includes one male and one female port. The S-parameters of the flush thru are also simply stated with $S_{21} = S_{12} = 1$, $S_{11} = S_{22} = 0$. From Eq. (3.5), the measured values of the flush thru become

$$\begin{aligned} S_{11}^{FlushThruM} &= EDF + \frac{ERF \cdot ELF}{(1 - ESF \cdot ELF)}, & S_{22}^{FlushThruM} &= EDR + \frac{ERR \cdot ELR}{(1 - ESR \cdot ELR)} \\ S_{21}^{FlushThruM} &= \frac{ETF}{(1 - ESF \cdot ELF)} + EXF, & S_{12}^{FlushThruM} &= \frac{ETR}{(1 - ESR \cdot ELR)} + EXR \end{aligned} \quad (3.29)$$

from which the values of ETF and ELF can be determined, if the 1-port terms and crosstalk terms have already been determined, as

$$\begin{aligned} ELF &= \frac{(S_{11}^{FlushThruM} - EDF)}{[ERF + ESF(S_{11}^{FlushThruM} - EDF)]}, & ELR &= \frac{(S_{22}^{FlushThruM} - EDR)}{[ERR + ESR(S_{22}^{FlushThruM} - EDR)]} \\ ETF &= (S_{21}^{FlushThruM} - EXF)(1 - ESF \cdot ELF), & ETR &= (S_{12}^{FlushThruM} - EXR)(1 - ESR \cdot ELR) \end{aligned} \quad (3.30)$$

From this it is clear that the load match term is determined from the flush thru by applying a 1-port error correction to the measured match of the flush thru. And the transmission tracking is similarly a measurement of the thru response, after subtracting the crosstalk and accounting for the source and load match interaction.

3.3.3.3 Thru Standards: Non-insertable Thru

Whenever a pair of test ports uses the same-sex coaxial test connector, such as for a DUT with SMA female connectors on each port, which is quite common in RF work, the calibration is

known as the *non-insertable* case. The most typical case is a pair of male test ports that require a female-to-female thru. Unfortunately, built-in calibration kits for most VNAs have definitions only for the flush, male-to-female thru, and so do not have thru standards defined for male-to-male or female-to-female DUT port pairs. Failure to account for the delay of the thru (and to a lesser extent, the loss of the thru) is one of the most common causes of error in RF measurements. If a non-zero-length thru is used during calibration, the effective measured load match will be phase shifted by the actual delay of the thru and can cause substantial error.

Figure 3.15 shows a measurement of a 15 cm airline where a 2 cm female-to-female thru was used in a 12-term calibration, and the delay and loss of the thru was ignored. The ripple seen is due to the error in the measurement of the load match beating against the raw source match of the VNA; this error can cause the corrected or residual load match to be larger than the *raw* load match, by as much as 6 dB. The other, more insidious error is the fact that the loss of the thru is not accounted for, so the loss of the DUT appears slightly less than it really is. In the case of a low-loss filter, this error allows a DUT to pass a test line limit even if the actual loss is out of spec.

In modern VNAs, this problem is seldom encountered because the unknown thru method (see Section 3.4.3) provides for highly accurate calibrations in the non-insertable case, as shown in the lower trace (the trace with Marker 2 active) of Figure 3.15. For modern instruments, the unknown thru method is the default method.

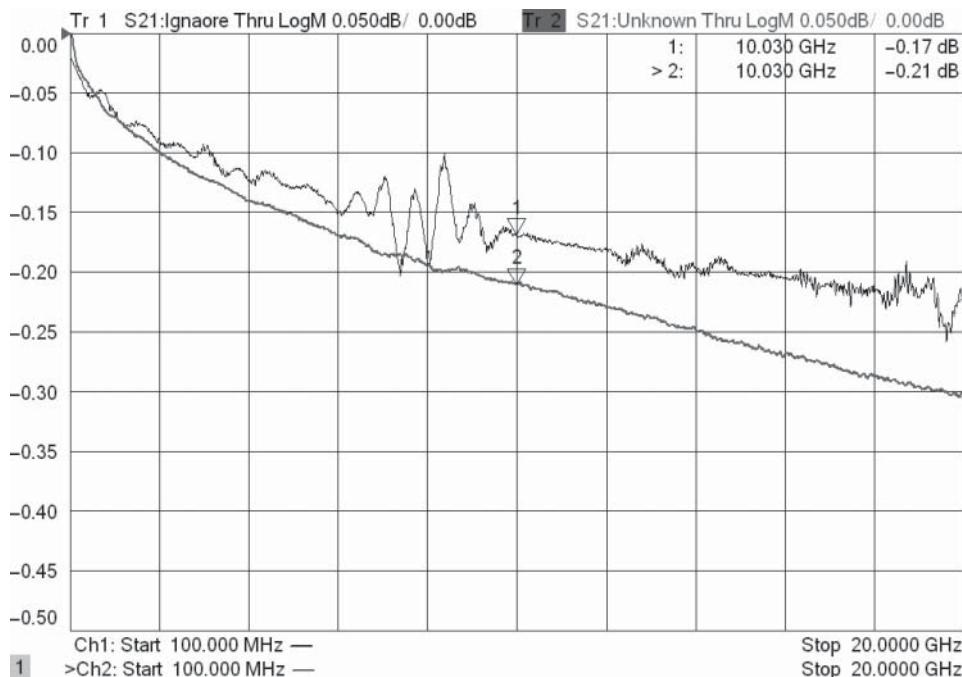


Figure 3.15 Error due to ignoring the length of a non-insertable Thru, compared with the Unknown Thru calibration.

3.3.3.4 Swap-Equal Adapter

In the earliest days of modern VNAs, a common method for handling the non-insertable case was to add an additional male-to-female adapter to the test port. One could purchase a matched set of adapters that included a male-to-female, male-to-male, and female-to-female thru adapter. These were designed to have good return loss and equal electrical length. One performed the calibration by adding an additional male-to-female adapter on one of the ports, typically port 2, during the 1-port calibration steps. Then during the thru step, the adapter was swapped for one that would mate with port 1, most commonly a female-to-female adapter. After the calibration was complete, the original male-to-female was replaced on the test port, and fully calibrated measurements could be performed. While this was the typical way a calibration preceded, a better choice is to do the thru step of the calibration first so that there is no need to disconnect and re-connect the male-to-female adapter between the 1-port calibration and the thru measurement. This reduces the error due to connector repeatability. In modern VNAs, the use of the unknown-thru (UT) calibration eliminates the need for swapping equal adapters.

3.3.3.5 Defined Thru

In most VNAs it is a simple task to add a defined thru standard, set the connector sex to mate to the sex of the DUT connectors, and then complete the calibration using this defined thru standard instead of the flush thru standard. This is perhaps the simplest way of dealing with the simple non-insertable case.

The offset delay and offset loss must be determined for the connector, but the offset loss for a family of connectors is essentially the same for every standard and is typically listed as part of the offset delay for the open or short standards in the calibration kit. For 7 mm connectors, it is about $700 \text{ M}\Omega \text{ s}^{-1}$, and for 3.5 mm connectors it is $2.36 \text{ G}\Omega \text{ s}^{-1}$. The most common way for determining the offset loss is to compute it from the loss at 1 GHz, from the formula

$$L_{Offset(\text{Gohms/sec})} = \left(\frac{\ln(10)}{10} \right) \cdot \left(\frac{S_{21(dB_loss)}}{\sqrt{f_{(GHz)}}} \right) \cdot \left(\frac{Z_{Offset}}{S_{21(delay_in_ns)}} \right) \quad (3.31)$$

where the frequency is given in GHz, the loss of S21 in dB (as a positive number), and the resulting value is in $\text{G}\Omega/\text{s}$. This is the loss for a coaxial line; the loss for waveguide is computed using different models and must be included particularly for mm-wave or higher frequencies.

For a measurement of a matched (50Ω) offset line at 1 GHz, this simplifies to

$$L_{\text{Gohms/sec}} = \frac{11.5 \cdot S_{21(dB_loss_1GHz)}}{S_{21(delay_in_ns)}} \quad (3.32)$$

Take, for example, the mated-pair of connectors used in Figure 1.24. These have a loss of 0.24 dB at 26.5 GHz and a delay of about 0.22 ns for the pair. From Eq. (3.32) this computes as

$$L_{\text{Gohms/sec}} = \left(\frac{\ln(10)}{10} \right) \cdot \left(\frac{0.24}{\sqrt{26.5}} \right) \cdot \left(\frac{50}{.22} \right) = 2.44 \text{ G}\Omega/\text{sec} \quad (3.33)$$

which is close to the expected value of $2.36 \text{ G}\Omega \text{ s}^{-1}$.

For a defined thru, the mathematics are essentially identical to (3.30), with the modification of replacing the one port error terms with modified versions that include the de-embedded effects of the thru characteristics as described by Section 3.5. That is, the ELF and ETF terms are computed by measuring the defined thru and applying Eq. (3.30) but first modifying the values of the EDF, ESF, and ERF terms by way of de-embedding the actual (modeled) defined thru S-parameter values from those error terms; a similar process applies to the reverse error terms. De-embedding techniques are discussed in detail in Chapter 11.

3.3.3.6 Adapter Removal Calibration

The non-insertable calibration case includes the common case of the same sex connector on each port of a DUT but also includes the case of different connector families on each port. In such a case, an adapter (with different connector types on each end) must be used to connect between the ports, and it is not easy to determine the delay of this adapter. This is also the case when making a measurement on a DUT with one waveguide and one coax port. To resolve this problem, a method for adapter removal has been developed that typically requires two 2-port calibrations, each one on either side of the adapter. From the error terms of the two 2-port calibrations, the characteristics of the adapter are computed. Forward reflection error terms from the port 1 side of the adapter of the first calibration (performed at the flush thru mating of the port 1 side) are combined with the forward reflection error terms from the second calibration performed at the port 2 side of the connector to determine the S-parameters of the adapter as

$$\begin{aligned} S_{11} &= \frac{(EDF_2 - EDF_1)}{[ERF_1 + ESF_1 \cdot (EDF_2 - EDF_1)]} \\ S_{21} = S_{12} &= \frac{\sqrt{ERF_2 \cdot ERF_1}}{[ERF_1 + ESF_1 \cdot (EDF_2 - EDF_1)]} \\ S_{22} &= ESF_2 + \frac{ESF_1 \cdot ERF_2}{[ERF_1 + ESF_1(EDF_2 - EDF_1)]} \end{aligned} \quad (3.34)$$

where subscript 1 refers to the error terms of the calibration on the port 1 side of the adapter and subscript 2 refers to the error terms on the port 2 side of the adapter. In both cases, the error terms are the forward 1-port terms, so this method for determining the S-parameters of an adapter could be used with only one port calibrations just as well. In this computation, it is assumed that the adapter is reciprocal and passive. The computation of S21 contains a square root of a complex number. This will yield two values based on whether the positive or negative root is chosen. Put another way, the phase of the two results differs by 180°. For a single point computation, one cannot know the proper phase to choose. However, if one knows the approximate delay of the adapter, then the phase can be compared to the expected phase of the adapter for each frequency, and the phase value closest to the expected value is chosen. See Chapter 11 for more information about determining complex roots.

Once the S-parameters of the adapter are determined, it can be removed from the port 1 side of the second 2-port cal using de-embedding, as discussed in Chapter 11, to create a new 12-term calset that contains the 1-port terms from the first cal for port 1, the 1-port terms from the second calibration for port 2, and a modification of the transmission tracking and load match terms for both directions determined by the de-embedding math.

Most modern network analyzers allow adapter remove as an option for non-insertable calibrations, but it has been essentially replaced by the unknown thru calibration method.

3.3.4 12-Term to 11-Term Error Model

The 12-term error model shows 12 apparently independent error terms. But other analysis (Marks 1997) shows that there is an interdependency between the terms so that there are only 11 degrees of freedom; if we ignore crosstalk terms, there are only 9 degrees of freedom for the 10 error terms. This interdependency is represented by the relationship

$$ETF \cdot ETR = [ERR + EDR(ELF - ESR)] [ERF + EDF \cdot (ELR - ESF)] \quad (3.35)$$

Though calibration acquisitions don't normally take advantage of this result, it implies that only 9 measurements are required to obtain the 10 error terms.

3.4 Determining Error Terms: Cal-Acquisition for 8-Term Models

The 8-term error models are determined through a different series of measurements, some of which rely more on transmission standards and do not rely on the details of known reflection standards. As described in Section 3.2.3, the measurements require knowledge of both incident and scattered signals at each test port. Because of the nature of the 8-terms models, there is not necessarily explicit computation of the 1-port error terms, but rather they are computed as part of the solution to the 2-port error model. Normally, the effect of crosstalk is ignored.

3.4.1 TRL Standards and Raw Measurements

The standards used in thru-reflect-line (TRL) calibrations are a thru, a reflect, and a line. Another version of TRL is sometimes called LRL for line, reflect, line, but they are essentially the same, as most versions of TRL provide for non-zero-length thru standards.

TRL calibrations are often described as the most accurate form of calibration. This is principally true in the case of connectorized Coax and Waveguide calibrations, but TRL may not provide the best calibration for in-fixture calibrations. The reason that TRL calibration is considered most accurate is that the quality of calibration depends almost entirely on the precision of the knowledge of the line impedance, or, in the case of waveguide, in the quality (in terms of low reflection) of the line section.

TRL specifications only apply to metrology-grade calibration standards, such as bead-less short airlines, and which are seldom used in practice. These metrology kits are designed to be used with specialized insertable calibration lines, rather than normal thru lines. One can create a TRL calibration using some short high-quality adapters, but the impedance or S11 of the adapter will set the source and load match of the system (along with the quality of the open/short standards).

For in-fixture TRL calibrations, it is common to create the standards on separate PC boards; in such cases, differences in the coax to PC board launches (typically SMA) create errors in the calibration. Details of PC board calibration are discussed in Chapter 11. Important attributes of the TRL standards are described next.

While TRL calibrations are commonly called eight terms, it was shown earlier that there are only seven independent error terms. And since it is common to convert the 8-term model to the 12-term model for applying the error correction, 2 additional unknown error terms, Γ_F and Γ_R are required, making a total of 9 unknown error-terms. Thus, at least nine independent equations are required to solve the TRL calibration problem.

3.4.1.1 Thru Standard

The thru standard is the simplest standard and consists of simply a flush thru. As such, its reflection is zero, and its transmission is unity (1). Just as for defined thru calibrations in the 12-term model, the thru standard provides a well-known standard. For coaxial connectors, the thru is established by simply making a male-to-female connection, excepting of course the sexless connectors such as 7 mm.

In the case of waveguide connectors, the thru standard is simply a mating of the two waveguide flanges. The loss and delay of this thru, like the flush thru of coaxial connectors, are zero. One note for waveguide calibrations: the orientation of the thru, when mated, must be maintained when measuring a DUT or the phase will experience a 180° phase shift.

In the case of on-wafer calibrations, it is not possible to make a direct connection from probe to probe, so the thru must have some length. In this case, the calibration is more correctly called an LRL calibration. The LRL calibration has additional characteristics in that this first line might have some loss and some offset impedance. Normally, the impedance of the thru or the first line in LRL is presumed to be the system Z_0 .

The reference plane is typically set at the center of the thru standard. In the case of a zero-length thru, this is typically the best choice. Most TRL calibration acquisition algorithms provide for setting the reference also at the reflect standard. This is a good choice if the thru standard is of substantial length or if a flush short is used as the reflect standard.

During the thru standard measurement, the four raw S-parameters are measured, along with two additional parameters, which are sometimes called the *switch terms*. The switch terms are measurements of a_2/b_2 with the source at port 1, and measurements of a_1/b_1 with the source at port 2. Some VNAs have limitations to the specific ratios that can be measured, so the measurement of a_2/b_2 might be accomplished with a measurement of a_2/a_1 with the source at port 1, from which the a_2/b_2 measurement can be computed by realizing that the raw S_{21} measurement contains b_2/a_1 , which can be combined with the a_2/a_1 to obtain an a_2/b_2 measurement.

There are six raw measurements for the thru portion of the calibration.

$$\begin{aligned} S_{11ThruR} &= E_{00} + \frac{E_{10}E_{01} \cdot ELF}{(1 - E_{11}ELF)}, & S_{22ThruR} &= E_{33} + \frac{E_{23}E_{32} \cdot ELR}{(1 - ELR \cdot E_{22})} \\ S_{21ThruR} &= \frac{E_{10}E_{32}}{(1 - E_{11}ELF)}, & S_{12ThruR} &= \frac{E_{01}E_{23}}{(1 - ELR \cdot E_{22})} \\ \left. \frac{a_{1M}}{b_{1M}} \right|_{a_2_active} &= \Gamma_R, & \left. \frac{a_{2M}}{b_{2M}} \right|_{a_1_active} &= \Gamma_F \end{aligned} \quad (3.36)$$

$$\text{where } ELF = E_{22} + \frac{E_{32}E_{23}\Gamma_F}{1 - E_{33}\Gamma_F}, \quad ELR = E_{11} + \frac{E_{10}E_{01}\Gamma_R}{1 - E_{00}\Gamma_R}$$

Note that the two additional unknown terms, Γ_F and Γ_R are explicitly measured as part of the thru-standard measurement. In fact, since these are related to the difference between the source and load match at the test port and this difference is usually stable and does not change over time, it is possible to acquire these values one time for the VNA and not repeat this measurement. Some lower-cost versions of VNAs have 3-receiver architectures that do not directly support the Γ_F and Γ_R measurements, but their values can be found using standard short-open-load-thru (SOLT) techniques, where the difference between the source and load match is saved; this is often called the *delta-match* calibration, and performing one on a 3-receiver VNA allows it to then be used for any of the 8-term error model corrections.

From Eq. (3.36), there are nine error terms and six independent equations. At least three more independent equations are required to solve for the error terms.

3.4.1.2 Line Standard

The line standard provides the key attributes to generate a good calibration. The quality of the impedance of the line standard sets the quality of the calibration. The one critical factor of the line standard is that its length must be different from the length of the thru, such that the phase shift through the thru is different from the phase shift of the line by at least 20° and not more than 160° . While this is often given as a hard rule, in fact the range can be extended somewhat, at the risk of degradation due to noise and other errors in the measurement system. The difficulty comes about from the necessity to have unambiguous results for the raw S-parameters of each standard. If the line standard were to be exactly 180° , the raw measurements of the thru and the line would be identical (not including a small effect of loss), and there would not be sufficient independent measurements to determine the error terms.

In the case of metrology standards, the line standards are produced with precise machining, often created in two parts (outer conductor and inner conductor) as an airline that has no support beads. These airlines are measured using some of the most precise dimensional measurement methods, and the impedance is computed based on these measurements. Often, the computations include accounting for the skin effect of the line and even include effects such as the sag of the line's center conductor due to gravity effects. As such, the knowledge of the line's impedance is perhaps one of the most precisely known attributes in RF and microwave engineering.

In most cases, the line impedance is presumed to be the same as the system impedance, and errors in the line impedance become residual errors in directivity, source-match, and tracking. But for metrology applications, the precise value of the airline is entered into the calibration kit definition, and the TRL calibration algorithm is modified to use the line's actual impedance instead of the system impedance in determining the error terms.

In most cases, the loss of the line is defined in the calibration standard, but in some cases the loss is not well known. During the solution of the TRL acquisition, both the length of the line and the loss of the line can be determined, and this is sometimes called *LRL auto-characterization*. Auto-characterization can be used when the thru and the line standards have the same offset loss and impedance.

A key area of concern for in-fixture or on-wafer calibration is variation between the impedance of the thru and the line standards. Since both are typically manufactured in a photo-lithographic process, errors in creating the width of the line (due to over-etching, for

example) or differences in the dielectric constant of the substrate material below the thru versus the line will create residual errors in the calibration.

The line standard measurements are

$$\begin{aligned} S_{11LineR} &= E_{00} + \frac{E_{10}E_{01}S_{21UT}^2 \cdot ELF}{(1 - S_{21UT}^2 E_{11} \cdot ELF)}, \quad S_{22ThruR} = E_{33} + \frac{E_{23}E_{32}e^{-2\gamma L} \cdot ELR}{(1 - e^{-2\gamma L}ELR \cdot E_{22})} \\ S_{21ThruR} &= \frac{e^{-2\gamma L}E_{10}E_{32}}{(1 - e^{-2\gamma L}E_{11}ELF)}, \quad S_{12ThruR} = \frac{e^{-2\gamma L}E_{01}E_{23}}{(1 - e^{-2\gamma L}ELR \cdot E_{22})} \end{aligned} \quad (3.37)$$

where $ELF = E_{22} + \frac{E_{32}E_{23}\Gamma_F}{1 - E_{33}\Gamma_F}$, $ELR = E_{11} + \frac{E_{10}E_{01}\Gamma_R}{1 - E_{00}\Gamma_R}$, and γL represents the response due to the length of the line.

Here four more equations are added, but two more unknown values are also added, the delay and loss associated with γL . Thus, the total number of independent equations has increased to 10, but the total unknowns have increased to 11.

3.4.1.3 Reflect Standard

The reflect standard is the simplest standard, and the only criteria is that it provides some non-zero reflection equally at each of the ports. While this does sound simple, it presents some problems in the case of coaxial connectors because the short standard for the male and female ports must be different physical components. In such a case, the standards must be electrically identical. Thus, the reflect standards may be unknown, but they must be the same.

For probes, it is common to leave the port unterminated and simply lift the test probes. This provides an open reflection that is presumed to be the same on each port. While this may work for probes, it is not advised for coaxial connections, particularly for a male test port (female DUT) because the male pin and the nut may act as an antenna and radiate such that the reflection at some frequencies is no longer constant. This is one of the reasons that shorts are recommended for TRL reflect standards.

For waveguides, an open cannot be used for similar reasons. An open waveguide provides approximately 12 dB return loss due to radiation, and this value can change substantially due to re-reflections from the environment. For waveguides, shorts are also recommended as the reflect standard.

In some cases, the reflect standard is used to set the reference plane of the ports. The VNA calibration usually provides for a choice of using the center of the thru or reflect standard to set the reference plane. In the case of the reflect standard, the phase of the standard must be approximately known, in a similar manner as the phase of the line standard. Often, it is defined to be zero delay, and this is one case where an open is sometimes used. An example is a PC board TRL cal-kit where the open is simply the PC board fixture with no DUT attached. Taking caution to avoid the effects of radiation from the open should be re-iterated here.

The measured values for the reflect standard are

$$S_{11Refl} = E_{00} + \frac{E_{10}E_{01} \cdot \Gamma_{Refl}}{(1 - E_{11}\Gamma_{Refl})}, \quad S_{22Refl} = E_{33} + \frac{E_{23}E_{32} \cdot \Gamma_{Refl}}{(1 - \Gamma_{Refl}E_{22})} \quad (3.38)$$

In Eq. (3.38), two more independent equations are available, but one more unknown is added: the reflection coefficient of the reflect standard.

Thus, there are now 12 unknowns and 12 independent equations, which can be solved to find the error terms. The solution is rather complex but has been presented in several forums and is available in the references at the end of this chapter (Engen and Hoer 1979).

3.4.2 Special Cases for TRL Calibration

3.4.2.1 TRM Calibration

The TRM calibration stands for thru-reflect-match and is really a degenerate case of the TRL calibration, where the L could be considered as a lossy, infinite length line. The calibration match standard is the same as a load standard for the SOLT calibration of Section 3.3.2.3 and is typically considered to have a zero reflection. However, the mathematical expressions for the case of TRM can also allow an arbitrary match condition so long as the reflection coefficient is known.

TRM is in fact almost always used as part of the TRL calibration for low frequencies, with a common transition frequency being between 2 and 3 GHz. The measured values for the match standard are

$$S_{11MatchF} = E_{00}, \quad S_{22MatchR} = E_{33} \quad (3.39)$$

It might seem that the TRM calibration eliminates the four measurements of the line and replaces them with only two independent equations. But the while the line standard provides four independent equations, it also adds two more unknowns; thus, the match standard of TRM provides the same order of new independent equations as the line standard.

3.4.2.2 Other TRL Considerations

While TRL has widespread use, there are some special cases that require additional care to avoid poor results.

In waveguide calibration, the line standard loss is determined as part of the auto-characterize, but if the thru is not zero length, its loss and delay must be known. A special case of the TRL calibration provides for the dispersion of the thru so that the phase shift error from waveguide dispersion is properly accounted for. In such a case, the cal-kit must specify that the thru standard is waveguide. Further, it is standard practice to set the system Z_0 and the line Z_0 to 1Ω for waveguide calibration. The concept of impedance is ill-defined in waveguides, and setting the value to 1 provides for a common reference. Some older VNAs required both values to be set on the instrument, but newer VNAs provide these values as part of the calibration kit definition.

A similar situation occurs for microstrip lines, which are somewhat dispersive. Unfortunately, there are no means in most VNAs for providing for the dispersion of microstrip lines, so the best approach is to use the delay value at the center of the band of interest for the thru offset delay. The offset loss will be computed as part of the TRL self-calibration process. Some methods of LRL provide for a means to automatically characterize the loss and delay of the line standards, and so may be beneficial in this case.

For conditions where the line impedance varies greatly from the system impedance, special care must be used. An example of this scenario is a case commonly found when testing high-power transistors, which often have a very low input or output impedance. The low impedance and high power output necessitates a wide trace. Thus, an impedance transformer, often in the form of a tapered line, is used to translate from $50\ \Omega$ to the transistor impedance. A TRL calibration of this wide line can have some difficulties as the width of the line might support higher-order planer waveguide modes or orthogonal TEM modes, or even act as a patch antenna. In essence, the line becomes so wide that it appears to be a separate pair of transmission lines connected perpendicularly to the main line.

Eqs. (3.36), (3.37), and (3.39) are sufficiently independent to be solved for the error terms; the solution is beyond the scope of this book but may be found in numerous references.

3.4.3 Unknown Thru or SOLR (Reciprocal Thru Calibration)

The unknown thru (UT) calibration, or short-open-load-reciprocal (SOLR) calibration as it is sometimes known, has become the preferred calibration method for most VNA measurement scenarios (Ferrero and Pisani 1992). The UT calibration is also based on a combination of the 8-term error model and the 12-term error model. It requires the computation of the switch terms similar to the TRL calibration and requires the same set of calibration standards as the 12-term calibration but allows different degrees of freedom than either the SOLT or the TRL calibration.

3.4.3.1 Unknown Thru Standard

The thru standard for the UT calibration has only one requirement: it must be reciprocal in transmission, that is, $S_{21}=S_{12}$. This requirement holds true for all passive devices except isolators or circulators. In practice, the loss of the UT must be sufficiently small to avoid numerical difficulties in the computation of the error terms. Different vendors of VNAs use different methodologies, with some allowing loss up to 40 dB and still maintaining good calibration integrity.

In the UT calibration, the error terms associated with the directivity, source-match, and reflection tracking are determined based on the same methods as a simple 1-port calibration or the 1-port portion of the 12-term calibration, as in Section 3.2.1. The raw measurements for the unknown thru are

$$S_{11UT_R} = EDF + \frac{ERF \left(S_{11UT} + \frac{S_{21UT}^2 ELF}{(1 - S_{22UT} \cdot ELF)} \right)}{\left[1 - ESF \cdot \left(S_{11UT} + \frac{S_{21UT}^2 ELF}{(1 - S_{22UT} \cdot ELF)} \right) \right]}$$

$$S_{21UT_R} = \frac{(S_{21UT} \cdot ETF)}{(1 - S_{11UT} \cdot ESF) \cdot (1 - S_{22UT} \cdot ELF) - ESF \cdot S_{21UT}^2 \cdot ELF}$$

$$S_{12UT_R} = \frac{(S_{12A} \cdot ETR)}{(1 - S_{11UT} \cdot ELR) \cdot (1 - S_{22UT} \cdot ESR) - ESR \cdot S_{21UT}^2 \cdot ELR}$$

$$S_{22UT_R} = EDR + \frac{ERR \left(S_{22UT} + \frac{S_{21UT}^2 ELR}{(1 - S_{11UT} \cdot ELR)} \right)}{\left[1 - ESR \cdot \left(S_{22UT} + \frac{S_{21UT}^2 ELR}{(1 - S_{11UT} \cdot ELR)} \right) \right]} \\ \left. \frac{a_{1M}}{b_{1M}} \right|_{a_2_active} = \Gamma_R, \quad \left. \frac{a_{2M}}{b_{2M}} \right|_{a_1_active} = \Gamma_F \quad (3.40)$$

where

$$ELF = E_{22} + \frac{ERR \cdot \Gamma_F}{1 - EDR \cdot \Gamma_F}, \quad ELR = E_{11} + \frac{ERF \cdot \Gamma_R}{1 - EDF \cdot \Gamma_R}$$

where similarly to the TRL calibration, the values for Γ_F and Γ_R are determined by measurements of the a_1/b_1 and a_2/b_2 terms during the thru step of the calibration.

Equation (3.40) gives six independent equations but also adds three additional unknowns: S_{11UT} , S_{22UT} , and the product $S_{21UT} \cdot S_{12UT} = S_{21UT}^2$, realizing that the unknown thru must be reciprocal in transmission.

There are 12 total unknown, just as in the TRL case: 7 independent TRL errors, 3 unknown-thru parameters, and Γ_F and Γ_R . These are solved in a simultaneous way to generate all the necessary error terms and give the values of the unknown thru as a bonus.

This calibration requires identical steps to the SOLT calibration but does not require that the thru standard be defined. The applications for this greatly extend the flexibility of the SOLT calibration, allowing a vast array of thru devices to be used that can greatly simplify complex calibration tasks. The quality of the UT calibration is derived from the quality of the reflection standards. If there is substantial loss between the test port and the measurement plane, noise effects will degrade the measurement of the reflection standards, and thus the transmission measurements will degraded to a greater extent than in the case of a defined thru calibration. Thus, the quality of the UT calibration may be better or worse than the SOLT calibration, depending upon the relative quality of the reflection standards compared to the defined thru standard of the SOLR calibration.

Equation (3.40) provides the set of equations from which the unknown error terms are computed, the solution of which is beyond the scope of this book, but may be found in many references.

3.4.4 Applications of Unknown Thru Calibrations

Some common cases where unknown thru calibrations are beneficial are described here.

3.4.4.1 Non-insertable Coaxial Calibration

For most components found in RF and microwave work, the connectors on each port are identical and typically female. The SMA female DUT connector is probably the most common connector used, perhaps followed by the Type-N female. Most cables have male connectors, so many devices must be tested with male test port connectors on each port. As such, a female-to-female thru standard must be used during calibration. In the 12-term

SOLT calibration, the attributes of this connector must be defined prior to calibration, as noted in Section 3.3.3.5, or determined through a second step of calibration, as described in Section 3.3.3.6. This is not necessary at all for the UT calibration. This flexibility can lead to many ingenious ways to reduce and reuse calibration steps to achieve complex calibration scenarios.

3.4.4.2 On-Wafer Calibrations

In some cases of on-wafer calibrations, the test probes are not aligned in a way that they allow a direct, straight-through line to connect between the probes. In on-wafer probing, it is common to refer to the probe positions as *east* and *west* when they are on the left and right sides of the wafer, and *north* and *south* when they are located above and below the wafer. In cases where the device must be probed with a set of probes that do not line up, such as east and north or south and west, the through-line on-wafer standard must contain a 90° bend, as shown in Figure 3.16. It is quite difficult to know exactly the impedance and length of the bend, so utilizing an unknown thru calibration removes the necessity of providing a precision designed on-wafer thru standard. The calibration method requires a 1-port calibration at each probe tip, but for the final step, any thru device may be used, including the DUT if it is reciprocal.

3.4.4.3 Fixed-Port Calibration

Another common situation for the unknown thru calibration is a case where the test ports are fixed and cannot be connected. This often happens in waveguide-to-waveguide situations, where the waveguides are hard-mounted to a frame or bench and cannot be easily moved. In the past, this presented a difficult calibration situation and required moving the waveguides using extra waveguide bends and line sections to provide for a flush thru calibration or characterizing a waveguide section that could be used as a defined thru standard. Even in the latter case, getting alignment of the defined waveguide thru to the fixed waveguide ports presented challenges.

With the unknown thru calibration, all that is needed is to perform a 1-port calibration at each port and then provide any thru connection, including using a flexible cable with coax-to-waveguide adapters on each end. The only requirement is that the flexible cable,

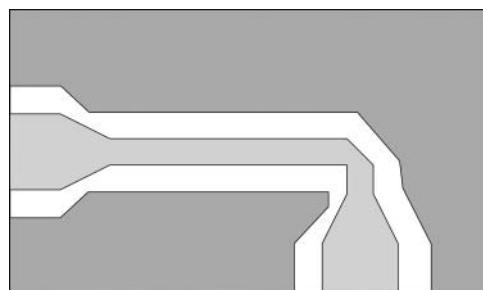


Figure 3.16 Using UT cal to provide a 90° on-wafer calibration.

or other means to create a thru, is stable over the time it takes to acquire the six thru ratios (four S-parameters plus the two switch terms). The 1-port calibrations are typically performed using short/offset-short/load/offset-load.

Using a flexible unknown thru standard between the test ports has the potential to minimize significant errors that may be caused by using flexible test ports. In some cases, semi-rigid transmission lines can be used in the place of flexible cables, positioned so that they connect directly to the DUT. This eliminates the need to move them together to provide the normal defined thru connection step.

3.4.4.4 Switched-Path Calibration

A common problem for calibration occurs when a switch matrix is added in front of the network analyzer ports (as shown in Figures 2.25 and 2.26) to provide the ability to measure a multiport device one port pair at a time. For an N-port device, there are $(N - 1)(N)/2$ possible paths, each one of which requires a 2-port calibration. The unknown thru calibration can be used to measure a thru between one port pair, with this same data being used as an unknown thru for all the other port pairs. From this observation, it becomes clear that all that is needed to calibrate all the port pairs of a multiport test set is a 1-port calibration on each port, as well as one measurement of an unknown thru on any single port pair. From this, the calibration of all the other port pairs can be determined, as discussed in Section 3.9.

3.4.5 QSOLT Calibration

The quick-open-short-load-thru (QSOLT) calibration is a different blend between the SOLR (unknown thru) calibration and the SOLT calibration (Ferrero and Pisani 1991). As the name implies, it requires an open/short/load 1-port calibration and a defined thru, but it is quick because the 1-port calibration needs to be performed on only one of the test ports. In fact, any 1-port calibration method can be used (such as offset short, or even Ecal) on just one of the test ports, and then a defined thru measurement is performed between ports 1 and 2. In this way, a full 2-port calibration is easily performed on an insertable path using a cal-kit or Ecal that has only female or only male standards.

One may understand the underlying principle of the QSOLT calibration by realizing that, like the TRL calibration, there are seven unknown terms plus the switch terms. The switch terms can be found similarly to TRL during the thru measurement. The three 1-port terms are determined during the 1-port calibration, and there are four more equations that can be formed from measurements of the four S-parameters of the thru. Since the actual values of the thru are known, the four measured values yield four more independent equations, providing a means to compute the remaining four unknown error terms.

The QSOLT calibration is convenient in cases where 1-port calibrations are difficult. One use of the QSOLT calibration is in a multiport system. If a multiport DUT has N ports, all of the same sex, one can create an $N + 1$ test system, with a flexible cable on the extra port that matches the DUT connector. A simple 1-port calibration on this extra port, plus a thru connection to each of the other ports, provides for a full $N + 1$ port calibration. In this way, none of the other ports needs to move or even have calibration kits for them.

3.4.6 Electronic Calibration (ECal™) or Automatic Calibration

Electronic calibration (Ecal), or automatic calibration, was first introduced by HP in partnership with ATN, which became Agilent and is now Keysight, for use with the HP 8510 in about 1995, and the term ECal is now associated with Keysight. In this work, Ecal (small *c*) is used loosely to refer to any of the electronic calibration modules. Since this first introduction, the quality, capability, and convenience of Ecal has been greatly enhanced, and now its use may exceed that of the mechanical calibration kits. Examples of some Ecals are shown in Figure 3.17. The upper set are the high-performance Ecals with built-in heaters. The lower set are the newer economy Ecals, without heaters but with temperature compensation. While prices vary, the lower-cost Ecals are typically priced quite closely to mechanical calibration kits.

The first electronic calibration modules consisted of a transmission line shunted by pin diodes. If the diodes were reverse-biased, the transmission line provided a thru connection from port 1 to port 2. If the diodes closest to the ports were forward biased, an offset short would be created with a short offset. If the diodes farther from the port were shorted, a short with a longer offset was created. The configuration allowed a variety of offsets to be created such that for any frequency in the specified range, a good spread of reflection coefficients for three of the states could be obtained.

At the time of manufacture, the actual value of each of these reflection states is measured, as well as the S-parameters of the thru state, and these values are loaded into an on-board memory on the module. From these values, a 12-term error correction could be acquired



Figure 3.17 Ecal modules are available in a variety of port configurations, connectors, and frequencies.

using the formation of Eqs. (3.25) and (3.30) after de-embedding the thru S-parameters from the 1-port calibration.

These older versions of electronic calibration could not be used at low frequencies because of the limitation of the length of the offset shorts used. More modern versions of electronic calibration use custom GaAs IC switches, which provide an embedded nominal open, short, load, and thru. These custom ICs may contain multiple short states in addition to the open state to ensure that a wide phase difference between standards is maintained over the entire frequency range. An example of these states over a small frequency range is shown in Figure 3.18.

The open state (Marker 1) has a quite high reflection coefficient, near the edge of the Smith chart. The short (Marker 2), which is limited by series resistance in the solid-state FET device, does not provide nearly as large of a reflection, but is of opposite phase. The load (Marker 3 in the figure) has the lowest reflection but is not nearly as good as a mechanical load standard.

Because Ecal modules contain solid-state electronic switches, they are repeatable and stable. Most have an internal heater that maintains a constant temperature of the unit, typically around +31 C, which ensures it is at a constant temperature over the commonly specified range for VNAs of 20–26 C or sometimes 20–30 C. Recently, some lower-cost Ecal modules have been introduced that do not have the heater circuits but maintain their performance

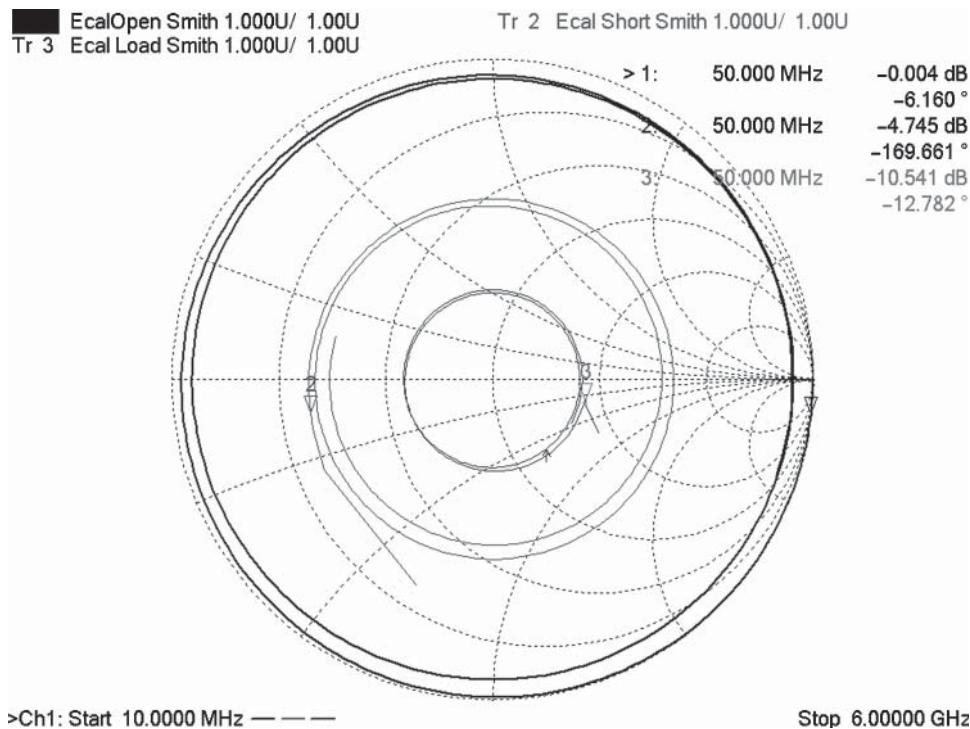


Figure 3.18 Measurement of the internal standards on an Ecal™.

through an internal temperature sensor such that the effects of environmental change can be compensated for. Because they use electronic switches, they can compress when driven with high power, so it is important to avoid driving the modules above their rated operational power, even though the damage level may be much higher.

3.4.6.1 Calibration Types for Electronic Calibration Modules

Ecal modules can support many of the calibration acquisitions described earlier, but the default type is unknown thru. Because each of the standards in the Ecal is a characterized device and must be measured, the uncertainty of the standards is not the same for each standard. For the most part, reflection measurements have lower uncertainty than transmission measurements and so the reflection standards have less uncertainty in their characterization than the transmission standards. The UT calibration does not depend upon the characterization in the Ecal thru but uses only the 1-port standards' characterization data. This method is referred to as *Ecal thru as unknown thru* because the Ecal unit is connected to both ports at the same time (to perform the 1-port calibrations), but the thru connection does not rely upon the S2P data saved for the thru standard of the Ecal.

The 3-receiver architecture of some lower-cost VNAs does not support unknown thru; in such a case only the SOLT with defined thru is used for Ecal. In other cases, where there is a great deal of loss after each test port, a defined thru may provide less error than an unknown thru when used as an enhanced response calibration and thus might be preferred (see Chapter 6 for an example). Also, in the case where a QSOLT calibration is desired, the defined thru of the Ecal module might be utilized as well.

Another mode of operation for Ecal provides the ability to separate the 1-port calibrations from the thru calibrations, such that a flush thru can be used instead of the Ecal thru, in cases where the test ports are insertable, or form a mated pair. In a related situation such as the fixed connector case, an Ecal may not be able to be connected to both test ports at the same time. In such a case, the Ecal can be used for the 1-port calibrations, but another thru, such as a cable, can be used for the unknown thru step. This is simply called an *unknown thru* rather than an *Ecal as unknown thru* and divides the normally single-step Ecal into three steps: Ecal at port 1, Ecal at port 2, and unknown thru.

Finally, Ecal modules can be used as calibration kits alongside mechanical calibration kits, where the Ecal ports don't match all the required DUT ports. An example case might be a coax-to-waveguide transition, where the Ecal might be used for one port of a calibration of the coax side, a mechanical waveguide kit used on for the 1-port calibration on the waveguide side, and a waveguide to coax adapter used for the unknown thru step.

3.4.6.2 User Characterization of Ecal Modules

Another useful feature of Ecal modules is the ability to add adapters, cables, or fixtures, and characterize the Ecal with the adapter in place. This characterization consists of creating a 2-port calibration (four ports in the case of 4-port Ecal modules) in the desired connector type and then adding an adaptor with that connector to each port of the Ecal and running a characterization function. During the characterization, each internal standard of the Ecal is measured by the calibrated VNA in the new connector type. The values of these standards are

then downloaded into the Ecal, or saved on the hard-drive of the VNA, to be used later, perhaps in a production test. Current Ecals can support up to 12 user characterizations internally and an unlimited number on the hard drive.

One convenient feature of user characterization is that it allows the Ecal to be extended beyond its normal frequency range. For example, the 3.5 mm Ecal is limited from the factory to 26.5 GHz frequency operation. However, the internal standards and the connector are substantially mode free until perhaps 33 GHz. It is possible perform a user characterization of this Ecal using 2.92 mm connectors and a mechanical cal-kit, such that the Ecal's useful frequency range can be extended. While this calibration is not guaranteed by the manufacturer, the quality can be traced through the mechanical calibration kit.

Finally, it is possible to embed an Ecal into a switching matrix and perform a user characterization at the end of each of the switch-matrix ports, such that the user characterization can provide calibrations at each of the ports. Figure 3.19 shows an example of a custom calibration test set that has an integrated Ecal module, power sensor module, and noise source module. The Ecal characterization is done at the test ports, and the loss to the common power sensor is determined from the difference in the reflection tracking term between the internal Ecal factory characterization and the external test port characterization, presuming that the power divider has equal loss. This kind of custom calibration test set allows multiple measurements, such as noise figure, power, intermodulation distortion (IMD), and S-parameters, to be calibrated with a single test port connection. Further, from the figure one can see that the test set

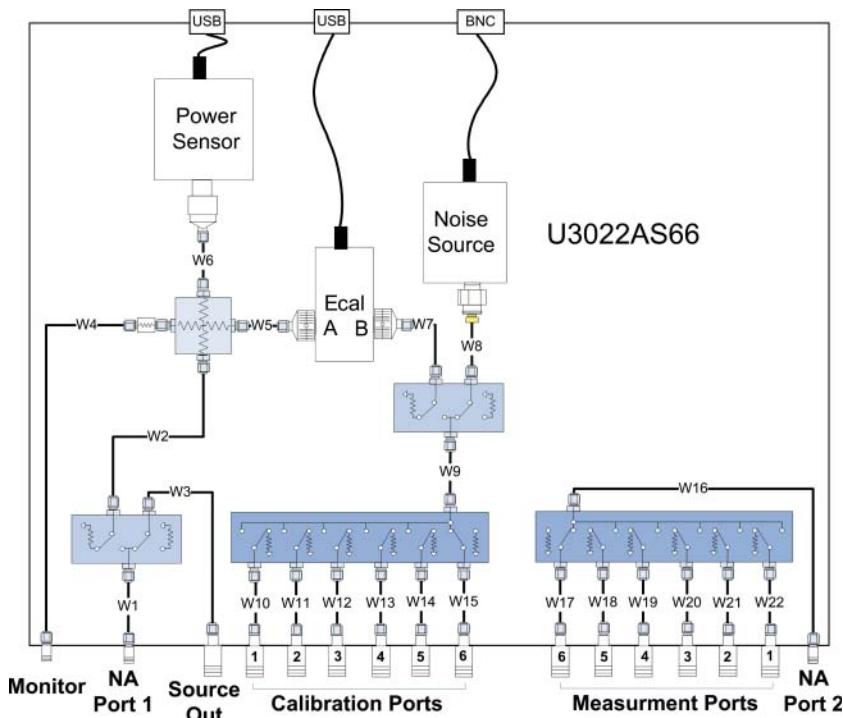


Figure 3.19 Custom multiport calibration test set including Ecal, noise figure and power calibration.

includes both the calibration unit as well as a matrix switch to generate six test ports from the VNA port 2. This calibration system was designed to provide calibration for a 1×6 matrix test system used to test a 1×6 DUT.

3.5 Waveguide Calibrations

When the DUT ports are waveguide, the calibrations that are performed using waveguide calibration kits require some special care to obtain optimum results. The first difference is that a waveguide left with an open circuit does not have a well-defined reflection; thus, open standards are not used. Instead, the cal-kits contain what is called a quarter-wave shim, which represents a 90° phase shift in the center of the waveguide band. The quarter-wave shim over the nominal waveguide band has a phase shift from about 50° to 120° . In the case of using the shim as part of an offset short calibration, the phase shift at each frequency must be precisely calculated using the waveguide dispersion formula.

$$\phi_f = \frac{360f}{c} \sqrt{1 - \left(\frac{f_c}{f}\right)^2} \quad \text{degrees/meter} \quad (3.41)$$

For example, the case of a Ka band waveguide (using WR-28, sometimes called R-band), the waveguide band is 26.5–40 GHz, the center of the band is at 33.25 GHz, and the cutoff frequency for the waveguide is 21.081 GHz. A shim that is 90° at 33.25 GHz would have a length of 2.9168 mm and would have a phase shift of 56° at the lower band edge and 119° at the upper band edge.

The calibration process using a quarter-wave shim usually proceeds with a short, load, offset short (using the shim). This provides enough standards to perform the traditional 1-port calibration. In the newest VNAs, an additional standard may be used, the offset load standard. In this case, the same shim that is used with the short is also used with the load. This should be used if the impedance and phase shift of the shim is known with better uncertainty than the impedance of the load. This is almost always the case with waveguide components, and so the offset load cal reduces the error caused by a non-ideal load down to the quality of the shim used.

Another aspect of waveguide calibration is that the TRL calibration is perhaps the best and simplest for a 2-port system. It requires only a reflect standard (usually a short), a thru, and a line. The quarter-wave shim makes an ideal line as it easily meets the criteria of providing a phase shift of between 20° and 160° . In this way, it is simple to create a calibration kit and use TRL to measure the exact phase shift of the quarter-wave shim.

For multiport waveguide calibration, the QSOLT calibration is well suited provided that the waveguide ports can be moved in such a way to directly mate. In this case, the 1-port cal will use four standards: the short, offset short, load, and offset load. The remainder of the calibration is to connect a flush thru between ports, but this can be assigned in almost any order, and for N ports, only N-1 thrus are required. For example, consider a DUT with balanced waveguide input ports (e.g. 1 and 3) and balanced waveguide output ports (2 and 4). A multiport calibration can proceed by performing a 1-port cal on port 1 and then connecting it to port 2 to measure a 1–2 thru, followed by connection of port 2 to port 3, for a 2–3 thru,

and finally connection port 3 to port 4 for a 3–4 thru. This order might minimize movement of the waveguide ports. While this minimizes the 1-port calibrations required, the residual errors will be cascaded to the other ports. Another approach is to use a 1-port calibration on each pair of ports, calibrate the opposite port, and then use unknown thru for the final port pair. For example, consider the case of ports 1 and 3 on the left and ports 2 and 4 on the right. A 1-port calibration on ports 1 and 4 can then allow a QSOLT from 1 to 2, QSOLT from 3 to 4, and unknown-thru from 1 to 3, from 2 to 4, from 1 to 4, or from 2 to 3. Using a mixture of calibrations methods on multiport situations can provide a dramatic improvement in the efficiency of the calibration steps.

In the past, waveguide standards assumed that the loss of the waveguide was negligible. In most RF and microwave bands, that is a reasonable assumption, but in mm-wave and above, the loss of the calibration standards can produce a noticeable offset in the measured results. Recently, some VNAs have provided the capability to account for loss following a formula that computes loss based on offset delay, and the dimensions (width/height) of the waveguide. The computation follows the formula

$$L_{Offset_WG} = \frac{60\pi \ln(10)}{10} \frac{S_{21(dB)}}{S_{21(delay)}} \sqrt{\frac{f_c}{f}} \left[\frac{\sqrt{1 - \left(\frac{f_c}{f}\right)^2}}{1 + 2\left(\frac{h}{w}\right) \cdot \left(\frac{f_c}{f}\right)^2} \right] \quad (3.42)$$

The use of some features on a VNA, such as port extension or electrical delay, also account for the dispersion of the waveguide in computing the effective physical length. Some capabilities, such as the time-domain transform, require special consideration when applied to waveguide measurements.

3.6 Calibration for Source Power

Traditionally, VNA measurements have been restricted to measurements of ratios of a and b waves. The value of the source power was not important to the first order, as the ratios of waves that generate the S-parameters, as well as the S-parameters themselves, have always been considered as linear terms, that is, terms that don't depend upon the absolute power levels applied. But many devices are mildly to strongly non-linear, and their performance is specified at particular input powers.

Older VNAs, such as the HP 8510, used external sources, routed through the reflectometer test set. The source power setting of the source was substantially different than the incident power from the test port, sometimes by as much as 10–20 dB. Since then, almost all sources for VNAs are integrated, and a factory calibration ensures that the source power is at least nominally correct. But with the addition of cables, connectors, fixtures, and switch matrixes, the actual loss to the DUT can be quite large and must be accounted for if accurate source power is to be applied.

In the remaining discussion, the details are given for the forward condition, recognizing that the reverse measurements follow similarly.

3.6.1 Calibrating Source Power for Source Frequency Response

Starting with the HP 8720A and HP 8753D, the first VNAs with a built-in source and test set, the source power setting on the front panel referred to the incident source power at the DUT. These analyzers also had the ability to control and read a power meter connected via the GP-IB interface. They contained the first built-in firmware that performed source power calibration. The process for source power calibration was a point-by-point measurement and correction of the source power. The calibration process starts by connecting the power meter's sensor to the test port, setting the source to the first point of the frequency sweep data, and acquiring a power meter reading. The source power is then adjusted up or down, according to the offset in measured versus desired power. The offset is recorded and becomes the source calibration factor (SCF). Ideally, this is essentially the STF term in the signal flow graph shown in Figure 3.2. However, in practice, the source is not-linear and can have errors in addition to the test set loss so that the SCF becomes

$$SCF = \Delta Src \cdot STF \quad (3.43)$$

where ΔSrc accounts for any difference in the source setting value and the value incident to the test set. If there are no reflections at the test port,

$$a_{IS} = \Delta Src \cdot a_{Vs} |_{\Gamma_{Load}=0} \quad (3.44)$$

where a_{Vs} represents the value of the VNA source power setting. Factors affecting ΔSrc are uncompensated loss between the source and test set, and non-linear response of the source output power to different source settings.

Figure 3.2 is enhanced in Figure 3.20 to more closely represent the actual source situation, where reflections from the test port make their way back to the source and reflect off the power source match to affect the incident power. For the simplest case of source power calibration,

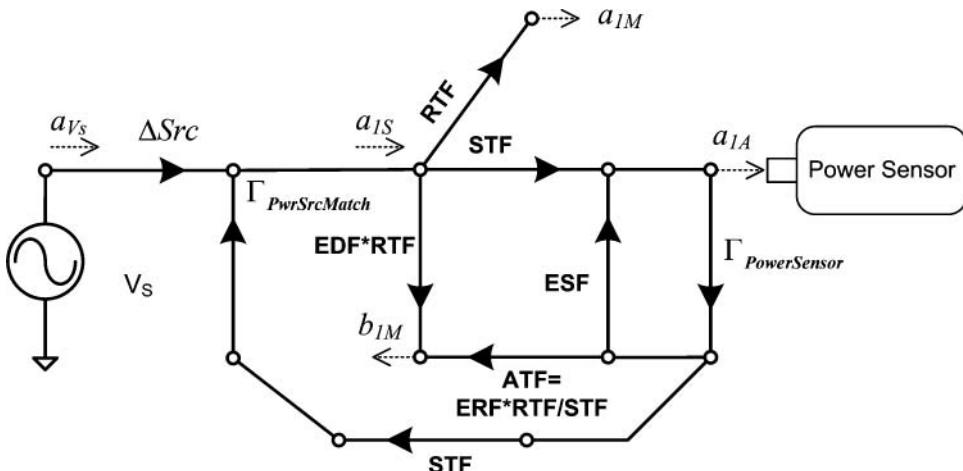


Figure 3.20 Signal flow diagram during source power calibration.

assuming no reflection from the power meter measurement, the STF term is simply

$$STF = \frac{P_{Meas}}{a_{1S}} \Big|_{\Gamma_{Load=0}} \quad (3.45)$$

where a_{1S} is the power incident to the test set, which can be monitored on the reference receiver. In general, the value of STF cannot be found directly without an independent measurement, but it is used in correction in a ratio with RTF. The SCF can be found as

$$SCF = \frac{P_{Meas}}{a_{Vs}} \Big|_{\Gamma_{Load=0}} = \frac{\Delta Src \cdot P_{Meas}}{a_{1S}} \Big|_{\Gamma_{Load=0}} = \Delta Src \cdot STF \quad (3.46)$$

which indicates that the source correction takes on two components: (i) the loss of the test set, and (ii) the offset between the source power setting and the actual source power provided to the test set. Normally, the source correction factor is expressed in dB as

$$SCF_{dB} = P_{Meas(dB)} - a_{Vs(dB)} \Big|_{\Gamma_{Load=0}} \quad (3.47)$$

Since the ΔSrc changes with power level, it may be necessary to iterate the source power to achieve some particular value. For normal source power calibration, a cycle of measurement to find SCF is repeated by a user-specified factor, typically 3–10 times. Unlike S-parameter error-correction, this is not a post-processed correction; rather, it is a calibration in the sense that the offset values are used to adjust the source settings before the data from the DUT is acquired. If the source is linear, the value for SCF is constant, but sometimes the source is not linear; that is, changing the setting of the source setting by 1 dB does not change the value of a_{Vs} by exactly 1 dB. Because of this, the value of ΔSrc takes up the non-linearity of the source and is only precisely defined at a single power level.

3.6.2 Calibration for Power Sensor Mismatch

From the signal flow graph, it is clear that the match of the power meter can affect the power measured for a given source setting and thus the extraction of the value for STF and SCF. The value measured on the power meter is essentially a_{1A} and is related to the VNA-programmed source power, a_{1S} as a function of STF, the source-match ESF, and the power sensor as

$$a_{1A} = \frac{a_{1S} \cdot STF}{1 - ESF \cdot \Gamma_1} = \frac{a_{1M}}{RTF} \frac{STF}{(1 - ESF \cdot \Gamma_1)} \quad (3.48)$$

where Γ_1 is the match of the power sensor. The power meter calibration factor includes a correction for mismatch loss of the sensor such that the power displayed on the power meter actually indicates the incident power to the sensor rather than the power absorbed by the sensor.

$$P_{Meas} = \frac{P_{Absorbed} \cdot \Delta PM}{1 - |\Gamma_{Sensor}|^2} = a_{1A} \cdot \Delta PM \quad (3.49)$$

Here, ΔPM is the difference between the power applied and the power meter reading, usually caused by errors in the power meter calibration factor. From Eqs. (3.48) and (3.49), the

match-corrected value for the source test set loss, STF with a power meter that has a non-zero match, can be computed as

$$STF = \frac{P_{Meas}}{a_{1S}}(1 - ESF \cdot \Gamma_{PwrSensor}) \quad (3.50)$$

In Eqs. (3.45) and (3.46), the values for source correction assume the power sensor is well matched; however, this is not always true in which case the SCF will be in error. Ignoring the second order terms, the source power can be approximated as

$$a_{1S} \approx \frac{a_{Vs} \Delta Src}{1 - \frac{(\Gamma_{PwrSrcMatch} \cdot STF^2 \cdot \Gamma_1)}{(1 - ESF \cdot \Gamma_1)}} \quad (3.51)$$

where a_{Vs} is the power setting from the source and $\Gamma_{PwrSrcMatch}$ is the match associated with the source. Γ_1 is the return loss presented to port 1 by the DUT. The value of a_{1S} can be monitored at the reference receiver, and so while the power level may change with mismatch, it can be precisely known.

Note that, in general, one does not know the value for $\Gamma_{SrcPwrMatch}$, and this remains a source of uncertainty in the output power. Since the power meter is typically well matched, these errors can be quite small. Modern VNAs also measure the raw match of the power meter during the source power calibration. The value for STF is computed based on Eq. (3.48) where a_{1S} is the power incident to the test set. If the dB loss of STF is large (meaning there is substantial loss between the reference coupler tap and the test port), the value of a_{1S} is nearly constant and does not change with changes in DUT match. This is the case when some source attenuation has been added, normally in order to provide a lower minimum power in the case of power-sensitive devices. However, if the loss of STF is small and the reflection of the load on port 1 is large, then the error described in Eq. (3.51) can be quite large. Figure 3.21 shows an example of the variation of incident power as the load match is varied through 360° for the case of $\Gamma_1 = 1, 0.5$, and $0.1 (0, -6, -20 \text{ dB})$. The upper plot shows the load contours for the three reflection states. The lower plot shows the incident power, a_{1A} , variation as the load varies. For the largest impedance, we can say that the peak-to-peak variation (1 dB in this case) is exactly the source VSWR, computed as match from

$$PowerSourceMatch = 20 \cdot \log_{10} \left(\frac{1 - 10^{\frac{P-P}{20}}}{1 + 10^{\frac{P-P}{20}}} \right) = 20 \cdot \log_{10} \left(\frac{1 - 10^{\frac{1}{20}}}{1 + 10^{\frac{1}{20}}} \right) = -24.8 \text{ dB} \quad (3.52)$$

Thus, the power source match will cause the power of the a1 signal to vary as a result of the DUT loading. Advanced techniques will be described in Chapter 6 that can dramatically reduce this effect, using the a1 receiver as a leveling reference.

From this, three key points should be remembered, listed here:

1. The source power calibration provides the proper offset of the source setting to achieve the desired output power into a matched load.
2. Mismatch at the port can affect the incident port power, a_{1A} , applied to the load.
3. The reference receiver measurement, a_{1M} , can be used to exactly monitor the incident power, provided that it is properly corrected.

Section 3.7 describes the proper methods for correcting the receivers to measure power.

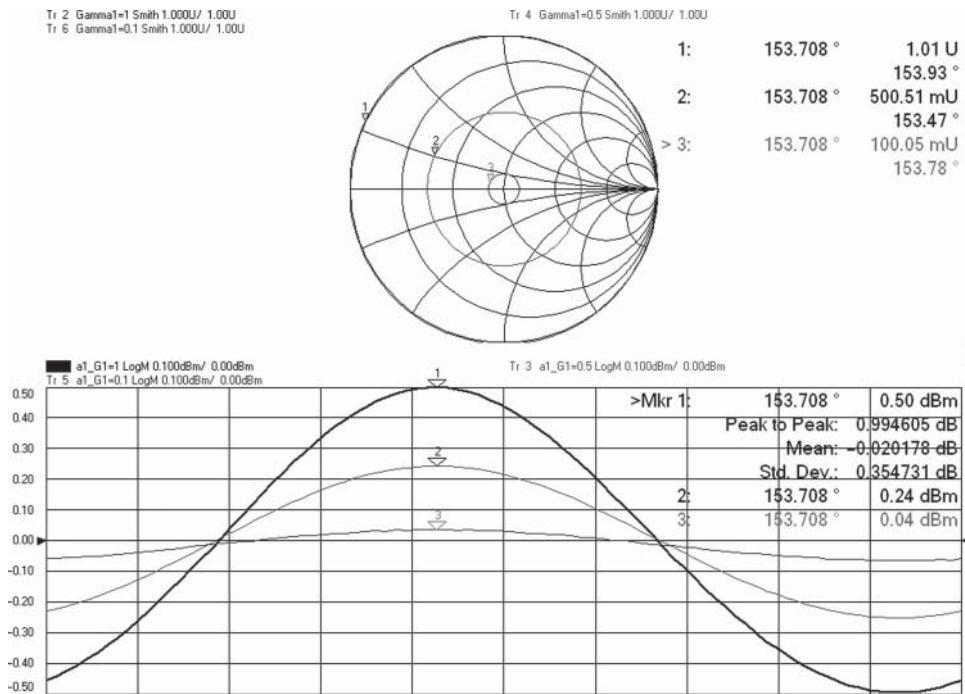


Figure 3.21 Variation in a_1 due to mismatch on port 1.

3.6.3 Calibration for Source Power Linearity

The source power from a VNA can have, in addition to errors caused by frequency response, errors due to source power linearity. Linearity describes the accuracy of the source power output with respect to programmed changes in the source power level. Since a VNA receiver is typically more than an order of magnitude better in linearity (0.02 dB typical) than the source (0.2–0.5 dB typical), it is an easy measurement to determine the linearity of the source. Linearity is usually defined as an error in the measured power relative to the set power, relative to the preset power. That is, the linearity error does not include the source flatness level. The measurement can be performed by normalizing the reference channel power reading at the preset power level, over the frequency span of the analyzer, by using the data into memory and data/memory functions. Next, the source power is set to a new level (usually the maximum or minimum specified leveled power), and the power level is read relative to the nominal offset. For example, if the preset power level is –5 dBm and the measured power on the reference receiver is –5.5, the trace is normalized to read 0 dB. If the power is now changed to –30 dBm, the reading should be at –25 dBc to the original power. Any error from this reference value is a linearity error, as shown in the upper trace of Figure 3.22.

The lower trace shows a similar measurement but this time with a +13 dBm change in input power from the reference –5 dBm.

It is normally the lower-power regions that have greater error because the ALC detector diodes are operating at much lower signal levels, and any DC drift or offset voltage will have

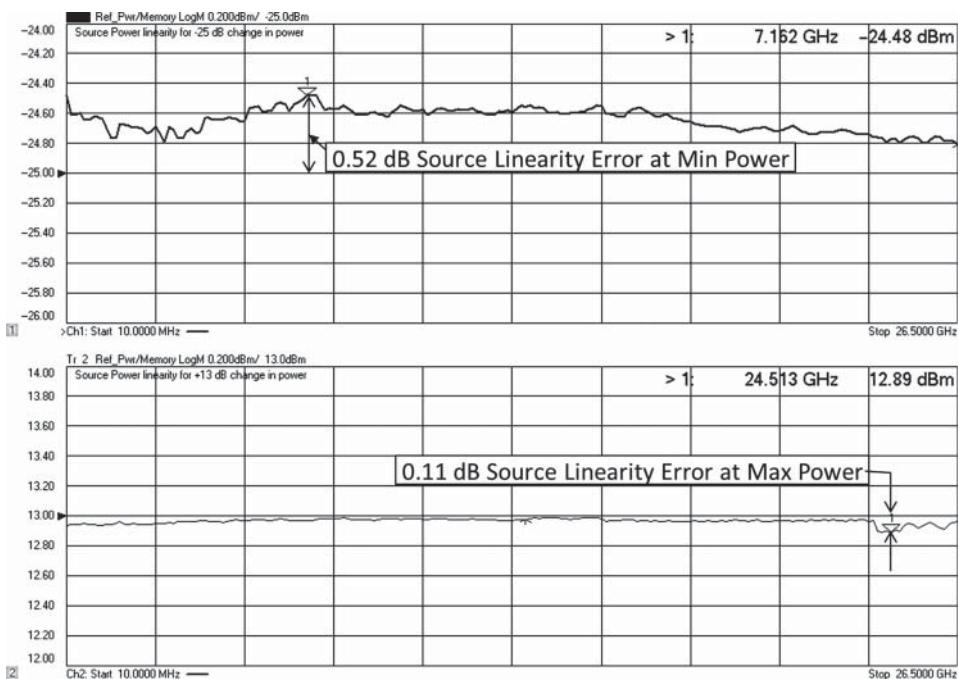


Figure 3.22 Results from a linearity error measurement for -25 dB (upper) and $+13\text{ dB}$ (lower) changes in power.

a proportionately larger effect. The power level stability at lower power levels also tends to be worse than at higher power levels. At high power levels, the source harmonics can contribute to linearity errors. The example from Figure 3.22, from a Keysight N5242B PNA-X, shows little error at high power as the harmonic performance of this VNA is good, typically better than -60 dBc .

Normally, linearity errors are not corrected for unless the source power calibration is performed in a power sweep mode. More recently, some manufacturers allow a 2-dimensional power sweep that calibrates the source across both frequency and power. This is particularly needed for the case of mm-wave extenders, which often have high-order multipliers in the source path. These have a non-linear input/output power transfer function, so a 1 dB change in input power can have a $6\text{--}10\text{ dB}$ change in output power. To properly compensate for this, a 2-dimensional source power calibration is performed, in which the user may specify the step size for the power steps. It is normal to use quite small steps sizes such as 0.25 dB or less to ensure a good measure of the mm-wave extender input/output curve. Recently the source power calibration has become a bit smarter where it will step in big steps until it detects a larger than expected power step, after which it will change to smaller steps.

Measurements at other powers will be susceptible to the linearity error. However, advanced power control methods can remove even this error, as discussed next.

Finally, new methods of calibrating the reference receiver and using it as a leveling receiver instead of the internal ALC detector effectively eliminates the need for source

power calibration altogether. The details of this capability are discussed in Chapter 6 under the topic of receiver leveling. The accuracy of this method of source power control depends entirely on the accuracy of the receiver calibration, discussed next.

3.7 Calibration for Receiver Power

3.7.1 Some Historical Perspective

The capability of a VNA to measure power has always been present in even the earliest VNAs. However, the power readings always contained the frequency response and roll-off of the VNA test port coupler and the VNA first-converter frequency response. In VNAs comprised of several individual instruments, such as the HP 8510A, no attempt was made to provide a well-known source power, nor was any calibration available to provide accurate power readings from the VNA. In the VNAs with integrated sources and receivers, such as the HP8753A and HP8720A, the source power was defined at the test port, and the receivers were provided with a factory calibration (called *samplet-cal* or *mixer-cal*) that roughly corrected for the frequency response of the VNA receivers.

These factory calibrations tended to be sparse in frequency (perhaps as few as 10 points across the frequency range) and therefore did not account for the fine-grain response of the receivers. As a result, the typical receiver flatness performance was on the order of $+ - 0.5$ dB. The complete error model for receiver measurements is shown in Figure 3.2.

Receiver calibrations were basic, essentially consisting of simple response normalizations of the receiver to a 0 dBm reference power. The essential steps were

1. Calibrate the source power to be exactly 0 dBm.
2. Connect the source to directly the receiver.
3. Normalize the receiver.

In the case of a reference receiver, step 2 consisted of simply leaving the power meter connected to the test port of the source so the match would be consistent.

This process has several drawbacks, listed here:

Any error or drift in the source power calibration became an error in the receiver calibration. Since only simple normalizations were available, the receiver calibration could be performed only at 0 dBm source power.

Any mismatch between the source and the power sensor, or the source and the receiver, would cause an error.

If any adapter was used to connect the source to the receiver, its loss and mismatch would be a direct error in the receiver calibration.

3.7.2 Modern Receiver Power Calibration

In more modern versions of VNA, the receiver response correction is enhanced by recognizing that during a source power calibration, the incident power is well known. The reference

receiver can be calibrated at the same time as the source power calibration occurs. This is sometimes called *guided power calibration*.

When the VNA port is terminated into a well-matched load, the reference-receiver forward tracking term, or RRF, is found by

$$RRF = \left. \frac{a_{1M}}{a_{1A}} \right|_{\Gamma_{Load}=0} = \frac{RTF}{STF} \quad (3.53)$$

When a power meter is used to measure the incident power, the mismatch is quite small, so Eq. (3.53) can be approximated as

$$RRF \approx \frac{a_{1M}}{P_{Meas}} \quad (3.54)$$

For a power measurement at the transmission test receiver, sometimes called the *B receiver*, the B transmission forward tracking (BTF) is defined as

$$BTF = \frac{b_{2M}}{b_{2A}} \quad (3.55)$$

The normal method for finding BTF is to perform a source power calibration at port 1, connect port 1 to port 2, and assume that $a_{1A} = b_{2A}$. If the reference receiver has been calibrated, then BTF term can be approximated as

$$BTF \approx \left. \frac{b_{2M}}{(a_{1M}/RRF)} \right|_{S21_Thru=1} \quad (3.56)$$

Here, the effects of the power sensor mismatch and port1/port2 mismatch and any loss in the thru-connection are not accounted for. However, until recently, this was the best estimate for receiver tracking that was available in commercial VNAs. But recently power calibration has been added as an option to the guided-power calibration (Keysight calls this the “SmartCal”) and so one seldom needs to perform an independent receiver power calibration.

3.7.2.1 Calibration for Power Sensor Mismatch

In the past, a simple source power measurement was the normal way in which the reference receiver tracking was determined. But from the signal flow graph, one can account for the match of the power meter as well to find the exact value of reference receiver tracking. From the signal flow graph, it is clear that the match of the power meter can affect the power measured for a given source setting and thus the extraction of the value for RRF. The value measured on the power meter is essentially a_{1A} and is related to the VNA source power incident to the test set, a_{1S} , as a function of STF, RTF, the source-match ESF, and the match of the power sensor as

$$a_{1A} = \frac{a_{1S} \cdot STF}{1 - ESF \cdot \Gamma_1} = \frac{a_{1M}}{RTF} \frac{STF}{(1 - ESF \cdot \Gamma_1)} \quad (3.57)$$

where Γ_1 is the match of the power sensor. The power meter calibration factor includes a correction for mismatch loss of the sensor such that the power displayed on the power meter

actually indicates the incident power to the sensor rather than the power absorbed by the sensor.

$$P_{Meas} = \frac{P_{Absorbed} \cdot \Delta PM}{1 - |\Gamma_{Sensor}|^2} = a_{1A} \quad (3.58)$$

Here, ΔPM is the difference between the power absorbed and the power meter reading. This is usually called the *power meter calibration factor* but also includes any offsets such as reference offset in the power meter.

From the previous two equations, the reference receiver tracking term can be computed exactly in the presence of mismatch as

$$RRF = \frac{a_{1M}}{P_{Meas} \cdot (1 - ESF \cdot \Gamma_{PwrSensor})} \quad (3.59)$$

Note that with this method, if the reference receiver and the power meter measurement are recorded at the same time, it does not make any difference to the result what the actual value of the a_1 source signal is. This removes the requirement for receiver calibration that an accurate source power calibration be performed first. In fact, this computation of receiver calibration can be done at the same time as the source power cal.

A 1-port error-corrected measurement of the power meter match can be acquired at the same time as the power sensor reading to determine $\Gamma_{PwrSensor}$.

This completes the calibration acquisition of error terms for the reference receiver measurements.

3.7.2.2 Response Correction for the Reference Receiver

In legacy VNAs and in most modern VNAs until recently, receiver correction applied the receiver tracking term in a simple way.

$$a_{1A_RcrvCal} \equiv \frac{a_{1M}}{RRF} \quad (3.60)$$

This simple response calibration does not provide an answer that exactly equals the actual incident power due to the source-match and DUT input match interaction.

To fully understand these effects, consider the block diagram shown in Figure 3.23. The directional-coupler is used at the end of the test port to sample the incident power. The match at the end of the coupler is changed while monitoring the incident power on the test port channel and the power reported in the reference receiver. Differences in these power values represent uncompensated mismatch errors. For proper evaluation, one must also consider the loss of the coupler, as described in Section 2.2.2.

Figure 3.24 shows the effect of changing the match of the DUT at port 1, on the measured value of the actual incident power, a_{1A} , along with the apparent or measured a_1 receiver power reading of the reference channel with a receiver correction, a_{1M}/RRF . In this case, the source ALC reference is not at the same point as the reference channel tap, so the value of a_{1S} varies as the load match changes and so does the value of a_{1M} . The reference channel reading properly captures the portion of variation of a_{1A} due to the DUT port 1 mismatch with the power source match, $\Gamma_{PwrSrcMatch}$. However, the reference reading *does not* properly capture the error due to the test set source-match, ESF, interacting with the DUT port 1 mismatch;

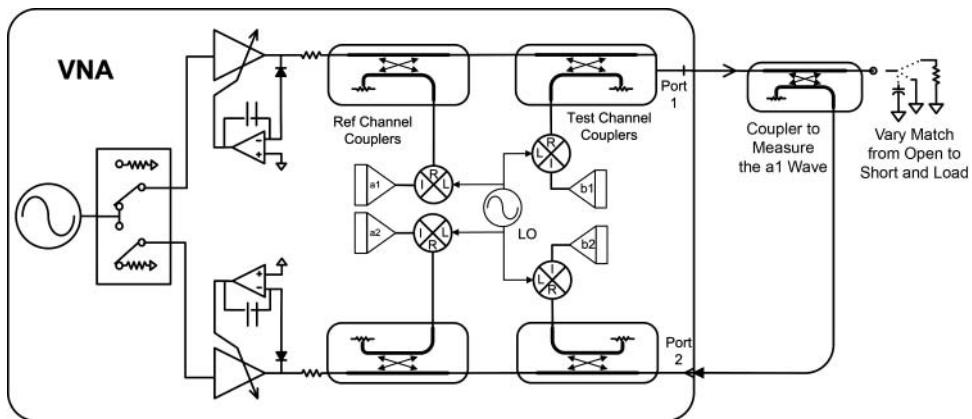


Figure 3.23 Block diagram for characterizing incident power mismatch.

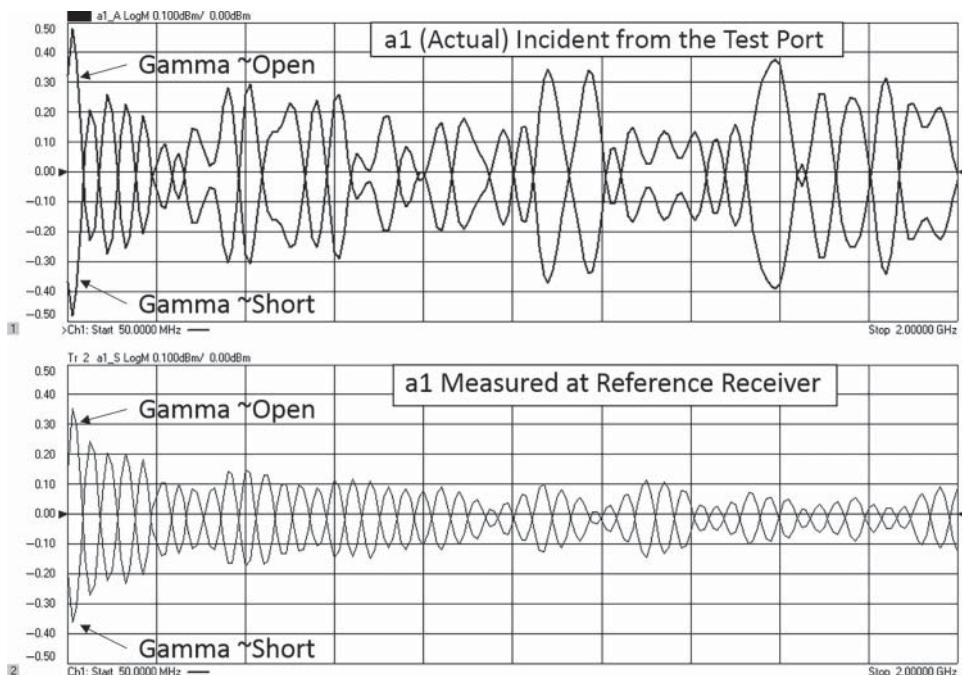


Figure 3.24 Ripple in incident power ($a1_a$) and measured source power ($a1_s$) due to DUT mismatch at port 1.

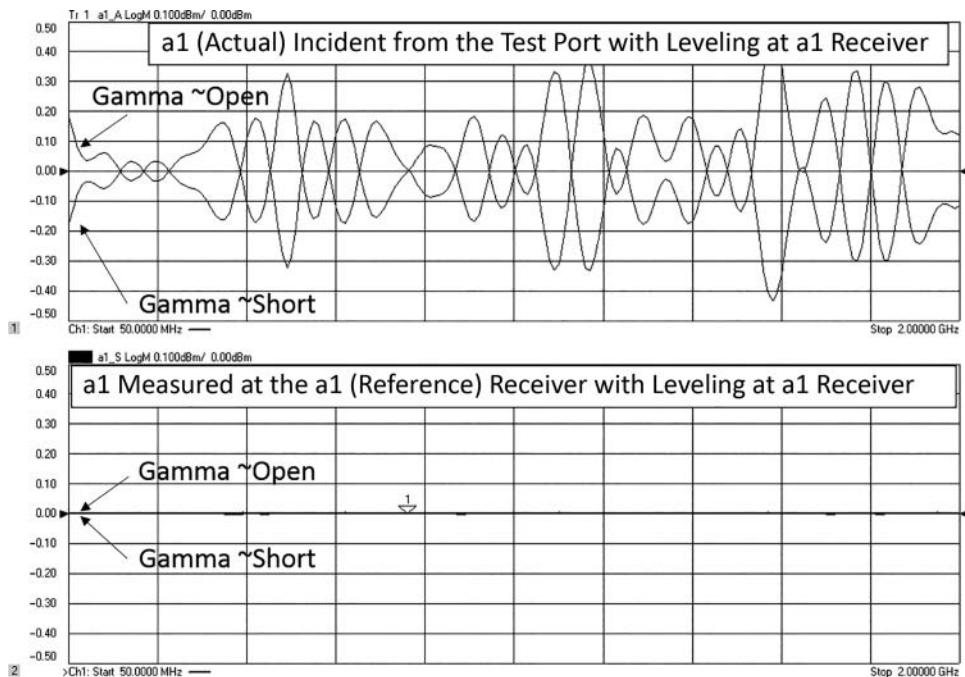


Figure 3.25 Ripple in the actual incident power (upper) when the ALC reference point is the same as the reference channel tap, causing a_{1S} to be constant (lower).

thus, the measured value of $a_{1A} \neq a_{1S} \cdot STF$ and, in general, a_{1A} (the actual incident power from the test port) has larger ripple although in some regions the source ripple can cancel the test port mismatch ripple.

In Figure 3.25, the ALC control is changed so that its reference and a_{1M} reference channel tap point are the same. As such, the value of a_{1S} does not change with the DUT match, which is reflected in the fact that the reference channel receiver measured power is perfectly flat (lower window). Any mismatch from the DUT port 1 to the source is detected by the ALC loop and compensated for. However, the power meter monitoring the incident power from the test port still has ripples indicating the effect of source-match after the reference coupler is not being removed. Here, it is clear that even with a response calibration, the reference receiver, which shows a flat trace, does not properly capture the true or actual incident wave applied to the DUT, as shown in the upper window of the figure. Thus, although the reference channel power with receiver calibration always reflects the error in source power because of the power source match, it does not reflect the error due to the test set ratio source match, ESF, as described in Section 2.2.2. This demonstrates that a simple response correction of the reference receiver is not sufficient to determine the incident power applied to a DUT because of the mismatch effect from the VNA source-match and the S11 effect of the DUT. In fact, since the ripple shown in the figure changes with every DUT, it cannot be pre-computed and removed. However, modern VNAs now employ match-corrected receiver calibrations, as discussed in 3.7.3.2.

3.7.3 Response Correction for the Transmission Test Receiver

For the testing power at port 2, the B test receiver response error correction becomes

$$b_{2A_RcrvCal} \equiv \frac{b_{2M}}{BTF} \quad (3.61)$$

However, this receiver calibration accuracy is still limited by the source and receiver mismatch terms. Until recently, this was the only calibration method available for power measurements in a VNA.

Figure 3.26 shows the measurement of an amplifier's input and output power (lower window), as well as S21 (upper window, light trace) and power gain computed as measured output power over measured input power (upper window, dark trace). The fact that there is substantial ripple on the input power in the lower window is because of the mismatch at the input of the DUT. The fact that there is some ripple on the output power is a natural consequence of the ripple of the input power, as well as the error due to mismatch between the amplifiers output impedance and the load match of the test system. The power gain term in the upper window accounts for the effect of the input power variation to show directly the variation of output power due only to the mismatch of port 2; that is, the difference between the S21 trace and the PowerGain trace is from an error in the measurement of the output power, due to mismatch ripple. Fortunately, most modern VNAs can now apply a full match correction to the output power to remove this effect, as discussed next.

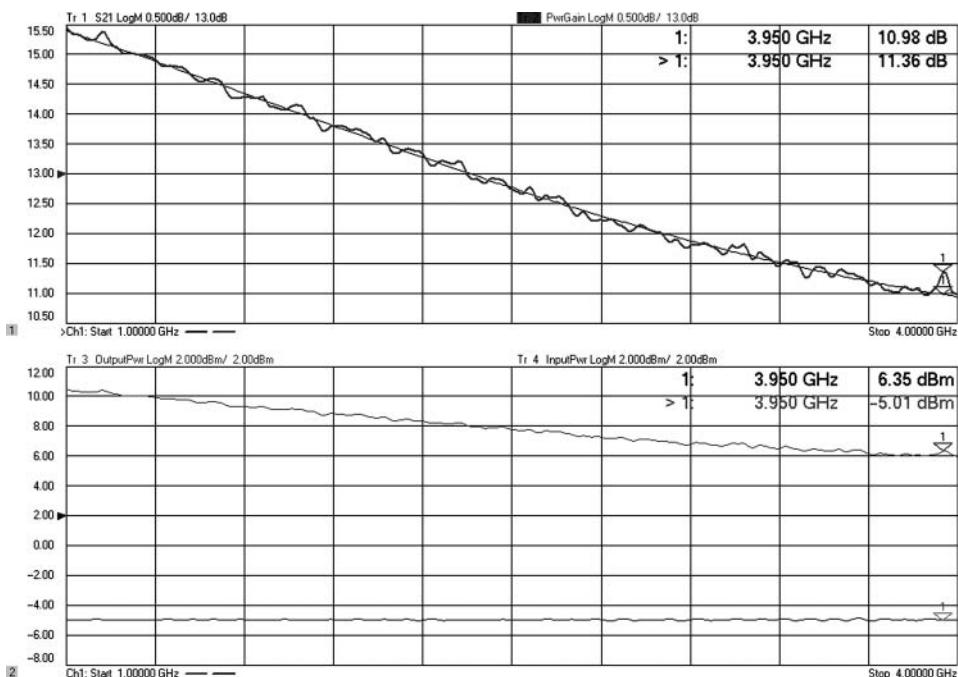


Figure 3.26 Power measurement of an amplifier after a receiver response calibration.

3.7.3.1 Enhanced Power Calibration with Match Correction

Power calibration in a VNA has traditionally been of lower quality than the S-parameter calibrations, and in many test systems, serious power measurements required the additional use of a power meter for only that purpose. Because power sensors have relatively good match and can be directly connected to the output port of a DUT, they provided a reasonable if inconvenient solution to the problem of accurate power measurements. The VNA provides nearly perfect measurements of gain, which is the key characteristic of a component, and direct measurement of power was seldom part of the VNA test process.

However, starting around the year 2000, VNAs were being increasingly used to measure mixers and frequency converters. Since there did not exist good methods for gain measurements of converters based on ratio measurements, the use of unratioed power measurements became necessary. This resulted in the development of match-corrected input and output power measurements for frequency converters. In late 2010, the full support of match-corrected power measurements for non-converting devices was introduced in a commercial VNA in the Keysight PNA-X. The details of mixer measurements will be described more fully in Chapter 7 but is based on the enhanced power calibration techniques described next.

3.7.3.2 Match Corrected Incident Power Calibration

The formulation of the measured reference receiver shown in Eq. (3.57) presents the key to the solution for the problem of mismatch correction for incident signal measurement. In the acquisition of the calibration, the RRF term is computed from this equation and the measured power on the power meter.

A solution for a fully corrected incident source power becomes apparent as

$$a_{1A_MatchCor} = \frac{a_{1M}}{RRF \cdot (1 - ESF \cdot \Gamma_{1M})} \quad (3.62)$$

where Γ_{1M} is the match presented at port 1 during the measurement of the incident power. The main source of uncertainty in the incident source power calibration is the error between the measured source power during calibration and the actual source power; this error is directly attributable to the errors in the power meter: power meter cal factor, power meter drift and noise, power meter reference cal error. For a high-quality power sensor, these are on the order of about 0.15 dB.

3.7.3.3 Match-Corrected Output Power Calibration Acquisition

Similar techniques can be used for the output power computation, to take into account the effects of the load match at the VNA output port. The acquisition and computation of the BTF term in Eq. (3.56) ignore source and load mismatch effects and require an ideal thru. The evaluation of the BTF term taking into account mismatch is

$$BTF = \frac{b_{2M}}{a_{1A}} \frac{[(1 - S_{11T}ESF)(1 - S_{22T}ELF) - (ESF \cdot ELF \cdot S_{21T} \cdot S_{12T})]}{S_{21T}} \quad (3.63)$$

Taking the case where the thru is well matched, so $S_{11} = S_{22} = 0$, this simplifies to

$$\begin{aligned} BTF &= \frac{b_{2M}}{(a_{1M}/RRF)} \left. \frac{[1 - (ESF \cdot ELF \cdot S_{21T} \cdot S_{12T})]}{S_{21T}} \right|_{S_{11T}=S_{22T}=0} \\ &= \frac{S_{21M} \cdot RRF \cdot [1 - (ESF \cdot ELF \cdot S_{21T} \cdot S_{12T})]}{S_{21T}} \end{aligned} \quad (3.64)$$

For a flush thru, this simplifies further to

$$BTF = S_{21FlushThruM} \cdot RRF \cdot [1 - (ESF \cdot ELF)] \quad (3.65)$$

The computation of BTF can be performed directly, but it can also be computed without additional measurements if a 2-port calibration has been performed, by recognizing that Eqs. (3.65) and (3.29) can be rewritten as

$$BTF = ETF \cdot RRF \quad (3.66)$$

In fact, Eq. (3.66) holds true regardless of the calibration standards used for the full 2-port calibration. Thus, if a reference receiver calibration and a full 2-port calibration have been performed, any of the test receiver power calibration factors can be computed. Some manufacturers combine the source power calibration, reference receiver calibration, and S-parameter calibration in a single guided power calibration function.

3.7.3.4 Match-Corrected Output Power Calibration Application

Equation (3.66) provides a good estimate of the B tracking error term and will greatly improve the response correction given in Eq. (3.61), removing the effects of port 1 and port 2 mismatch during the acquisition of the error term. However, during measurement of an DUT, the mismatch caused by the S_{22} of the DUT and the load match of port 2 will not be properly compensated by the simple response calibration.

The match corrected power measurement is computed as

$$b_{2A_MatchCor_1-Port} = \frac{b_{2M}}{BTF} \cdot (1 - ELF \cdot \Gamma_2) \quad (3.67)$$

where Γ_2 is the output impedance of the DUT, presented at port 2 of the VNA. A simple one-port corrected S_{22} measurement is sufficient to determine the value of Γ_2 , provided of course that the device is linear.

One might suspect that the output power of a DUT can be fully corrected if the input power to the DUT is known, by recognizing that the power into a matched load is simply the incident power times S_{21} ; and the match-corrected incident power is found by Eq. (3.62) so that

$$b_{2A_MatchCor_Z_0} = a_{1A_MatchCor} \cdot S_{21_Full_2-Port_Cor} \quad (3.68)$$

This is subtly different from the upper equation in that it will report the power as though the source and load were perfectly matched. Equation (3.67) gives the power incident from the

DUT without altering it for the mismatch error at the input. So while (3.68) gives that power that would have been measured if the source would have been 50Ω , Eq. (3.67) gives the power that is measured from the existing system's source-match. Some advanced techniques described in Chapter 6 use the reference receiver to level the input power so that the effective source-match is fully corrected for.

For non-linear devices, such as amplifiers in compression, for example, more complex non-linear analysis is required for a complete solution, details of which are described in Chapter 6 under the topic of load pull and X-parameters and active (hot) S-parameters.

3.7.3.5 Measuring Match-Corrected Reflected Power

To complete the measurements of all powers, the match-corrected reflected power can be found similarly to the match corrected output power by

$$b_{1A_MatchCor} = a_{1A_MatchCor} \cdot S_{11_Cor} \quad (3.69)$$

where S_{11_Cor} is the 1-, 2- or N-port corrected reflection at port 1.

3.7.4 Power Waves vs. Actual Waves

There is one subtlety in power correction that becomes important when considering power measurements where the intended load is not the reference impedance (50Ω), as in the case of a load pull measurement. The mismatch correction of Eq. (3.67) provides correction for both the tracking error of the receiving coupler and the mismatch between the non-ideal load of the test port and the output match of the DUT. Implied in the mismatch correction is also that the directivity of the test coupler is corrected for so that a reflected wave from the internal load of port 2 does not leak into the port 2 receiver through the test port isolation.

The correction provides the power reading that one would have seen had port 2 of the VNA been exactly 50Ω . If one were to measure the DUT on different test systems, with different non-ideal mismatches, this correction allows them to measure identical powers out of the same DUT. Without this correction, different mismatches on each system would yield different power measurements. Thus, we say this measures the power waves coming from the DUT.

But, these power waves are not the actual wave emanating from the DUT. The actual wave may be higher or lower (the higher the port 2 impedance, the higher the voltage wave, and the lower the current wave). Thus, with different port 2 impedances, different voltage waveforms at the DUT output will be created. This is the essence of a load-pull system, and in such a case, one wants to measure the actual wave, not the power wave, and the mismatch correction of Eq. (3.67) is not used. Instead, only the coupler loss and the reflected power leaking through the coupler directivity term are corrected for. The mismatch of the load becomes part of the measurement condition. Newer VNAs allow selection of the receiver power wave correction method.

3.8 Calibrating Multiple Channels Simultaneously: Cal All

Older VNAs had two or at most four channels, so calibration of the VNA across all the channels could be a little tedious, but with at most four channels, it was quite reasonable to manually perform the two to four calibrations. But modern VNAs can have up to 500 independent channels, with each channel having differences in frequency range, measurement type, number of points, power level, attenuator levels, and any of a hundred different settings. Additionally, the multiple channels could even represent different measurement classes such as standard (S-parameter), gain compression (2-D power/frequency sweeps), IMD, noise figure, mixer measurements, and differential or IQ measurements.

A recent advance in instrumentation control provides the capability to make a single connection of each calibration standard, at each port, and then measure and compute the error terms for all of the channels. This is called Cal All and available from Keysight Technologies.

Cal All is much more sophisticated than simply sweeping each of the channels for each of the calibration standards; rather, it creates a master calibration channel that consolidates all the channels' frequency content into a single master list and then sweeps through that list one time for each calibration standard at each of the measurement ports. It can use any method for performing the calibration such as TRM, unknown thru (SOLR), or QSOLT. However, during the thru portion of the calibration sweep, the VNA automatically takes an additional sweep for any change in any source or receiver attenuation between channels or for any change in switch path.

Figure 3.27 shows the initial setup dialog for the Cal All function. In this case there are six active channels, but only five of them are selected for calibration. Sometimes, one of

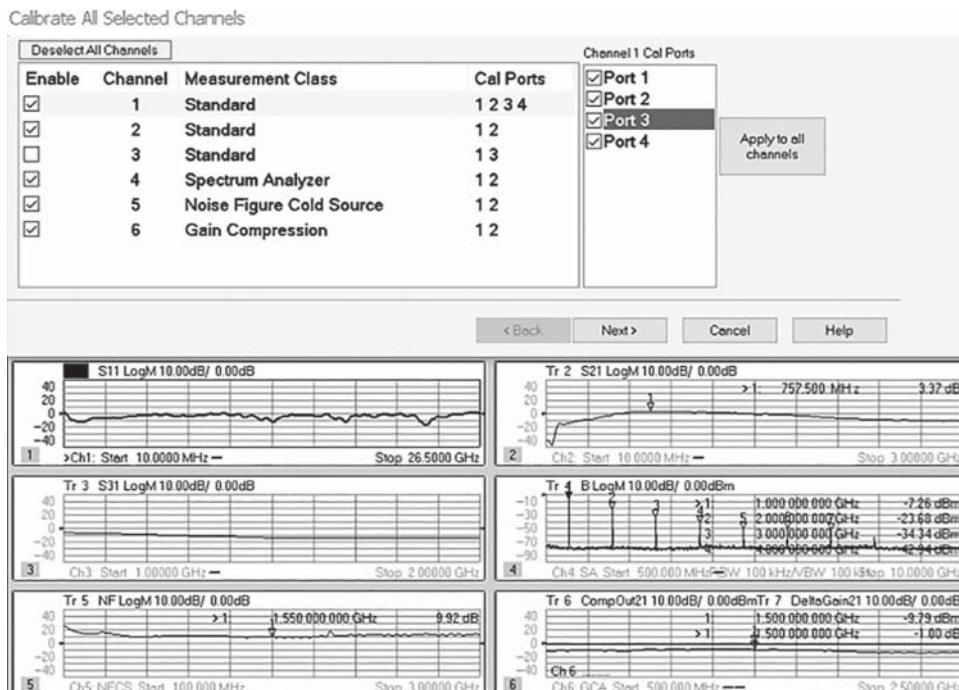


Figure 3.27 Setup for the Cal All function.

the channels is used for some alternative measurement and is not needed to be calibrated, and in this case Channel 3 is not selected for calibration. Channel 1 is selected for a 4-port calibration, and the remaining channels are selected for only 2-port calibrations. Notice that the Cal All function allows a single calibration to be performed to support a variety of different measurement classes, such as gain compression and noise figure; at the same time, it is used to calibrate the standard channels. The Cal All application also provides a means to select various parameters for calibration setup that might be associated with individual measurement classes such as choosing the noise calibration noise source in the noise figure channel or choosing the number of calibration points in the spectrum analyzer calibration.

Once the calibration channels are chosen, the Cal All application allows setting attributes for the master calibration channel. This allows, for example, the calibration IF BW or source power to be set for best performance of the calibration, even when the measurement channels might have completely different settings. A common occurrence is when the DUT to be measured is a high-gain amplifier requiring very low source power. In such a case the calibration will be quite noisy due to the low power at the reference and test receivers. Cal All provides a means to set the source power for calibration independently from that for the measurement, as illustrated in Figure 3.28. Here, each of the source and receiver attenuators may be set for optimum performance during the calibration. Best performance typically occurs with some source attenuation set if the power drive for most channels is below 0 dBm. As described in Chapter 2, the source-match of the system is normally set by the directivity of the reference coupler, but setting some source attenuations provides twice the attenuator's set value in isolation of the source-match, makes the power source match much closer to the ratio source match, and typically makes the power somewhat more accurate, without the use of receiver leveling.

Further, if there are different settings of attenuation for different channels, the change in source-match is minimized between channels if the master channel has some small (5 or 10 dB) source attenuation set. During the Cal All process, the source-match for each attenuation state is determined and corrected for, but the residual error will be less if the raw difference in source-match is minimized for the Cal All setup. The dialog shows two additional choices, one for Noise Reduction and one for Mechanical Devices. The Noise Reduction tab allows the user to optimize the IF BW for the calibration (default is 1 kHz, which is usually a good choice) as well as allowing one to add sweep-to-sweep averaging to the calibration

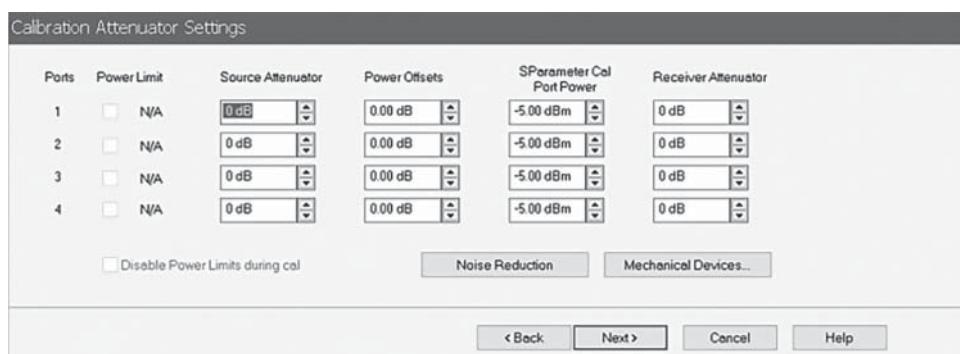


Figure 3.28 Dialog for setting the master channel power and attenuator values.

	> Cal All: Hold	Ch 1 : Hold	Ch 2 : Hold	Ch 3 : Hold	Ch 4	Ch 5 : Blocked	Ch 6 : Blocked
Meas Class	Standard	Standard	Standard	Standard	Spectrum Ana	Noise Figure	Gain Compres
Port 1 Bypass	Thru	Thru	Thru	Thru	Thru	Thru	Thru
Port 2 Bypass	Thru	Thru	Thru	Thru	Thru	Thru	Thru
Port 3 Bypass	Thru	Thru	Thru	Thru	Thru	Thru	Thru
Port 4 Bypass	Thru	Thru	Thru	Thru	Thru	Thru	Thru
Port 1 Atten	0 dB	Auto 0 dB	Auto 5 dB	Auto 0 dB	Auto 0 dB	20 dB	0 dB
Port 2 Atten	0 dB	Auto 0 dB	Auto 5 dB	Auto 0 dB	Auto 0 dB	0 dB	0 dB
Port 3 Atten	0 dB	Auto 0 dB	Auto 5 dB	Auto 0 dB	Auto 0 dB	20 dB	0 dB
Port 4 Atten	0 dB	Auto 0 dB	Auto 5 dB	Auto 0 dB	Auto 0 dB	20 dB	0 dB
Rcvr A Atten	0 dB	0 dB	0 dB	0 dB	0 dB	0 dB	0 dB
Rcvr B Atten	0 dB	0 dB	0 dB	0 dB	0 dB	0 dB	10 dB
Rcvr C Atten	0 dB	0 dB	0 dB	0 dB	0 dB	0 dB	0 dB
Rcvr D Atten	0 dB	0 dB	0 dB	0 dB	0 dB	0 dB	0 dB
Noise Tuner Sw	Internal	Internal	Internal	Internal	Internal	Internal	Internal
Noise Rcvr Sw	Normal	Normal	Normal	Normal	Normal	Noise	Normal
Port 1 Ref Sw	Internal	Internal	Internal	Internal	Internal	Internal	Internal
Port 2 Source	Src1Out2	Src1Out2	Src1Out2	Src1Out2	Src1Out2	Src1Out2	Src1Out2

Trigger... OK Cancel Help

Figure 3.29 Cal All Mechanical Devices dialog.

data acquisition sweeps. Sweep-to-sweep averaging can help to reduce variation caused by environmental factors such as vibration of the test port cables, or other interference. IF BW reduction in effect increases the data averaging at each point, to reduce noise, but slow-rate vibration, or other stability issues, is not removed if the IF BW measurement (typically 1 ms) is shorter than the variation occurrence (maybe 1–10 s). In these cases, over the short term the variation is correlated and is not removed by IF BW reduction, whereas in sweep-to-sweep averaging the long time period between sweeps provides a means to de-correlate the variation, thus enabling the averaging function to reduce its effect.

The Mechanical Devices dialog, shown in Figure 3.29, displays the values and positions of mechanical attenuators and various path configuration switches for all the channels. Further, it shows the current status of the channels, whether in hold mode (Channel 1 through Channel 3), currently sweeping (Channel 4), or blocked (Channel 5 and Channel 6). A channel will be blocked from sweeping if its mechanical switching selection doesn't match the active channel. This prevents the continuous cycling of mechanical switches that will wear out the switches prematurely. Also shown is the Cal All master channel setting in the first column; it is always a standard channel. The settings default as shown in the Cal All channel, but if many or all of the other channels have some switch in a different position, one can simply click on the appropriate table entry to change the value of that setting in the Cal All channel. This dialog, also available from the setup menu, is a convenient way to see what all the various channel settings are. While it is best to make the Cal All master channel match the common settings of other channels, it is not really required because for any difference in settings, the Cal All channel will acquire an additional sweep to fully characterize the change in state and provide a complete calibration of that state.

Finally, all the channel settings are consolidated into the master channel, which provides a full, guided power calibration, shown as the first row in Figure 3.30, with only the unique frequency points. Overlapped and duplicate points are removed from the master calibration channel. This can be a huge savings in time.

The screenshot shows a software interface titled 'Cal All'. At the top left is a button labeled 'Cal All'. Below it is a table with the following data:

Contributor	Channel	Ports	Frequency	Points	IFBW
SParameter Cal	200	1 2 3 4	10.000 MHz - 26.500 GHz	793	1000
Standard	1	1 2 3 4	10.000 MHz - 26.500 GHz	201	
Standard	2	1 2	10.000 MHz - 3.0000 GHz	201	
Spectrum Analyzer	4	1 2	500.00 MHz - 10.000 GHz	202	
Noise Figure Cold So...	5	1 2	100.00 MHz - 3.0000 GHz	101	
Gain Compression	6	1 2	500.00 MHz - 2.5000 GHz	101	

Below the table are navigation buttons: '< Back' and 'Next >' on the left, and 'Cancel' and 'Help' on the right.

Figure 3.30 Cal All creates a master cal with only unique frequency points.

Consider an example of a frequency converter with wideband LO (say 1–20 GHz), characterized for RF-to-IF response over 201 points in a 500 MHz channel (2.5 MHz per point), at each LO frequency (perhaps every 100 MHz). This would result in about 191 channels, each channel having 201 points. The calibration time would be extremely long and not at all practical. Even if all the standards are measured only once, there would be about 40 000 data points to measure. Cal All will remove overlapped settings and reduce the number of unique points to just about 7500 and require a single sweep per calibration standard. A normal 2-port calibration takes from 15 to 20 sweeps for a mixer measurement, depending on cal type, and, at a 1 kHz IF BW, will take about eight seconds per sweep for a total calibration time of two to three minutes. Without Cal All, doing a normal calibration process, the calibration time would be on the order of 10 minutes for the sweeps alone.

The Cal All settings also allow one to choose the port for the initial power calibration; the initial calibration uses a power meter to establish an absolute power reference on one of the receivers. The S-parameter calibration is used to transfer the absolute power and relative phase of one receiver to all the other receivers.

During the S-parameter calibration each different attenuator and RF switch path setting is properly and exactly characterized and accounted for. This is perhaps the most innovative part of Cal All. Normally, if a channel is calibrated in one attenuator setting and then the attenuator value is changed, the resulting calibration will not be exactly correct, as the VNA code simply shifts the values by the nominal attenuator setting and does not accommodate for the change in match, or any frequency response or ripple in the attenuator. This typically leads to about a 0.5 dB error in the calibrated response. However, with Cal All, during the thru-line acquisition process, the calibration engine takes additional sweeps for every different attenuator setting and switch path. Further, since the Cal All computation engine knows the Cal All channel attributes of power level, attenuator setting, and IF BW, it can automatically adjust these settings to accommodate properly for the change in attenuator value. For example, if the Cal All channel is set to a 10 dB attenuation state and one of the other channels has a 30 dB attenuator state, the Cal All channel will do an additional thru measurement of the 30 dB state, but change the IF BW to 100 times smaller, to reduce the noise on the measurement receiver associated with the change in power due to the larger attenuation.

The exact characterization of a switched path is achieved through a measurement and computation similar to the QSOLT method. The QSOLT method requires a 1-port calibration at one of the test ports and a known thru (sometimes called a *defined thru*) measurement. During the initial 2-port calibration on a path, an unknown thru calibration is performed in the Cal All master channel configuration. This provides two 1-port calibrations at each port, and at the end of the calibration, the characteristics of the unknown-thru are determined and can be de-embedded from one of the ports to provide a flush thru type of characterization at either port. When an attenuator or path change occurs, say a changing source attenuator at port 1, the unknown thru is de-embedded from port 2 to provide a 1-port calibration mating with port 1. The thru is now defined as a flush thru, and an additional QSOLT thru sweep is acquired. Thus, there are now a 1-port cal and a defined thru cal for this new state, which is enough to compute all the new error terms, source-match, load match, transmission tracking, and all the response terms such as the a1 tracking or b1 tracking term. In a similar way, if a receiver path is switched on port 2, the unknown thru is de-embedded from port 1 to make a flush thru interface at port 2. An additional QSOLT thru port sweep is performed to find the error terms associated with the new port 2 receiver attenuation state. These added thru measurements are acquired for each path change, and at the end, all the combinations of paths, which represent each channel's unique configuration, are computed and saved as independent, channel-specific calsets.

Thus, Cal All can provide an exact calibration, over a nearly unlimited number of channels and configurations, with only a single step for each calibration standard.

3.9 Multiport Calibration Strategies

Multiport test systems, as described in Chapter 2, require calibrations just as the 2- or 4-port systems do, but as the number of ports grows, the calibration complexity grows as the square of the number of ports, if the traditional approaches are used. New calibration techniques can greatly reduce both the number of steps and the total time to perform calibrations for systems with a large number of ports.

3.9.1 $N \times 2$ -Port Calibrations: Switching Test Sets

One of the most common scenarios for multiport characterization is an $N \times 2$ -port configuration of a switching test set, as described in Chapter 2. These provide a switch tree from each 2-port VNA to any N-ports and can support many paths, up to the maximum number of paths for an N-port system, which is

$$\text{Number of Paths} = \frac{N \cdot (N - 1)}{2} = \left(\frac{N^2 - N}{2} \right) \propto N^2 \Big|_{\text{Large } N} \quad (3.70)$$

Many times a switch tree is configured to allow connections from the left side of a DUT (inputs) to the right side of a DUT (outputs) where you might have K inputs and L outputs. If each path is to be measured from any input to any output, then there exists K times L paths.

Note that this is fewer paths than the total paths where one might measure from one input to another input, and the total paths would follow Eq. (3.70) where $N = K + L$.

$$\text{TotalPaths} = \frac{(K + L) \cdot (K + L - 1)}{2} \quad (3.71)$$

So for a left/right or input/output only configuration, the path scales as the product of inputs and outputs and increases linearly with each; for a full cross-bar configuration, the path scales as the square of the number of ports.

In either configuration, either full cross-bar or switch tree, utilizing standard methods, there are a tremendous number of paths to calibrate. Consider a real example from a large aerospace company, measuring a device with approximately 30 inputs and 30 outputs. This company utilized standard methods and had to perform 900 full 2-port calibrations to measure each required path. (In fact, if every possible path were measured, it would have required 1800 paths.) The measurement was performed inside the thermal chamber, and each path was calibrated using a mechanical calibration kit and the full 2-port calibration method. Each calibration took about five minutes to complete, so a complete system calibration required about three weeks to complete, considering breaks, re-measurement for poor or mistaken correction steps, etc. Even with an Ecal, the total calibration time was cut about in half only, as an unknown thru method using a relatively long thru cable was required, so the Ecal needed to be connected and disconnected 900 times.

However, new understanding of calibration programming and advanced de-embedding, along with some convenient new firmware features, reduced the total calibration requirement to just 60 1-port calibrations and 1 thru calibration. These new methods depend upon the fact that there are a few common reflectometers (with the associated receivers) that are switched to different paths through a switching test set. As such, the error terms of any one path are related to error terms of every other path through a simple 2-port S-parameter characterization of the path difference. The idea behind these new methods is to characterize the path to each port, along with one 2-port characterization, and then compute all the 2-port calibrations for any path pair.

There are two different implementations that provide about the same improvement. Here are the outlines of the improved methods:

```
Method 1: Load redundant 1-port calibration Eterms, and redundant thru
for each port n of N
    Set the switch to connect the VNA to the port
    Perform a single 1-port calibration on each port, save each
    calset
    Make a single, unknown thru connection through any path from port 1
to port 2
    For each port pair J and K
        Send a command to load the Eterms for ports J and K for the
appropriate calset
        Take the unknown thru measurement
        Compute the calset
```

With this method, the 1-port calibration data is reused for every port pair. All that is required is repeating the same unknown thru step for each pair. Since the unknown thru can be through

any path, it is not necessary to change the thru standard to connect to the ports being calibrated, only that the VNA port 1 is connected in some way to VNA port 2.

The second method relies upon de-embedding math to simplify the calibration steps. Refer to Chapter 11 for more details on de-embedding.

Method 2: De-embed Path Differences

```

Perform a 2-port Calibration on 1-port pair, n1 and n2, switched
through to VNA ports 1 and 2. Call this the primary calset
For every other port n3 through N
    Switch the test set path to each port and perform a 1-port cal
    Compute the de-embedding network from the primary calset to
this 1-port calset and save the S2P file associated with the path difference
For every port pair J and K
    Apply the primary calset, then de-embed the Jth and Kth path
difference from the primary calset

```

This method requires a separate channel for each de-embedding network to be applied. An alternative is to use the remote calset de-embedding function, which allows one to programmatically chose a calset and a de-embedding network and create a new calset with the path difference accounted for. For example, in the case of the 30-input/30-output DUT described earlier, one can first create 30 input calsets where the 30 input paths are removed. For each of these 30 input calsets, one can now create 30 more input/output calsets where each the output path is removed from every one of the 30 calsets, generating a total of 900 calsets for the 900 paths. Thus, one can create 900 paths of 2-port calibration through a switching test set with only 60 1-port calibrations and one thru step.

3.9.2 *N*-port Calibration: True Multiport

In the case of a true-multiport VNA, where each port has a full reflectometer with reference and test receivers, the methods in the previous sections cannot be utilized as there is not a common set of reflectometers from which to derive path differences through a switching matrix. Instead, each port has its own unique set of error terms, and every port pair has its own unique set of transmission tracking terms. It is not possible, in general, to compute the error terms of one path from terms other paths. But there is a shortcut that can be done for true-multiport calibrations. All that is needed is to be able to link from one port to every other port, through some 2-port calibration, to enable the computation of all *N*-port path terms. There are two general methods for generating the *N*-port calibrations needed.

3.9.2.1 Multiport Unknown Thru

The first method is simply a guided calibration using the unknown thru method, where a 1-port calibration is required at each port, and a set of linked unknown thru connections. Having linked connections means the set of thru standards used allows one to traverse from one port to any other port through a series of thru connections. A simple example is connecting from port 1 to every other port. Since port 1 is common to all the ports, the transmission tracking from port 1 to any port, e.g. port N or port M, is measured. To compute the transmission

tracking from port N to port M, one simply combines the tracking terms Etf_{N1} and Etf_{1M} and Erf_{11} to obtain Etf_{NM} .

$$Etf_{MN} = \frac{Etf_{N1} \cdot Etf_{1M}}{Erf_{11}} \quad (3.72)$$

An alternative to having one common port is to link through multiple ports. Generally, each link adds some additional uncertainty, so using a single common port typically gives the lowest uncertainty (the exact uncertainty depends upon the match of each particular port and port pair).

Figure 3.31 illustrates multiple linked paths by way of an example of a 12-port probed system. This configuration of six ports on the left and six ports on the right is a common configuration when evaluating the coupling of pairs of differential lines. The six ports on each side represent three differential pairs. This configuration allows one to measure the differential cross talk between nearest neighbors and next-nearest neighbors. Here we presume that the probe cards have already been calibrated against on-wafer open/short/load standards that match the probe pitch. Since the probe configuration is fixed, it is difficult to make standards that can connect port 1 to every port. But it is possible, through only two probe positions, to make a linked list of thru connections. In the first probe position, thru paths for port pairs 1 and 2, 3 and 4, 5 and 6, 7 and 8, 9 and 10, and 11 and 12 are measured. In the second position, ports 3 and 2, 5 and 4, 7 and 6, 9 and 8, and 11 and 10 are measured. Thus, there are 11 paths measured, and we can link every port to every other port.

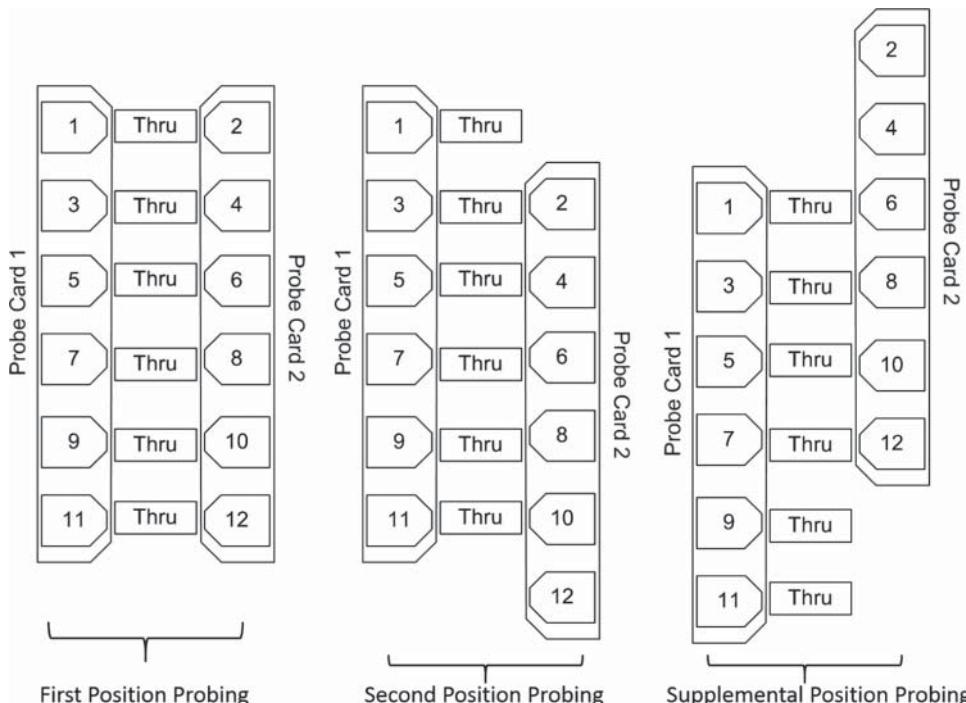


Figure 3.31 Using multiple Thrus to link ports.

But the linkage becomes quite long for ports 1 and 12, where the transmission tracking link includes all 11 measured thru paths to get a tracking term from 1 to 12. In fact, this is typically acceptable in the case of measuring nearest neighbor crosstalk because each differential pair has only three links at most (e.g. to link the second differential pair, ports 5, 6, 7, and 8, we need to have the thrus from 5 to 6, from 6 to 7, and from 7 to 8). The added uncertainty on the ports that are farther spaced (e.g. 1 and 12) likely does not matter as the loss factor will be quite high for those measurements.

Note that adding a supplemental probe position, as shown in the figure where port 1 is positioned to measure a thru to port 6, the number of thru connections is greatly reduced for far-away ports. Now 1 and 12 are linked through the 1–6 path, the 6–7 path (of the second position), and the 7–12 path. The 1–11 path requires one more link, the 11–12 link (from the first position).

Since the probe movement is only in the vertical direction, it is quite reasonable to have additional supplemental path measurements to further reduce the linkages needed. The choice of thru paths to use can be set in the calibration wizard by the user.

But, in reality, if the calibration standards are good, the uncertainty of linkages is small. Thus, if one has an N port multiport system, one needs N 1-port calibrations and N – 1 unknown thru measurements.

3.9.2.2 Multiport QSOLT

Generating an N-port calibration requires a lot of 1-port calibrations, and in some systems these might be tedious to perform even with the use of an Ecal module. Another method, utilizing the QSOLT algorithm, greatly reduces the total number of connections and can make a quite simplified calibration.

QSOLT requires a 1-port cal on one port and then a known thru (defined thru or flush thru) to every other port. For most connector types, this means a mating pair (male on one port, female on the other), but most devices have the same connector on each port (typically female). If one defines the calibration according to the DUT connectors, the calibration process will normally not show the QSOLT calibration as an option, so one must “fool” the system in one of two ways.

The first way is to do a full 2-port unknown thru calibration between the QSOLT reflection port (call it port 1) and any other port. Do the unknown thru step last, and at the end of the calibration, measure the unknown thru and save it as an S2P file. Then modify the calibration kit associated with the connector to add a data-based 2-port standard, call it a memorable name such as “QSOLT-Cable,” and save the kit. The data-based standard has defined connectors, in this case female to female, and uses the measurement of the S2P file to define the thru parameters. Now that particular connector type cal-kit has a female-to-female defined thru standard, and it will allow a QSOLT calibration method to be defined for each thru path. Go through the calibration wizard to redefine each thru to be a defined thru, and the cal steps will be very easy: the cal steps become a 1-port cal on port 1, and a thru from port 1 to every other port. In the case of a large N-port setup, there are only N connections (a 1-port connection for the one-port Ecal, and N - 1 thru connections with the QSOLT cable).

An alternative way to accomplish a similar calibration is to define the DUT port 1 as a male connector (presuming all the DUT ports are female, which means all the VNA test ports are

male) and add cable with a female connector at the end to mate with all the other ports on the test system. Now from the new port 1, which is the end of the cable, to all the other ports, we have a male-to-female flush thru connection.

Perform first a 1-port cal at the end of the cable (using male standards or male Ecal, on the female end of the port 1 cable that was added) and then a QSOLT flush thru calibration to each other port. You now have an N-port calibration at each port, but port 1 includes an extra cable. To remove the port 1 cable effects, remove the cable from port 1, and perform another calibration (this time using female standards at the true port 1 interface, which itself might be a cable) and save the one-port cal. From this 1-port cal and the 2-port QSOLT cal from port 1 to any other port, you can compute the 2-port characteristics of the port 1 QSOLT cable using the cal-plane manager available in some VNAs or using the de-embedding math described in Chapter 11. De-embed this cable from port 1 (the cal-plane manager allows saving this as a new calset), and you have a complete N-port calibration, at the proper reference planes, without modifying a cal-kit.

3.10 Automatic In-Situ Calibrations: CalPod

All the previous discussions with respect to calibration presumed that the user first calibrates the VNA system (with cables, switch paths, etc.) and then connects to the DUT and makes the required measurements. In some cases, the stability of the system, such as the flexing of long cables or poor repeatability through layers of switching, causes the calibration to degrade quickly. In some cases, just the process of moving the cables from the calibration standards to the DUT can degrade the calibration by a significant amount. This condition is often seen where long test port cables, up to 10 m or more, are used to connect the VNA to a DUT and may be placed in a thermal chamber or even a thermal-vacuum (TVAC) chamber. The connection to the DUT is done one time, and the thermal chamber is cycled. The cycling of the chamber, often over as much as -40°C to $+80^{\circ}\text{C}$, quickly degrades the calibration as the cables flex change. For these conditions, the normal calibration techniques are completely inadequate.

Recently, some instrument manufacturers have created automatic, in-situ calibration modules (e.g. Keysight CalPods) to act as a kind-of in-situ Ecal, and they can recover the test system from a variety of degradations due to the issues mentioned.

The test setup for using CalPods in a switched-path, multiport configuration in a TVAC chamber is shown in Figure 3.32. The control cables can be multiplexed up to 48 ways to support large number of ports in the chamber. In some cases, multiple devices are tested at the same time, so allowing them all to be cycled through thermal and vacuum cycles together can save significant time.

In the configuration shown, the switch matrix can be any arbitrary configuration from a simple switch tree to a full cross-bar switch. Each path of the switch is connected through a CalPod, which will connect to a port of the DUT in the chamber.

While this may seem like a lot of hardware to throw at a stability problem, the difficulty in getting good RF measurements inside a chamber that is controlled over such a wide temperature range cannot be discounted. In at least one commercial satellite test situation, without CalPods, every cable used inside a thermal chamber was required to be pre-tested over the entire temperature range, and data at each test temperature for the cable was recorded as an

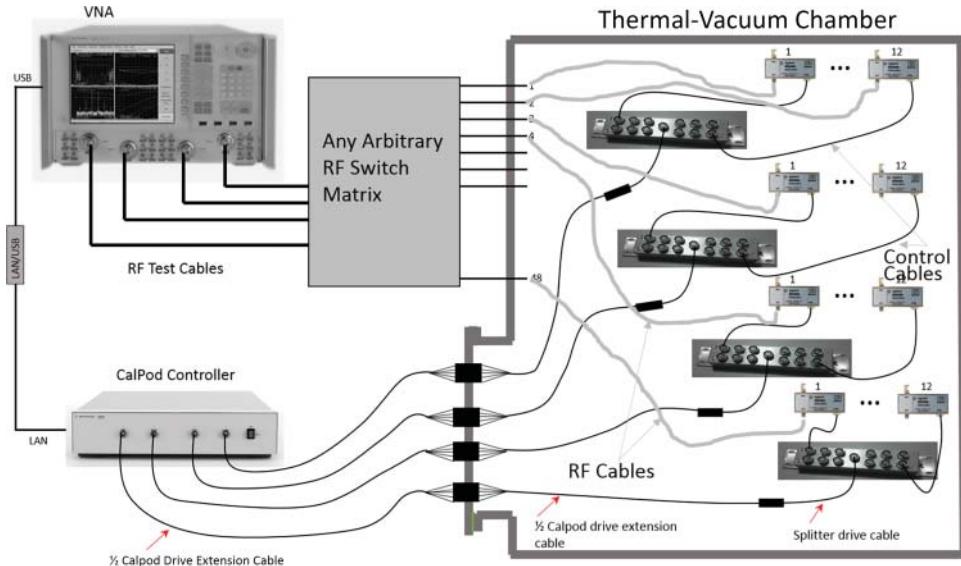


Figure 3.32 Multiport TVAC test setup with CalPods in the chamber.

S2P file. This testing took up to six weeks to complete (there were about 100 cables). Utilizing CalPods removed the need for any pre-testing, data capture, and post-processing, in addition to giving much better RF results.

3.10.1 CalPod Initialization and Recorrection

The basic concept for CalPods is that they contain three or more switchable reflection standards, along with a bypass through state. These are used for recorrection, but before doing a recorrection, the standards must be measured against a known calibration, which should be performed at the DUT reference plane beyond the CalPod. This is known as the *initialization* step. The CalPod is connected to the end of the test cable, just before the DUT, and the instrument is set for one or more measurement types and calibrated channel by channel or using Cal All (see Section 3.8). This calibration is done with the CalPod in-line and provides a calibrated reference plane at the output of the CalPod. If the CalPod is not able to be connected directly to the DUT (perhaps the connectors are too close), a short semi-rigid cable can be added, and the initialization calibration is performed at the end of the semi-rigid cable. Immediately after the calibration is complete, the CalPod standards should be measured in a process known as *initialization*. Here, using the 1-port calibration of the reference plane, the three or more internal states are characterized and saved in the instrument. Now, if there is any change, drift, or even when replacing a test port cable between the VNA test port and the CalPod module, the system can be recorrected by remeasuring these internal standards and computing the complex S2P difference array associated with the change. This difference is automatically de-embedded from the current calset to produce a new calset with the change removed.

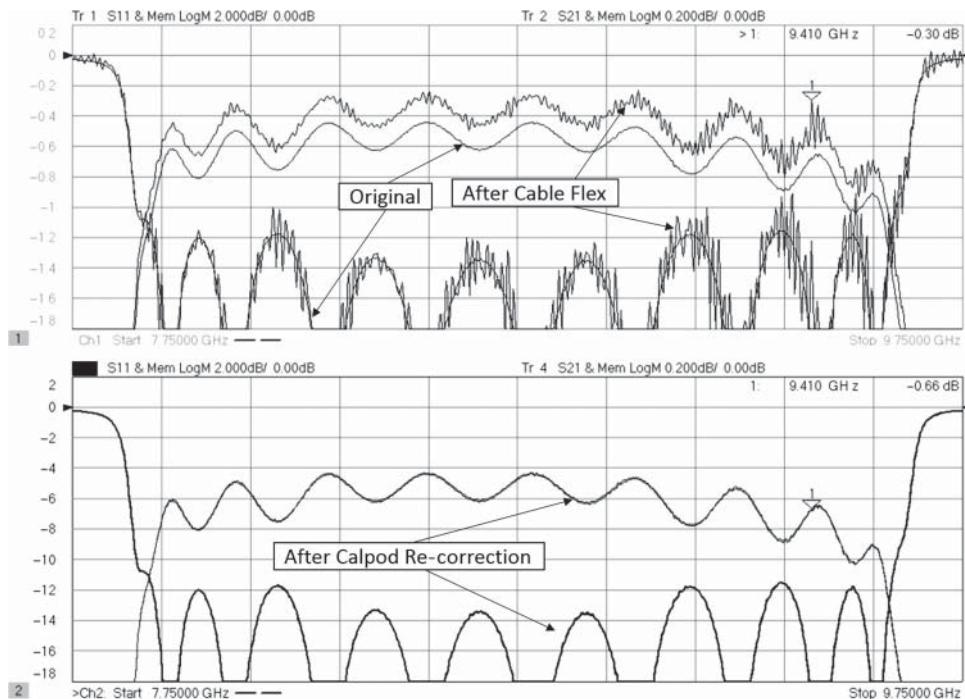


Figure 3.33 (Upper) After flexing the test cable, (Lower) after re-correction.

For thermally compensated CalPods, the internal states are characterized over a wide temperature range (typically -30°C to $+80^{\circ}\text{C}$, but it may be extended as far as -40°C to $+90^{\circ}\text{C}$ or more) such that the change or drift in the internal reflections are known as a function of temperature (the CalPod also includes an internal temperature sensor). In this way, re-correction can be achieved over wide temperature ranges.

An example is shown in Figure 3.33 of a measurement of a filter, connected to the VNA with two high-quality, 10 m long cables (each cable has about 10 dB loss at these frequencies). This is a rather extreme case, but it does come up when testing in TVAC-type conditions. The filter is measured in the passband to demonstrate the CalPod facility. In the upper window, the darker trace is the original S21 and S11 response of the filter immediately after calibration. The light gray response is the measurement after the moving the 10 m cables. Even though the cables were carefully moved, their change in phase and response due to the movement was enough to add significant ripple to the measured response. Further, the system was left overnight, and drift in the S21 level is also noted.

The lower window shows the result after doing a CalPod re-correction. Before re-correction, the S21 had excess ripple of over 0.1 dB and drift of nearly 0.2 dB, and the S11 had added more than 2 dB of ripple. After in-situ re-correction, there is almost no discernable change in the response from the original. Careful measurement shows the S11 is nearly identical, and the S21 trace is within 0.01 dB of the original response.

While this case utilized quite long cables, they were of good quality. A similar degradation can occur if shorter, but lower-quality, cables are used. This kind of effect can also occur if there is a series of switches between the VNA port and the DUT test port. And of course, if the cables are inside a temperature chamber, they will see substantially more change.

3.10.2 CalPod-as-Ecal

The CalPod recorrection is based on initializing the setup with a normal calibration, such as an Ecal, and then using the initialization function to measure the internal CalPod standards against the Ecal. In this way, the CalPod is really a kind of transfer standard, whose value is not particularly known but is stable.

One newer capability added recently is the ability to apply the Ecal user characterization function to a CalPod, as an in-situ device. This process follows two simple steps: (i) connect the CalPod to VNA test port (don't use any flex cables) and do a 1-port calibration using Ecal or mechanical standards at the CalPod output port, and (ii) run the Ecal characterization function on setting the control as a CalPod-as-Ecal. In this way, the internal standards are measured against the Ecal-calibrated output port and saved in the VNA memory. After this characterization, the CalPod can be placed anywhere and used instead of an Ecal, with one caveat: it is a 1-port Ecal referenced to the output port, not a 2-port Ecal.

To use the CalPod-as-Ecal in a 2-port (or multiport) calibration, one must have a CalPod-as-Ecal on each port where a 1-port calibration is required. An unknown thru is used between the CalPods. This provides an accurate and stable calibration and may be the best method for making low-loss measurements. If the DUT has a loss of less than about 10–20 dB, it can be used as the unknown thru standard. In such a case, the CalPods are connected to the DUT, and the CalPod is used as an Ecal for the 1-port portion, the DUT is used as the unknown thru for the through response portion, and at the end the DUT is measured, all without making a single disconnection and reconnection. This means no cable movement and connector repeatability error.

3.11 Devolved Calibrations

The calibrations described in earlier sections provided full correction for all the normal error terms found in a measurement. However, in some instances the full 2-port correction will be inconvenient and, in other cases, will provide poor results depending upon the particular configurations of the measurement system. In these cases, a devolved, or lower-order, calibration may prove a better choice. Some VNAs allow you to directly devolve an N-port calibration to a lower-order calibration. This may be convenient when making a test system with N-port capability, for example, an eight port system, but where you want the measurements to be sets of two ports; an example of this is a so-called multisite fixture where (in the case given) four identical devices are to be measured at the same time. This case is often found in high-speed manufacturing where multisite part handlers are common.

Figure 3.34 shows a dialog that allows the user to directly choose the ports to devolve (here it is called a *subset*) a higher-order calibration to a lower-order calibration. In this case, a 4-port calibration has been set to create a 2-port subset from port 2 to port 3. Also shown in this dialog is a method to choose the power wave correction type, as discussed in Section 3.7.4.

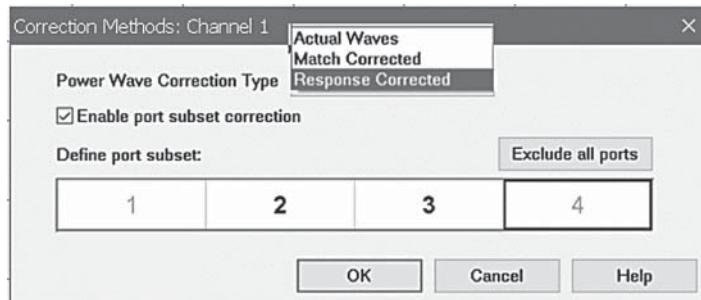


Figure 3.34 Dialog for selecting wave correction and devolving ports.

Some important considerations of correction method for transmission (e.g. S21) measurements are discussed next.

3.11.1 Response Calibrations

One instance where a full 2-port calibration is impractical in a transmission measurement is where 1-port calibrations standards might not exist to provide for a full 2-port calibration, and so a response calibration will provide a transmission correction even if some errors are not fully corrected for. An example of this is some in-fixture measurements, measurements using new or unusual connectors, or antenna measurements. For these cases, a response-only calibration (sometimes called *normalization*) may sometimes be performed.

The errors in a response calibration come from the source and load match interaction during the calibration, and their interactions with the DUT during the measurement. For response calibration, the errors can be minimized by improving the source and load match of the VNA system. Often, precision attenuators, which have very good match, are added to the test ports to improve the source and load match. Without mismatch correction, the errors in response calibration can be substantial. The measurement of a transmission thru follows Eq. (3.5), where the actual response includes effects of the S-parameters of the thru standard.

Response calibrations are often used for “quick-and-dirty” calibration, but the quality of the calibration is frequently misunderstood. Often, one simply uses a “data into memory” and “data/memory” math function to accomplish the normalization, which is equivalent to a response calibration. After the response calibration is complete, the response of the thru standard will appear to be a perfectly flat trace. This may give the impression that the calibration is of good quality when in fact it simply means that the measurement after calibration matches the measurement before calibration. Figure 3.35, in the upper window, shows a measurement of a nearly ideal airline after calibration using a full 2-port calibration and a response-only calibration.

The response shows substantial ripples, even though a thru response would show a flat line. This is because the errors in the response tracking exactly cancel the source and load match errors for the case where the DUT matches the thru, which occurs when remeasuring the same thru. But measuring an airline, which also has nearly 0 dB insertion loss but a different insertion phase, generates a substantial ripple. Half the ripple is due to error during the calibration

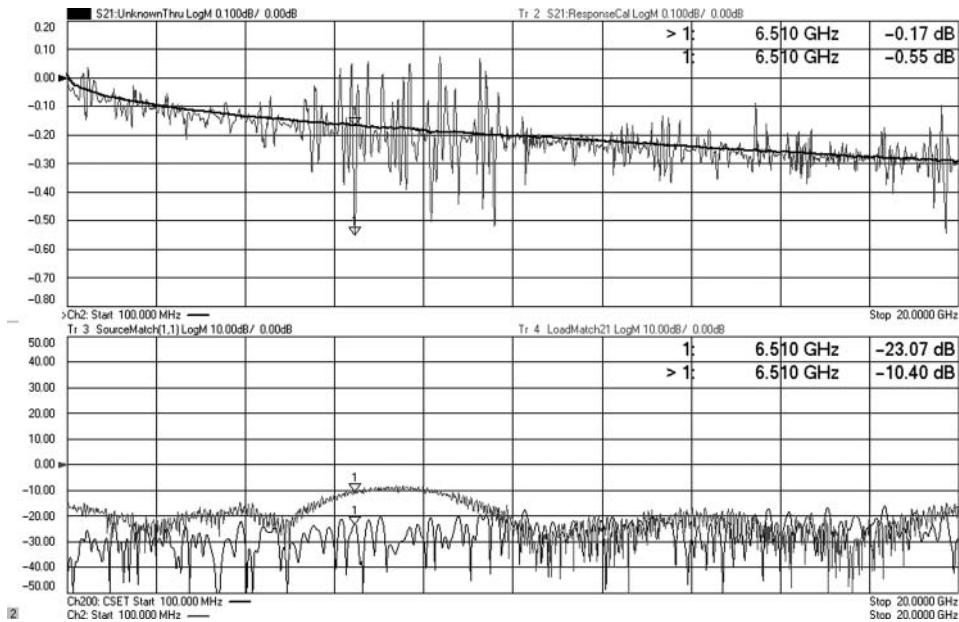


Figure 3.35 (Upper) Measurement of an airline with a response calibration and a full 2-port calibration (Lower) Source and Load match of the system.

acquisition, and half is from error during the measurement. The source and load match raw error terms are shown in the lower window. It is clear that the worst-case ripple occurs where the combination of source-match and load match are highest. At the position of the markers, the combination of source and load match is about -33 dB. The error due to this on the thru measurements is nearly equal to the mismatch error added to the thru, in the linear sense, and then converting to dB and multiplying by 2, or

$$S_{21}Err \approx 2 \cdot 20 \log_{10} \left(1 + 10^{\frac{ESF_{dB} + ELF_{dB}}{20}} \right) = 40 \cdot \log_{10} \left(1 + 10^{\frac{-33}{20}} \right) = .38 \text{ dB} \quad (3.73)$$

This exactly matches the error of the ripple at the marker point, which is the difference between the full 2-port correction (0.17) and the response correction (0.55), or exactly 0.38 dB. Thus, from knowledge of system source and load match, the error in a response measurement is straightforward to compute.

3.11.2 Enhanced Response Calibration

Another reason for not using a full 2-port calibration is when the configuration of the test system causes calibration acquisitions to have substantial error. A typical case of this is where a large loss is added in front of the port 2 test port. This case occurs when testing high-power amplifiers where the power is too great to be applied directly to the VNA test port, as shown in Figure 3.36. If the attenuation between port 2 and the DUT is greater than 10 dB, difficulties

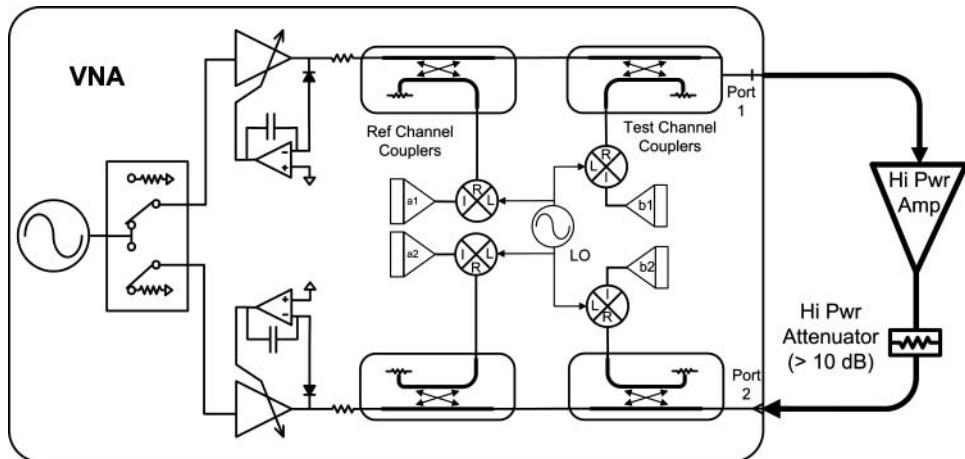


Figure 3.36 Using an external attenuator to reduce power to the VNA port 2.

with 2-port calibration math can occur. Chapter 6 describes several alternatives to the configuration of Figure 3.36 that will allow a full 2-port calibration for high-gain or high-power situations. But, for the configuration shown, a response calibration, or better yet the enhanced response calibration (ERC) discussed next, may provide a better result than a full 2-port calibration.

Another example where a full 2-port calibration may have difficulties is when long cables are used, which have substantial loss and degrade the ability to make reflection measurements at the ports. In such a case, the error terms associated with the port, such as directivity, load match, or source-match, can have substantial errors. Further, the loss causes the reflection tracking term to be small, and measured data becomes noisy. The noise and loss in the reflection measurements will result in much noisier transmission measurements that can have substantial errors. In the case of measuring a high-gain amplifier, the reverse measurement, S₁₂, can have substantial noise, or even be entirely noise, due to the high isolation of the DUT S₁₂ and the loss due to attenuation of the test port. Thus, if S₂₂ and S₁₂ have substantial errors and noise in their measurements, through the application of Eq. (3.7), they can cause noise in the corrected S₂₁ result. Some specific examples are given in Chapter 6.

In these cases, doing a full 2-port calibration can do more harm than good. Further, some test systems, particularly at high mm-wave or THz regions, do not support S-parameter measurements in both directions. For these cases, the ERC provides a useful addition to the error correction algorithms.

The acquisition for ERC proceeds identically to that of the forward full 2-port calibration but requires a defined-thru calibration procedure. In the ERC acquisition, a 1-port calibration is performed on port 1, and then a defined thru is measured. The best case is a flush thru, but for non-insertable devices, such as the common SMA female-to-female case, a short female thru is used. The length of this thru is often ignored, but much better results are obtained if the delay and loss of the thru calibration standard are modified to account for the physical device being used. During the thru measurement, the load match of port 2 is also measured so that the complete characterization of the forward path is accomplished. As such, the error terms

for ERC are identical to the forward full 2-port terms, and there is no increase in uncertainty at this point.

During the measurement of the DUT, the raw input match and transmission are measured. The input match is corrected with a 1-port calibration, as in Eq. (3.8), and the transmission is corrected with a modified forward response calibration as

$$S_{21A} = \frac{(S_{21} - EXF)}{ETF \cdot \left(1 + \frac{(S_{11M} - EDF)}{ERF} \cdot ESF \right)} \quad (3.74)$$

which is identical to the result one would obtain by setting the forward load match term, ELF, to zero, in Eq. (3.7). Figure 3.37 shows the measurement of an airline with an ERC (light gray trace, trace 1) and with a full 2-port cal using the unknown thru method (dark trace, trace 2). The ripple in this measurement is much smaller than that of a response-only calibration (Figure 3.35). This is because half of the ripple in Figure 3.35 is from errors in the response tracking term, and the rest of the ripple is caused about equally from the input mismatch and output mismatch, with some caused by the round-trip mismatch of $S_{21}^*S_{12}^*ELF^*ESF$. With the ERC, four of these five terms are compensated. Using the 1-port calibration to compensate for Γ_{In} includes effects of the DUT and load as seen from port 1 so that only the term of S_{22}^*ELF is not compensated for. For an airline, S_{22} is small. In the case of an amplifier measurement, where S_{12} is small, the difference between a full 2-port calibration and an ERC is the effect of the load match term interacting with the amplifier S_{22} .

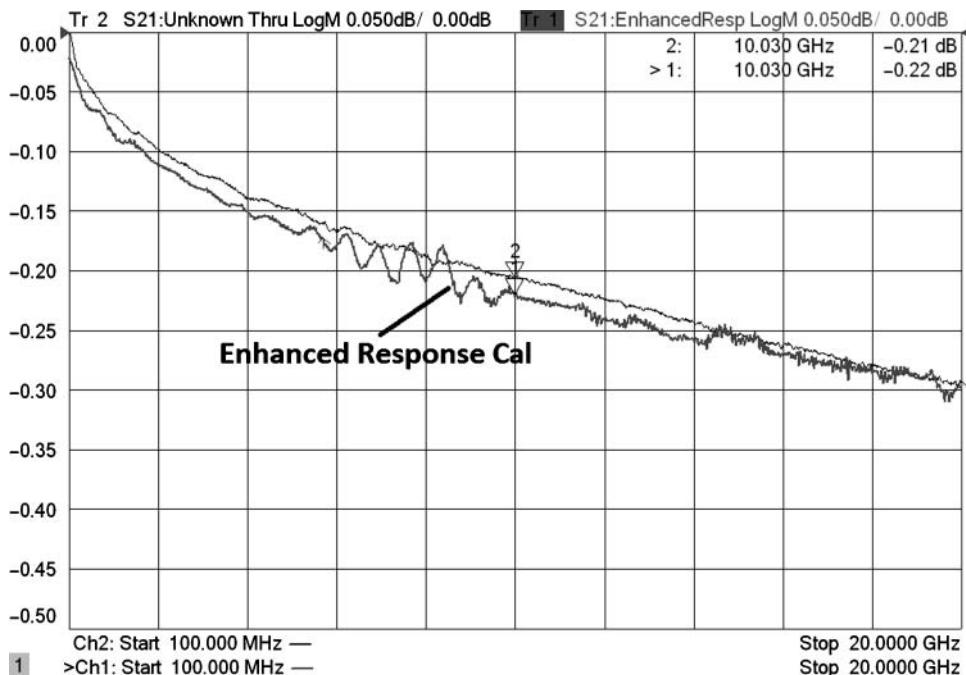


Figure 3.37 Measurement of an airline with Enhanced Response Calibration, with actual airline performance also shown.

When a large attenuator is added to port 2, in the case of measuring high-power amplifiers, the load match error term is very small, if one uses a well-matched attenuator. Thus, using the enhanced response calibration, which ignores the mismatch effect between the DUT S22 and the port 2 load match, can produce a smaller error effect than the noise in the S22 measurement for a full 2-port calibration. In this way, a lower-order calibration can provide less uncertainty and error than a full calibration.

3.12 Determining Residual Errors

3.12.1 Reflection Errors

After a calibration is complete, it is often desired to understand the quality of the resulting calibrated measurement. The error terms have been compensated for, but not perfectly, so inaccuracies in the acquisition of the error terms yield errors in the measurement.

The first test for calibration quality is to simply remeasure the calibration standards. This does *nothing* to tell one about the quality of the calibration but does show the repeatability and noise in the measurement system. If after reconnecting a standard and remeasuring, the results do not match exactly the model for that standard, then there is a stability or noise issue in the measurement that may dominate the determination of the residual errors.

3.12.2 Using Airlines to Determine Residual Errors

Remeasuring the calibration standards does no good in determining calibration quality because the systematic errors behave in such a way that they will exactly cancel when the calibration standards are re-applied; but for any other DUT, the errors will not cancel. For example, if there is an error in the acquisition on the directivity error term, EDF, which is almost always caused by the load return loss being non-ideal (i.e. not equal to zero), that error will propagate through the extraction of the source-match term and the reflection tracking term in such a way that when the short standard or open standard is remeasured, the error in source-match and error in load match exactly cancel for that particular standard measurement. However, if an offset short or offset open is measured, the errors add instead of cancel and will yield a large ripple, the peak deviation of which is essentially the combination of source-match error and load error. If an offset load is measured, the error is essentially twice the residual directivity error. The best method for establishing the offset is to use a bead-less airline (see Figure 1.13).

For these methods to be effective, the quality of the offset airline must be better than the quality of the calibration standards. Fortunately, airlines provide about the best possible impedance reference, and their values are well known.

3.12.2.1 Determining Directivity

Using an airline to determine the residual directivity can proceed in one of two ways. The first method relies on creating a ripple envelope, the peak value of which is twice the residual directivity.

The measurement of any load in the presence of residual errors is

$$\begin{aligned} S_{11A_Load} &= EDF_R + \frac{ERF_R \cdot \Gamma_L}{(1 - ESF_R \cdot \Gamma_L)} \approx EDR_R + ERF_R \cdot \Gamma_L \cdot (1 + ESF_R \cdot \Gamma_L) \\ &= EDF_R + ERF_R \cdot \Gamma_L + ESF_R \cdot \Gamma_L^2 \end{aligned} \quad (3.75)$$

where the approximation assumes the residual directivity and source-match are small and the residual tracking is almost exactly one. For even moderately good loads, the measured S11 can be simplified to

$$S_{11A_Load} = EDF + \Gamma_L \quad (3.76)$$

The residual directivity is the difference between the measured load and the actual load value. In the case where the model of the load is ideal, in that its reflection is presumed to be zero, the measured directivity term is defined as

$$EDF_M = S_{11M_Load} \quad (3.77)$$

The residual directivity is

$$\Delta EDF = EDF_M - EDF \quad (3.78)$$

Combining Eqs. (3.76), (3.77), and (3.78), the residual directivity is found as

$$\Delta EDF = \Gamma_L \quad (3.79)$$

Figure 3.38 is created by measuring the load used for calibration at test port 1, at the end of an airline. The ripple in the figure is caused by the load having different impedance than the airline, which should be at the system impedance. If this load is the load used for calibration, the load's own return loss can be estimated from the airline measurements, and thus the residual directivity can also be estimated. At the frequencies where the length of the airline is a multiple of one-quarter wavelength, the impedance of the load is transformed to

$$Z_{L-\lambda/4} = \frac{Z_0^2}{Z_L} \quad (3.80)$$

The calibration of the system sets the actual system reference to the load reference so that the peaks of the load return loss represent twice the load's reflection coefficient (6 dB higher), referred to the desired system impedance. As an example, consider a case where the load has an impedance of 51Ω , for a residual directivity of -40 dB . The effective Z_0 reference for the system for zero reflection after calibration is 51Ω . But at $\lambda/4$ points, the impedance presented to the test port by the offset load is

$$Z_{L-\lambda/4} = \frac{50^2}{51} = 49.02 \text{ ohms} \quad (3.81)$$

The reflection coefficient of the load offset by airline, when measured after calibration is

$$\rho = \left| \frac{Z - Z_{0A}}{Z + Z_{0A}} \right| = \left| \frac{49.02 - 51}{49.02 + 51} \right| = \frac{1.98}{100.02} = .0198 \text{ or } -34 \text{ dB} \quad (3.82)$$

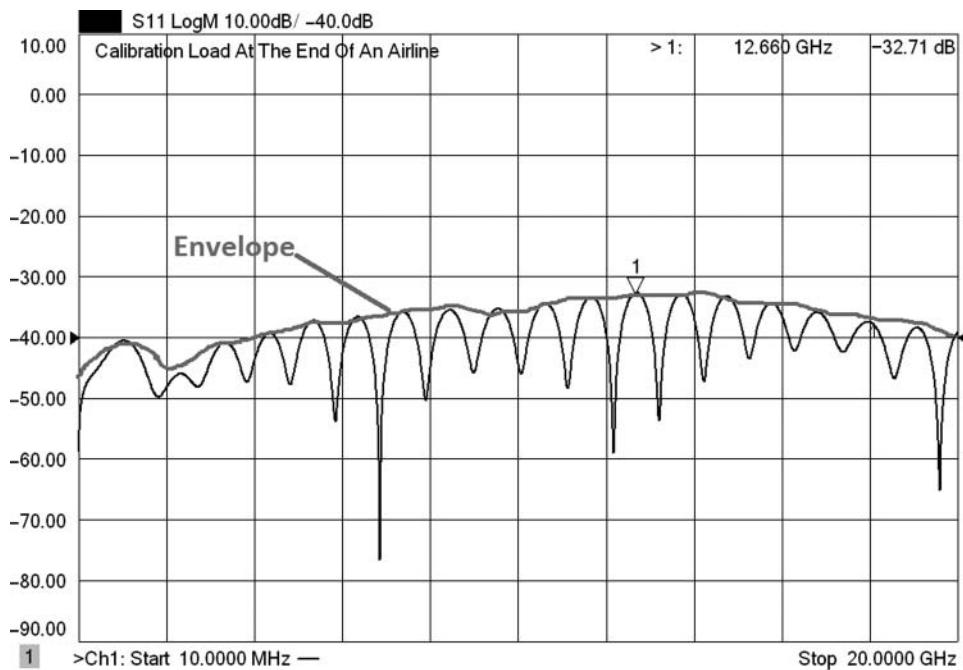


Figure 3.38 Ripple envelope of the calibration load at the end of an airline.

Here, the actual reference impedance is that of the calibration load, or $Z_{0A} = 51$. The actual residual directivity of the system based on the 50Ω impedance is -40 dB. The peaks of the ripple are 6 dB above the actual residual directivity value. Note that this method provides values in a simple way, but the estimate of the residual directivity is only at the peaks of the ripples. It is common to draw an envelope of the ripple pattern, presuming that the errors in the loads change slowly with frequency, as shown by the gray line in Figure 3.38. In this case, where the marker is set, the measured peak is -32.7 dB, and thus the estimated directivity is approximately 38.7 dB. The source-match and reflection tracking errors are negligible when measuring a load of even moderately low return loss. Of course, this method does not work if some characterized device, such as Ecal, is used for the calibration standards, or if a method such as TRL is used.

An alternative method makes use of the time-domain transform to separate the reflection of the offset load from the directivity error term. For this method, one must have a sufficiently long line to separate the input error (directivity) from the reflection at the end of the airline. An example is shown in Figure 3.39, where the response of a measurement of an airline is shown, after calibration, where the load element has about a 40 dB return loss. The upper plot, light trace, shows the S11 of the airline terminated in a load. The lower window shows the time domain response of the S11, along with the gated response around the input connector. The darker trace in the upper window shows the gated frequency response, which is the directivity.

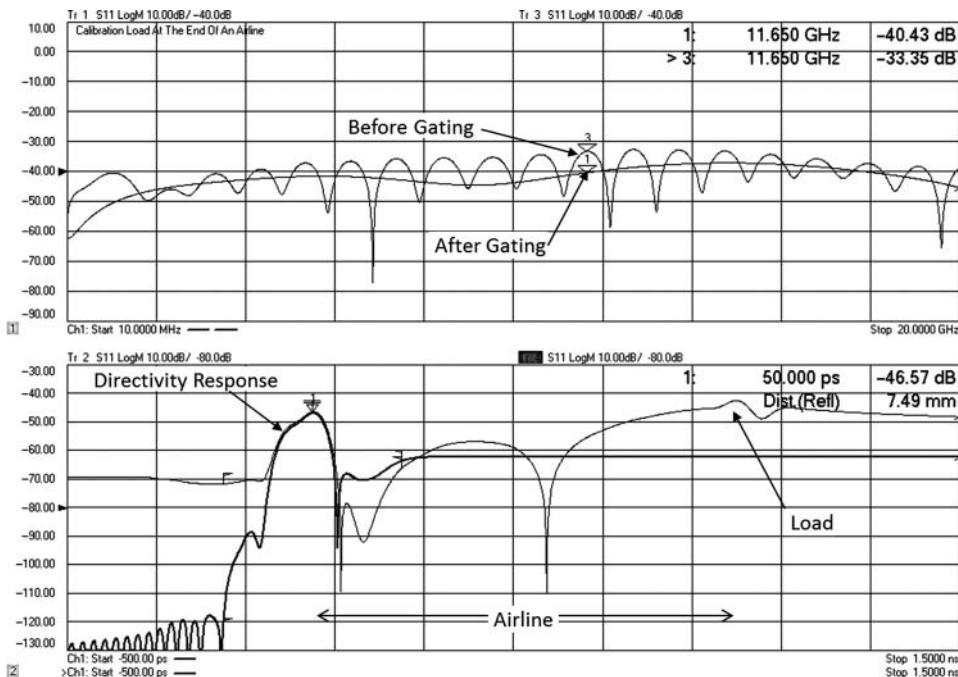


Figure 3.39 Determining directivity with time-domain gating.

3.12.2.2 Determining Source-Match and Reflection Tracking Residual Errors

Using either an open or a short at the end of an airline creates a ripple pattern in the S11 trace. An example is shown in Figure 3.40 where both the open and short responses are shown. The light trace, labeled “Mismatch,” shows the peak-to-peak response (in dB) of the error, after removing the effect of the loss of the airline. This is simply done by taking the open data divided by the short data and expressing the value in dB. The loss curve (trace 3 in the figure) is created by taking the square root of the open data times the short data and matches very well to twice the through-loss of the airline. If the loss of the airline is known, this can be compared to the measurement result, and from this, errors in the reflection tracking can be inferred. However, the reflection tracking errors are typically quite small and are difficult to quantify with airline techniques.

The envelope of the ripple represents a combination of the reflection tracking error, the source-match error, and the load match error. The measurement of an open after calibration is

$$S_{11M_open} \approx EDF_R + \frac{ERF_R}{[1 - ESF_R]} \approx EDF_R + ERF_R + ESF_R \quad (3.83)$$

where here it is assumed that the open and short reflections have a magnitude of 1, and the residual error terms (represented by the subscript R) are relatively small; that is, directivity and match are nearly zero, and reflection tracking is nearly one. From the signal flow diagram

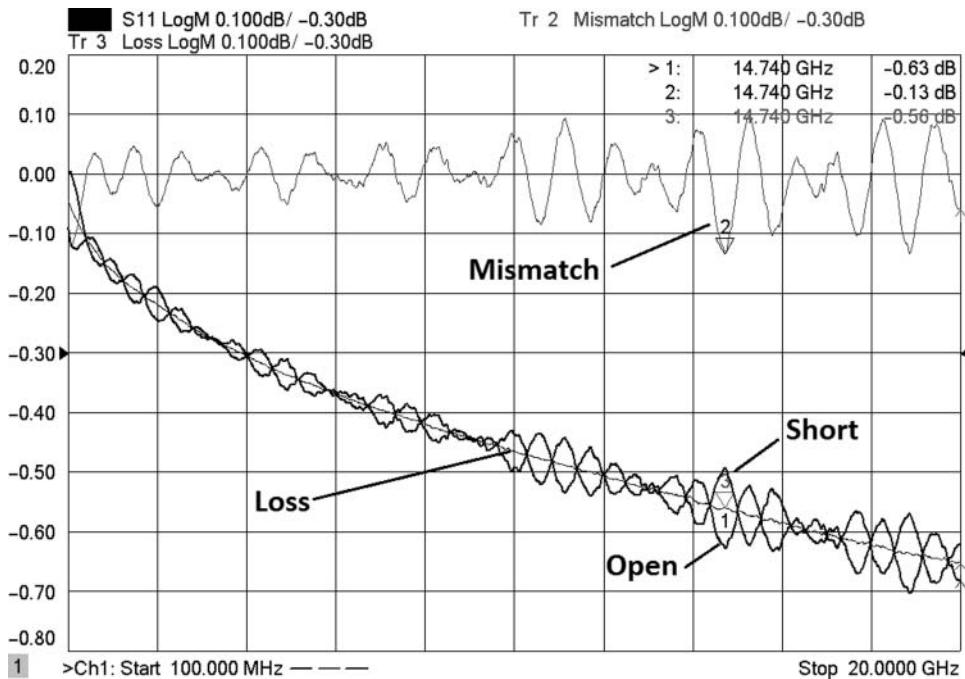


Figure 3.40 Ripples from an open and short at the end of an airline.

of Section 3.3.1, the source-match is a combination of the errors in the open/short difference and the residual directivity so that

$$ESF_R \approx EDF_R + \frac{(\Delta O - \Delta S)}{2} \quad (3.84)$$

where ΔO and ΔS are the differences between the model of the open and the actual open, and the model of the short and the actual short, respectively. In practice, the open and short are well known, so the residual source-match is almost entirely due to the residual directivity. Thus, the load response is the main source of error in the calibration kit. Errors in the load response will appear in the source-match term for using SOL calibrations, forming a floor in the source-match value. That is one reason why techniques such as TRL or even Ecal (which is characterized with a TRL traceable calibration) have better theoretical specifications than SOLT calibrations. A sliding load calibration also derives its directivity error from the quality of the airline used in the load and so is close to the quality of a TRL calibration. In reviewing corrected system specifications, one will note that the source-match is usually equal to or worse than the directivity. In some newer high-frequency calibration kits, such as the 1.85 or 1.0 mm, several standards are used in to provide an over-determined solution; in such a case the simple rules for estimating residual error sources don't apply.

The source-match error term can be estimated by knowing the directivity error, and the additional ripple must be due to a combination of reflection tracking and source-match. An

upper limit on the source-match error can be determined if the entire mismatch error is assumed to be the source-match. Using both a short and an open helps to define the envelope of the ripple.

From the figure, the worst-case ripple is 0.13 dB, which represents the VSWR of the source match. The equivalent source-match in dB can be computed for the marker value of 0.13 dB as

$$ESF_{Residual_dB} \approx 20 \cdot \log_{10} \left(\frac{1 - 10^{\frac{OS_{Ripple}(dB)}{20}}}{1 + 10^{\frac{OS_{Ripple}(dB)}{20}}} \right) = 20 \cdot \log_{10} \left(\frac{1 - 10^{\frac{0.13}{20}}}{1 + 10^{\frac{0.13}{20}}} \right) = -42 \text{ dB}$$
(3.85)

The value can be computed directly on the VNA if the equation editor function is used, recognizing that the open data is in the Tr1 data and the short data is in the Tr1 memory. The trace values are stored internally in the linear form and the results can be computed in terms of the linear source-match, the equation to use is simply

$$ESF = \left(\frac{1 - |Tr1/Tr1.Mem|}{1 + |Tr1/Tr1.Mem|} \right)$$
(3.86)

This is shown in Figure 3.41 in the lower window. One might notice that the residual source-match appears to rise at the low-frequency portion of the trace. This is a measurement artifact arising from the fact that the airline impedance actually increases because of skin effect, so its use to characterize source-match is limited. Also, the lower trace represents

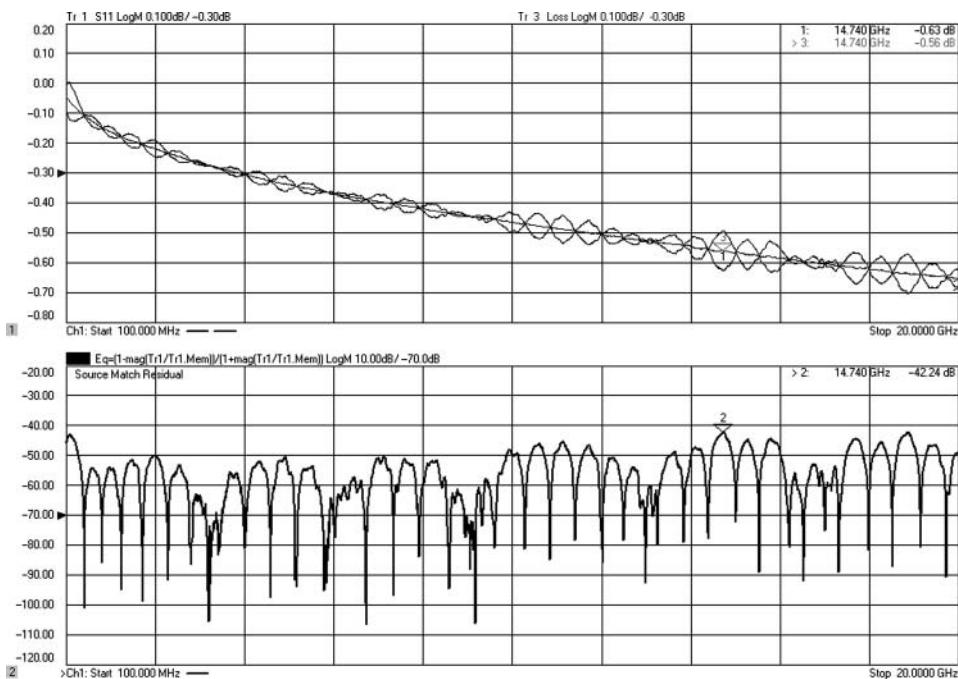


Figure 3.41 The computed residual source-match shown in the lower plot.

the combination of source-match and directivity errors, along with errors in airline itself. Using the envelope of the peaks of the ripples gives a good conservative estimate of the source-match. In this case, -42 dB is consistent with the previous measurements of directivity error.

A final note: if the calibration kit is not a precision kit but created by a user with its own open and short (for example, an on-wafer cal-kit), any error in the modeled delay and phase of the open or the short will go directly to a source-match error. This is a case where the load does not dominate the source-match error.

3.12.2.3 Reflection Tracking Residuals

The reflection tracking, from Eq. (3.24), can be considered as the dB average of the open and short response or the geometric mean of the linear terms. The evaluation of the residual error can be computed by considering the case where the reflection tracking is defined as 1. The actual open differs from the ideal open by ΔO , so the actual open is $(1 + \Delta O)$ and the short differs by ΔS so the actual short is $-(1 + \Delta S)$. The derivation here follows the common practice when dealing with residuals to neglect products of error terms (which are small). It is instructive to go through the derivation for the error in ERF.

$$\begin{aligned}
ERF_R &= \frac{ERF_{Computed}}{ERF_{Actual}} = \frac{ERF_{Computed}}{1} = \frac{-2[(1 + \Delta O) - EDF][(-1 - \Delta S) - EDF]}{(1 + \Delta O)(-1 - \Delta S)} \\
&= \frac{-2[(1 + \Delta O)(-1 - \Delta S) - EDF \cdot (-1 - \Delta S) - EDF \cdot (1 + \Delta O) + EDF^2]}{1 + \Delta O - (-1 - \Delta S)} \\
&= \frac{-2[(-1 - \Delta O - \Delta S - \Delta O \Delta S) + EDF \cdot (\Delta S) - EDF \cdot (\Delta O)]}{2 \left(1 + \frac{(\Delta O + \Delta S)}{2}\right)} \\
ERF_R &= \frac{\left(1 + (\Delta O + \Delta S)\right)}{\left(1 + \frac{(\Delta O + \Delta S)}{2}\right)} \cdot \frac{1 - \frac{(\Delta O + \Delta S)}{2}}{1 - \frac{(\Delta O + \Delta S)}{2}} = \frac{1 + \frac{(\Delta O + \Delta S)}{2} - \cancel{\frac{(\Delta O + \Delta S)^2}{2}}}{1 - \cancel{\left(\frac{(\Delta O + \Delta S)}{2}\right)^2}} \\
&= 1 + \left(\frac{\Delta O + \Delta S}{2}\right)
\end{aligned} \tag{3.87}$$

From this result, one can infer several attributes of the reflection tracking term; first among them is that for small residual errors, the error in the load does not contribute to the reflection tracking error. Also, it is the average of the open/short errors that contribute to the reflection tracking; so if the short is a little long and the open is a little short, the errors will cancel leaving the reflection tracking with no error. It is only when the error in the open and the short are in the same direction that the reflection tracking picks up the error.

A similar analysis of the source-match error due to open and short error yields the result that in addition to a contribution of the load error, the source-match error is proportional to the difference between the open error and the short error; thus, if the open is a little long and the short is a little short, the source-match error is large. In a similar way, if the open and short are both a little long, the source-match error is not affected.

In practice, the magnitude of the open and short, especially for mechanical standards, is well known, so the reflection tracking error is negligibly small; however, the phase of the open and short depend directly on the fact that their length matches the model for each, so the phase error of the standards translates directly to a contribution to source math error and can contribute to reflection tracking error. Examination of the residual errors specified for calibration kits will show that reflection tracking can almost always be ignored in considering the overall error contribution in reflection measurements.

In a normal situation, where the residual reflection tracking is nearly 1, the residual source-match is typically a little worse than the residual directivity. But in some situations, where the coupler raw directivity is poor, such as when a lot of loss is added after the directional-coupler perhaps in the form of a long cable or attenuator pad, the residual directivity can be much worse than the residual source-match due to drift in the test system after calibration. In a normal analysis of the error terms, one would find no difference with this test configuration, as only the load match of the calibration load is considered. But in the case of substantial loss after the coupler, the drift and noise in the raw directivity become the limiting factors in the directivity error term.

Consider the case of a large attenuator on port 1, which has been corrected with a careful low-noise calibration. Drift in the test port coupler will give a large change in directivity but almost no change in source-match. In fact, the raw source-match in this case might be quite good as well.

For the most part, uncertainty computations ignore drift in the system, but for many test scenarios, it is the dominant error. In these cases, the source-match is often good as the attenuator or lossy cable present a good match to the DUT. When such a situation occurs, using a devolved calibration such as enhanced response cal or even response-only calibration, with no mismatch correction, may provide a better result. This is due to a poor or unstable EDF term that in turn will yield a bad value for the DUT S11 and, from that, a bad value for the match correction.

3.12.2.4 Load Match Residual Error

The load match error term is measured as part of the transmission thru measurement. In the case where the thru standard is a flush thru, the error in the load term is essentially the same as the error in the directivity term. From Eq. (3.75), the source-match and reflection tracking are both multiplied by the raw load match.

$$\begin{aligned} ELF_R &= ELF_{Cal} - ELF_A = (EDF_R + ERF_R \cdot ELF_A + ESF_R \cdot ELF_A^2) - ELF_A \\ &\approx \underline{ELF_A(ERF_R - 1)} + EDF_R + \underline{ESF_R \cdot ELF_A^2} \\ ELF_R &\approx EDF_R \end{aligned} \quad (3.88)$$

Since the raw load match is typically reasonably low and the reflection tracking is typically near 1, the first product in the previous equation can usually be neglected. The source-match term, while typically larger than the directivity error term, is multiplied by the square of the load match term and can also be neglected. Thus, the residual load match is essential equal to the residual directivity of the other (mating) test port. A quick example illustrates this: consider a VNA system with 15 dB raw load match (0.18 linear), 40 dB residual directivity (0.01

linear), 30 dB residual source-match (0.032 linear), and 0.1 dB residual reflection tracking (1.012 linear). Converting all the terms to linear, Eq. (3.88) becomes

$$\begin{aligned} ELF_R &= (0.18)(1.012 - 1) + (0.01) - (.032)^2(.18) \\ ELF_R &= (0.0022) + 0.01 + (0.0002) \\ ELF_R &\approx 0.01 \end{aligned} \quad (3.89)$$

Verifying the residual load match is quite similar to verifying the directivity, using a similar method. The directivity is verified by adding the calibration load to the end of an airline and looking at the peak value. Alternatively, the time-domain gated response of the airline plus load can be determined as discussed in Section 3.12.2.1 to give the directivity. Once this is done, the airline can be connected to port 2, and the response can be viewed. Since the directivity error is known, the peaks of the response represent the sum of the directivity and the load match. Typically, this result matches the result of using the same calibration load to terminate the airline as the load match and directivity error terms should be similar. However, if the load port is connected to the VNA with a test port cable (the most common case), then cable flexure will add to the measured load match and must be considered a part of it.

For the case where a non-flush through is used during the calibration, such as a defined thru, then an additional error source occurs if the match of the defined thru is not zero or if the delay and loss of the defined thru does not match the model in the calibration kit. In legacy VNAs, before the introduction of the unknown thru method, it was common to have large errors in the load match, principally because the DUT had non-insertable connectors and a thru adapter was used to connect the ports, but the cal-kit was not modified to add the delay and loss of the thru adapter. In these cases, the phase of the load match is incorrectly determined. The error effect of this depends upon the length of the adapter, but if it approaches one-quarter wavelength in the frequency of interest, the error due to the adapter is in fact up to 6 dB larger than if no port match correction was done at all. For the case of low-loss filter measurement, this was the most common reason for customer-support calls complaining about ripple in the transmission measurement. Fortunately, modern VNAs provide more techniques, such as Ecal and unknown thru, to avoid this issue. Some VNA calibration engines (for example, the PNA SmartCal™) requires the user to enter the connector types and sexes for DUT ports and will not allow a defined thru calibration if the cal-kit does not have a thru adapter defined that matches the connectors of the DUT. In these cases, if a thru standard is desired to be used, one must modify the calibration kit to add a thru with the properly defined ports, most typically a female-to-female thru. The delay of the thru must be entered and can be reasonably estimated by a number of means including measuring the physical length if it is an airline thru. Alternatively, using one-half of the S11 reflection delay measurement of the thru standard, after a 1-port calibration, can be used as well, but one must consider the open fringing of the thru if it is not terminated (also a common practice) or account for the delay of the termination (short or open) if it is terminated.

To illustrate the measurement of the load match residual, consider the two cases shown in Figure 3.42. The upper trace shows a ripple in the S11 from a measurement of an airline, after an SOLT-defined thru calibration (dark, Marker 1). Here the thru adapter is ignored (the thru adapter in this case is one found in many calibration kits for the purpose of the “swap equal adapter” and is about 27 mm long). Also shown is a good unknown thru calibration (light,

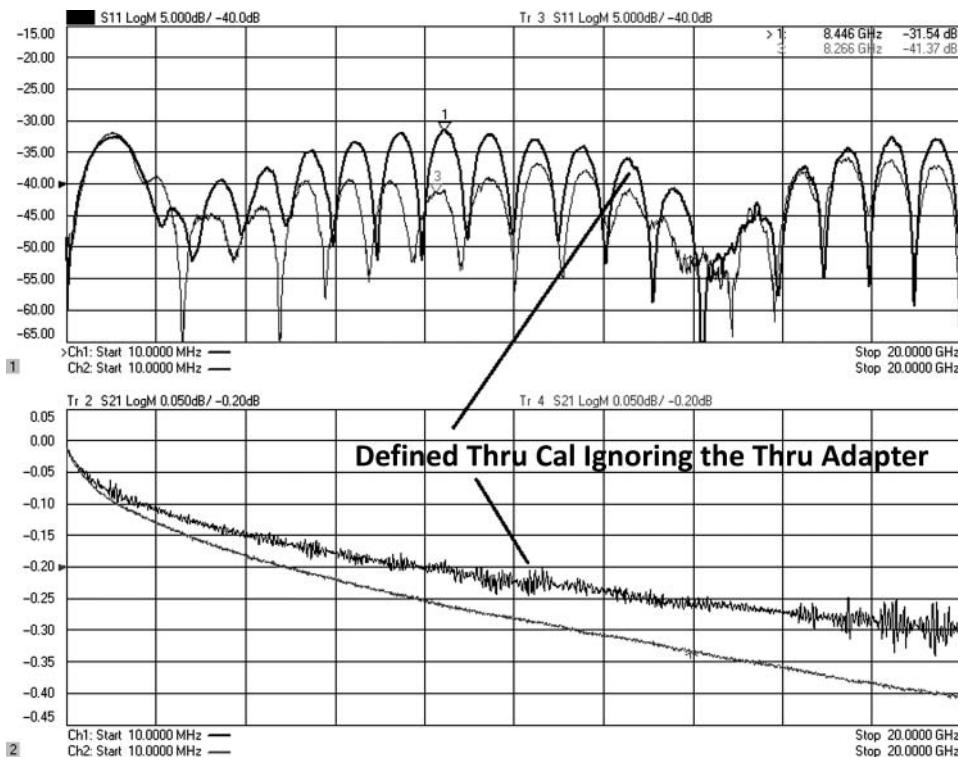


Figure 3.42 Measurement of a test port load match, upper is with a good calibration, lower is with an erroneous defined Thru cal.

Marker 3). The figure shows much worse S11 in the case of the defined thru, because the load match is not properly corrected for. The error represents a residual load match as bad as 30 dB compared to better than 40 dB for the unknown thru calibration. The lower window shows the equivalent S21 traces. The darker trace is from the defined thru calibration, ignoring the thru, and it has both lower loss (incorrectly showing lower loss because the loss of the thru adapter is not properly accounted for) and higher ripple, due to a poor load match acquisition.

A somewhat unexpected result is that in normal systems, where the raw terms are not too large, the residual errors for reflection depend *only* on the calibration kit and do not depend upon the quality of the test system. If the test system is stable, then a system with poor directivity and one with good directivity would yield the same results after correction. From this point of view, there is no benefit in trying to make a well-matched test system with good directivity. However, if there is any drift in the system (and there always will be), the residual performance will degrade more quickly on a system with poor raw performance.

3.12.2.5 Transmission Residual Errors

Transmission errors are evaluated in a manner quite similar to the load match evaluation. In this case, a transmission measurement of the airline is used, and any ripple on the S21

measurement is associated with transmission tracking error. The expected residual tracking error can be computed from Eq. (3.5) by considering the effects of the other residual values on the computation of the tracking error. For a flush thru calibration, the value of ETF is found from the measured S₂₁ value and the values for the other error terms.

$$\begin{aligned} S_{21M_Cal} &= \frac{(S_{21A} \cdot ETF)}{(1 - S_{11A} \cdot ESF) \cdot (1 - S_{22A} \cdot ELF) - ESF \cdot S_{21A} \cdot S_{12A} \cdot ELF} + EXF \\ S_{21M_Cal} &= \left. \frac{ETF}{1 - ESF \cdot ELF} \right|_{S_{11A}=S_{22A}=0, S_{21A}=S_{12A}=1} \\ \therefore ETF &= S_{21M_Cal} \cdot (1 - ESF \cdot ELF) \end{aligned} \quad (3.90)$$

The actual values for ESF and ELF are not used when computing ETF; rather, the extracted values (or measured values) of the match terms are used in the computation. The distinction is that the measured values are only estimates of the actual value. The residual tracking can be found by taking the ratio of the computed tracking over the actual tracking.

$$\begin{aligned} \Delta ETF &= \frac{ETF_M}{ETF} = \frac{S_{21M_Cal}(1 - ESF_M \cdot ELF_M)}{S_{21M_Cal}(1 - ESF \cdot ELF)} = \frac{[1 - (ESF + ESF\Delta) \cdot (ELF + \Delta ELF)]}{(1 - ESF \cdot ELF)} \\ &= \frac{(1 - ESF \cdot ELF - ESF \cdot \Delta ELF - ELF \cdot \Delta ESF - \cancel{\Delta ESF \cdot \Delta ELF})}{(1 - ESF \cdot ELF)} \\ &\cdot \frac{(1 + ESF \cdot ELF)}{(1 + ESF \cdot ELF)} \\ &= \frac{[1 - (ESF \cdot ELF)^2 - (ESF \cdot \Delta ELF + ELF \cdot \Delta ESF) - (ESF \cdot ELF)(ESF \cdot \Delta ELF + ELF \cdot \Delta ESF)]}{1 - (ESF \cdot ELF)^2} \end{aligned} \quad (3.91)$$

In the derivation, it is assumed that the values of the source and load match are substantially less than one, and these error terms are larger than their residual terms. In the simplification, higher-order terms are neglected where the residuals are considered to be on the order of the square of the associated error term. This means that the square of the product of error terms is a fourth-order term, and the product of an error term and a residual is a third-order term. Simplifying with these considerations yields the residual tracking term as

$$\Delta ETF = 1 - (ESF \cdot \Delta ELF + ELF \cdot \Delta ESF) \quad (3.92)$$

This equation is most significant and represents an important result in understanding the errors in a test system. Unlike the residual errors in the 1-port terms, which depend only upon the quality of the calibration standards, the transmission tracking terms depend both on the quality of the calibration standards *and* on the raw source and raw load match of the test system. The corrected load match term has a similar dependency. This is in contrast to the result for 1-port terms, where the raw system performance did not affect the residual error terms, ignoring drift and stability effects.

Because of this, modifying the test system to improve the match at the DUT will reduce the residual transmission tracking error. This is typically accomplished by adding an attenuator pad to the test port cable, right at the DUT reference plane. While this will improve the transmission tracking, the loss after the directional-coupler will cause any drift in the directivity of the system to be amplified by twice the attenuator's value. Further, source power is lost as well as receiver sensitivity. But, in many cases, a small attenuation (on the order of

3 or 6 dB) can reduce the residual errors in a transmission measurement without adversely affecting other measurements.

The importance of Eq. (3.92) can be seen if one looks at a manufacturers' data sheet for residual errors after calibration. In the cases of mechanical calibration kits, the transmission tracking error nearly always computes to be exactly the value predicted in Eq. (3.92) if the source and load match of the system, and the residuals of the cal-kit, are used.

These are two important points to remember about airline ripple techniques:

1. They provide estimates of magnitude errors but not estimates of the phase or delay errors. Some phase errors may be estimated as described in Section 3.15.
2. Imperfections in the airlines, and in particular in the end connectors of the airlines, will limit the level to which verification of residual errors can be determined.

3.13 Computing Measurement Uncertainties

Now that methods have been described to determine the residual errors of the measurement system, the overall uncertainty of any measurement can be determined. It's important to note that these errors apply to all RF measurements, whether they are on sources, spectrum analyzers, power meters, or VNAs. With the exception of VNAs, the effect of source-match and load match errors in the measurements cannot be removed with error correction (as the errors' effects depend on both the magnitude and phase of the interaction of the errors and the signals to be measured, and only the VNA measures the phase of both), so the resulting uncertainty will depend on the raw match terms. For VNAs, the uncertainty depends upon both the raw match and the residual match, as will be discussed next.

In this discussion, the term *uncertainty* is sometimes used interchangeably with the term *measurement error*, but strictly speaking, the uncertainty is a derived value that bands the actual error. In many cases, several error terms are present so that at any particular frequency, some might cancel, and the actual difference or error becomes small, even while the uncertainty remains as a computed limit band of the errors.

For reflection, uncertainty is defined as a difference (in linear terms) between the reading and the true value. This may then be added and subtracted to the measurement to come up with the upper and lower linear limits, each of which maybe then be converted to a dB uncertainty limit.

For transmissions terms including source power, receiver power, and S21, the uncertainty is defined as the ratio (in linear terms) between the reading and true value. This can be converted to a dB error by taking $20\log_{10}$ of the ratio.

3.13.1 Uncertainty in Reflection Measurements

The uncertainty of a reflection measures can be derived by comparing the S11 corrected with the actual error terms to that corrected with the estimated error terms. From Eq. (3.8), the difference between the actual S11 and the error corrected S11 can is defined as

$$S_{11A} - S_{11_1PortCal} = S_{11A} - \frac{(S_{11M} - EDF_R)}{[ERF_R + (S_{11M} - EDF_R) \cdot (ESF_R)]} \quad (3.93)$$

From this, it is clear that only the residual error terms play a part in the uncertainty of a 1-port corrected measurement and that the raw error terms play no part in the overall uncertainty. While true in the abstract, in fact the raw error terms do play a significant role in the overall uncertainty if there is any system drift (cable drift, coupler drift, receiver conversion loss drift) over the measurement period. These drifts can dominate the overall uncertainty if the raw error terms are large, even if the residuals are quite small.

3.13.2 Uncertainty in Source Power

For any system using a source, for example, driving a complex modulated signal into a receiver from a vector signal generator or driving a power amplifier from a VNA, the uncertainty of the source power depends upon the three errors: the source tracking error, STF; the source-match, ESF; and the effective input match of the DUT, Γ_1 . The uncertainty (in dB) can be computed directly from (3.48) in Section 3.6 as

$$\Delta\text{SourcePower} = \left| 20\log_{10} \left(\frac{a_{1A}}{a_{1S}} \right) \right| = \left| 20\log_{10} \left[\frac{\text{STF} \cdot [(a_{1S}/a_{1R}) \cdot \text{SourceLinearity}]}{1 - |\text{ESF} \cdot \Gamma_1|} \right] \right| \quad (3.94)$$

where a_{1S} is the source power setting, a_{1A} is the actual source power applied to the DUT, and a_{1R} is the reference power for the source amplitude specification or measurement. Source linearity represents the error that occurs when the source power is changed for a calibration or reference value and is often represented as a dB error for particular dB change in value and must be converted to linear terms for use in the previous equation. Here the absolute value of the product of $\text{ESF} \cdot \Gamma_1$ is used to give worst-case uncertainty, and the absolute value is taken for the whole equation as uncertainty is traditionally expressed as an absolute value.

In the case of a complex modulated source, the modulation produces power over a range of frequencies, and the match of the DUT and the source may vary over the that range of frequencies. If the signal modulation is narrow, the source-match and load match are likely slowly moving, and the resulting source power error occurs similarly to entire modulation envelop, yielding only an error in the modulation power. However, if the signal is wideband, then the variation of the source power due to mismatch can cause amplitude and phase uncertainties in the wideband signal.

The error term STF is essentially the error in the source power when it drives into a perfect 50Ω matched load. For signal sources, this error can be caused by amplitude offset error and by source flatness and source linearity errors. If a power meter is used to calibrate the source power, then the STF becomes the residual source tracking error, from Eq. (3.50),

$$\Delta\text{STF} = \frac{\Delta P_{\text{Meas}}}{a_{1S}} (1 + \Delta\text{ESF} \cdot \Gamma_{\text{PwrSensor}} + \text{ESF} \cdot \Delta\text{EDF}) \quad (3.95)$$

where the ΔP_{Meas} is the error associated with the power meter, and the error associated with the measurement of the power sensor match is essentially the directivity error of the system. These same errors apply to the source power for a DUT driven by a VNA source. In the case of the VNA, the ESF term is not strictly correct as that is usually associated with the ratio source-match, where the error depends upon the power source match as defined in Section 3.6.

However, in the case of using enhanced power calibration with receiver leveling, the error terms in Eq. (3.94) become residual error terms. In general, source power errors don't affect any of the ratio measurements (e.g. S-parameters) in a VNA measurement but certainly will affect gain measurements when using a signal source and other receiver such as a power meter or spectrum analyzer.

3.13.3 Uncertainty in Measuring Power (Receiver Uncertainty)

Just the analysis of Section 3.6 provided the basis for computing source uncertainty, so too can the analysis of receiver calibration in Section 3.7 be used to compute the uncertainty of receiver measurements. Similar to source uncertainty, the receiver uncertainty computation applies equally to all types of receivers including power meters, spectrum analyzers, and VNA receivers. Also, for all receivers except VNA receivers, the uncertainty in power readings depend upon the raw output match of the DUT, Γ_2 ; the raw input match of the measuring receivers, ELF; and the receiver tracking, BTF. In power meters, the receiver tracking is a combination of the reference calibration error, the calibration factor (or Cal factor) accuracy, and the power meter linearity. For a spectrum analyzer, it depends upon the amplitude flatness calibration (which in itself may depend upon the match and accuracy of the calibration source) and SA linearity. The uncertainty of a measuring receiver can be computed from Eq. (3.67) as

$$\Delta RcvrPower = \left| 20\log_{10} \left(\frac{b_{2A}}{b_{2M}} \right) \right| = \left| 20\log_{10} \left\{ \frac{(1 - |ELF \cdot \Gamma_2|)}{BTF} \left[\left(\frac{b_{2M}}{b_{2R}} \right) \cdot R_{DA} \right] \right\} \right| \quad (3.96)$$

where b_{2A} is the actual power at the receiver (in this case, the b2 receiver), b_{2M} is the measured power, R_{DA} is the receiver dynamic accuracy, and b_{2R} is the reference power for the receiver dynamic accuracy. In practice, receiver dynamic accuracy is given as a dB error for some dB change in power and so must be converted to linear form to be used in the previous equation. In practice, for modern VNAs, the receiver dynamic accuracy can be very small; for example, the Keysight PNA-X has less than 0.01 dB error over 80 dB of power change and so generally can be neglected; a typical power sensor has an error of 0.004 dB over a 10 dB range, plus a range-error offset for each range that is traversed.

For power meter and SA measurements, the errors are all in terms of raw or actual system errors. For VNA measurements, the BTF term is always a residual term, after error correction, and is computed as in Eq. (3.65). For response-only corrections, the ELF term is an actual (raw) load match; for match-corrected power measurements, the residual value of the ELF term should be used.

3.14 S21 or Transmission Uncertainty

While there are many contributors to S21 uncertainty, it is dominated by the source and load match, both raw and residual, of the test system. The uncertainty of an S21 measurement can be derived from the results of Eq. (3.5) as

$$\Delta S_{21} = \left| 20\log_{10} \left(\frac{S_{21M}}{S_{21A}} \right) \cdot \left[\left(\frac{b_{2Cal}}{b_{2M}} \right) \cdot R_{DA} \right] \right|$$

$$= \left| 20\log_{10} \left[\frac{ETF}{(1 - |S_{11A} \cdot ESF|) \cdot (1 - |S_{22A} \cdot ELF|) - |ESF \cdot S_{21A} \cdot S_{12A} \cdot ELF|} \right] \right. \\ \left. \cdot \left[\left(\frac{b_{2Cal}}{b_{2M}} \right) \cdot R_{DA} \right] \right| \quad (3.97)$$

where b_{2M} and b_{2Cal} are the power levels at the test receiver during measurement and calibration, respectively, and R_{DA} is the receiver dynamic accuracy. If this is an uncorrected S21 measurement, or a gain measurement using a signal source and power sensor or SA as a receiver, then the error terms are the raw error terms. If this is a VNA measurement and a calibration has been performed, then the error terms are residual error terms. In many cases, a response calibration is performed, ignoring mismatch, so the transmission tracking term becomes a residual term, but the other match errors that remain are the raw match terms.

The residual transmission tracking term was computed in Eq. (3.92). From this it is clear that the transmission tracking error relies on both raw and residual terms. Note that this is a different result from that of uncertainty in reflection measurements, which is dependent only on the residual errors. Following a similar argument, the S21 uncertainty for a 2-port calibrated measurement can be determined. In this case, the uncertainty is computed for a low-loss device so that receiver dynamic accuracy doesn't contribute much error. To compute the uncertainty for a 2-port calibration, the measured values for the error terms are substituted for the actual error terms

$$\Delta S_{21_2PortCal} = \frac{S_{21A}}{S_{21Corr}} \\ = \left[\frac{\frac{ETF}{(1 - |S_{11A} \cdot ESF|) \cdot (1 - |S_{22A} \cdot ELF|) - |ESF \cdot S_{21A} \cdot S_{12A} \cdot ELF|}}{\frac{ET\hat{F}}{(1 - |\hat{ESF} \cdot S_{21A}|) \cdot (1 - |\hat{ELF} \cdot S_{12A}|) - |\hat{ESF} \cdot S_{21A} \cdot S_{12A} \cdot \hat{ELF}|}} \right] \quad (3.98)$$

where $ET\hat{F}$ is the measured (but not actual) transmission tracking error, and similar meanings for the other error terms. This can be further simplified for the case of a matched device where S11 and S22 are negligible (a common assumption as attenuators are often used as reference verification devices) to become (in linear form)

$$\Delta S_{21_2PortCal} = \frac{S_{21A}}{S_{21Corr}} = \left[\frac{\frac{ETF}{1 - |ESF \cdot S_{21A} \cdot S_{12A} \cdot ELF|}}{\frac{ET\hat{F}}{1 - |\hat{ESF} \cdot S_{21A} \cdot S_{12A} \cdot \hat{ELF}|}} \right] \\ = \frac{ETF}{ET\hat{F}} \cdot \frac{(1 - |(ESF + \Delta ESF)(ELF + \Delta ELF)S_{21A} \cdot S_{12A}|)}{(1 - |ESF \cdot S_{21A} \cdot S_{12A} \cdot ELF|)} \\ = \Delta ETF \frac{1 - |S_{21A} \cdot S_{12A} \cdot (ESF \cdot ELF + ESF \cdot \Delta ELF + \Delta ESF \cdot ELF + \Delta ESF \cdot \Delta ELF)|}{(1 - |ESF \cdot S_{21A} \cdot S_{12A} \cdot ELF|)}$$

$$\begin{aligned} &\approx \Delta ETF \left[1 - \frac{|S_{21A} \cdot S_{12A}(ESF \cdot \Delta ELF + \Delta ESF \cdot ELF)|}{(1 - |ESF \cdot S_{21A} \cdot S_{12A} \cdot ELF|)} \right] \\ &= \Delta ETF(1 - |S_{21A} \cdot S_{12A}(ESF \cdot \Delta ELF + \Delta ESF \cdot ELF)|) \end{aligned} \quad (3.99)$$

For the case of a low loss device, where S21 and S12 are nearly one, this simplifies to

$$\begin{aligned} \Delta S_{21_2port} &= \Delta ETF(1 - |ESF \cdot \Delta ELF + \Delta ESF \cdot ELF|) = \Delta ETF^2 \\ \Delta S_{21_2port(dB)} &= 20\log_{10}(\Delta ETF^2) = 2 \cdot \Delta ETF_{dB} \end{aligned} \quad (3.100)$$

Just like the ETF term, the uncertainty of an S21 measurement is primarily due to the source and load match, raw and residual.

An intuitive way of thinking about this is to recognize that when one does error correction, the mismatch terms are subtracted from the flow graph equations through their characterization. But the characterization of the source and load match is not perfect, and some residual error is left. Thus, when a reflection off the load occurs, the reflection is compensated for by the estimated load match, but the residual load match is free to reflect off the load and re-reflect off the raw source. Since this residual re-reflection is unknown by the error correction math, it is not corrected for by the source-match term, and thus the error becomes the product of the residual load match times the raw source-match. Similarly, when there is a reflection off the load match that re-reflects off the source-match, it is corrected for by the estimated source-match, but there is a remaining portion of the reflection that is not compensated for, and it is equal to the raw load match term times the residual source-match.

These residuals occur in two stages: first there are the residuals during the acquisition of the error terms, yielding the residual tracking term; this is the calibration residual. Next there is the residual due to the measurement of the DUT, and for low-loss DUTs, it is identical to the calibration residual. In fact, if the phase of the DUT matches the phase of the calibration thru, these residuals will cancel, and there will be no error in the S21 measurement. That is why remeasuring the thru standard tells one nothing about the quality of the calibration. However, measuring an airline of a different length causes phasing of the residual terms, and the envelope gives a good estimation of the uncertainty, as was demonstrated previously in Figure 3.42.

3.14.1 General Uncertainty Equation for S21

For completeness, in the case where S11 and S22 are not zero, their residual effect can be added as a simple input and output mismatch error. The loop term is ignored as it contains a product of residuals and will be quite small. Thus, the general uncertainty of a 2-port calibrated S21 measurement is

$$\begin{aligned} \Delta S_{21(dB)} &= 20\log_{10} \left[\frac{1 + (|ESF \cdot \Delta ELF| + |\Delta ESF \cdot ELF|)(1 + |S_{21A} \cdot S_{12A}|)}{(1 - |\Delta ESF \cdot S_{11A}|)(1 - |\Delta ELF \cdot S_{22A}|)} \right] \\ &\cdot \left[\left(\frac{b_{2Cal}}{b_{2M}} \right) \cdot R_{DA} \right] \end{aligned} \quad (3.101)$$

the last bracketed quantity represents the dynamic accuracy error and can be ignored if S21 of the DUT is nearly 1.

A final note on S21 uncertainties: it is clear from Eq. (3.101) that the uncertainty of an S21 measurement depends upon the actual characteristics of the DUT and not only on the test system. This is an uncomfortable result for some engineers, who would like to have simple number for the uncertainty of a measurement. In fact, one could compute a worst-case uncertainty by taking all the S-parameters to equal 1 (not physically realizable) and find that this would give a limit to S21 uncertainty for any device. In fact, this can be extended to any arbitrary test system by presuming the raw source and load match as unity so that the S21 uncertainty limit for any device on any test system is determined only by the residual source and load match. For example, an arbitrary test system calibrated with 40 dB residual source and load match would have a worst-case limit of S21 uncertainty of about 0.5 dB given a presumed worst-case raw source and low match of 0 dB.

3.14.2 Dynamic Uncertainty Computation

The computation of the uncertainty of a device depends directly on the DUT characteristics, as well as the test system characteristics. Recently, based on a new method of error correction algorithms that maintain correlation between sources of error (Garelli and Ferrero 2012), a dynamic uncertainty calculator has been introduced.

Dynamic uncertainty uses a more sophisticated computation of vector error correction, which while substantially more complicated (much too complicated to repeat here), does have the advantage of maintaining the correlation of errors, which allows a better estimation of the uncertainty for a particular DUT. The uncertainty presented in Eq. (3.101) represents a worst-case addition of errors. When the errors are uncorrelated, it is more common practice to add the individual contributors in a root-sum-square (RSS), which will reduce the overall uncertainty value. The new method takes this a step further by maintaining the proper correlation. Some errors are completely uncorrelated (source-match and noise), and some errors are strongly correlated (source-match and directivity). How the correlation manifests itself depends upon the DUT response as well.

The new method produces results based on the proper statistical computation of uncertainty expressed as a variance from an expected value. So that the error in any calibration standard is not expressed as an absolute, or worst-case, value but rather as what is known as a *standard uncertainty* value. For example, the standard uncertainty of the phase of an open might be expressed as “1° standard uncertainty.” Standard uncertainty is always in terms of a 1 sigma or 1 standard deviation value. In fact, it does not put a worst-case limit to the value, but only expresses the limits in terms of confidence intervals. So we can say at 1 standard deviation we are 67% confident the actual value of the phase is with 1° of the stated value, and at 2.58 sigma we are 99% confident. Different manufacturers specify different confidence values when they publish their specifications. Typical values range from 95% to 99.5%, representing either 2 sigma or 2.77 sigma values.

Figure 3.43 shows the result of the Keysight dynamic uncertainty display on an S21 trace, where the user may define the desired coverage factor. The 2.58 sigma value might represent the specified (guaranteed) performance, and the 1 sigma value might specify the range of the typical performance.

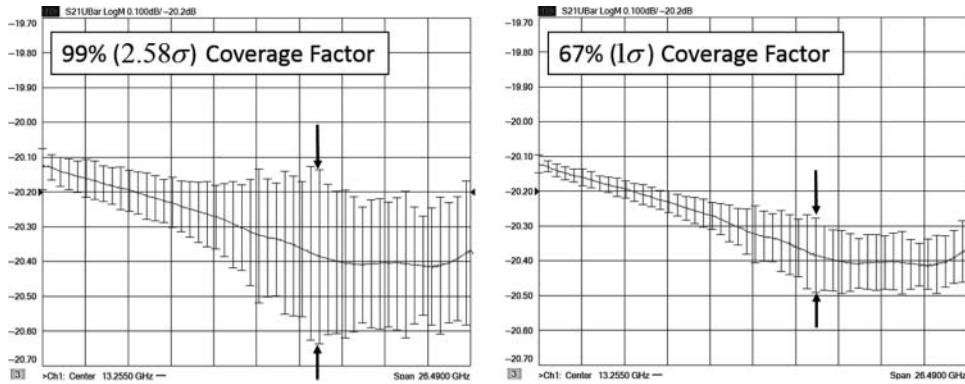


Figure 3.43 Two examples of S21 uncertainty with different coverage factors.

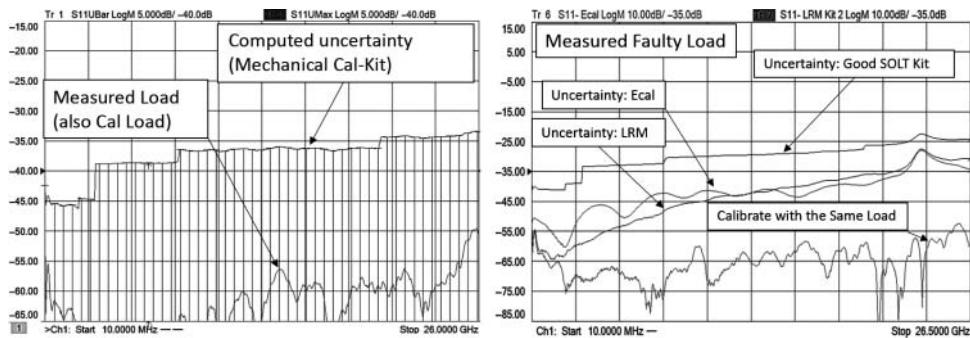


Figure 3.44 Uncertainty depends upon the Calkit quality.

Note that the uncertainty changes with frequency because of many factors: the match of the test system and the match of the DUT changes, the noise floor changes, and the error in the calibration standards change with frequency. Only the dynamic uncertainty gives a proper estimate of the true uncertainty at the DUT reference plane. This is because it uses the actual error terms drawn from the measurement channel's calset, as well as the actual DUT measurements, and the specified uncertainty on the cal-kit from the cal-kit definition.

The largest contribution to uncertainty is the error in the calibration kit, and different kits can cause different errors. Figure 3.44 shows the effect of the cal-kit on the uncertainty. In the left plot we see the measurement of the load, which is the same one used to calibrate the system, so we expect a perfect result. Also shown is the computed upper bound for S11 (S11 plus uncertainty) that the load could be due to the uncertainty of the SOL kit used. We can see the discrete band breaks where the specifications on the load change.

The plot on the right illustrates just the uncertainty (it is the upper bound minus the measured S11) for three different cases. The reference trace of an LRM calibration was performed, and the same load as the match value is measured, labeled “Calibrate with the same load.” This trace shows the measured S11 (essentially the connector repeatability). The load in this

case was a little faulty and has a higher than normal reflection, but as it was used in the LRM calibration it does not appear to be bad.

Another LRM calibration with good standards was performed and the load measured, along with an Ecal calibration and an SOLT calibration. The LRM shows the lowest uncertainty, followed by the Ecal, and the mechanical calibration kit (SOLT) shows the worst uncertainty. Interestingly, at high frequencies, where the measured load is poor, the uncertainties are more closely clusters. This illustrates that both the cal-kit and the actual DUT characteristics both contribute to uncertainty.

Finally, other effects such as noise and connector repeatability can be added as they will affect the overall measurement uncertainty. Figure 3.45 shows the uncertainty of a filter with noise floor effects included.

The uncertainty of the S21 becomes greater in the stop band for two main reasons: first, the input match becomes very poor (about 0 dB) and so mismatch error is greater, and second, as the loss of the filter becomes greater, the noise floor adds significant uncertainty. Once the filter response goes below the noise floor, the uncertainty on the upper bound becomes the noise floor (that is the filter S21 must be lower than that) and the filter has -200 dB (essentially negative infinity) as a lower bound as the minimum value cannot be distinguished at all. It is interesting to note the peaks of the noise closely follow the upper bound. It is possible to see a point or two of noise exceed the upper bound, as the upper bound is statistically derived

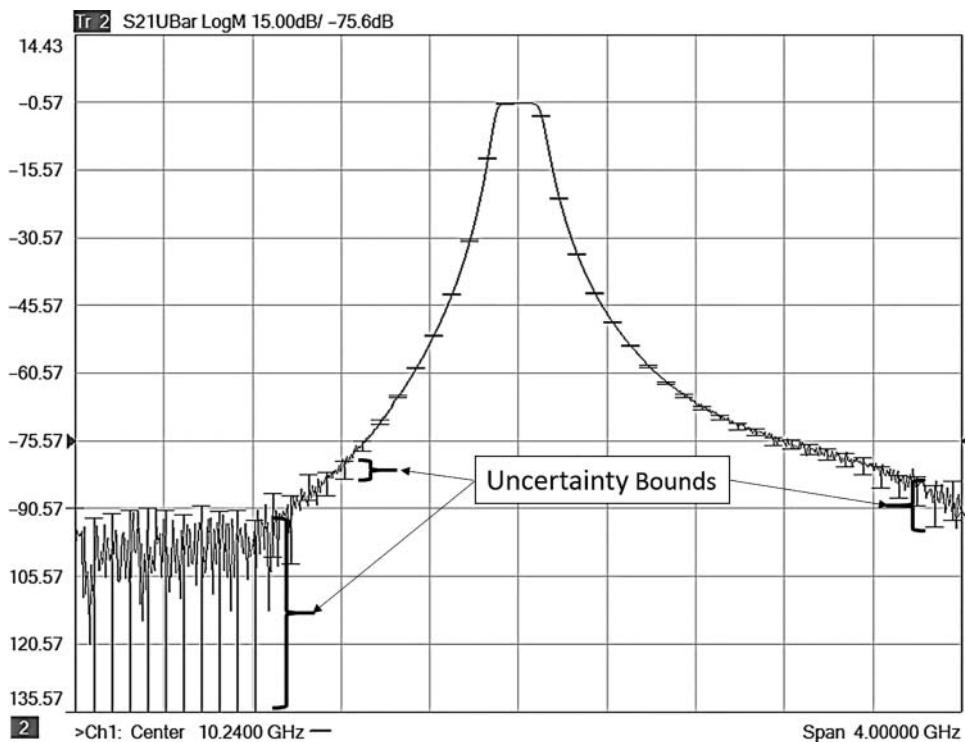


Figure 3.45 Uncertainty changes with DUT loss.

over a confidence interval, so if it is set to 2 sigma, we expect about a 93% confidence, that the true measure is inside the uncertainty bounds; or 7% expectation of a point falling outside the bounds.

Dynamic uncertainty has changed the game when it comes to evaluating complex test systems, especially ones that include external components and switch matrixes. In the past, computing the uncertainty was a huge task where each individual external element must be evaluated and its effect added to the instrumentation uncertainty. But now, using dynamic uncertainty, the uncertainty of any measurement can be directly displayed with no additional work by the user.

3.15 Errors in Phase

The error computations given thus far have been vector in nature, so magnitude and phase errors could be computed as well. However, a simplified method for computing phase errors given an amplitude error can be determined by assuming that the error term represents a locus of points around the true value, as seen in a vector plot. The magnitude error occurs when the error adds or subtracts in-phase with the measured value. The phase error occurs when an error terms adds in quadrature with the measured value, as shown in Figure 3.46.

From this construction, one can see that the phase of the error is the arc-sine of the ratio of magnitude of the error and the signal, or

$$\begin{aligned}\Delta dB &= 20 \log \left(\frac{\text{Signal} + \text{Error}}{\text{Signal}} \right), \quad \text{Error} = \text{Signal} \cdot 10^{\frac{\Delta dB}{20}} - \text{Signal} \\ \Delta\phi_{\text{deg}} &= \frac{180}{\pi} \cdot \arcsin \left(\frac{\text{Error}}{\text{Signal}} \right) = \frac{180}{\pi} \cdot \arcsin \left(10^{\frac{\Delta dB}{20}} - 1 \right)\end{aligned}\quad (3.102)$$

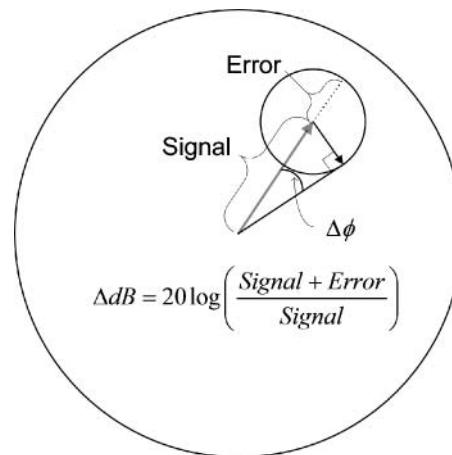


Figure 3.46 Phase error as a result of an error signal.

For small errors, the ratio of phase error to dB error approaches a constant value of $6.6^\circ/\text{dB}$; the analysis to demonstrate this is shown in Eq. (3.103)

$$\begin{aligned}\frac{\Delta\phi_{\text{deg}}}{\Delta\text{dB}} &= \frac{180}{\pi} \cdot \frac{\arcsin\left(10^{\frac{\Delta\text{dB}}{20}} - 1\right)}{\Delta\text{dB}} \approx \frac{180}{\pi} \left(\frac{10^{\frac{\Delta\text{dB}}{20}} - 1}{\Delta\text{dB}}\right) \Big|_{\arcsin(x) \approx x \text{ for small } x} \\ &= \frac{180}{\pi} \cdot \lim_{\Delta\text{dB} \rightarrow 0} \left(\frac{10^{\frac{\Delta\text{dB}}{20}} - 1}{\Delta\text{dB}}\right) = \frac{180}{\pi} \lim_{\Delta\text{dB} \rightarrow 0} \frac{d}{d\Delta\text{dB}} \left(10^{\frac{\Delta\text{dB}}{20}} - 1\right) \\ &= \frac{180}{\pi} \frac{\lim_{\Delta\text{dB} \rightarrow 0} \left(\frac{10^{\frac{\Delta\text{dB}}{20}} \log_e 10}{20}\right)}{1}\end{aligned}$$

since $\lim_{x \rightarrow 0} (10^x) = 1$,

$$\frac{\Delta\phi_{\text{deg}}}{\Delta\text{dB}} = \frac{9}{\pi} \log_e(10) = 6.6, \quad \text{or} \quad \Delta\phi_{\text{deg}} = 6.6 \cdot \Delta\text{dB} \quad (3.103)$$

This useful result applies to errors from unknown signals, from mismatch, from calibration errors, from noise, or from any other situation where an error signal adds in a vector manner with the desired signal. It allows one to predict with confidence the error or ripple in phase due to some error signal by observing only the dB level of the error or ripple. Inspection of the calibration residual presented in many data sheets will show that the phase errors in tracking are approximately 6.6 times the dB error, as an example.

3.16 Practical Calibration Limitations

While the details of calibration and error correction have been thoroughly derived and described, the results presented so far apply to an idealized situation that considers only the systematic errors, that is, ones that can be characterized and removed. This is the common situation found in most discussions of microwave metrology. These conditions include attributes such as calibrating at the optimum power level, using narrow low-noise IF bandwidths, making measurements only at exactly the same frequencies as calibrated, using only the test ports of the VNA without cables or adapters, and, in particular, only specifying performance when using a particular connector with particular high-performance calibration kits.

In the real-world experience of the RF and microwave engineer, it is often necessary to diverge from these idealized conditions for reasons of measurement speed, device limitations, or other practical applications. In this section, the effects of these real-world conditions are discussed. Some of the errors are random errors, such as noise, stability, and connector repeatability, which must be treated in a statistical manner. Other errors are related to systematic errors, and while not compensated for, error bounds might reasonably be assigned to them.



Figure 3.47 A good dog and a good cable: how long will they stay good?

3.16.1 Cable Flexure

Cables are a bane to the existence of the RF and microwave engineer. By way of a simile, one can say that cables are like dogs; either they are bad, they've been bad, or they are going to be bad, and when they're good, they only stay good with great care (Figure 3.47).

In almost all measurement cases, it is necessary to use a flexible or semi-flexible cable to connect between the measurement instrument and the DUT. For S-parameter and related measurements, the cables add mismatch, group delay, and loss to the measured results. These additions to the systematic errors are assessed during the measurement calibration process and for the most part are removed. However, in some cases, the cable effects remain the dominant source of error in measurements due to drift and flexure in the RF cable. The length and loss of the cable contribute directly to the size of the error.

Metrology grade cables have thick outer jackets that restrict the cable bending radius to try to defend against kinking or other damage to the cable.

In old VNA systems, the RF cables had specifications for phase tracking because of limited calibration capabilities. In all modern VNAs, since integrated error correction has been added, the phase matching of cables is entirely unnecessary. What is critically important is the cable stability, for both magnitude and phase, and in both return loss and transmission. Often semi-rigid cables are used as a lower-cost alternative and give good stability performance. Flexible braid cables are used for lower cost and more flexibility in connecting to DUT interfaces.

The loss of the cable has the effect of degrading the directivity stability of the system. Because the loss occurs after the directional-coupler, the raw directivity (which is the directivity error term divided by the reflection tracking) is a combination of the coupler directivity minus twice the cable loss. If the VNA has a coupler directivity of 30 dB, followed by a cable with a loss of 12 dB, the effective directivity will be only 6 dB. In this sense, directivity is the difference between a full reflection (open or short) and a perfect load. The stability of the cable match is also degraded by the loss. Adding a fixed attenuator at the end of the test cable improves the match of the cable but further reduces the effective directivity and stability.

Some details of characterizing system and cable stability are discussed in Chapter 11.

3.16.2 *Changing Power after Calibration*

The source or port power setting is one of the stimulus settings associated with the calibration state. In older VNA systems, the calibration would display an indication that the stimulus settings had changed; this was an on-screen annunciator “C?” This annunciator led many users to believe that the calibration was questionable, but for the most part this was not the case. Later VNAs use the annunciator “CΔ” to indicate a change to the calibration settings. The idea behind identifying a change in power as a possible condition that would degrade the calibration comes about due to the receiver linearity of the VNA, and the notion that changing the power after calibration would cause the receiver to be operated in a different portion of its linearity curve and add a dynamic accuracy error to calibration. However, for almost all conditions where the power is changed, this is not the case.

The first thing to note is that the receivers for the test signal will be at a different power level than the calibration power level for any device that is not exactly 0 dB S21 gain. The common reason for changing power level is to avoid overdriving an amplifier during test. The usual way to go about calibrating is to set the source power to the desired (lower) test power level for the amplifier and then perform the 2-port calibration at that level. However, for an amplifier test, that level may be very low and so the calibration will have substantial noise due to low drive power. When the amplifier is connected, the power to the test receiver is changed by the amplifier gain, so the dynamic accuracy error occurs on the output receiver. The noise in the calibration may be removed with averaging and IF bandwidth reduction, but this often leads to unacceptably long calibration and measurement times.

Consider instead if the calibration is done at a higher power where the noise is less of a problem. After calibration, the power is lowered. The reference channel receiver sees the change in power level, and the dynamic accuracy error of the reference receiver adds to the calibration error. But for most VNA systems, the reference receiver is padded down 5–10 dB lower than the test receiver, just to ensure that it stays out of compression and so its dynamic accuracy performance is better than the test receiver. Further, when the amplifier is connected, its gain will bring the signal at the transmission test receiver up closer to the level that it was during calibration so that the dynamic accuracy error of the test receiver is reduced or eliminated. From this analysis, it is clear that changing the power level has essentially no effect on the calibration accuracy provided the signal levels remain within the linear-gain range of the DUT.

For a low-loss device such as a filter, where the test channel power is nearly equal to the calibration power, there is a possibility for reduced accuracy due to dynamic accuracy error as the test and reference receiver both will see a different power level if the source power is changed. But there is seldom reason to change the power level when measuring low-loss passive devices.

The final reason for changing power level is to investigate the non-linear behavior of a device. The DUT is often measured at a low-power state, then the power level is increased, and the change in S21 is recorded as device compression. In these cases, the dynamic accuracy error will be introduced into one of the two power measurements. It is often the case that the calibration will be performed at the lower power level and then the power level raised for measurement, but this is exactly opposite of the best practice, as the lower power level will likely mean higher noise and a poor calibration. Experience shows that the dynamic accuracy

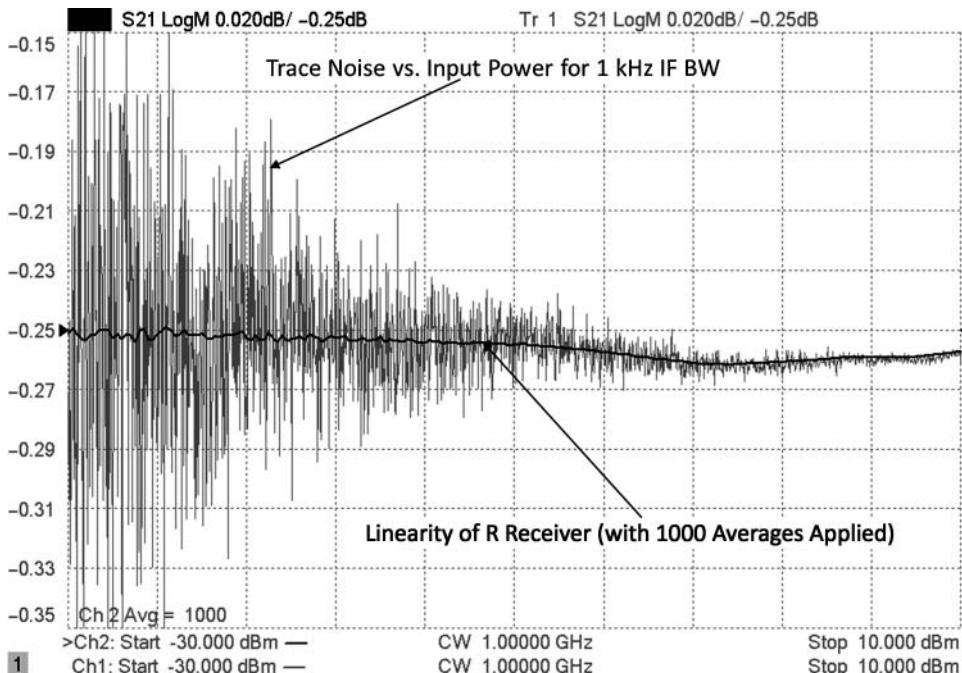


Figure 3.48 Noise and dynamic accuracy error versus drive.

error for a change in power is almost always less than the error due to increased noise in the calibration. Another idea is to calibrate at each power level, but again the error due to noise in the low power level will likely overwhelm the dynamic accuracy.

This is especially true on more modern VNAs, where improved circuitry and methods have yielded nearly perfect dynamic accuracy especially at lower power levels. Figure 3.48 shows a measurement of S21 for a thru connection, where the test receiver is padded so that it is measuring in a very linear region and essentially characterizes the linearity and noise of the reference receiver as the power level is changed. There are two traces, both taken at a normal IF bandwidth of 1 kHz, and one taken with a very large number of averages (1000). These demonstrate the relative effects of noise and dynamic accuracy on the reference channel receiver. For this example, a Keysight PNA-X is used. The specified compression at maximum power is 0.1 dB, but it is clear that the actual receiver compression is much less, on the order of 0.01 dB (note the scale is at 0.02 dB/div). Even at 1000 averages, at the lower-power portion, trace noise and compression are equivalent.

This is from the reference receiver. The test receiver will normally have more compression at the high power, as it is not padded down as the reference receiver is. From these measurements, it is clear that if one calibrates at a power where the receiver is not in compression, normally 10 dB below the maximum specified power, and then moves to any other power, the error in calibration due to both noise and dynamic accuracy is minimized. In fact, the rule for modern network analyzers can be changed from “don’t change power after calibration” to “calibrate at the power that gives the best result and then move to any other power level.”

This applies as long as the source attenuator is not changed. What happens if the attenuator is changed? Please carry on to the next section.

3.16.3 Compensating for Changes in Step Attenuators

Many VNAs included integrated step attenuators between the reference and the test couplers. These allow extending the power range of the VNA to much lower levels. The placement between the reference and test couplers means that changes in the power due to the step attenuator are not seen by the reference receiver, and so the calibration appears to be invalidated if the step attenuator is changed. This placement is intentional to allow the reference-channel receiver to maintain a relatively high signal level, and thus lower noise, even when the signal to the DUT is set to a very low level.

Many VNAs compensate for nominal changes to the test set source attenuator by modifying the values displayed on the reference receiver by the nominal value of the attenuator change. In this way, the readings of the reference channel always represent the source power applied to the DUT. However, the step attenuator is not perfect, and so it often has between 0.25 and 0.5 dB of variation from the nominal value. Further, the step attenuator has different port matches at different attenuator states, with the largest change being from the 0 dB state to any other state. In the 0 dB state, the source and load match are determined by quality of components behind the step attenuator, such as the directivity of the reference channel coupler. For any attenuator step greater than 0 dB, the loss of the attenuator effectively isolates the source-match from the test port, meaning the error from changing the step attenuator between non-zero states will be much smaller.

Also, the match of the attenuator is better than the source-match of the system behind it, so a best practice is to set the step attenuator to at least the first position (5 or 10 dB) before calibration, unless the source power available is not sufficient for the tests to be performed. If the initial calibration is performed with the source attenuators at each port set to the first position, lowering it to other positions will have smaller effect on the mismatch.

The recent introduction of Cal All (see Section 3.8) allows one to set up a measurement channel with each potential attenuator setting and do a single calibration (at high power level) that properly measures the attenuator if another channel has set it to a different level. This completely eliminates the error associated with attenuator setting changes. But sometimes, one does not think ahead to set up multiple channels with different attenuators; still, it is possible to compensate to a great extent the offset caused by an attenuator change.

There is one simple way to characterize the attenuator states and remove most of the attenuator switch effect from the measurement. After a 2-port calibration, leave the thru standard in place, or put a thru standard from port 1 to 2. If it is a non-insertable DUT, use a non-insertable thru and measure the thru's four S-parameters and save this as an S2P file. Then de-embed the thru from port 2 (here one must reverse the sense of S11 and S22 in the S2P file of the thru, if the VNA requires port 1 of the de-embedding file to face test port 2). After de-embedding, the result should be a perfectly flat S21 and very low S11. Switch the attenuator to each of its states, and record the state as an S2P file. This represents the set of S-parameters that is the difference or offset in the original attenuator setting and the new setting. De-embedding these S-parameters from the original calibration will remove most of the error from the attenuator switch. To measure a DUT, turn off the de-embedding on port 2 (as the thru is not used

anymore), turn on the de-embedding for the appropriate attenuator state S2P file on port 1, and measure the DUT with the new attenuator setting.

This removes most but not all the errors associated with the attenuator change. When the attenuator is switched, both the source-match and the load match of the test port change, but in slightly different ways. The S2P file of the switch state that is measured captures the change in load match, but not the change in source-match. For the most part, this is a small error, on the order of 0.05 dB or less if the original calibration is performed in a non-zero attenuator state and the new state is also not the zero attenuator state. The residual error even in the zero state is on the order of .10 dB, which is quite good for most applications. Figure 3.49, upper window, shows the S11 trace after calibrating with the attenuator in the 5 dB state (memory) and changing to the 10 dB state (data). The lower window shows the same result, but for the S21 response. The DUT is an airline, the ripple of which highlights the quality of the calibration.

When the de-embedding method as described is applied to the measurement, the S11 trace is improved dramatically such that it is indistinguishable from the 5 dB state, and all of the offset for the S21 trace is also removed (Figure 3.50). The memory trace shows the data of the

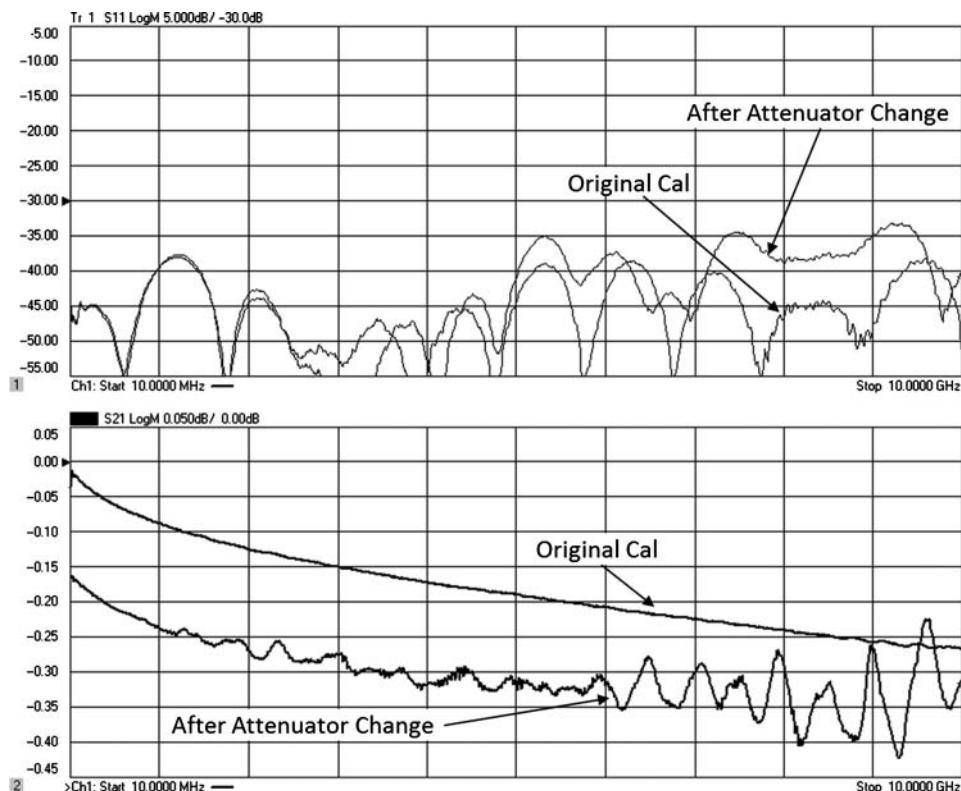


Figure 3.49 (upper) Error in S11 changing the attenuator difference for 5 and 10 dB; (lower) error in S21 due to attenuator change, measuring an airline.

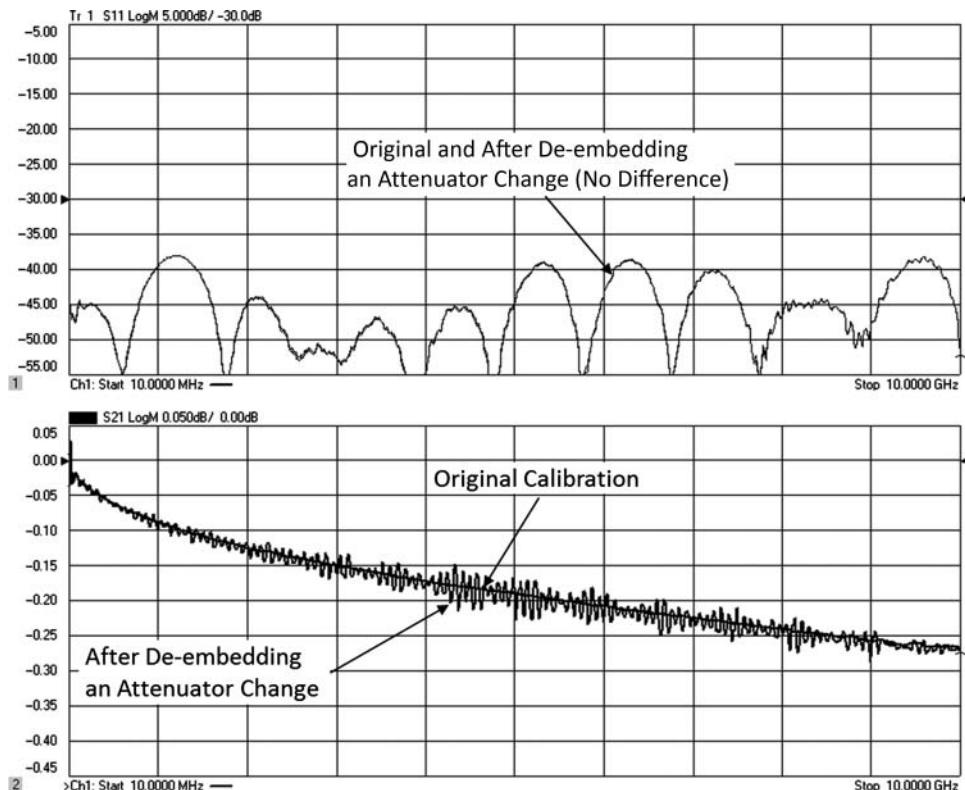


Figure 3.50 Attenuator offset applied as a de-embedding after the attenuator change.

5 dB state, and the data trace shows the result after changing attenuation and de-embedding the attenuator offset. The fine-grain ripple on S21 is a result of differences between the source and load match at port 1, and the simple de-embedding cannot compensate for that, but the worst case error here is less than 0.05 dB.

Changing the attenuator after calibration is seldom recommended, but it can actually produce less error than calibrating at a low power level, where noise dominates. Using the de-embedding technique can reduce the error to a negligible amount, which is especially useful when drive signals must be at a very low level, for example, below -60 dBm.

3.16.4 Connector Repeatability

Connector repeatability is closely related to the problem of cable flexure, in that it is a non-repeatable error and cannot be removed with normal error correction process. Unlike cables, however, the location of the connectors and their associated mismatches is well known. The use of time-domain gating, as discussed in Section 3.12.2.1, can be used to remove the residual connector repeatability error provided the connector is far enough from the DUT so that the time-gate resolution can fully isolate it.

While, classically, connector repeatability referred to only the change that occurs when a connector is de-mated and re-mated, it can also be applied to the situation commonly found in using PC board fixtures, where different standards are realized in different PC board layouts of each standard on a calibration PC board. The standards, opens, shorts, loads and thru's, are PC board traces with a separate coax-to-microstrip launch, typically an SMA connector. Even when identical connectors are used, differences in mounting or soldering can create differences in match between the connectors. This has the same effect as connector repeatability where a connector is de-mated and re-mated for each step of the calibration.

Good connectors typically have repeatability an order of magnitude better than their mismatch, or greater. Precision connectors using slot-less contacts, such as 3.5 mm, have repeatability on the order of 65 dB or so. Commercial-grade Type-N connectors can have repeatability as low as 40 dB. PC board coax connectors have connector-to-connector repeatability on the order of 30 dB up to about low microwave frequencies of 3 GHz or so, and 20 dB up to 20 GHz. An odd reality for PC board connectors is that the repeatability between connectors can be worse than the connector return loss. Consider if one connector is low in impedance by a few Ω and another is high, the difference in the match is greater than the mismatch of either connector. If this case exists, it is actually more detrimental to calibrate on the PC board than to ignore the connector effect. The connector-to-connector repeatability can be determined on PC board standards using time-domain gating techniques; this is discussed in Chapter 11.

3.16.5 Noise Effects

Noise effects have been somewhat discussed already in the context of setting and changing source power after calibration. Noise comes in two forms that affect calibration: noise floor and high-level trace noise. At lower power levels, they are the same thing, and the trace noise apparent on the measurement trace can be entirely attributed to the contribution of the noise floor. The low-level noise floor will affect the measurement trace noise as

$$N_{Trace(dB)} = 20 \log \left(10^{\frac{Signal(dB)}{20}} + 10^{\frac{N_{Floor}(dB)}{20}} \right) - Signal_{(dB)} \quad (3.104)$$

For a -60 dB signal measured on a system with a -100 dB noise floor, the trace noise will be about 0.1 dB. The noise floor of the VNA is set by the IF bandwidth used. In the case of calibration, if source power is set very low, say -60 dBm, then some standards such as the load standard will be right at the noise floor, causing inaccuracies in the determination of the error terms.

From the power sweep trace of Figure 3.48, it is clear that the trace noise increases with lower power levels; it does so at the rate of three times more noise for each 10 dB change in power level relative to the noise floor. One simple way to lower the noise floor of a measurement is to lower the IF bandwidth. The noise floor goes down 10 dB for each 10 times decrease of the IF bandwidth, and trace noise goes down 3 times for each 10 times change in IF bandwidth. Unfortunately, changing the IF bandwidth lowers the sweep speed in the same ratio. In some VNAs, the source is swept at wider bandwidths and automatically changes to stepped sweep at lower bandwidths. This sometimes causes errors due to the delay of the DUT (this effect is discussed in detail in Chapter 5). This effect can be avoided if calibration is always performed in a stepped mode; many VNAs have an explicit stepped mode frequency sweep selection.

In some cases where the highest speeds are needed, such as filter tuning, the same effect of noise reduction can be accomplished using sweep-to-sweep averaging. Using averaging maintains the same sweep dynamic but averages a number of responses to generate the final measurement response. Usually, sweep averaging displays all the intermediate average results, and unless the sweeps are stopped after the max average value is reached, the average value will improve somewhat during subsequent sweeps even after the max average number is reached. This is because the sweep averaging function operates like a 2-tap infinite-impulse-response (IIR) filter. Rather than accumulate N sweeps of data and average all of them together, the averaging function in most VNAs follows the formula

$$A_N = \left(\frac{N-1}{N} \right) Data_{old} + \left(\frac{1}{N} \right) Data_{new} \quad (3.105)$$

Thus, on the $N+1$ or even $N+100$ sample, data from the first reading will be a small portion of the result.

Once the signal to noise gets to 80 dB, the noise floor effect on the trace noise is less than 0.001 dB. However, at high signal levels one may find that the trace noise does not further diminish. That is because at high signal levels, the noise on the signal may come from the source phase noise rather than the receiver noise floor. Once one is in the high-level trace noise region, raising the power to a higher level has no effect on the level of the trace noise. For older VNAs, which often had relatively poor trace noise, the transition point from low-level to high-level noise was between -30 and -20 dB power at the receiver. For more modern VNAs, which have phase noise comparable to a signal generator, the high-level trace noise may not become apparent until 0 to $+10$ dBm. In Figure 3.48, the effect of power level and noise floor on the trace noise is very apparent in the light gray trace. Here the test receiver was padded down to maintain good compression, so the receiver power is 35 dB below the x-axis source power displayed. However, at the high-power end, it is difficult to ascertain the properties of the noise signal.

One may more readily see the effects of high-level noise by taking the noise trace of Figure 3.48, which is a measurement of a thru, and computing the added noise by the formula

$$(S_{21} - 1) \cdot a_1 = \left(\frac{b_2 + N_{Added}}{a_1} - \frac{b_2}{a_1} \Big|_{Thru} \right) \cdot a_1 = N_{Added} \quad (3.106)$$

This will show the effective noise floor for all power, and it is clear from the lower plot in Figure 3.51 that the noise power added is flat over much of the x-axis range but then increases with power of the source and rises above the receiver noise floor for source powers above -5 dBm. This was measured on a PNA-X, which has rather good phase noise. And the absolute level of the noise floor depends upon the IF BW used (here a wide bandwidth is used). Older VNAs such as the HP 8753 or HP820 show this effect at power levels as much as 20 dB lower. This is most likely due to the source phase noise rising above the receiver noise floor when the source has sufficiently high gain, that is, at the higher powers.

3.16.6 Drift: Short-Term and Long-Term

A common question from VNA users is “How long is my calibration good for?” And a common answer is “Until it goes bad!” Most VNA manufacturers are silent on the question, as

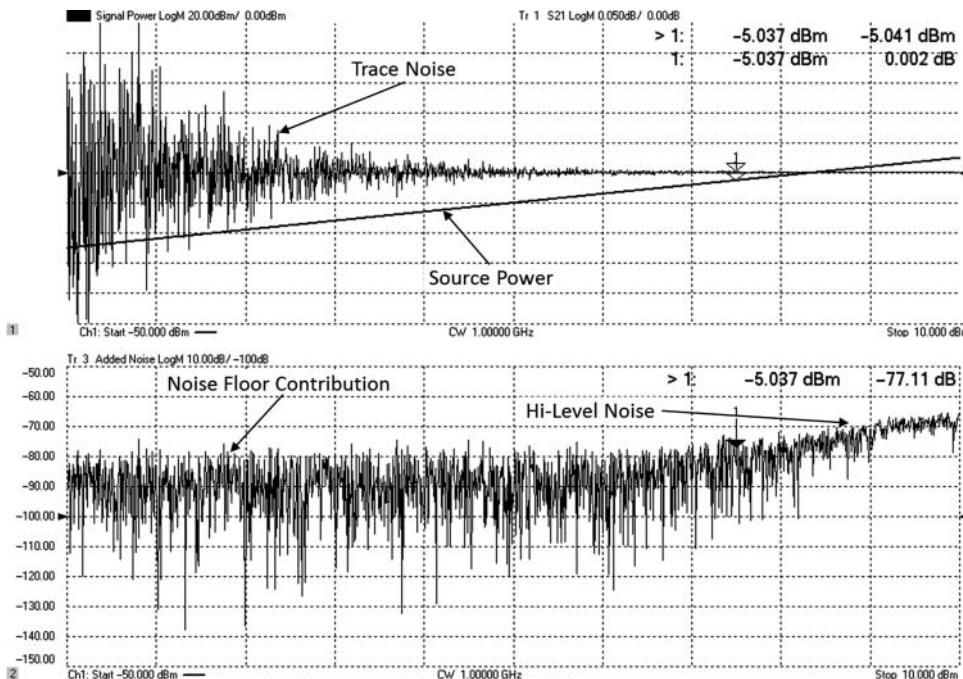


Figure 3.51 Noise added in a signal trace vs. source power.

the answer depends strongly on the particular conditions of the VNA in use. If a VNA is used in a temperature-controlled environment, with very stable test port cables, a calibration can last many days or weeks. In almost all cases, the drift of the VNA itself is less than the drift of the interconnect cables and the repeatability of the test connectors.

The local environment has a strong influence on the stability of a test system. If the temperature varies between day and night (a common occurrence in an office building), then the expansion and contraction of components, particularly cables inside and outside the VNA, can degrade the quality of the calibration, and small ripples will start to appear in the calibrated measurements.

Even in a temperature-controlled lab, the heating and cooling systems can cause many degrees of localized temperature changes, and care must be taken to isolate the air flow of the heating and cooling systems from the test equipment.

Short-term drift can occur over several minutes of measurement time and can be related to aspects such as relaxation in test port cables (a phenomena whereby after flexing, a cable takes some time to return to a previously quiescent state) and slowly varying response due to subtle heating and cooling caused by external environmental factors, as well as slight internal heating in the VNA itself. This internal heating can occur if a VNA's internal construction is comprised of several modules that turn on and off as the VNA sweeps across bands of frequency or from one port to another. For the best of metrology measurements, these drifts can be avoided by always ensuring that the swept data is taken with the same dynamics. For example, to ensure that the delay on a retrace from one sweep to another is consistent, it is

sometimes best to take groups of sweeps rather than single sweeps. Single sweeps can have arbitrary delay between a first and second sweep for forward and reverse sweeps. But a group of sweeps have the same dynamic on the second and subsequent sweeps. These effects are very small indeed and only need to be attended to in the cases where the highest metrology is required.

As described in Section 3.10, the use of in-situ calibration refresh modules (CalPods) will completely eliminate the drift effects of cables, connectors, and other external components.

3.16.7 Interpolation of Error Terms

Whenever calibration topics are discussed, it is almost always assumed that the VNA will be set to exactly the desired frequencies before the calibration is performed, and the measurements will be made only at those frequencies. Older analyzers such as the HP 8510A would turn off the calibration if the frequencies were changed at all. Starting with the HP 8753A, interpolation of error terms was provided, which allowed one to change the number of points, and start or stop frequencies as long as the final frequency set was within range of the original calibration. Later, the Wiltron 360 and the HP 8510 introduced the concept of “zoom-cal” that allowed changing the start and stop frequencies but would reset all the frequencies to land exactly on the original calibration points, reducing the overall number of points. This provided a way to zoom the display although no more data resolution was provided.

Interpolation of error terms has often been controversial, and even experts at the same measurement company have not always agreed on its utility, but if the limitations are well understood, interpolation of error terms can be set up to generate reasonably small errors, often less than the specified uncertainty of the VNA. Lowering the span or changing the number of points slightly will cause the VNA to interpolate the error terms, and one can look for any changes in the VNA trace after interpolation to see if errors occur.

Error terms such as transmission tracking and reflection tracking have slowly varying functions with respect to frequency and are usually more amenable to interpolation than match terms. Match terms often have responses that vary dramatically with frequency, because they are composed of mismatch elements that are spaced some distance apart, such as mismatch at each end of a cable. Since the functions vary quickly with frequency, it is more difficult to interpolate between data points. While simple interpolation of complex numbers interpolates the real and imaginary parts separately, a preferred method is to interpolate the magnitude and phase separately. Note that mismatch elements separated by transmission lines form circles on the Smith chart. Recognizing this, some VNAs utilize circular interpolation to generate an improved interpolation result. Circular interpolation uses three points to define a circle and then computes the interpolated result by linearly interpolating between the angles of the two points surrounding the targeted point, based on frequency spacing. Figure 3.52 (upper) shows an example of the circular interpolation and a linear interpolation of load match of the same points, along with a calibration done on the exact points. In this case, calibration was performed over a 50 MHz span with 6 points (10 MHz spacing) and then interpolated to 201 points. A second calibration was performed at 201 points. The 6 point cal indicates the trajectory that linear interpolation would apply, while the other smoother traces are the 201-point circular interpolation and the 201-point calibration (nearly indistinguishable on the Smith chart). The lower window shows the difference between the interpolated load match

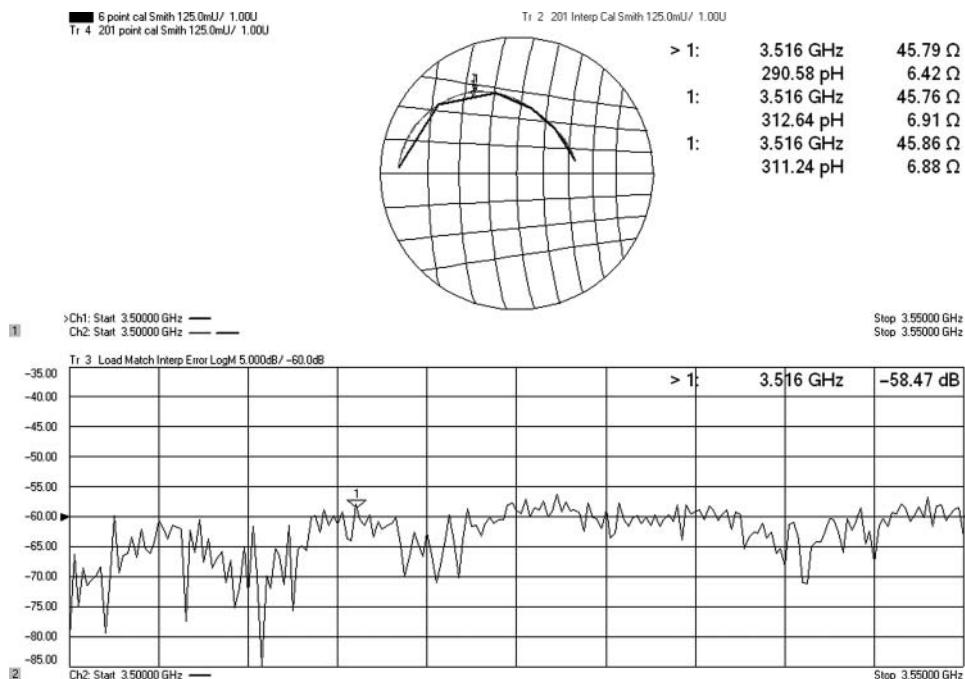


Figure 3.52 Circular interpolation of the load match term.

and the 201-point measured load match. Over most of the band, the interpolation error is less than -55 dB, which is on the same order as the error in the calibration kit.

Of course, if the frequency point density is not sufficient to capture three points of a circle on the Smith chart, the interpolation will result in large errors. One can estimate the number of points required by determining the lengths associated with the mismatch error terms. Usually, this will be the somewhat longer than the length of the test port cable. For good interpolation, the point density should be set to something on the order of 25 points for each wavelength so that the phase shift between points is less than 15° . For a typical test situation, where the test port cable is approximately 1 m, a point spacing of about 5 MHz usually gives good results.

An example of interpolation results is shown in Figure 3.53, where a calibration was performed with point spacing of 10 MHz, over a 1–10 GHz frequency range. After the calibration, the span was changed to a range of 10 MHz to 1 GHz with 1 MHz-point spacing. The resulting S11 and S21 plots of an airline measurement show the error due to interpolation artifacts. The error from interpolating is less than 0.035 dB over the entire trace for S21 and less than –45 dB residual match for S11.

Pay particular attention to the points at low frequencies. One issue with interpolation is the assumption that the error term functions are smooth. But many VNAs operate in octave bands of frequency, due to source stimulus or LO implementations, and often at these band breaks there can be a step discontinuity of the error term. At low frequencies there are many band breaks. Furthering the problem at low frequencies is the coupler roll-off, which makes the tracking responses change rapidly near the bottom of the VNA frequency range. Left alone,

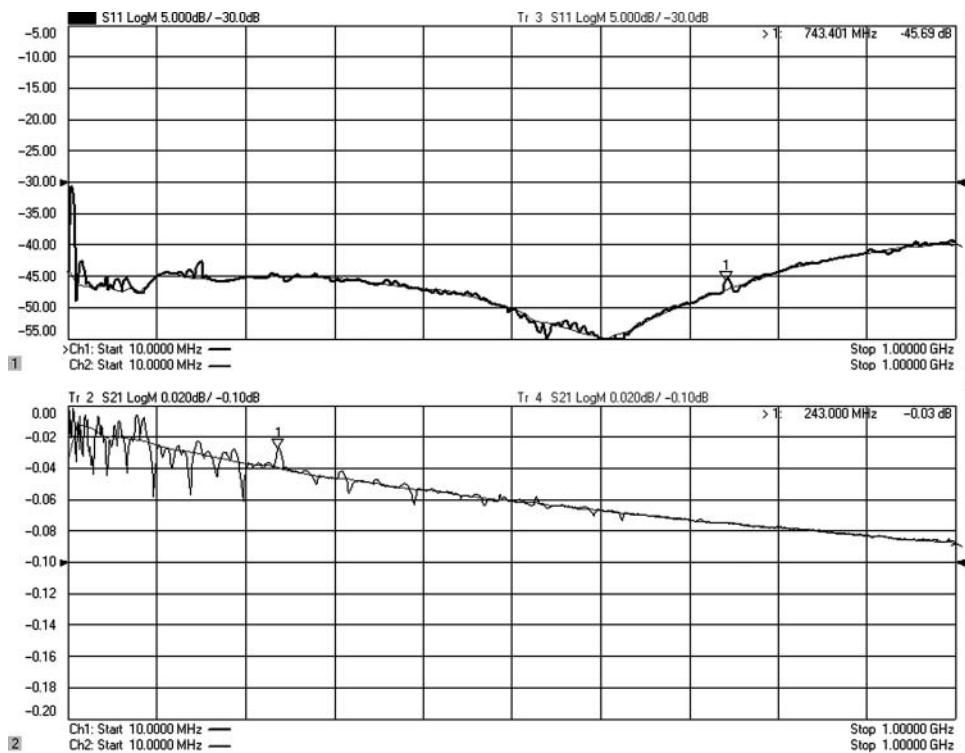


Figure 3.53 Interpolation results with various point spacing.

these steps would cause large errors in interpolation. For that reason, some VNAs incorporate a factory-generated receiver calibration, the purpose of which is to correct the raw receiver response so that it provides approximately correct values and no step discontinuities. Having this receiver calibration in places means the error terms based on receiver measurements will not have step discontinuities either but should maintain a smooth response.

One common point of confusion is whether the DUT characteristics have any effect on whether interpolation can be used or its quality; it does not. The DUT characteristics have no effect on the interpolation, even very narrow band DUTs, DUTs with long delay, or DUTs with complex frequency response may be measured with interpolation. It is not the measured data that is interpolated, but only the error terms. If the error terms are smooth over the range of interest, interpolation will work quite well regardless of the DUT response.

3.16.8 Calibration Quality: Electronic vs. Mechanical Kits

From a specification or theoretical standpoint, the best TRL mechanical calibration kits provide the highest quality of calibration. Next are Ecal standards using the best Ecal modules, followed by SOLT with sliding loads. Fixed-load SOLT calibrations typically have the poorest performance.

Table 3.3 S21 uncertainty vs. calibration kit

Uncertainty computed at 20 GHz with Keysight N5242A option 423 VNA, 3.5 mm		
Kit Type	Worst Case S21 Uncertainty (dB)	RSS Uncertainty (dB)
85052C TRL short line	0.047	0.029
N4291B Ecal	0.081	0.057
85052B SOLT sliding load	0.134	0.103
85052D SOLT fixed load	0.192	0.166

In the earliest versions of Ecal, the module stability and calibration methods made the calibration quality worse than the SOLT calibrations, but with unknown thru calibrations, modern Ecal modules perform better than any but the best metrology TRL cal-kits.

In fact, if RF cables are used, the error from the cable flexure will undoubtedly cause sufficient errors in the TRL calibration to degrade its quality below that of the Ecal. If one includes the likelihood of human handling error and added connector repeatability for mechanical calibrations, then there is little question that in practice, Ecal modules almost always provide superior calibration to mechanical calibration kit methods. In some cases, it is best to modify the default Ecal method to use separate unknown thru (rather than the Ecal as unknown thru) to minimize the cable flexure after cal. For low or moderate-loss DUT, using the DUT itself as the unknown thru can produce the best results. This gives the Ecal quality to the one-port error terms and improves the stability of the thru measurement.

Table 3.3 shows the S21 uncertainty (worst case, and RSS with coverage factor of 2) for various calibration kits used with an Keysight PNA-X VNA. Choosing a different VNA with different source and load match will change the value of the uncertainty, but the relationship between the relative quality of the calibration with each cal-kit will remain the same.

Reference

- Keysight Application Note 154, (2014) Available at <https://www.keysight.com/zz/en/assets/7018-06743/application-notes/5952-1087.pdf?success=true>
- Davidson, A., Jones, K., and Strid, E. (1990). LRM and LRRM calibrations with automatic determination of load inductance. In: *36th ARFTG Conference Digest-Fall*, vol. **18**, 57–63. ARFTG.
- Engen, G.F. and Hoer, C.A. (1979). Thru-reflect-line: an improved technique for calibrating the dual six-port automatic network analyzer. *IEEE Transactions on Microwave Theory and Techniques* **27** (12): 987–993.
- A. Ferrero and U. Pisani, “QSOLT A new fast calibration algorithm for two port S parameter measurements,” in 8th ARFTG Dig., San Diego, CA, Dec. 5–6, 1991.
- Ferrero, A. and Pisani, U. (1992). Two-port network analyzer calibration using an unknown ‘Thru’. *IEEE Microwave and Guided Wave Letters* **2** (12): 505–507.
- Fitzpatrick, J. (1978). Error models for system measurement. *Microwave Journal* **21**: 63–66.
- Garelli, M. and Ferrero, A. (2012). A unified theory for s-parameter uncertainty evaluation. *IEEE Transactions on Microwave Theory and Techniques* **60** (12): 3844–3855.

- Marks, R.B. (1997). Formulations of the basic vector network analyzer error model including switch-terms. In: *50th ARFTG Conference Digest-Fall*, vol. **32**, 115–126. ARFTG.
- D. Rytting, “An Analysis of Vector Measurement Accuracy Enhancement Techniques,” RF & Microwave Symposium and Exhibition, 1980.
- D. Rytting, “Network Analyzer Error Models and Calibration Methods,” RF & Microwave Measurements for Wireless Applications (ARFTG/NIST Short Course Notes), 1996a.
- D. Rytting, “Network Analyzer Error Models and Calibration Methods,” RF &24 Microwave Measurements for Wireless Applications (ARFTG/NIST Short Course Notes), 1996b.

4

Time-Domain Transforms

4.1 Introduction

For most engineers, their introduction to electrical circuits is based on the time-domain response of the signals and circuits. The first experiments in college-level electrical engineering courses are measuring sine wave signals on an oscilloscope. The first lab exercise using electrical networks is often finding the Bode plot response of a filter by measuring the output sine wave from the filter on an oscilloscope while varying the frequency. The better labs include measuring the phase from the sine wave zero crossing. From this, the concept of the frequency response of a network is often developed, with the frequency domain being demonstrated to be the Fourier transform of the time-domain impulse response, and from then on, the electrical engineer works and thinks in the frequency domain. The penultimate instrument for frequency domain measurements is the vector network analyzer (VNA), and it provides unrivaled accuracy in obtaining the frequency-domain response. For the new radio frequency (RF) or microwave engineer, the frequency-domain response becomes second nature, and the time domain response of networks is sometimes considered an anachronistic approach, useful as a learning tool and then forgotten.

However, for the microwave engineer working on distributed circuits, components separated by cables, transmission lines, or waveguides, the time-domain response provides unique insight into the attributes of the circuit as well as methods for improving measurement results by removing artifacts caused by test fixturing and equipment that is separated in time from the device to be measured. The details for using time domain in a variety of situations are discussed later in this chapter, but first it seems appropriate to put down in one place sufficient details so that the time-domain response as shown on a VNA is clearly understood.

First, a frequently asked question is “Why doesn’t the VNA time domain transform match the result I get if I run an fast Fourier transform (FFT) on the data?” It should be understood that the time-domain response is not at all the same as the FFT (or more correctly, inverse FFT) of the frequency response and has many subtleties that can cause confusion in the casual observer. To that end, the next few sections give the exact mathematical details of the time-domain transform used in many modern VNAs. The VNA time domain transform is developed here from the definition of the Fourier transform, by applying the various limits and compensations that are necessary when used in the VNA. By far, this is the most rigorous use of mathematics in this book; the author apologizes in advance for copious use of integrals.

4.2 The Fourier Transform

While a network is mathematically characterized by its transfer function, the frequency response of a network provides the physically measurable response of a network, utilizing sinusoidal signals as the stimulus and measuring the response as magnitude and phase changes in the stimulus signals. Fourier analysis is ideally suited to represent the physical response and can provide for useful analysis of a network. However, measurement systems are limited to measuring finite frequency points over specified bandwidths, so any interpretation of the measurements must include these limitations. This chapter provides some important details of Fourier analysis as applied to device measurements using VNAs. Since the data is measured in the frequency domain, the transformation we are most interested in is the inverse Fourier transform (IFT), generating the time-domain response from the frequency-domain data. Most statements about the IFT have corollaries in the forward transform.

4.2.1 The Continuous Fourier Transform

The Fourier transform can be interpreted as a Laplace transform with the special case of $s = j\omega$. Many of the significant theorems of the Fourier transform are quite similar to their Laplace counterparts, and those that are particularly useful are presented here. When data is measured in the frequency domain, an IFT is used to determine the time-domain response of the device-under-test (DUT), such as a filter or transmission line. If the data represents the frequency response of the filter, then the inverse transform represents the impulse response of the filter. Since the Fourier transform plays such a key role in the VNA time domain transform, it is appropriate to review some of its details, as well as standardize the nomenclature. The Fourier transform pair (forward and inverse) is defined as

$$\mathbf{F}(f(t)) = F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt \quad (4.1)$$

$$\mathbf{F}^{-1}(F(\omega)) = f(t) = \frac{1}{2\pi} \cdot \int_{-\infty}^{\infty} F(\omega) e^{j\omega t} d\omega \quad (4.2)$$

applied to analytic functions $f(t)$ and $F(\omega)$ over all time and all frequency, respectively (Glass 1976). The careful reader will note that nomenclature used in the forward transform by electrical engineers differs slightly from the commonly defined transform where the value for frequency used (Bracewell 1986) is $\omega = 2\pi s$.

4.2.2 Even and Odd Functions and the Fourier Transform

Functions are even if $F(\omega) = F(-\omega)$, and they are odd if $F(\omega) = -F(-\omega)$. All functions can be represented as a sum of an even function and an odd function. Evenness, oddness, and other types of symmetry can simplify calculating transforms and is often assumed for cases of some transforms. A function $f(t) = e(t) + o(t)$ has the Fourier transform

$$F(\omega) = 2 \int_{-0}^{\infty} e(t) \cdot \cos(\omega t) dt - 2j \int_{-0}^{\infty} o(t) \cdot \sin(\omega t) dt. \quad (4.3)$$

where $e(t)$ and $o(t)$ are even and odd functions, respectively. From this result many Fourier transform relationships can be deduced. For modeling physical functions, a key relationship is that for a pure-real time functions, $f(t)$, the Fourier transform must be in the form of

$$\mathbf{F}[f(t)] = E(\omega) + jO(\omega) \quad (4.4)$$

That is, the Fourier transform of a pure real time function has an even real part and an odd imaginary part.

4.2.2.1 Hermitian Functions

The transform of functions such as those described in Eq. (4.4), which an even real part, and an odd imaginary part are called *hermitian*. This can also be written as $F(\omega) = F^*(-\omega)$. Functions which are real and symmetric (even), that is, which exist from negative to positive infinity, with the response in negative time equal to that positive time, have pure real transforms but do not exist in nature. Time functions that represent real networks are pure-real and non-symmetric (that is, $F(t)$ for $t < 0$ is 0) and must have hermitian transforms. Note that all physically realizable networks have non-symmetric real impulse responses, due to causality (the output response of a network to an impulse at time zero must be zero for all negative time), and thus must have hermitian Fourier transforms. A consequence of this is if one knows the frequency response in the positive frequency domain, one also knows the negative frequency response, since $F(\omega) = F^*(-\omega)$.

4.2.3 Modulation (Shift) Theorem

Many filter derivations and communication analysis are based on low-pass to band-pass transformations. This represents a shift in frequency. It is sometimes useful to use a similar transformation in time domain analysis, if the DUT is band limited, such as a filter. The shift or modulation theorem can be derived from the definition of Fourier transform.

$$\text{if } \mathbf{F}^{-1}(F(\omega)) = f(t), \text{ then } \mathbf{F}^{-1}(F(\omega + \Delta\omega)) = f(t) e^{-j\Delta\omega t} \quad (4.5)$$

Note that the resultant time function is in general a complex function, so pure shift in frequency is not physically realizable. To transform a low-pass prototype to a realizable band-pass filter, one must replicate a positive shifted response and a negative shifted response. Thus, if $H_{LP}(\omega)$ is a low-pass filter's frequency response, and

$$H_{BP}(\omega) = H_{LP}(\omega + \omega_0) + H_{LP}(\omega - \omega_0) \quad (4.6)$$

is the band-pass filter frequency response, then the inverse transform of this is

$$h_{BP}(t) = h_{LP}(t) e^{-j\omega_0 t} + h_{LP}(t) e^{+j\omega_0 t} \quad (4.7)$$

Expanding the complex exponential, we find that

$$h_{BP}(t) = h_{LP}(t) \cos(\omega_0 t) - j\underline{h_{LP}(t) \sin(\omega_0 t)} + h_{LP}(t) \cos(\omega_0 t) + j\underline{h_{LP}(t) \sin(\omega_0 t)} \quad (4.8)$$

with this result

$$h_{BP}(t) = 2 \cdot h_{LP}(t) \cos(\omega_0 t) \quad (4.9)$$

The sum of these two shifts results in the imaginary terms canceling. The real portions add, and the result is that if $h_{LP}(t)$ is the low-pass prototype time (or impulse) response, the correct band-pass impulse response of the band-pass filter will be a cosine wave at the center frequency of the band pass, with an envelope of two times the low-pass prototype's impulse response. However, this band-pass time response is *not the same* as the response obtained from the band-pass mode of a network analyzer time domain transform, which will be discussed in some detail.

4.3 The Discrete Fourier Transform

Since measured frequency response of networks consists of discrete data, it is appropriate to discuss the discrete version of the IFT to determine the associated time response. The inverse discrete Fourier transform (IDFT), which is defined only at discrete time points, for a discrete frequency data set, is

$$f(\tau) = \sum_{n=0}^{N-1} F(n) e^{j2\pi(n/N) \tau} \quad (4.10)$$

where (n/N) is analogous to frequency in samples per cycle, τ is the discrete time increment, and $F(n)$ is the discrete frequency data set. The inverse fast Fourier transform (IFFT) is an efficient way to compute $f(\tau)$ over the entire discrete time set. It might appear that the conversion of VNA frequency-domain data to the time domain can be simply accomplished with an IFFT for computational efficiency. However, the IFFT limitations on the flexibility of the data (time) output can hide important effects that occur between calculated time samples, as described next. Further, much more conditioning is done in the VNA transform to enhance its applicability to practical problems.

4.3.1 Fast Fourier Transform (FFT) and Inverse Fast Fourier Transform (IFFT)

The FFT and IFFT are well-known algorithms for calculating the Fourier transform pair of a discrete data set, as described in Eq. (4.10). If the discrete data set is generated from a sampled data set of a frequency response and the data is sufficiently sampled as described in this section, then the IFFT generates the time response of the network associated with the sampled data. FFTs and IFFTs have the attribute of greatly reducing the numbers of computations needed to compute a Fourier transform but are limited in the data that is used and presented. One common limitation on FFT/IFFT transforms is that the sampled data and transformed data must have the same number of points. Some transforms also require that the number of points be in the form of 2^n , and all IFFTs distribute this limited number of points across the entire range of the time transform, which is the inverse of the frequency spacing of the VNA data.

4.3.1.1 Fine Structure Response

If an IFFT is applied to a frequency response, the resulting time response must have the same number of points, and the time intervals must evenly span the time period. A consequence of this is that fine-grain time response is not necessarily evident in the IFFT data, and this fine-grain response is where much of the intuitive understanding of the network can be found. The IFFT is equivalent to the analytic IFT sampled evenly over a time period with the number of sample points equaling the number of frequency response points. Thus, any time domain response information that is present between these points is not evident in the IFFT data. This is also true of the FFT of a time-sampled signal.

It is illustrative to use a familiar example to demonstrate this fact. Take a time function shown in Figure 4.1 (left) consisting of a cosine signal of known frequency such as

$$V(t) = 1 \cdot \cos(8.5t) \quad (4.11)$$

If several cycles of the time signal are sampled at higher than twice the highest frequency, it is sufficiently sampled to avoid aliasing. One might naively assume that the FFT of this time signal should return the original frequency of the time waveform. However, if the frequency of the signal is not synchronous with the sampling, the FFT does not have an output “bucket” at the frequency of the cosine, and the FFT appears to have two main output signals as the spectrum of the sampled time waveform shows in Figure 4.1 (middle), neither of which is the correct amplitude of 0.5 based on the time function. Thus, the fine-grain nature of the signal is not revealed by the FFT. A Fourier transform can be performed at discrete frequencies over the range of the two largest-valued FFT outputs, using the same time data set, from which the correct magnitude of the original frequency of the signal is revealed, as shown in Figure 4.1(right). In fact, since the time data is a finite set of discrete sampled points, the frequency response must be a periodic and continuous function with an infinite response to represent the transitions at the start and end of the data set. The FFT is exactly a sampled version of this continuous frequency function. The non-zero values of the FFT for all the other frequencies is a consequence of the taking the time data over a finite time, essentially turning the sine wave signal into a pulsed sine wave, with values assumed to be zero before

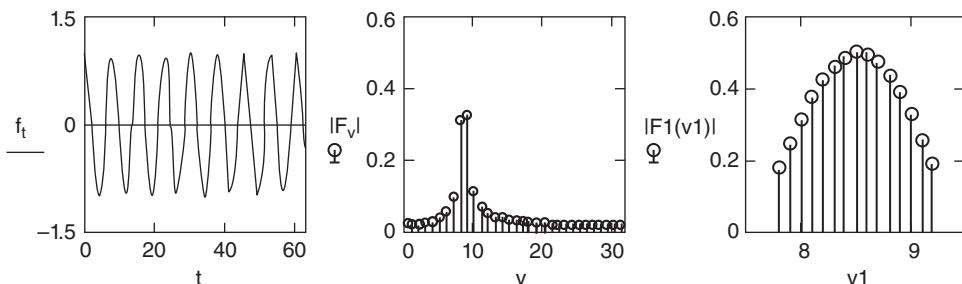


Figure 4.1 Cosine of frequency 8.5 Hz (left), FFT of the waveform in the left plot (middle), Fourier transform of the waveform around the frequency 8.5 Hz (right).

and after the sampled data. This has the effect of spreading the signal over several frequency buckets. Reducing this effect is a key attribute of the VNA time domain transform.

To get faster computation speeds, FFTs are often used instead of direct calculations of the discrete Fourier transform (DFT). However many commercially available signal analysis tools take further shortcuts in calculating the FFT. One common shortcut is to assume that the time response is real. From this, the frequency response must be hermitian, and therefore only half the FFT needs be calculated to obtain the full frequency response. With an IFFT, it is common to assume a hermitian frequency response input and calculate only the real portion of the output time signal. Thus, the IFFT is simply two times the IFFT of the positive half of the real part of the input frequency response, plus the zero frequency (DC) term. However, there are several cases, such as the band-pass transforms, where it is useful to consider frequency responses that are not hermitian, and in these cases, care must be used in considering the shortcuts that are permissible when calculating IFFTs.

4.3.2 Discrete Fourier Transforms

The fine structure of a time response can be determined if an IDFT is used, in which the time axis can be arbitrarily small. If a small time spacing were used for the FFT, an extremely large number of frequency points would have to be used as the input, greatly slowing measurement time to generate the frequency response terms. However, the DFT takes considerably longer to calculate than the FFT and therefore is also not satisfactory where real-time transformation is needed. Fortunately, if the transform is needed over a relatively small portion of the time response, faster methods of calculation are available that provide finer resolution in time, without sacrificing speed.

Most faster-transforms methods require equally spaced time points. The time spacing will be the time span divided by the number of points minus one. However, there are no restrictions on the start or stop times. If the start and stop times are chosen to be $t_{start} = 0$ and $t_{stop} = 2\pi/\Delta\omega = 1/\Delta f$, and the number of time points equals the number of frequency points, then the DFT will return the same values as the IFFT.

4.4 Fourier Transform (Analytic) vs. VNA Time Domain Transform

The limitations of the IFFT as applied to microwave measurements required other techniques for analyzing these networks. The time-domain transform of vector network analyzer measurements was first introduced in 1974 (Hines and Stinehelfer 1974) and has been widely used since its real-time commercial introduction with the HP 8510A (1984), which allowed increased accuracy and real-time gating (Sharrit n.d.; Rytting 1984). This VNA provided the capability to calculate the time domain response of the frequency-domain data, using a form of the IFT. However, there were several modifications that are important to note, which causes the time domain response of the VNA to be different than the actual IFT of the frequency response of a network, that is, different from the impulse response of the network being measured. These differences come from the mode of the VNA transform (low-pass step, low-pass impulse, or band-pass impulse), data windowing and truncation, window renormalization, and data gating. For much of this time, the principal use was in low-pass impulse mode for fault location, and much has been written about the interpretation of the low-pass step mode

time domain response. Recently, the time domain response has been applied to solving the problem of filter tuning, using the band-pass mode (Dunsmore 1999).

A rigorous analysis comparing this time-domain mode to the analytically derived impulse response may be obtained by applying appropriate functions to the analytical frequency response until the IFT of this modified response exactly matches the VNA time-domain response. Each of these functions applied to the frequency response can be evaluated in the time domain, and their associated time-domain effects can be individually determined. This approach differs from others (Hines and Stinehelfer 1974) in that Hines and Stinehelfer develop the time-domain response from assuming a periodic time function, the Fourier transform of which reproduces the measured frequency response. However, in the treatment presented next, a continuous analytic frequency response is assumed, and modifications are applied to account for discrete frequency sampling and windowing to directly obtain the VNA time-domain transform in terms of the original frequency response and these modifications.

4.4.1 Defining the Fourier Transform

The IFT of a function provides directly the impulse response of that network and provides the same time-domain result as driving the network with an impulse, $\delta(t)$, and determining the time response. Figure 4.2 shows the analytically derived transform of S_{11} of a three-pole Butterworth filter (meaning the reflection frequency response is calculated using standard network theory, and the IFT from Eq. (4.2) is calculated to get the time response), along with a VNA time-domain transform of the same function. Clearly, they are not the same though they have some similarities in their structure. The differences will be reconciled in the following sections by describing the way in which each aspect of the VNA measurement must be accounted for with the appropriate mathematical transformation to achieve the same result as the IFT.

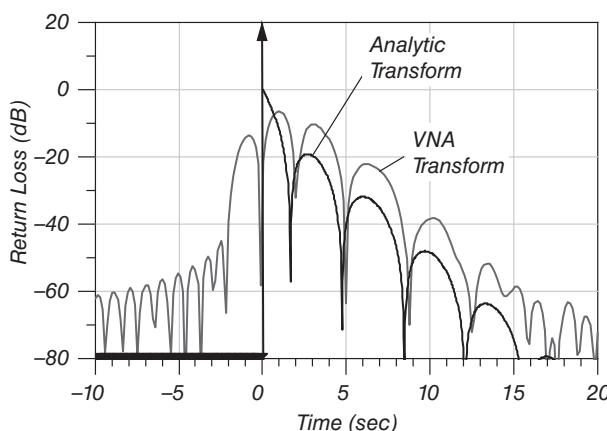


Figure 4.2 Analytically derived impulse reflection response versus VNA time-domain response for a three-pole Butterworth filter.

4.4.2 Effects of Discrete Sampling

The Fourier transform operates on continuous functions, while the VNA time-domain transform must operate on measured (discrete) data. One approach is to assume the measured data is a sampled version of a continuous analytic frequency response. Since the data applied to the time-domain transform is discrete, the time-domain transform must differ from the analytically calculated IFT of the network, but an equivalent discrete representation of an analytic function can be obtained by a mathematical representation of the sampling process. Note that such a time function would be identical to one determined in (Hines and Stinehelfer 1974), but this approach is more readily applicable to the problem of comparing the VNA time-domain transform to the analytic impulse response of a network.

A frequency sampling function can be represented as $III(\omega)$, which is defined as

$$III(\omega) = \Delta\omega \cdot \sum_{n=-\infty}^{\infty} \delta(\omega - n \Delta\omega) \quad (4.12)$$

and can be visualized as a collection of delta functions with $\Delta\omega$ spacing. The effect of discrete data in the measured frequency response can be analyzed by forming a sampled function composed of the analytic frequency response multiplied by the sampling function, such that its value is zero between measured points, and the scaling factor of the delta function at each frequency is the measured value of the frequency response. The IFT of the sampled function, $f_s(t)$, can now be represented analytically by multiplying the original frequency response function by a sampling function and taking the IFT of the result:

$$f_s(t) = \mathbf{F}^{-1}(F_s(\omega)) = \frac{1}{2\pi} \cdot \int_{-\infty}^{\infty} F(\omega) \cdot \Delta\omega \cdot \sum_{n=-\infty}^{\infty} \delta(\omega - n \Delta\omega) \cdot e^{j\omega t} d\omega \quad (4.13)$$

or, through the operation of the integral on the delta function:

$$f_s(t) = \mathbf{F}^{-1}(F_s(\omega)) = \frac{1}{2\pi} \cdot \sum_{n=-\infty}^{\infty} F(n\Delta\omega) \cdot \Delta\omega \cdot e^{jn\Delta\omega t} \quad (4.14)$$

This operation can also be understood by noting that multiplication of two functions in the frequency domain is the same as convolving the inverse transforms of functions in the time domain. Convolving a function by a delta function returns the original function, at the origin of the delta function. Thus, the inverse transform of the sampling function returns another sampling function,

$$III(t) = \frac{1}{\Delta\omega} \cdot \sum_{n=-\infty}^{\infty} \delta\left(t - n \frac{1}{\Delta\omega}\right) \quad (4.15)$$

Sampling in the frequency domain is the same as convolving the original time-domain response by the sampling function $III(1/\Delta\omega)$. Therefore, the transform of an analytic function can be related to the transform of the discrete sampled version by convolving the inverse impulse response of the original function with the sampling function of Eq. (4.15). The effect of discrete data sampling can be seen to create images of the original function (sometimes called *aliases*) spaced at the inverse of the sampling spacing. The time range of $\pm\pi/\Delta\omega = \pm 1/2\Delta f$ is referred to as the alias-free range of the inverse transform for

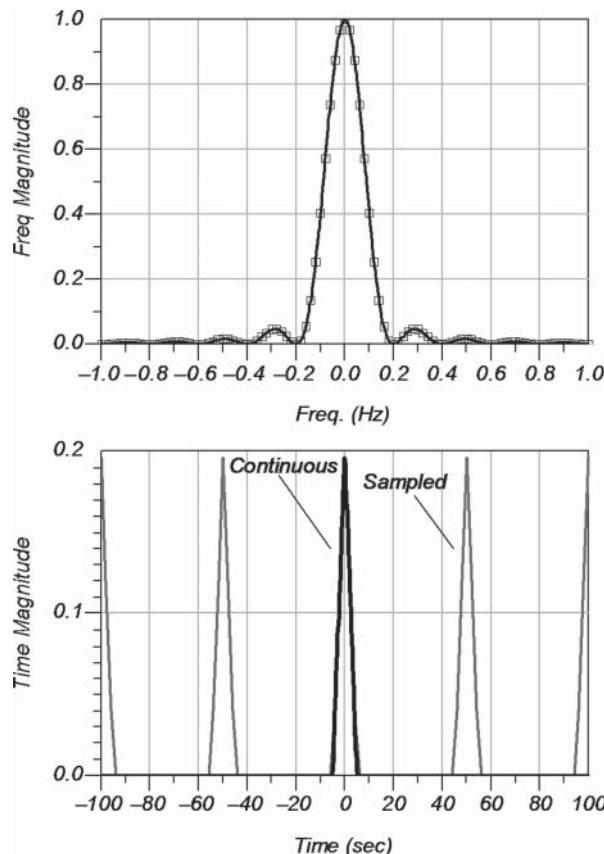


Figure 4.3 Sinc-squared frequency response continuous and sampled with a sampling frequency of 0.02 Hz; time-domain response continuous and showing repeated time responses due to sampling (lower).

sampled data. Many commercial products display a maximum range of $\pm 1/\Delta f$. If the impulse response of the original function does not tend to zero by $\pm 1/2\Delta f$, then the appearance of the transformed sampled function in the alias-free range will be distorted by effects from previous and subsequent images. Figure 4.3 (upper) shows a frequency plot of the sinc^2 function, $(\sin(x)/x)^2$, the transform of which is well known analytically to be a triangle pulse, along with the sampled data points. The lower plot shows the IFT of the continuous function, as well as the IFT of the sampled function.

The inverse transform of a sampled frequency response must have an infinitely repetitive (periodic) time response. Even if the frequency response is discrete, the time response may still be continuous. Only if the frequency response is discrete and periodic will the time response be discrete. Since any *real* sampled frequency response must be sampled over a finite frequency span (and thus cannot be periodic), the time response associated with any measured frequency response will be continuous and periodic. That is, every time-domain response displayed on a VNA represents a periodic time function.

4.4.3 Effects of Truncated Frequency

Another consequence of taking a transform of measured data is that the frequency response must be truncated, rather than extend to plus and minus infinity. That is, all VNAs are limited in their range of response measurements, and the sample frequency data will not have an infinite response. For transmission responses, this does not present much of a problem, as the response of most networks act as filters and becomes arbitrarily small at high frequencies; the high-frequency contribution to the inverse Fourier integral is negligible. However, for reflection responses, the value of the response remains large at high frequencies. In fact, these responses are not strictly Fourier transformable, as they do not satisfy

$$\int_{-\infty}^{\infty} |f(\omega)| d\omega < \infty \quad (4.16)$$

However, most reflection functions can be represented with the help of the generalized function, $\delta(t)$. But, if the response is truncated and the response data is finite, then the Fourier transform of the data strictly exists. In fact, if functions are derived from an accurately specified physical quantity, it is a sufficient condition for the existence of a transform. Or put another way, if a thing is real, its time-domain response must exist.

Truncation of the frequency response data of a network is mathematically equivalent to multiplying the data by a rectangular window. In the time domain, this can be represented as convolving the impulse response of the network with a $\sin x/x$ function, which is the inverse transform of the rectangular window. In this way, the inverse transform of truncated data will always have a response with “side lobes” if the original data does not go to and remain zero sometime before truncation occurs. These side lobes can be so large as to obscure the impulse response, and much work has been done to reduce this effect.

For the most part, side lobes, or *ringing* as it is sometimes called, can be controlled through the appropriate use of windowing. Taking the IFT of the product of the original function and a rectangular window can represent the effect of truncated data in the VNA time-domain transform. Referring to Eq. (4.13), this truncation is equivalent to redefining the limits of the integral to be the endpoints of the measured data. Figure 4.4 (upper) shows an example of a one-pole filter response (upper, gray trace) with the analytic function $F(s) = 1/(s + 1)$, or $F(\omega) = 1/(1 + j\omega)$ where $s = j\omega$, along with its truncated frequency response, in Figure 4.4 (upper, black trace). Figure 4.4 (middle) shows the time-domain response of the truncation function, which is the IFT of a rectangular window, which in turn is a $\sin x/x$ function. The filter response has an analytic time response of $f(t) = e^{-t} \cdot U(t)$ (where $U(t)$ is the unit step function), as shown in Figure 4.4(lower, gray trace). The truncation effect on the analytic time response can be obtained by convolving the IFT of the original function with the $\sin x/x$ function, as shown in Figure 4.4 (lower, black trace). From this, the original transform is almost not recognized due to the distortion of the side lobes caused by the truncation effect.

For a sampled data set, over the range of $\omega = -N\Delta\omega$ to $+N\Delta\omega$, the IFT becomes

$$f_s(t) = \frac{\Delta\omega}{2\pi} \cdot \sum_{n=-N}^N F(n\Delta\omega) \cdot e^{jn\Delta\omega t} \quad (4.17)$$

Equation (4.17) might be called the sampled inverse Fourier transform. Note the similarity to the IDFT of Eq. (4.10). The sampled inverse Fourier transform of Eq. (4.17) can be used to calculate the inverse transform for any particular time t , so there are no limits on the time span or point spacing in time, as there is with the FFT.

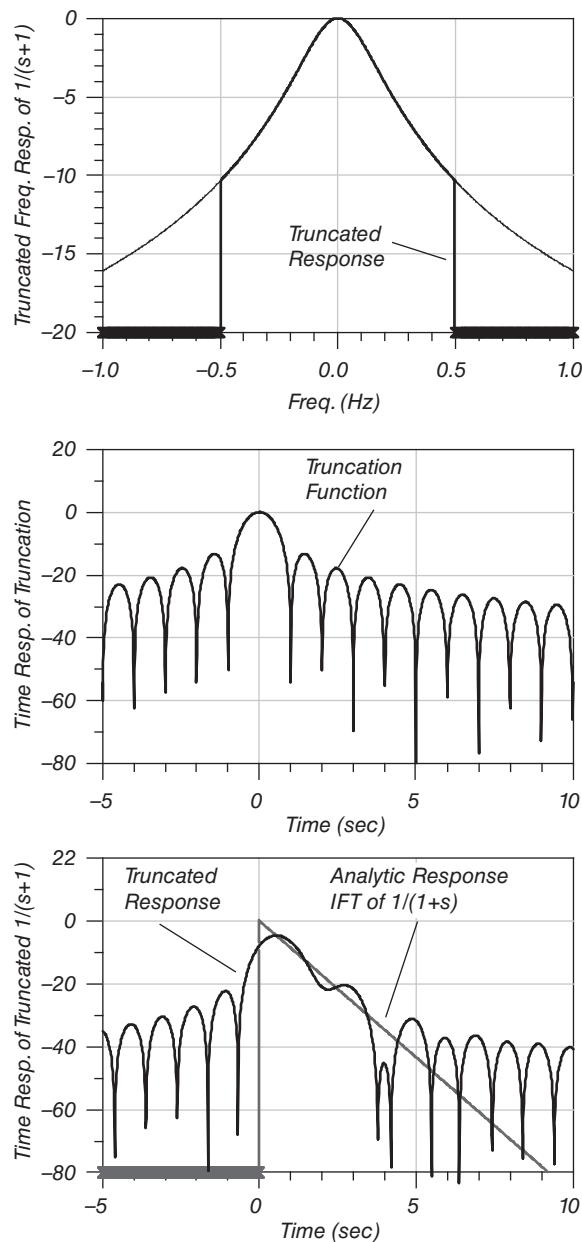


Figure 4.4 One-pole filter frequency response with and without truncation.

4.4.3.1 Causality

A consequence of truncation is that the response appears to be non-causal, that is, the side lobes for the impulse response occur at times before zero. While this is not desired, it is a mathematical fact. With some other processing, the causality violation can be reduced to acceptable levels, as discussed in the next section. Also note that the peak of the function does *not* occur at time $t = 0$ but is somewhat delayed. Again, this is a consequence of truncating the frequency response.

4.4.4 Windowing to Reduce Effects of Truncation

Data truncation is shown previously to have the effect of convolving the original transform with a $\sin x/x$ function. The side lobes of this function are quite high and continue for a substantial extent, often obscuring the desired response of the original function. The effects of truncation are minimized if the original function tends to zero at the frequency endpoints. A windowing function may be applied that gradually reduces the frequency response, thus controlling the side lobes created during the truncation process.

However, the windowing process tends to reduce the sharpness of the original response, spreading pulses, stretching out slopes, and thereby reducing the resolution of the transform and distorting the transitions of the original function. This makes it difficult to assess the true nature of the transformed function. Thus, there is a trade-off between side lobe height and resolution when determining the windowing function. Window functions, including Hanning, Hamming, cosine, and cosine-squared, have been extensively described, and each window function has benefits and drawbacks. Typically there is a trade-off between side lobe suppression and loss of rise time; a window function used commonly in commercial products is a Kaiser-Bessel with β or KB value to set the relative width of the window. A KB value of 0 gives no window, and a KB value of 6 is the normal value used in many commercial VNAs, with a maximum value of 12. For time-domain analysis of small reflections (such as looking at in-line connectors), using KB values as low as 3 provide improved resolution, and the small reflections mean side lobe values don't interfere with the results. If there is a large reflection (such as an open or short at the end of a line), then the side lobes from the large reflection may mask the reflection of interest. In such a case, wider windows will be required to remove the side lobe effect.

Figure 4.5 shows various window factors (upper), the factors applied to a one-pole filter response (middle), and the time response of windowed functions for KB = 0 and 6, along with the analytic impulse response (lower). This windowed version of the transform properly shows the shape of the analytic function, but the rise time is extended and the peak is diminished. Windowing further spreads the time response, adding to the appearance of being non-causal.

To reconcile the analytic impulse response with the VNA time-domain transform, the effects of finite frequency, sampling, and windowing on the analytic IFT can be mathematically represented as f_{SW} (for sampled, windowed)

$$f_{SW}(t) = \frac{\Delta\omega}{2\pi} \cdot \sum_{n=-N}^N F(n\Delta\omega) \cdot W(n\Delta\omega) \cdot e^{jn\Delta\omega t} \quad (4.18)$$

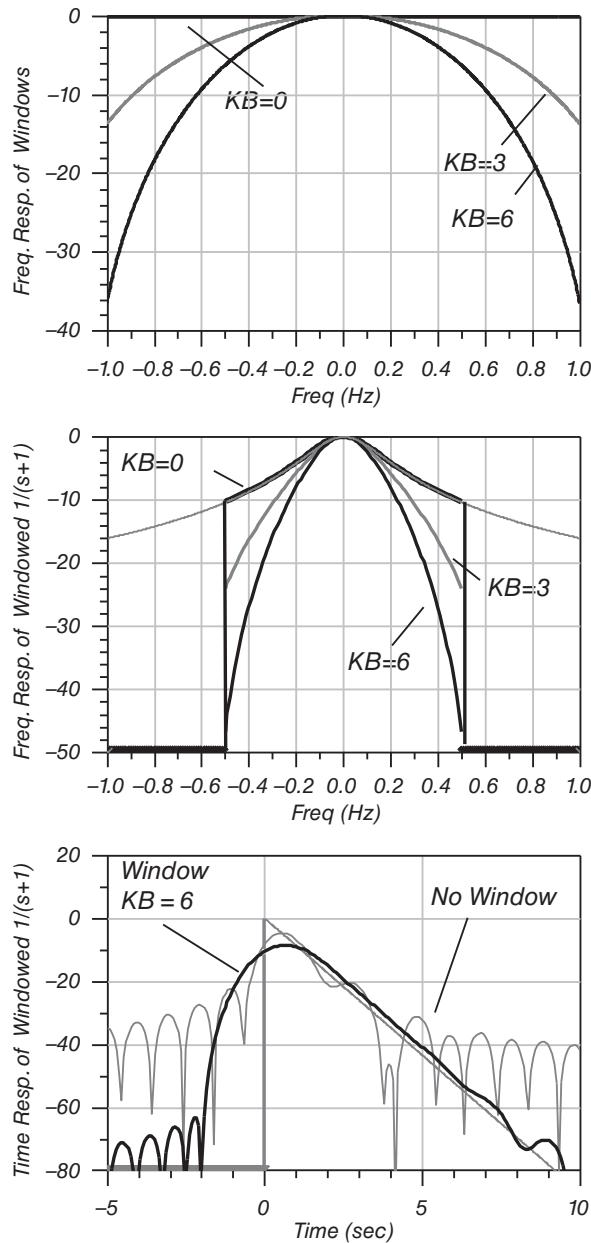


Figure 4.5 Windows for Beta factors 0, 3, and 6 (upper); windows applied to a one-pole filter (middle); time response of windowed trace (lower).

where $W(\omega)$ is the windowing function, and the function is sampled over $\omega = -N\Delta\omega$ to $+N\Delta\omega$. This response includes all the obvious changes to the analytic function, but there is one final modification that must be included such to completely match the VNA time-domain transform, as described next.

4.4.5 Scaling and Renormalization

The value of the time-domain transform has to be renormalized such that it retains its physical meaning. For example, the frequency response of the S_{11} of an ideal open circuit, with no delay, has a value of 1 for all frequency; its inverse transform is a delta function. However, when the data is sampled and windowed, the time-domain transform of the response of an open circuit will be spread by the windowing function and does not return an impulse of unity height. It would be preferable if the time-domain response of the open circuit had a value of unity at time $t = 0$. Taking the sum of the windowing factors provides the correct scaling factor for subsequent transforms

$$W_0 = \frac{\Delta\omega}{2\pi} \cdot \sum_{n=-N}^N W(n\Delta\omega) \quad (4.19)$$

and the renormalized transform becomes

$$f_{VNA}(t) = \frac{1}{W_0} \cdot \frac{\Delta\omega}{2\pi} \cdot \sum_{n=-N}^N F(n\Delta\omega) \cdot W(n\Delta\omega) \cdot e^{jn\Delta\omega t} \quad (4.20)$$

Note that this scales the transform to always return 0 dB for a unit frequency input, regardless of windowing factor. If the data that is being transformed already tends to zero at the band edges, the windowed response will appear higher after this normalization, when compared to an analytic time response. Since the window scaling always maintains a unity peak amplitude, regardless of how wide the window has made the response, it is in effect amplifying the DC and low-frequency responses. For some data, such as a low-pass filter response, this can result in a windowed response that is higher in amplitude than the corresponding analytic impulse response.

4.5 Low-Pass Transforms

Since measured data has a finite frequency sampling, some assumptions are made about the behavior of the sampled function. Vector network analyzers offer alternative assumptions, which yield two different modes of transformation: low-pass mode and band-pass mode.

4.5.1 Low-Pass Impulse Mode

The assumption for low-pass impulse mode is that the underlying frequency response is that of a real network. As such, the frequency response is hermitian, and the time-domain response is pure-real. Also, it is assumed that the network response becomes asymptotic at low frequencies; that is, the response at low frequencies is roughly constant, and the frequency response

beyond the measured frequency range contains no important information about the network. In other words, everything of interest occurs over the frequency of measurement. The data points must be linearly spaced over the range of $\omega = n\Delta\omega$ from $n = 1$ to N . Thus, the frequencies must be harmonically related. For this transform, the windowing function is centered at $\omega = 0$ and extends to the max frequency $\omega = N \cdot \Delta\omega$. From this, it follows that the complex sum in Eq. (4.20) becomes

$$f_{LP}(t) = \frac{\Delta\omega}{2\pi} F(0) \frac{W(0)}{W_0} + \frac{1}{W_0} \frac{\Delta\omega}{2\pi} 2 \cdot \operatorname{Re} \left[\sum_{n=1}^N F(n\Delta\omega) \cdot W(n\Delta\omega) \cdot e^{jn\Delta\omega \cdot t} \right] \quad (4.21)$$

Given a hermitian function, the imaginary parts of the negative and positive transform cancel, and the real parts double, so only the real portion is computed. Further, it is clear that a value must be determined for $F(0)$, which is done with DC extrapolation. From Eq. (4.21), it can be seen that the time-domain transform consists of sums of sines and cosines and that the highest frequency measurement point determines the highest frequency element. Thus, the rise time is determined by the maximum slope of the highest frequency measured. The transform will repeat itself at intervals determined by the frequency step value, which is the same as the lowest frequency point.

4.5.2 DC Extrapolation

In addition to being limited in upper frequency response, measurement equipment is limited to its minimum frequency response. However, the Fourier transform includes effects of the DC value on the frequency response. Since VNAs do not commonly measure the DC response, DC extrapolation is used. Some analysis programs allow DC to be directly entered. DC extrapolation requires the assumption that the network response approaches DC asymptotically, and different algorithms are used by different applications. The consequences of DC extrapolation are discussed next.

4.5.3 Low-Pass Step Mode

Up to now, the discussion has focused on the impulse response of a network. The step response of a network can be useful in directly determining the network characteristics, particularly in the case of concatenated transmission lines, and evokes the normal mode of operation of a time-domain reflectometer (TDR), which employs a stepped DC stimulus. The unit step-function $U(t)$ is defined as

$$U(t) = \begin{cases} 0 & \text{for } t < 0 \\ \frac{1}{2} & \text{for } t = 0 \\ 1 & \text{for } t > 0 \end{cases} \quad (4.22)$$

From this, its Fourier transform may be determined as

$$\mathbf{F}[U(t)] = \pi \delta(\omega) - j \frac{1}{\omega} \quad (4.23)$$

The time-domain step-response may be found by multiplying the Fourier transform of the unit step-function, as shown in Eq. (4.23), by the frequency response, $F(\omega)$, of a network and taking the inverse transform,

$$f_{Step}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) \cdot \left(\pi \delta(\omega) - j \frac{1}{\omega} \right) e^{j\omega t} d\omega = \frac{F(0)}{2} - \frac{j}{2\pi} \int_{-\infty}^{\infty} \frac{F(\omega)}{\omega} e^{j\omega t} d\omega \quad (4.24)$$

Taking the derivative of the step response yields the desired impulse response of the network.

The low-pass mode of the VNA time-domain transform has two forms: low-pass impulse, which is defined by Eq. (4.21), and low-pass step, which is essentially the integral of the low-pass impulse response, with respect to time and with some particular choice for the constant of integration. The step response of the VNA should retain the property that its derivative is the VNA time-domain impulse response, and since the frequency-domain sampling function creates a periodic time-domain response, with a period of $1/\Delta f$, the step response should retain this aspect of the periodicity, and low-pass data will be valid between $t = 0$ and $t = 1/\Delta f$.

Figure 4.6 shows the step-response stimulus (labeled “VNA Unit Step Response”) that meets the properties of having a periodic impulse response for its derivative. This response differs from the square wave response described by Hines and Stinehelfer, and from the plot, it is obvious that this function cannot have a Fourier transform. However, it may be written as the sum of two functions, the first one being periodic (labeled “Periodic Portion” in Figure 4.6) and the second being a ramp function (labeled “Ramp Portion” in Figure 4.6).

The time step-response can be determined from the network function and the unit step stimulus by applying the appropriate Fourier transform to the periodic portion, and some appropriate Laplace transform to the ramp portion. From Eqs. (4.21) to (4.23), the step-response for a sampled, truncated, windowed function can be proposed to be

$$f_{Step}(t) = \left\{ \frac{F(0)}{2} + \frac{\Delta\omega}{2\pi} \cdot 2 \cdot \operatorname{Re} \left[\sum_{n=1}^N \frac{F(n\Delta\omega) \cdot W(n\Delta\omega)}{jn\Delta\omega} \cdot e^{jn\Delta\omega \cdot t} \right] \right\} + F(0) \cdot \frac{\Delta\omega}{2\pi} \cdot t + C \quad (4.25)$$

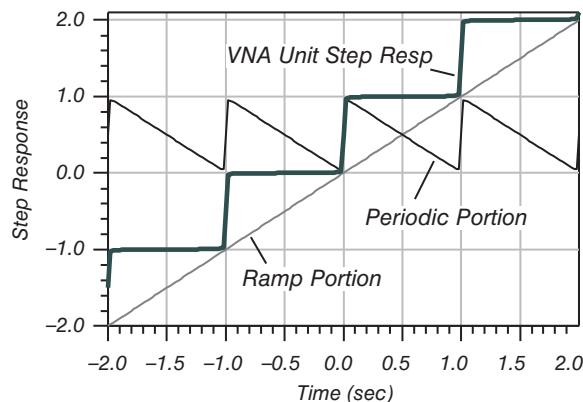


Figure 4.6 VNA unit step response comprised of a periodic portion (which is Fourier transformable) and a ramp portion.

Differentiation of Eq. (4.25) clearly results in Eq. (4.18). The second term in the equation (the ramp portion) is needed if the impulse response contains a DC term. The final constant of integration is chosen to give the proper response value of the transform at time $t = 0$.

Thus, the step-response can be obtained by taking the inverse Fourier transform of the frequency response divided by j times the step frequency, and adding to this a linear time ramp. The time-domain step-response is available only in the low-pass mode.

4.5.4 Band-Pass Mode

The band-pass mode provides an alternative method of the time-domain transform that may be used when the low-pass mode assumption of harmonically related frequencies cannot be met. This might occur, for example, in the measurement of a network that is band-pass or high-pass filtered. The output of a VNA measurement is typically an odd-numbered set of points, linearly spaced in the form of $\omega = \omega_c + n\Delta\omega$ from $n = -N/2$ to $N/2$, and ω_c is the center frequency of the data. The inverse Fourier transform is calculated only on the data points measured, rather than presuming the negative frequency response to be the conjugate of the measured data. That is, the band-pass mode does not presume a Hermitian frequency function and uses data as though the frequency response was positive only (of course, this cannot represent a real device but is useful). Windowing is applied, where the center for the windowing function is the center frequency of the data set. In contrast, the center of the windowing function in low-pass mode is centered on the DC term, or the first point of the data set. The inverse band-pass transform is defined by

$$f_{BP}(t) = \frac{1}{W_0} \frac{\Delta\omega}{2\pi} \cdot \sum_{n=-N/2}^{N/2} F_{BP}(\omega_C + n\Delta\omega) \cdot W(n\Delta\omega) \cdot e^{j(\omega_c + n\Delta\omega)t} \quad (4.26)$$

This is an important difference between the VNA band-pass mode and that described by Hines and Stinehelfer (1974), which results in a pure-real time-domain response. In contrast, the VNA band-pass response results in a complex time-domain response, and this choice of transform is key to the useful application of the band-pass mode response, one example being the application to filter tuning. To illustrate the band-pass transform mode, consider the frequency function of a band-pass filter. The frequency response tends to zero away from the center frequency, so the windowing function will have little effect on the transform. If the frequency response F_{BP} represents a band-pass version of a low-pass prototype response (Blinchikoff and Zverev 1976), such that $F_{BP}(\omega) = F_{LP}(\omega - \omega_C)$ and thus $F_{BP}(\omega_C) = F_{LP}(0)$, the relationship between the time-domain band-pass transform of band-pass filter and the low-pass prototype's frequency response can be established as

$$f_{BP}(t) = \frac{e^{j(\omega_c)t}}{W_0} \frac{\Delta\omega}{2\pi} \cdot \sum_{n=-N/2}^{N/2} F_{LP}(n\Delta\omega) \cdot W(n\Delta\omega) \cdot e^{j(n\Delta\omega)t} \quad (4.27)$$

or in terms of the low-pass time-domain response

$$f_{BP}(t) = e^{j\omega_c t} \cdot f_{LP}(t) \quad (4.28)$$

From this, it follows that the band-pass time-domain mode always returns a complex time-domain response. This effect is due to removing the assumption that the frequency

response contains negative frequency elements. The magnitude response of the band-pass transform is the same as the low-pass prototype.

$$|f_{BP}(t)| = |f_{LP}(t)| \quad (4.29)$$

Thus, the band-pass mode response of the time-domain transform is quite different from the analytic impulse response of the network. Consider a network, such as a filter, that has a low-pass response, $f_{LP}(t)$. If this filter is used as a prototype for a band-pass filter and is shifted to create a band-pass response, the band-pass filter will have an analytic impulse response of

$$f_{Imp}(t) = 2f_{LP}(t) \cdot \cos(\omega_C \cdot t) \quad (4.30)$$

which is pure-real as would be expected of an analytic transform of a real network. So, the band-pass mode transform has, in addition to the windowing, sampling, and frequency truncation effects, an additional effect due to the data being taken as though the network has a single-sided (positive frequency only) response. Also, since the windowing function is centered on the center frequency of the transform, it forces the function to zero at the lowest as well as highest frequencies; there is no point in extrapolating the DC term.

One consequence of the band-pass transform is that the resolution is half that of the low-pass transform. This can be seen from Eq. (4.26), which shows the maximum frequency in the complex exponential is one-half of the frequency span (since the data ranges from $n = -N/2$ to $N/2$). The alias-free range for this transform remains the same as the range of the low-pass transform.

With this introduction to the time-domain transformations used in VNAs, the concept of time-gated measurements can be better understood.

4.6 Time-Domain Gating

Time-domain gating refers to the process of selecting a region of interest in a portion of the time domain, removing unwanted responses, and displaying the result in the frequency domain. Gating can be thought of as multiplying the time-domain response by a mathematical function with a value of one over the region of interest, and zero outside this region. The gated time-domain function can then be forward transformed to display the frequency response without the effect of the other responses in time. The gating effects, however, are somewhat subtle in their response, and there are consequences of the gating function that are not readily apparent.

In practice, the gating is not a “brick-wall” function. This is because a sharp transition in the gate function causes a similar sharp transition in the gated time function. As such, the frequency response will have ringing associated with the sharp transition (as the frequency response is limited to the measured data region). To avoid this ringing, the gating function is windowed in the frequency domain before being transformed to the time domain. For a rectangular time gating function centered at $t = 0$, the Fourier transform can be calculated analytically, with the result that the gate frequency response will have a $\sin(\omega)/\omega$ or sinc(ω) function. The width of the sinc main lobe is inversely proportional to the width of the time gate. If the center of the gate time is not at $t = 0$, the resulting Fourier transform produces a response that corresponds to the sinc function multiplied by a complex exponential factor, namely, $\text{sinc}(\omega) \cdot e^{j\omega t_c}$. This is windowed in the frequency domain, which sets the maximum

gate transition slope (think of it as the gate rise time) in the time domain and reduces gate side lobes. The gate side lobes cause a phenomenon sometimes called *gate leakage*, which is observed when a large reflection signal gated-out near a small reflection signal. The gate leakage can still produce a significant response that will produce an erroneous gated response. The gate function is then transformed to the time domain and multiplied by the time-domain response to display the gated time-domain response. If the gate frequency response is desired, the gated time domain can be transformed back to the frequency domain. In practice, an alternative computation is used whereby the time-gated frequency response is computed by convolving the gate frequency response by the measured frequency response. This method reduces the number of transforms required and makes for faster processing, and this convolution interpretation of gating allows a more intuitive understanding of one of the subtle gating effects, described next.

4.6.1 Gating Loss and Renormalization

A curious effect of the gating function occurs at the endpoints of the time-gated frequency-domain response if applied as described previously: these endpoint regions are lower by 6 dB, as though the gating has caused some loss at the frequency endpoints. In fact, the gating is a kind of filtering function, and the loss is real. The 6 dB offset can be understood by comparing the center point and last point of a gated frequency response of a unit function $F(\omega) = 1$, that is, a frequency response that is constant (such as a thru line). The time-domain response will approach a delta function, $f(t) = \delta(t)$, but the frequency response is just a flat line. This function is the dark solid trace in Figure 4.7.

In the convolution process, the gated value at any frequency ω_1 can be determined by multiplying $F(\omega)$ by a frequency-reversed version of the gating frequency function centered at that frequency ω_1 and integrating (summing since it is discrete) the product, this being the definition of convolution

$$F_g(\omega_1) = \sum_{n=-N}^N F(n \cdot \Delta\omega) \cdot G(\omega_1 - n \cdot \Delta\omega) \quad (4.31)$$

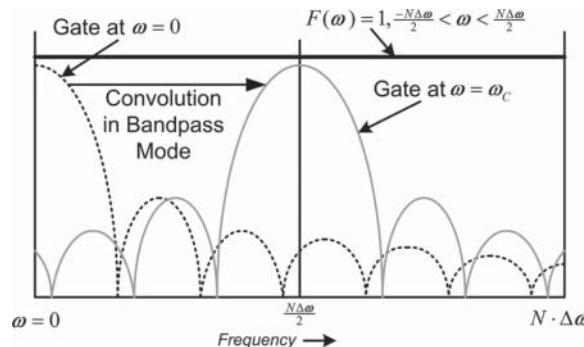


Figure 4.7 Convolution of the frequency gate response.

For the center frequency point in band-pass mode (or zero frequency, DC, in low-pass mode), the gated frequency response is a sinc function with the response centered at $\omega = \omega_C$, multiplied by the frequency response (which is just 1), and appears as the light gray trace in the figure. The integration of the light gray trace is the gated response at that frequency. For the case of the first (or last) frequency point, the sinc function is centered on the first (or last) data point, and half the gate function is multiplied by zero (for frequencies outside of the original response) and thus does not contribute to the sum, as shown by the dashed trace in the figure. So, the first data point will be one-half the value of the center point, or 6 dB lower. This can be seen graphically in Figure 4.7 as the area under the dashed line is one-half the area under the light gray line. Unfortunately, this creates the result that any gating will distort the first and last few points (last points only in low-pass mode) of the gated frequency response. This span of the distorted data is related to span (in frequency) of the gate; for wider time gates, the distortion appears over fewer of the points at the edges of the gated frequency response.

The VNA time domain compensates for this roll-off through a post-gate renormalization. The post-gate renormalization is determined by creating a frequency response that is unit magnitude. A pre-gate window is applied to this unit response that is the same as the pre-transform window applied to the normal frequency response data. This unity-magnitude frequency response is convolved with the gate frequency response to generate the final normalizing frequency response. The time-gated frequency response is divided by this function to remove the roll-off effects of the time gating. This normalizing function works perfectly for a unit time response at the center of the gate. If the gate is not symmetric around the time function, there will be some errors in the gated response when compared to the original frequency response.

It is instructional to view the actual gate shape in the time domain, which can be done using a function not normally available in commercial VNAs. The gate shape may be generated by creating a delta-like frequency response ($F(0) = 1$; $F(\omega) = 0$ for $\omega \neq 0$, but with some spreading due to windowing), applying gating, and transforming the result to the time domain to see the actual gate shape. This is useful in understanding how the gate shape affects the gated response.

Figure 4.8 (upper) shows the gating function for various gate center times, and Figure 4.8 (middle) shows a unit frequency response ($F(\omega) = 1$) in the time domain, with two of the gates applied the extremes of the gate offset times. Note that the peak of the time-domain response is nearly unchanged as the time gate fully encompasses the impulse at all three center times, but there is some difference in the side lobes for the shifted gates. Figure 4.8 (lower) shows the frequency response after gating. Here there is a substantial difference in the response at high frequency for the different gate center times. It is clear that normalization is optimal when the gate is centered on the response being gated.

The gated time response may be viewed in the time domain by taking the IFT and displaying the result. In fact, it is almost always required to first view the time-domain response to assign proper gating start and stop values: the transform function is turned on, the resultant time-domain response is displayed, and the gate start and stop are set. Next gating is turned on. Finally, the transform is turned off, and the time gated frequency response is shown in the frequency display.

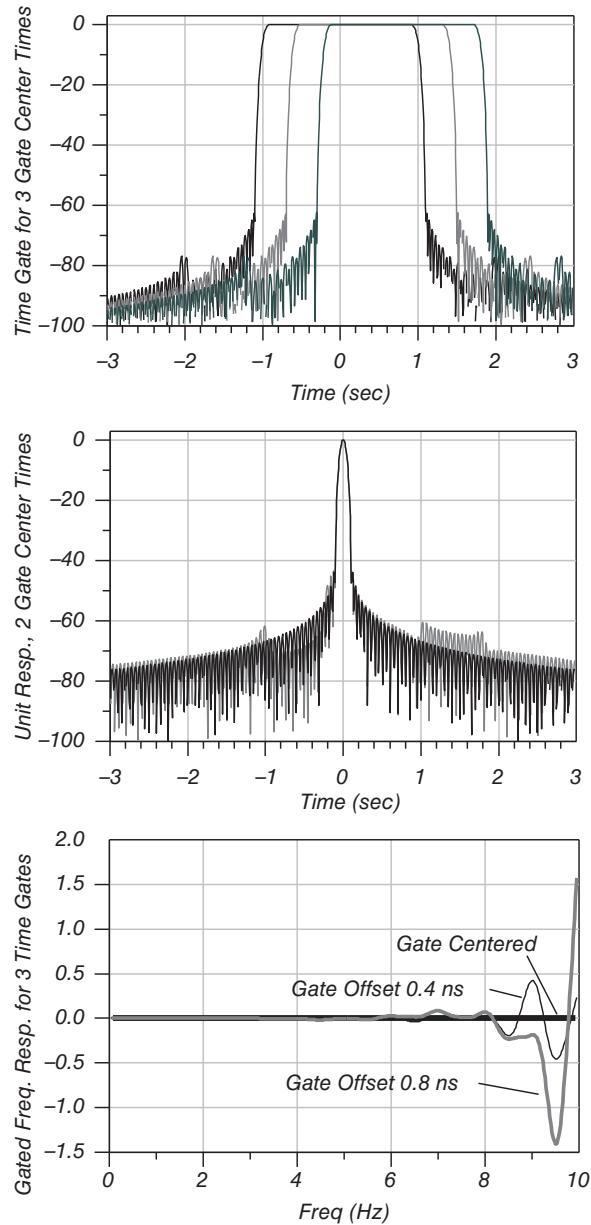


Figure 4.8 Time gates at three center times (upper); time response of gated unit response, with the first and last gate center (middle); gated frequency response of three gate centers, showing normalization error at the band edge (only positive frequency shown) (lower).

A study of the time-domain response of several examples of composite responses of several component elements will show how time gating can be used to separate the responses in time and display the individual frequency responses of the component elements, but with some distortion due to masking effects (Lu and Brazil 1993). From this, a method is developed that compensates for these effects.

4.7 Examples of Time-Domain Transforms of Various Networks

4.7.1 Time-Domain Response of Changes in Line Impedance

For ladder networks, that is, networks that consist of series-connected elements, the time-domain transform provides good insight into the nature of the discontinuities by which the frequency response is generated. As a first example, consider the network in Figure 4.9 (upper), consisting of a short length of Z_0 line followed by a $Z_0/2$ impedance line segment terminated in another Z_0 line. Note that there will also be re-reflections if the time scale is extended. There are two main reflections from the impedance steps at the beginning and end of the $Z_0/2$ line segment. The impedance value of a discontinuity caused by a step in impedance of a transmission line can be directly related to the time-domain step

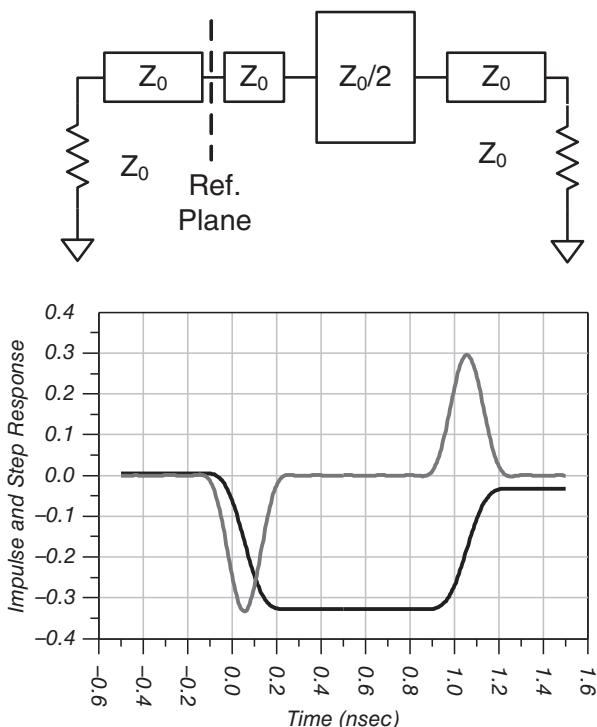


Figure 4.9 Model of concatenated lines of different impedances (upper), step (black), and impulse (gray) response of the lines in time (lower).

response, which shows reflections as a function of time. The reflections are relative reflection coefficient, so for a $50\ \Omega$ reference impedance, a 1% reflection relates to approximately $1\ \Omega$ change in impedance, as

$$\Gamma = \frac{Z - Z_0}{Z + Z_0}, \text{ and for } Z \cong 50, \Gamma(\%) \cong \Delta Z, \text{ where } \Delta Z = Z - Z_0 \quad (4.32)$$

Care must be used in this interpretation, as other factors such as loss in the transmission line, changes in line impedance, and previous reflections can affect the apparent reflection being investigated. For the lines in Figure 4.9 (upper), the step in impedance is apparent for the reflection being investigated, and the step in impedance is quite large, but the reflection coefficient of each step is the same, $|\Gamma_1| = |\Gamma_2| = 0.33$. However, the second apparent reflection coefficient, computed by the difference between the level of the $Z_0/2$ line and the final Z_0 line after the second transition, $\hat{\Gamma}_2$, is only 0.30, as shown in Figure 4.9 (lower). Also, the impulse response shows a similar “masking” effect in the second reflection response. This masking effect is a consequence of the response of the network and illustrates why the simple concept that a time-domain traces shows the impedance along a transmission line is not completely valid. In fact, the time-domain response shows the reflected voltage along the line, and the offset associated with masking can be determined from first principals, as discussed in the next section. Several more measurement examples using time domain and time gating are discussed in Chapter 5 on measurement of passive devices.

4.7.2 Time-Domain Response of Discrete Discontinuities

As a second example, concatenated transmission lines with discrete discontinuities between sections are evaluated with a time-domain transform of the frequency response, and the values of the various discontinuities are individually determined. Figure 4.10 (upper) shows a schematic of a Z_0 reference, followed by a first capacitive discontinuity, and followed by a Z_0 line and then a second identical capacitive discontinuity terminated in a Z_0 load.

The time-domain response of this network is shown in Figure 4.10 (lower). This is the low-pass step response, which shows capacitive discontinuities as negative dips in the time domain. The reflections of the discontinuities repeat at the spacing of the discontinuities, and these repetitive reflections should ideally continue on, at diminishing levels, for infinite time (though actually they get added to all the aliased responses). Also, note that even though the responses are caused by identical discontinuities, the response of the second discontinuity appears smaller than the first. The second response is somewhat masked by the first, though by a different amount than in the example of Figure 4.9, indicating a different masking mechanism.

4.7.3 Time-Domain Responses of Various Circuits

The time-domain responses of many elements are well known; the low pass mode in particular can be used to identify the type and relative value of these discontinuities. This is a particularly useful aspect in evaluating unwanted discontinuities in connectors, cables, and transmission line structures. A catalog of useful responses is shown in Table 4.1.

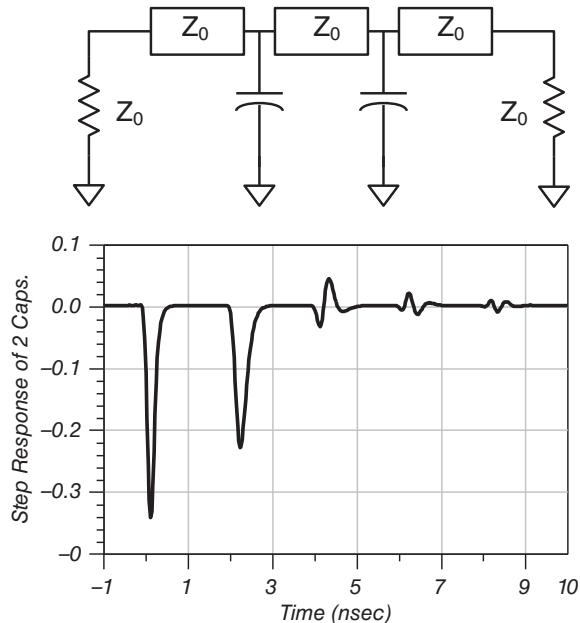


Figure 4.10 Model of 2 capacitive discontinuities (upper), step response of the S_{11} of two discontinuities (lower).

Table 4.1 Catalog of time-domain responses

Element	Step Response	Impulse Response
Open	Unity Reflection	Unity Reflection
Short	Unity Reflection, -180°	Unity Reflection, -180°
Resistor $R > Z_0$	Positive Level Shift	Positive Peak
Resistor $R < Z_0$	Negative Level Shift	Negative Peak
Inductor	Positive Peak	Positive Then Negative Peaks
Capacitor	Negative Peak	Negative Then Positive Peaks

4.8 The Effects of Masking and Gating on Measurement Accuracy

The concept of time gating mentioned refers to mathematically removing a portion of the time-domain response and viewing the result in the frequency domain. The intent is to remove the effects of unwanted reflections, say from connectors and transitions, leaving the desired response of the device being measured. This should improve the quality of the response; that is, the gated response should more closely resemble the device response as if it were measured with no other reflections. However, the effects of previous reflections can have an effect on the time-gated measurement. Previous work has reported on the compensating for the effect of loss (Bilik and Bezek 1998) but ignored the effect of previous reflections. Others have proposed an error associated with previous reflections but have not provided for compensation methods, or for errors associated with change in impedance. These effects are mathematically described next, along with new compensation methods, and with an uncertainty analysis on the time-gated frequency response, applied to several particular examples.

4.8.1 Compensation for Changes in Line Impedance

For the lines in Figure 4.9, the apparent reflection coefficient of the second transition, Γ_2 , is only about 90% of the actual value. The normal time-domain transform does not present the true reflection coefficient at that particular point. To understand this, consider that at the interface of the first reflection, the reflection coefficient is calculated as defined in Eq. (4.32). However, the signal that continues down the transmission line structure is changed by the transmission coefficient defined as (Pozar 1990)

$$T_1 = \frac{2 \cdot Z_1}{Z_1 + Z_0} \quad (4.33)$$

where Z_0 is the input line, and Z_1 is the second section of line. The reflection apparent at the input due to the second step in impedance, Γ_2 , is further changed by a second (reverse) transmission coefficient, T_2 as defined by

$$T_2 = \frac{2 \cdot Z_0}{Z_1 + Z_0} \quad (4.34)$$

The total apparent reflection, $\hat{\Gamma}_2$, due to the second step is now computed as

$$\hat{\Gamma}_2 = \Gamma_2 \cdot T_1 \cdot T_2 = \frac{(Z_0 - Z_1) \cdot (4Z_1 Z_0)}{(Z_1 + Z_0)^3} \quad (4.35)$$

or, for the example, $\hat{\Gamma}_2 = +0.30$, which precisely matches the measured value in Figure 4.9.

Further, for the case of a response following a change in line impedance, where the first line impedance is not the reference impedance, two compensations are required. First the reflection response must be compensated by dividing the apparent response by the transmission coefficient term product, $T_1 \cdot T_2$, to produce a reflection response relative to the line, $S'_{11} = S_{11}/(T_1 \cdot T_2)$ (derived from $\hat{\Gamma}_2$ as shown earlier). The second compensation is renormalizing the response by the impedance of the line just before the desired response. The frequency response assumes a reference impedance of the system impedance, typically 50Ω . The renormalization consists of converting the reflection response to an effective impedance Z_{eff} , using the line impedance Z_{line} just before the desired response (S'_{11}), as the reference

impedance. This is then reconverted from the resulting effective impedance to the effective reflection response ($S_{11(\text{eff})}$) using the system impedance (Dunsmore 2004).

$$Z_{\text{eff}} = Z_{\text{line}} \cdot \frac{1 + S'_{11}}{1 - S'_{11}}, \quad S_{11(\text{eff})} = \frac{Z_{\text{eff}} - Z_0}{Z_{\text{eff}} + Z_0} \quad (4.36)$$

4.8.2 Compensation for Discrete Discontinuities

Figure 4.10 shows the time-domain response of two capacitive discontinuities. The second discontinuity, which is caused by an identical element in the circuit, has a different time-domain response from the first element. The most noticeable aspect is the magnitude of the response is smaller, which is consistent with the first example. However, in this case, there is no change in reference impedance to account for the difference. Instead, the first reflection removes some of the energy from the forward (incident) wave, such that there is less energy available at the second discontinuity. A similar effect occurred in the first example and was accounted for by the transmission coefficients. For a localized discontinuity, with the same impedance on each side, the effect on the transmitted wave must be determined in a different manner.

From power conservation, the magnitude of the voltage wave incident on the second reflection, $|V_2^+|$ (assuming the first reflection is loss-less), is

$$|V_2^+| = |V_1^+| \cdot \sqrt{1 - |\Gamma_1|^2} \quad (4.37)$$

where V_1^+ is the incident voltage wave and Γ_1 is the first reflection. The magnitude of the reflected voltage, $|V_2^-|$, from the second reflection is

$$|V_2^-| = |V_2^+| \cdot \Gamma_2 = |V_1^+| \cdot \left(\sqrt{1 - |\Gamma_1|^2} \right) \cdot \Gamma_2 \quad (4.38)$$

The signal V_2^- reflects again off Γ_1 with a portion transmitted, V_3^- (the portion of the signal from Γ_2 that is actually measured at the input port), which is reduced in the same manner as Eq. (4.37) to yield the effective value of the second reflection as

$$|\hat{\Gamma}_2| = \frac{|V_3^-|}{|V_1^+|} = (1 - |\Gamma_1|^2) \cdot \Gamma_2 \quad (4.39)$$

This result applies only to the magnitude of the reflection as the power conservation argument does not apply the phase of the transmitted signal, and while consistent with the results described in other publications (Lu and Brazil 1993), it goes further to provide a means to remove the effects of the first discontinuity.

4.8.3 Time-Domain Gating

4.8.3.1 Gating the First of Two Discontinuities

The effectiveness of gating can be evaluated using the circuit from Figure 4.10. Figure 4.11 shows the original frequency response in light gray, with the characteristic ripple pattern

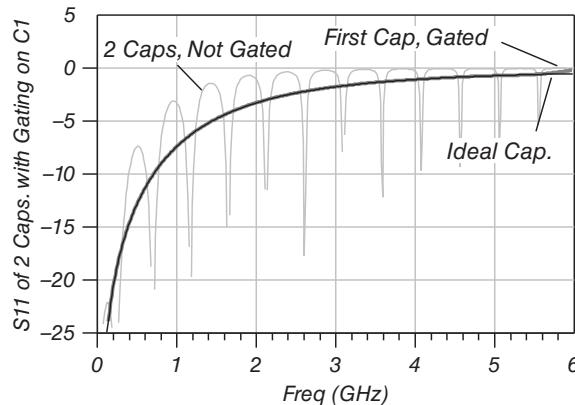


Figure 4.11 S_{11} response of two capacitive discontinuities (light gray) and gating around the first cap (dark gray). Also shown is the S_{11} of just the first cap (black).

found from two discontinuities separated by a length of line. The thin black trace is the result of computing the ideal S_{11} of just a single capacitive discontinuity, terminated in Z_0 . Gating around the first capacitive discontinuity yields a response (Figure 10.3, dark gray) nearly identical to the frequency response calculated for only the first discontinuity. The difference is seen only at the high frequency of the response, most likely due to the errors in the renormalization, as described in Figure 4.8 (lower). Clearly, gating about a first discontinuity terminated in Z_0 is very effective in removing effects of other elements. However, if the gate is applied to the second discontinuity, the response is not similar.

4.8.3.2 Gating the Second of Two Discontinuities

The time-gated response of the second discontinuity is quite different from the underlying response, as shown in Figure 4.12 (thin dark trace, labeled “2nd Cap, Gated, No Comp.”). The frequency response of the gated measurement of the second discontinuity may be compensated by taking the gated response of the first discontinuity, and using Eq. (4.39) creating a compensation for Γ_2 as

$$\Gamma_2 = \frac{|\hat{\Gamma}_2|}{(1 - |\hat{\Gamma}_1|^2)} \quad (4.40)$$

where $\hat{\Gamma}_1$ is the gated result of the first capacitor, and $\hat{\Gamma}_2$ is the gated result of the second capacitor. This compensation has been applied in Figure 4.12, with the result showing remarkably good compensation over most of the frequency range (thick black trace, labeled “2nd Cap, Gated & Comp.”). However, the band edge response deviates because the normalization does not completely compensate the error due to the gate, as described in Figure 4.8. This error was apparent as a slight upturn of the S_{11} trace of the first capacitor, after gating. The effect of this slight upturn on $\hat{\Gamma}_1$ moves it closer to 1, and the denominator of Eq. (4.40) moves it closer to zero, causing the apparent value of Γ_2 to increase more rapidly at the band edge. Increasing the gate window function will reduce the edge effect, or alternatively, the frequency response can be performed over a wider span, and the upper 10% can be disregarded as a gating effect.

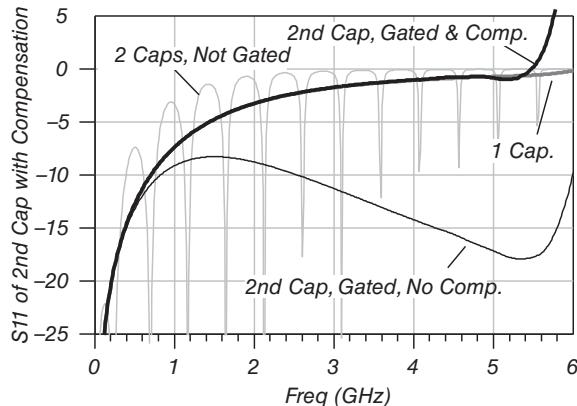


Figure 4.12 S_{11} of two capacitive discontinuities (light gray) not gated, S_{11} response of one capacitive discontinuity (dark gray), S_{11} of second discontinuity gated (thin black) with no compensation, and S_{11} of second discontinuity, gated and compensated (thick black).

Also shown in Figure 4.12 is the ungated S_{11} measurement of two discontinuities (light gray), and, for reference, the S_{11} of just a single discontinuity (dark gray, labeled “1 Cap.”). This shows the effectiveness of the compensation method, where there is substantial deviation from the ideal in the uncompensated gated response of the second capacitive discontinuity.

Also note that this is a large first discontinuity, having a frequency response return loss value of nearly 0 dB over much of the frequency range. A more normal situation is where the first discontinuity, often a connector interface from coax to PC board, has a much smaller reflection, and so the compensation method works even better.

Most VNAs do not have multiple section compensation, so it is up to the user to create the compensation function. However, many VNAs have equation editor functions that can allow real-time computation of these compensations.

4.8.3.3 Compensation for a Combination of Discontinuities and Line Impedance Changes

Many practical applications of time-gated measurements include effects of both discrete discontinuities and impedance steps. For example, the measurement of a connector at the far end of a cable is affected by the cable’s near-end connector and the impedance of the cable. Figure 4.13 (upper) shows a circuit diagram and (lower) time-domain response of a $55\ \Omega$ cable with a 20 dB return loss (at 1 GHz) input and output mismatch due to capacitive loading, terminated in $50\ \Omega$. In this example, the discrete reflections are much smaller than those shown in the previous example and more in line with a typical application. Here, the cable itself has a 5% reflection due to impedance error, and each discontinuity is approximately 10% at 1 GHz.

Figure 4.14 (upper) and (lower) shows the time-gated frequency response of the first and second discontinuity as a result of gating about each respectively. In the case of gating the first discontinuity, there is only an impedance step to be accounted for.

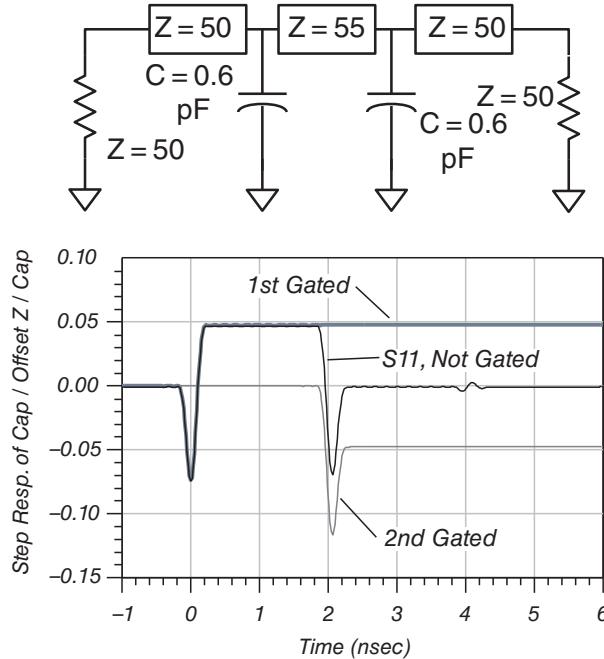


Figure 4.13 Circuit with 2 capacitive discontinuities, and an offset impedance line (upper); time-domain response of the circuit (black) with first cap after gating (wide gray) and gating around only the second cap (thin gray) (lower).

A compensation of the first case can be performed by recognizing in the step response that the value of the step after the gate represents an offset impedance in the termination. The effective reflection results in

$$\hat{\Gamma}_1 = \Gamma_1 + \cdot \frac{\Gamma_2}{(1 - \Gamma_1 \cdot \Gamma_2)} \quad (4.41)$$

which can be reduced to

$$\Gamma_1 = \hat{\Gamma}_1 - \Gamma_2 - \frac{\Gamma_1(\Gamma_2)^2}{(1 - \Gamma_1 \cdot \Gamma_2)} \approx \hat{\Gamma}_1 - \Gamma_2 \quad (4.42)$$

Here, Γ_2 is the reflection coefficient of the 55Ω cable that forms the termination after gating, and there is assumed no delay before Γ_1 . Figure 4.14 (upper) shows the first discontinuity, after gating, and gating plus compensation. If there is delay before the first discontinuity, the phase of Γ_2 must be shifted to account for the delay.

The masking of the second discontinuity is more difficult to account for. There are three effects: first, masking due to signal loss in the first discontinuity as described by Eq. (4.39); second, by the change in reference impedance as described by Eq. (4.35); and finally, by the change in terminating impedance, relative to the modified reference impedance, as described by Eq. (4.42). Thus, three compensations are required. First, the effective reflection after the first discontinuity is found by applying the first compensating equation, then the effect

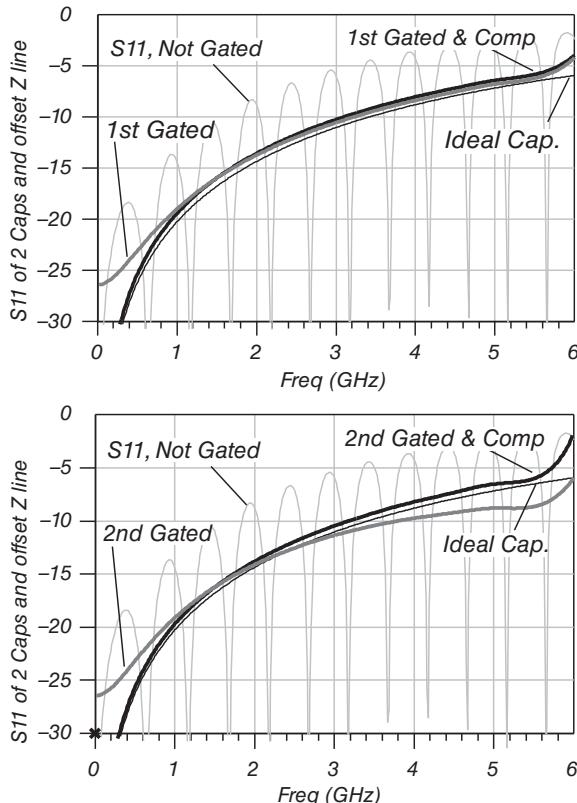


Figure 4.14 Time gated response of the first discontinuity (thick gray, “1st Gated”) (upper), and with compensation (thick black, “1st Gated & Comp”), along with original S_{11} (light gray) and ideal single capacitive discontinuity (thin black); similar to the upper trace but with gating and compensation applied to the second discontinuity (lower).

of the impedance transformation is compensated by applying Eq. (4.35) to the result of the first equation, which in this case is a small effect, and finally, compensating for the step in impedance at the termination. Also, as this compensation includes an additive element associated with the terminating impedance, phase becomes important. Both Γ_2 and Γ_3 are phase shifted by the delay of line Z_1 , which may be determined directly from the time-domain response. Thus, the effective impedance $\hat{\Gamma}_2$ is

$$\hat{\Gamma}_2 = \frac{\Gamma_2 \cdot (4Z_1Z_0)}{(Z_1 + Z_0)^2} \cdot (1 - |\Gamma_1|^2) \cdot e^{j\omega \cdot 2\tau_d(Z_1)} + \Gamma_3 \cdot e^{j\omega \cdot 2\tau_d(Z_1)} \quad (4.43)$$

where $\hat{\Gamma}_2$ is the gated response of the second discontinuity, Γ_2 is the reflection of only the second discontinuity, Γ_1 is the (gated) first discontinuity, Z_1 is the 55Ω line, Γ_3 is the reflection

from Z_1 to the terminating impedance, Z_0 , and $\tau_d(Z_1)$ is the delay corresponding to the length of the line Z_1 . The compensation is determined by solving Eq. (4.43) for Γ_2 .

Figure 4.14 (lower) shows the results of gating the second discontinuity and applying the compensations as described in Eqs. (4.42) and (4.43). Also shown is the frequency response of a single (ideal) discontinuity associated with the first capacitance as though it were on a matched line, and the original response of the two discontinuities with the offset impedance line in between. In Figure 4.14, for both the (upper) and (lower) plots, there is remarkable improvement in the gated measurement, especially at low frequencies, when evaluated against the ideal response. In this example, the compensated result is quite sensitive to the delay selected for the 55Ω line. This delay was determined by choosing the delay displayed at the peak of the second discontinuity. These compensations are appropriate for single discontinuities that are nearly loss-less and non-distributed, as might be found from a coax to PC board adapter.

4.8.4 Estimating an Uncertainty Due to Masking

The proposed compensations described previously may generate some error, in part due to lack of consideration of loss in the network and due to inability to totally separate responses. In some cases, it may not be necessary to actually compensate the network for gated response but rather to establish an estimate of uncertainty associated with the response that is gated out. This uncertainty can be derived from Eq. (4.39), for a second gated response on matched line. Additional uncertainty will come from non-matched lines leading up to the reflection of interest and following the reflection of interest. The magnitude of the uncertainties can be determined in a manner similar to Eqs. (4.35) and (4.43) with the resultant total uncertainty after gating determined as

$$\Delta\Gamma_2 = \frac{\Gamma_{2G}}{1 - |\Gamma_1|^2} \left(|\Gamma_{Z1Z0}|^2 + |\Gamma_1|^2 \left| \frac{(4Z_1Z_0)}{(Z_1 + Z_0)^2} \right| \right) + \Gamma_{Z2Z1} \quad (4.44)$$

where Γ_1 is the first (gated out) discontinuity; Z_1 and Z_2 are the lines before and after the desired reflection Γ_2 , respectively; Z_0 is the system impedance; Γ_{Z1Z0} and Γ_{Z2Z1} are the reflection between lines Z_1 and Z_0 and Z_2 and Z_1 , respectively; and Γ_{2G} is the value of the gated response.

4.9 Time-Domain Transmission Using VNA

Time-domain reflectometry is well known for evaluating components, but the time-domain transmission response is commonly used to evaluate the signaling quality of a digital signal at the end of digital transmission line, such as a USB cable, an Ethernet cable, or even a computer PC board back plane. In these instances, the quality of the digital path is not expressed in S-parameters or in impedances, but rather it is expressed as a time-domain interpretation of what an ideal signal would look like on an oscilloscope input, known as an *eye diagram*.

Recently, there have been software packages that can take S-parameter data and provide the equivalent of an oscilloscope measurement of the output signal, as though an ideal input signal has been applied, and present the result in an eye diagram. In fact, these software packages can process either time-domain data from an oscilloscope or S-parameter data from a VNA. In many cases, the signals are fully differential, such as low-voltage differential signal (LVDS). Keysight's Physical Layer Test Solution is one example of such a program that can utilize previously measured data to generate these equivalent eye patterns.

Eye diagrams represent the response from a repeated pattern of a digital pattern, which in the past was applied to the high-speed digital interconnect, the output of which was measured on an oscilloscope. While this works well at moderate speeds, at the Gbit speeds common now, effects such as connectors, interface cables, and minute skew in the test leads make Gbit speed testing of interconnects difficult or impossible, without a means of compensation. This has led to using VNAs to measure the S-parameters of the interconnect, using the best calibration and de-embedding methods and then combining the S-parameter results with a numerical representation of the bit pattern input to the interconnect network to generate a result that produces the same picture as a classic eye diagram.

Figure 4.15 shows an example of such an eye diagram created by measuring a high-speed interconnect cable in the frequency domain, transforming it into a time-domain transmission response, then creating a bit pattern in the time domain (including attributes such as rise time, jitter, and source noise), and finally convolving it with the time-domain transmission impulse response (Keysight Technologies 2019).

The key attributes labeled in the figure are described in Table 4.2.

For eye height, an ideal eye opening would be measured from the one level to the zero level. However, noise on the eye will cause the eye to close. The eye height measurement determines eye closure due to noise. Normally it is statistically determined from the variation of the signal in the eye window through the formula

$$EyeHeight = (OneLevel - 3\sigma) - (ZeroLevel + 3\sigma) \quad (4.45)$$

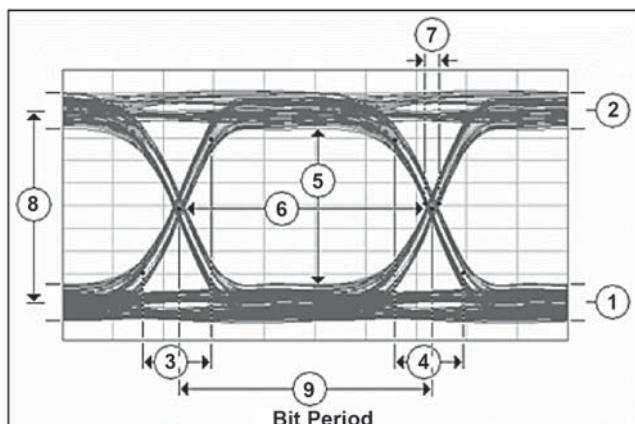


Figure 4.15 Eye diagram example with key attributes identified.

Table 4.2 Attributes of an eye diagram

Attribute Number (Figure 4.15)	Attribute Name	Attribute Description
1	Zero level	Zero level is a measure of the mean value of the logical 0 of an eye diagram.
2	One level	One level is a measure of the mean value of the logical 1 of an eye diagram.
3	Rise time	Rise time is a measure of the transition time of the data from the 10% level to the 90% level on the upward slope of an eye diagram.
4	Fall time	Fall time is a measure of the transition time of the data from the 90% level to the 10% level on the downward slope of an eye diagram.
5	Eye height	Eye height is a measure of the vertical opening of an eye diagram.
6	Eye width	Eye width is a measure of the horizontal opening of an eye diagram.
7	Deterministic jitter	Deterministic jitter is the deviation of a transition from its ideal time caused by reflections relative to other transitions.
8	Eye amplitude	Eye amplitude is the difference between the logic 1 level and the logic 0 level histogram mean values of an eye diagram.
9	Bit rate	Bit rate (data rate) is the inverse of bit period (1/bit period). The bit period is a measure of the horizontal opening of an eye diagram at the crossing points of the eye.

Eye width, ideally, would be measured between the crossing points of the eye. However, jitter may appear on the waveform and influence the eye opening. The exact determination in the presence of jitter or noise depends upon some preferences in the measurement setup.

Using a software package is convenient for research purposes where one can carefully characterize the high-speed interfaces in the lab, save the S-parameter data, and then use the offline program to do detailed analysis. But now eye diagram measurements are becoming important manufacturing characteristics that often have pass/fail criteria associated with them. This has led to a new set of applications available on VNAs: enhanced TDR. The enhanced TDR functions provide a real-time measure of the eye diagram while allowing a wide range of definitions and adjustments to the driving signals. Remember, in this case the driving signals are simulated bit patterns applied to the measured S-parameter response of the part under test. Typical parts tested in this way are high-speed USB connectors (such as USB-C), Firewire, HDMI cables, and some computer back planes and interconnect harnesses. Many of these devices have multiple differential lines and are tested for attributes such as near-end cross talk, far-end cross talk, and eye diagram. Multiport VNAs, such as shown in Figure 2.34, are often

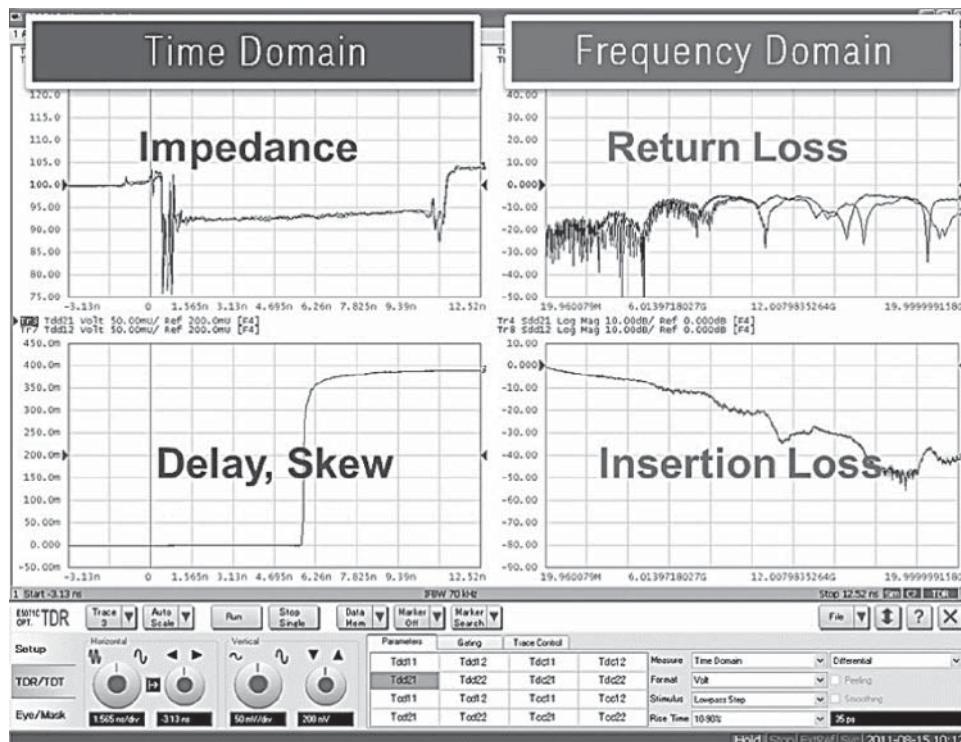


Figure 4.16 Enhanced TDR application on a VNA.

used in manufacturing sites rather than switching test sets for the high-speed measurements they enable.

Figure 4.16 shows the display of the enhanced TDR on a Keysight ENA. This provides many advanced features beyond basic time domain. One of the most convenient is the display of impedance directly on the y-axis. Chapter 5 discusses converting a time-domain response to impedance through the use of the equation editor, but the enhanced TDR does it as a matter of default display. Also shown is the time-domain transmission. These can be defined for single-ended or differential transmission lines.

The enhanced TDR mode also provides a direct eye diagram readout, in addition to the TDR traces and the frequency response traces, as shown in Figure 4.17.

The interface is easy to understand. The attributes of the source such as zero level, one level, data rate (1/bit period), and rise time are direct stimulus attributes. The jitter measurement shown on the right is what is known as *deterministic jitter*. This jitter (or widening of the eye crossing) is caused by multiple reflections in the DUT, which cause the digital signal to not overlap itself perfectly. If it gets bad enough, it will close the eye. Enhanced TDR applications are now available in VNAs with frequencies beyond 120 GHz.

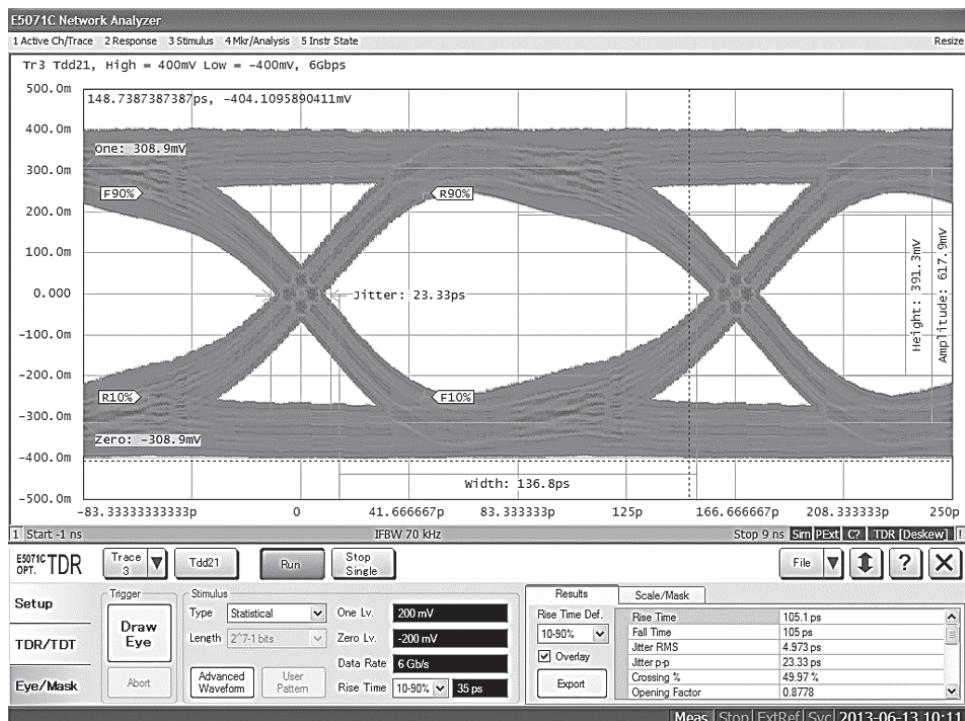


Figure 4.17 Eye-diagram on a VNA-based time-domain transmission test.

4.10 Conclusions

The exact relationship of the time-domain transform used in VNAs to the analytic impulse response of measured networks has been described in detail with mathematical rigor, and details of the time gating function have been presented. From this background, the effects of time gating on measured results have been explored, and a method was given for compensating for undesired masking effects, yielding superior time-gated measurement results. Additionally, the uncertainty of time-gated measurements has been quantified, and qualitative errors due to the subtle effects of gate properties and renormalizations have been presented. More examples of the use of time domain and gating will be shown in subsequent chapters, as they are applied to specific component measurements.

References

- Bilik, V. and Bezek, J. (1998). Improved cable correction method in antenna installation measurements. *Electronic Letters* **34** (17): 1637.
 Blinchikoff, H.J. and Zverev, A.I. (1976). *Filtering in the Time and Frequency Domains*. New York: Wiley.

- Bracewell, R.N. (1986). *The Fourier Transform and Its Applications* (Revised), 2e. New York: McGraw-Hill.
- Dunsmore, J. (1999). Advanced Filter Tuning using Time-domain Transforms. In: *Proceedings of the 29th European Microwave Conference*, 5-7 Oct., vol. 2, 72–75. Munich: IEEE.
- Dunsmore, Joel Phillip., The time-domain response of coupled-resonator filters with applications to tuning, PhD Thesis, University of Leeds, 2004.
- Glass, C. (1976). *Linear Systems with Applications and Discrete Analysis*. St. Paul: West Publishing Co.
- Hines, M. and Stinehelfer, H. (1974). Time-domain oscillographic microwave network analysis using frequency-domain data. *IEEE Transactions on Microwave Theory and Techniques* **MTT-22** (3): 276–282.
- Lu, K., Brazil, T., “A Systematic Error Analysis of HP8510 Time-Domain Gating Techniques with Experimental Verification”, 1993 *IEEE MTT-S Digest*, pg 1259.
- Keysight Technologies. (2019) Physical Layer Test System [Online] Available at: <http://na.support.keysight.com/plts/help/WebHelp/PLTS.htm>
- Pozar, D.M. (1990). *Microwave Engineering*. Reading, MA: Addison-Wesley. Print.
- Rytting, D., “Let Time-domain Provide Additional Insight into Network Behaviour”, Hewlett-Packard RF & Microwave Measurement Symposium and Exhibition, April 1984.
- Sharrit, D, “Vector network analyzer with integral processor”, US Patent No. 4,703,433, 1987.

5

Measuring Linear Passive Devices

Linear devices are at one time both some of the simplest of microwave components and some of the most difficult to characterize as the quality of measurements must be very good, especially for low loss devices. Filters, transmission lines, couplers, and isolators, to name a few, all are designed to be as nearly ideal as possible, and any error in the vector network analyzer (VNA) measurements can lead to production test failures and loss of yield. In this chapter, some particular details of the real-world problems and solutions to measuring these devices are presented.

5.1 Transmission Lines, Cables, and Connectors

Cable and connectors are perhaps the simplest of radio frequency (RF) components, usually having low loss and good match. However, making measurements on devices with such ideal performance requires careful calibration and measurement techniques so that the error in measurement doesn't overwhelm the devices' performance.

For short, low-loss lines, the main difficulties are dealing with the mismatch of the input and output connection and properly calibrating the source and load mismatch of the VNA. These devices might be interconnect cables with integrated connectors or might represent transmission line traces on PC boards. If the device includes connectors of standard type, then any mismatch associated with the connector is appropriately assumed into the measurement. On the other hand, if a device represents some transmission line structure on a PC board and the connectors are used only as a fixture to allow connection to the VNA, then the effects of the connectors should properly be removed from the overall measurement.

5.1.1 Calibration for Low Loss Devices with Connectors

If one reviews published specifications for most calibration kits, one often sees that thru-reflect-line (TRL) mechanical calibration provides the best performance, followed

by nearly identical performance from sliding load SOLT calibrations and Ecal¹ modules. However, in practice, Ecal almost always provides better calibrations because the quality of TRL and mechanical standards used in practice is not of the same quality as the metrology versions. However, Ecal modules are calibrated to a high quality, and they retain this high quality in any measurement application. Regardless of which type of calibration kit is used, the best calibration method is almost always using the short-open-load-reciprocal (SOLR) or “unknown thru” calibration, unless the device to be tested is an insertable device, that is, it has one male and one female connector of the same type.

For the system setup for low-loss devices, where even a small amount of trace noise can be significant, it is best to use a moderate intermediate frequency (IF) bandwidth and sweep-to-sweep averaging to reduce trace noise. While lowering the IF bandwidth to a very narrow value will reduce noise, it will also increase the sweep time, and very slow sweep times leads to instability in the calibration due to system drift that might occur over a long sweep time. One cause of system drift is a slowly varying temperature in a measurement lab, due to cycling of the air conditioning. Even if the lab is ostensibly held to $\pm 1^{\circ}\text{C}$, the instantaneous air temperate around test systems, and cables, can vary by several degrees over a short time. If the sweep rate is slow relative to this, there will be drift in the measurement during the sweep. If, however, sweep averaging is used on a faster sweep, the drift effect will average out over several sweeps. Before starting a measurement, evaluate the effect of trace noise. The simplest way to do this is to set up the frequency span and power and turn on about 10 sweep averages. After the sweep average completes, normalize the trace using the “data into memory” and “data/memory” functions, and turn the averaging off. One can use the trace statistics functions (often found under the analysis section of the VNA user interface) to determine the trace noise, which is usually expressed as a standard deviation or peak-to-peak. However, because it is a noise-like nature, the peak-to-peak on any particular sweep can sometimes have a single high noise point even if the average noise is low, so typically the mean-to-peak noise is specified as two or three times the standard deviation noise and the peak-to-peak is four to six times. Using averaging to store the memory trace eliminates double counting the noise as the deviations from the memory trace will add to the deviations of the data trace and gives a higher value for trace noise than would normally appear in the measurement. Alternatively, for slow sweeps, one can simply normalize the trace with no averaging, look at a new trace with data/memory, and recognize that the standard deviation noise will be approximately 1.41 times the actual noise due to the noise of the normalization memory trace. This should be done before any calibrations are performed.

One can reduce the trace noise by either decreasing the IF bandwidth, increasing the averaging, or increasing the power. A rule of thumb for improvement is the standard deviation trace noise will improve 3 times for every 10 \times changes in IF bandwidth or averaging and for every 10 dB increase in source power (Figure 5.1). One caution with changing power: if the power is changed after calibration, there will be an additional error if the power range or source attenuator value changes. This sometimes occurs if the power ranges are set to automatic.

¹ Ecal is a registered trademark for Keysight Technologies. Its use here should be understood to refer to both the Keysight implementation specifically, as well as the more general category of electronic calibration modules.

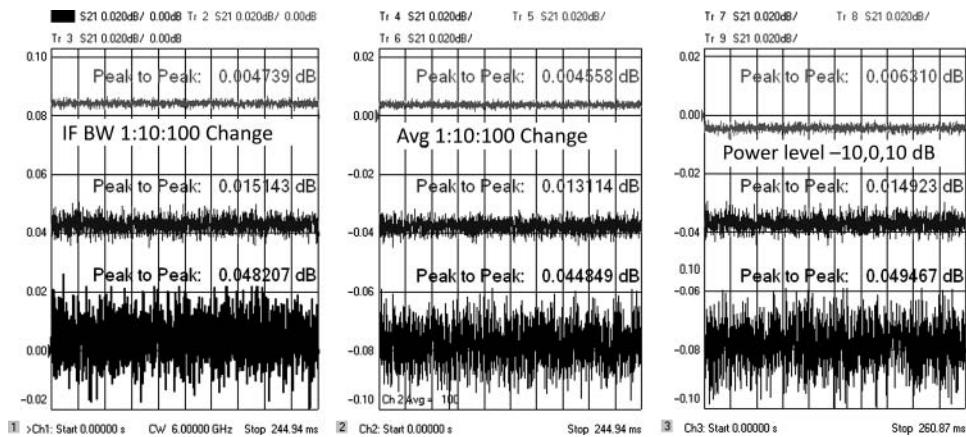


Figure 5.1 Measured trace noise with changes in IF BW, trace averaging, and changing power.

Calibration power for low-loss devices should be low enough to ensure the test receiver is not in any compression, on the order of 10 dB below the specified 0.1 dB compression point. On most VNAs, this is in the range of -10 to 0 dBm. At high enough power, the trace noise does not improve further due to high-level noise caused by the phase noise of the source (see Section 3.16.5).

For the very best measurements, it is helpful to improve the source and load match of the VNA. This can be done by adding a well-matched precision attenuator to the source and/or load ports. While this reduces the power (thus increasing the noise), it typically improves the raw source and load match by two times the attenuator value. The trade-off is that it will degrade the raw directivity by the same amount. If the test system is stable, it is a good trade-off; if the test system is not stable or the pad is at the end of a cable, it may degrade the S_{11} measurement even as it improves the transmission measurement. This will also improve the corrected load match. This is particularly important if one is unsure of the quality of the calibration kit. An attenuator of 6 dB on port 2 provides a good trade-off between increased noise and better match calibration.

Figure 5.2 shows two S_{21} measurements of a low-loss airline, first connected directly to the VNA (light trace) and then connected to a VNA with a 6 dB pad on port 1 and 10 dB pad on port 2 (dark trace). The improvement in S_{21} ripple is most likely due to the improved raw match. Also shown are two S_{11} traces under the same conditions. Note, since the S_{11} calibration-accuracy depends almost entirely on the quality of the 1-port standards and not on the quality of the raw test port, the improvement in port 2 raw load match has little effect on S_{11} .

5.1.2 Measuring Electrically Long Devices

Electrically long devices such as long cables or surface acoustic wave (SAW) filters (which have substantial delay) can cause some particular problems in characterization, especially with older VNAs. Many VNAs have both a swept frequency mode (where the source is

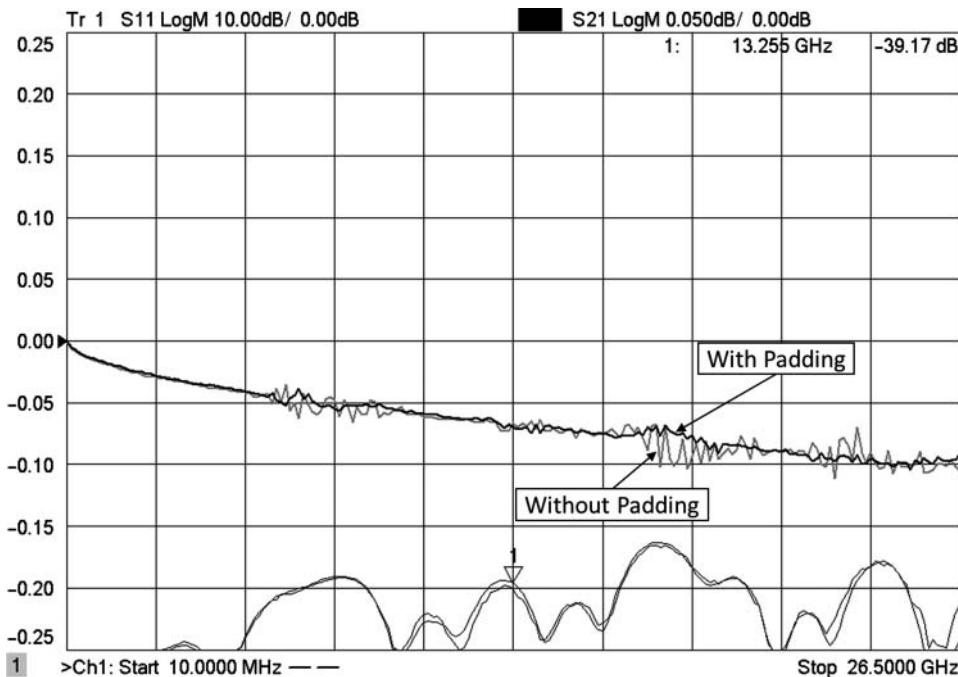


Figure 5.2 Measurement of an airline with normal calibration and with additional 6 dB pad on input and 10 dB on the output.

continuously swept across frequency band) and stepped frequency mode (where the source is stepped to discrete frequencies across a frequency band) (Keysight Technologies, Inc., 2002). The swept mode has the advantage of having no dwell time (or frequency settling time) at each point and therefore can be much faster in the fastest sweep modes. In one sense, the swept frequency mode provides a better characterization of the device-under-test (DUT) in that it does not skip any frequencies but rather ramps through all frequencies. If there is some particular frequency that has a dropout or resonance mode, it will be indicated by the swept-frequency measurement. However, in stepped mode, if the dropout occurs at a frequency that is not one of the stepped frequencies, there will be no indication on the trace.

However, for electrically long devices such as cables, even the swept mode can cause difficulties, due to an effect known as *IF delay*.

5.1.2.1 IF Delay

IF delay occurs when the source and receiver of a VNA are continuously swept (as opposed to stepped) and there is a long delay from the source to the receiver. Figure 5.3 illustrates the problem with measuring long cables. If the sweep generator is moving at a constant rate of change, $\Delta f/\Delta t$, the delay of the DUT will delay the measurement of the signal from the source while the receiver moves in frequency, causing an apparent shift of the signal in the IF of the receiver channel. This shift causes an apparent drop in signal, which results in more

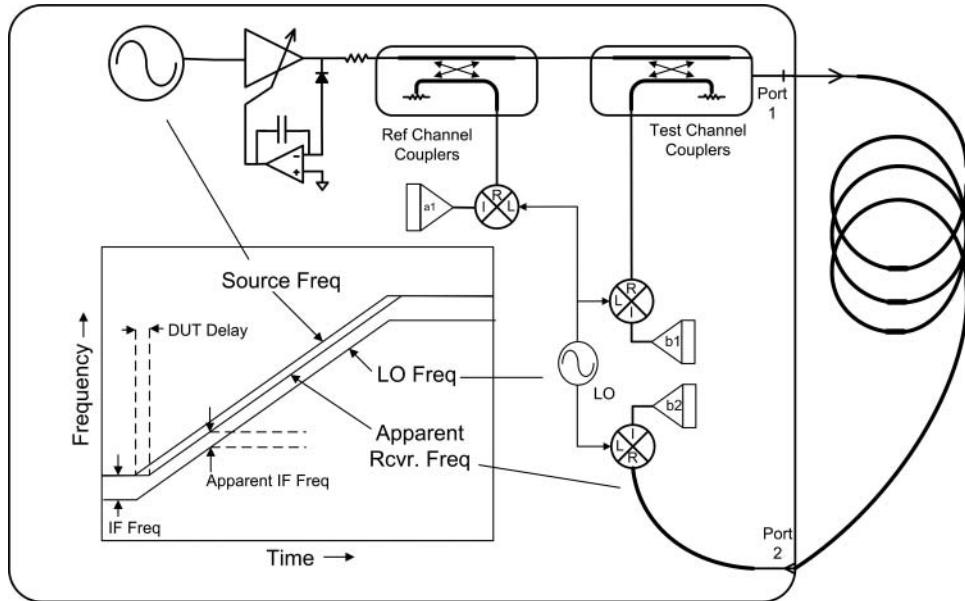


Figure 5.3 Illustration of IF delay for a long cable.

than expected loss. The amount of shift is computed as

$$IF_{Shift} = \left(\frac{\Delta f}{\Delta t} \right) \cdot DUT_delay \quad (5.1)$$

This IF delay effect occurs differently at different bands of the VNA as the sweep rate changes depending upon the source and phase lock architecture; for example, if the fundamental oscillator has a maximum rate of 300 GHz s^{-1} , then in the doubled band, it can have a rate of 600 GHz s^{-1} and show twice the IF delay effect. As an example, consider a system that sweeps at 37 GHz s^{-1} , which translates into about 270 ms sweep on a 10 GHz span. This corresponds approximately to a 3 kHz IF BW with a 801-point sweep. Suppose the group delay of the DUT is 10 ns (approximately the group delay of 3 m cable); then the IF shift would be

$$IF_{Shift} = \frac{10 \text{ GHz}}{.267 \text{ sec}} \cdot 10 \text{ nsec} = 0.37 \text{ kHz} \quad (5.2)$$

This corresponds to about 10% of the IF bandwidth, giving approximately 0.3 dB of error in the measurement. For a low-loss device, this is unacceptable.

The effect of IF delay is shown in Figure 5.4. Two measurements are made, one in stepped mode and one in swept mode. The calibration is done in the same mode as the measurement. The IF delay amplitude offset is somewhat compensated for by calibration, but for a DUT with a sufficiently long delay, the IF delay offset will become apparent.

In Figure 5.2, the measurement parameters are adjusted to get the faster possible sweep to illustrate the issue, in this case a sweep with only 51 points and with a 300 Hz IF BW. For such a low-point count, the sweep rate can be very high, and the narrow IF BW means

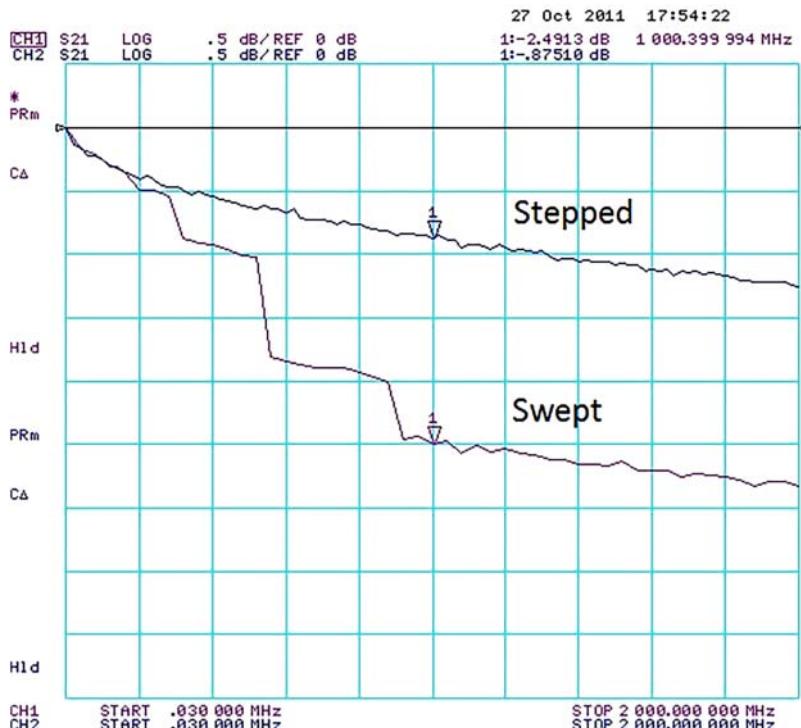


Figure 5.4 Comparing stepped with swept mode on a 3 m cable insertion loss measurement.

that any apparent frequency shift due to delay will cause a drop in amplitude. In the figure, the swept trace shows four discrete steps, each one corresponding to a different band break for the instrument, an HP 8753. This VNA used a sampler with a high harmonic order (low voltage-controlled-oscillator [VCO] frequency) and so it has many bands. If the number of points is changed to 201, a more normal value, or the IF BW is increased to a wider value, the difference between swept and stepped becomes much smaller.

IF delay effects also occur when the calibration is done in one sweep mode and the measurement is performed in another sweep mode. This can occur if the calibration is done in a different IF bandwidth from the measurement. Sometimes, the IF bandwidth is reduced to lower the trace noise of the measurement; or sometimes the calibration is performed at a lower IF bandwidth to reduce the noise effect of the calibration, and the IF bandwidth is widened to speed up the measurement trace. Either case can cause an IF delay shift, as demonstrated in Figure 5.5.

The IF delay becomes apparent in wideband measurement traces as a drop in amplitude (or shift in phase) at the VNA discrete band breaks. Because of some of the phase locking methods, the first point or first few points in a band may be stepped. The VNA determines the fastest method for sweeping, and this sometimes means using stepped points at the band edges. These stepped points become obvious as “spikes” in the trace, since they are stepped, and have no error relative to the calibration. They are apparent in the figure where the markers

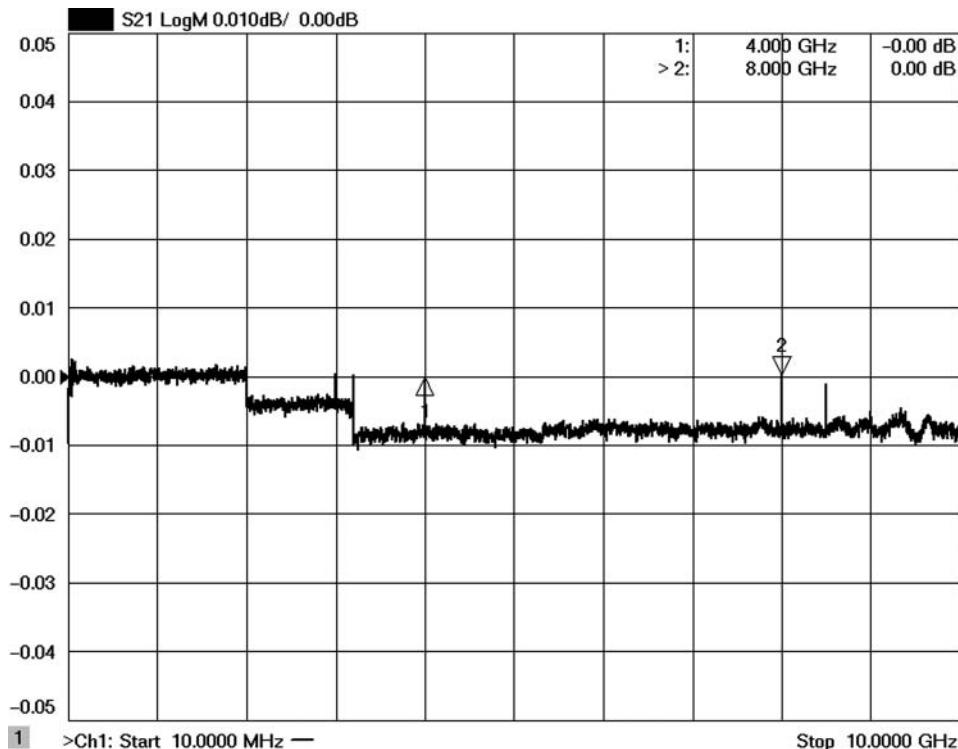


Figure 5.5 Calibrating in a stepped mode and measuring in a swept mode.

are posted. At these points, the band break occurs, but there is no change in sweep rate since these first points are stepped. Typically, the sweep rate increases at high bands as multipliers are used on the fundamental oscillator, giving the stair-step appearance. However, much effort goes into programming the sweep rate so that the steps are minimized. In the figure, the worst error due to swept-versus-stepped is about 0.01 dB.

Typically, swept mode is used when real-time response is needed for a tuning application, as in tuning a filter. For historic reasons, many VNAs default to swept mode. In the HP 8510, swept mode is sometimes known as *lock-and-roll* as the frequency is phased locked only at the first point and the rest of the sweep is performed in an open-loop mode. The HP 8510 provides a software switch to change from swept to stepped mode. One drawback in the swept mode is that the analyzer sweep accuracy may not be adequate to determine band edges in filters and similar components, so for the HP 8510, it was common to tune the filters in swept mode and then change to stepped mode for a final measurement. In the HP 8720, PNA, and ENA families, there is also a switch, but in a difference from the HP 8510, the HP 8720 maintains phase lock throughout the sweep. Some analyzers, such as the R&S ZVA, do not have a swept mode at all and provide only stepped sweep.

The HP 8753, which is used extensively in filter test, does not have an explicit mode for switching between swept and stepped mode; it switches into a stepped sweep in three conditions: (i) if the IF BW is less than 30 Hz, (ii) if the sweep time is greater than 15 ms

per point, and (iii) if the source power calibration is activated. These conditions also apply to the HP 8720.

For the Keysight PNA series, stepped sweep mode is automatically enabled for IF bandwidths of 1 kHz and below or for any of the advanced application channels or whenever source power calibration is enabled.

Realistically, one should always calibrate and measure electrically long devices with a stepped-sweep mode.

5.1.3 Attenuation Measurements

Attenuation measurements of cables are relatively straightforward measurements, essentially S21 versus frequency. As described previously, any cable measurements should be performed in stepped sweep mode. For long cables, where the loss is large, the IF BW should be set narrow enough to reduce trace noise. Figure 5.6 shows the attenuation of a 1 m cable used as a typical test port cable. These cables are chosen for their high quality and stability. Markers show the attenuation of the cable at various frequencies.

One major source of error in measuring cables is the input and output connectors used to mate from the VNA to the cable. If the cable is an assembly including connectors, then the only issue is to ensure a good calibration at the connector interface. However, it is common to measure a reel of cable that does not have connectors (such as a Community Antenna

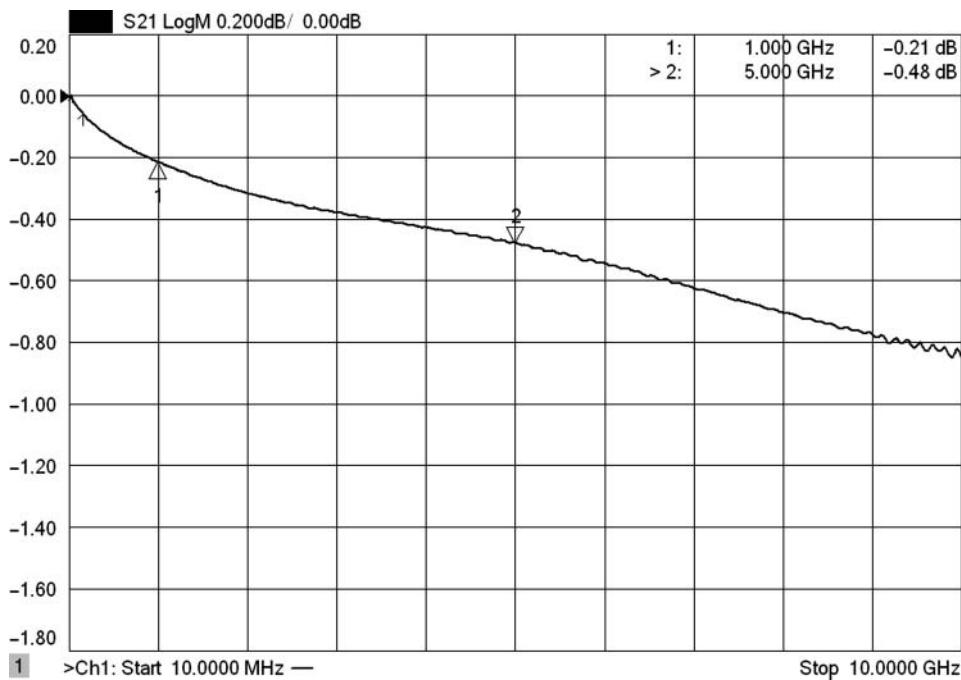


Figure 5.6 Attenuation measurement of a 1 m cable.

TeleVision [CATV] cable), or the connectors are used only for test purposes before shipping the cable. For long cables, the loss of the cable diminishes the mismatch effects, and input and output match correction of a full 2-port calibration compensate somewhat for the mismatch loss. For shorter cables, the mismatch between connectors can add significant ripple to the cable measurement.

5.1.3.1 Connector Compensation Using Port Matching

There are two techniques that may be used to reduce the effect of test connectors on the measurement of a long cable: connector compensation and time-domain gating.

Connector compensation involves creating a simple model of the input connector, typically a series inductive element and a shunt capacitive element. A compensating model is created using a port matching function in the VNA to compensate for the input connector. A common method is to generate a time-domain display of the S_{11} of the cable and adjust the series and shunt elements until the input mismatch is minimized. An example of this is shown for a measurement of a cable with poor input and output connectors, in Figure 5.7, with the S_{21} frequency response shown in the upper-left corner and the S_{11} and S_{22} frequency response shown in the upper right. The time-domain response for S_{11} and S_{22} is shown in the lower plot. In this case, the low-pass time-domain mode is used so that the type of discontinuity, capacitive or inductive, is readily apparent. For S_{11} , the first response shows a negative dip

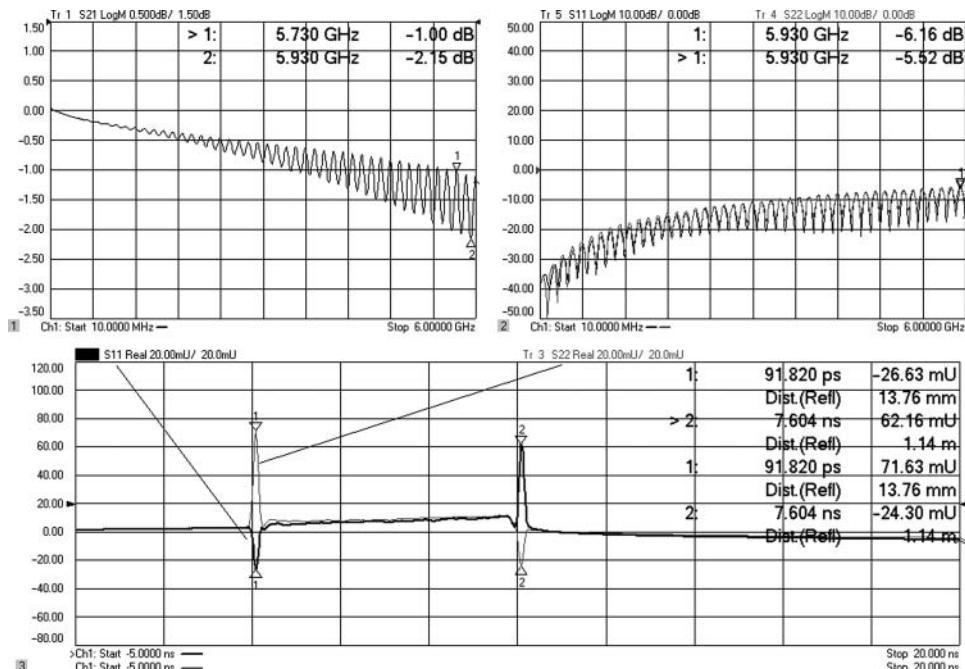


Figure 5.7 Cable measurement with poor connectors, frequency domain (upper), and time-domain response (lower).

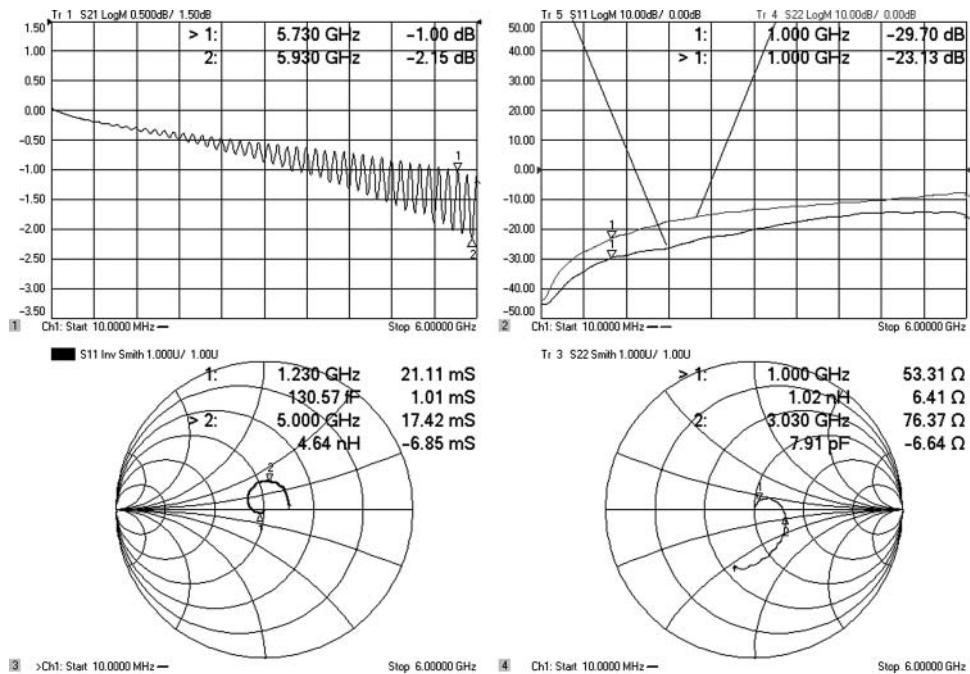


Figure 5.8 S_{11} and S_{22} after gating, indicating the vector error of each connector.

indicating a capacitive discontinuity in the connector. For the S_{22} trace, the first response shows a larger, positive peak, indicating a larger inductive discontinuity.

To help determine the compensation needed, a time-domain gate is applied to the input and output reflections, in Figure 5.8, and the resulting response is shown on as return loss log magnitude (upper, right) and in the Smith chart (lower). Since the input connector S_{11} time-domain response indicated a shunt capacitive discontinuity, the inverse Smith chart is used to display the gated S_{11} , so that the best estimate of shunt capacitance may be determined. The output-connector S_{22} response indicated an inductive discontinuity, so the normal Smith chart is used to display its response. From these responses, one can see the characteristic curve associated with a mismatch that has been rotated in phase from the reference plane (refer to Figure 2.35). The display in Figure 5.8 indicates that the mismatch in each connector is not located precisely at the VNA reference plane but is somewhat rotated in phase or displaced in distance along the connector. This is to be expected as one would assume the interface connector should have a reasonable response at the normal connector interface; it is the connection to the raw cable that will have some discontinuity.

From the Smith chart plots of Figure 5.8, one can see that the reactive element does not appear constant and in fact appears to change between capacitive and inductive; however, if port extensions are used to rotate the reference plane of each port, the actual value of the discontinuity, along with its location along the connector, can be determined. Figure 5.9 shows the result of applying port extension to each port, individually, until the reactive portion of Smith chart response is as constant as possible. For port 1, this occurs at about 49 ps of

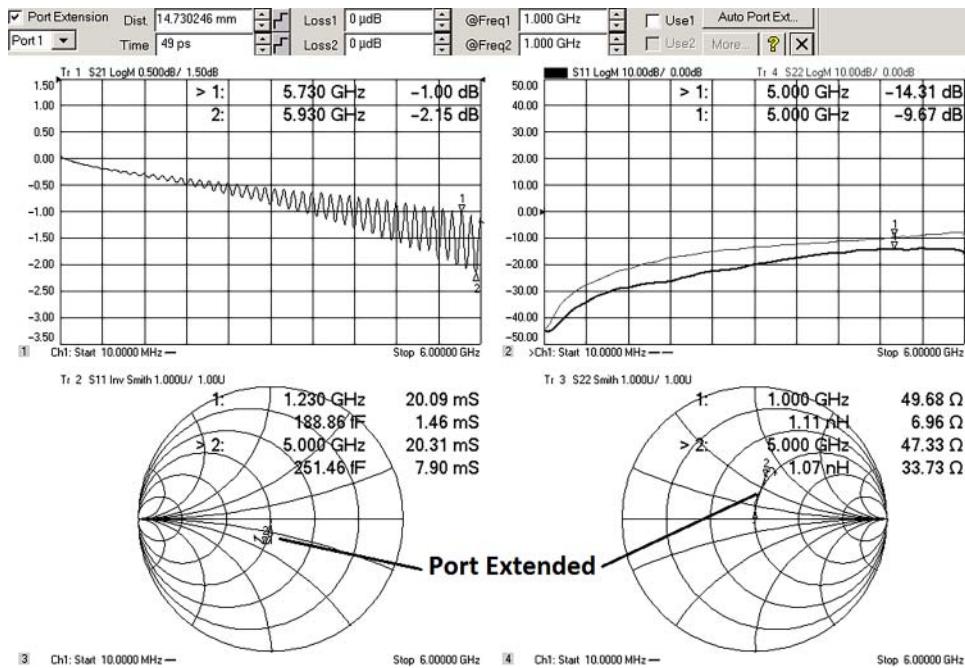


Figure 5.9 Port extension is applied to each port to determine the location and value of each discontinuity.

extension, which is adjusted until the difference in the reactive capacitance value between the marker at 1 and 5 GHz is minimized. For port 2, the value is about 40 ps of delay where the difference in the reactive inductance value is minimized. In this case, the port 2 discontinuity is more ideally inductive (has a more constant value) than the capacitive discontinuity of port 1.

From these Smith chart displays, an average value for the discontinuity is determined as a starting point for using the port matching feature of the VNA.

The port matching function in a VNA allows one to add a simulated or virtual matching network at the input or output port of the DUT, as illustrated in Figure 5.10. With the port extension used, the reference position for the added network is rotated so that the network's effect is at the desired point along the connector. Figure 5.10 also shows the dialog used for adding port matching of -0.220 pF of shunt capacitive (negative so that it compensates for the positive capacitance shown in Figure 5.9) to port 1. A similar simulated network adds -1.1 nH of series inductance to port 2. With the port matching on, the Smith chart traces are nearly fully compensated and appear close to the center of the Smith chart. The log-magnitude traces of the time-gated return loss in the upper-right corner of Figure 5.10 show that the effective input match has been improved to better than 30 dB. Finally, the upper-left plot shows the initial S_{21} response (including the ripple and loss effects of the bad connectors) and the S_{21} response with port matching. Clearly, the port matching removes nearly all the ill effects of the poor connectors.

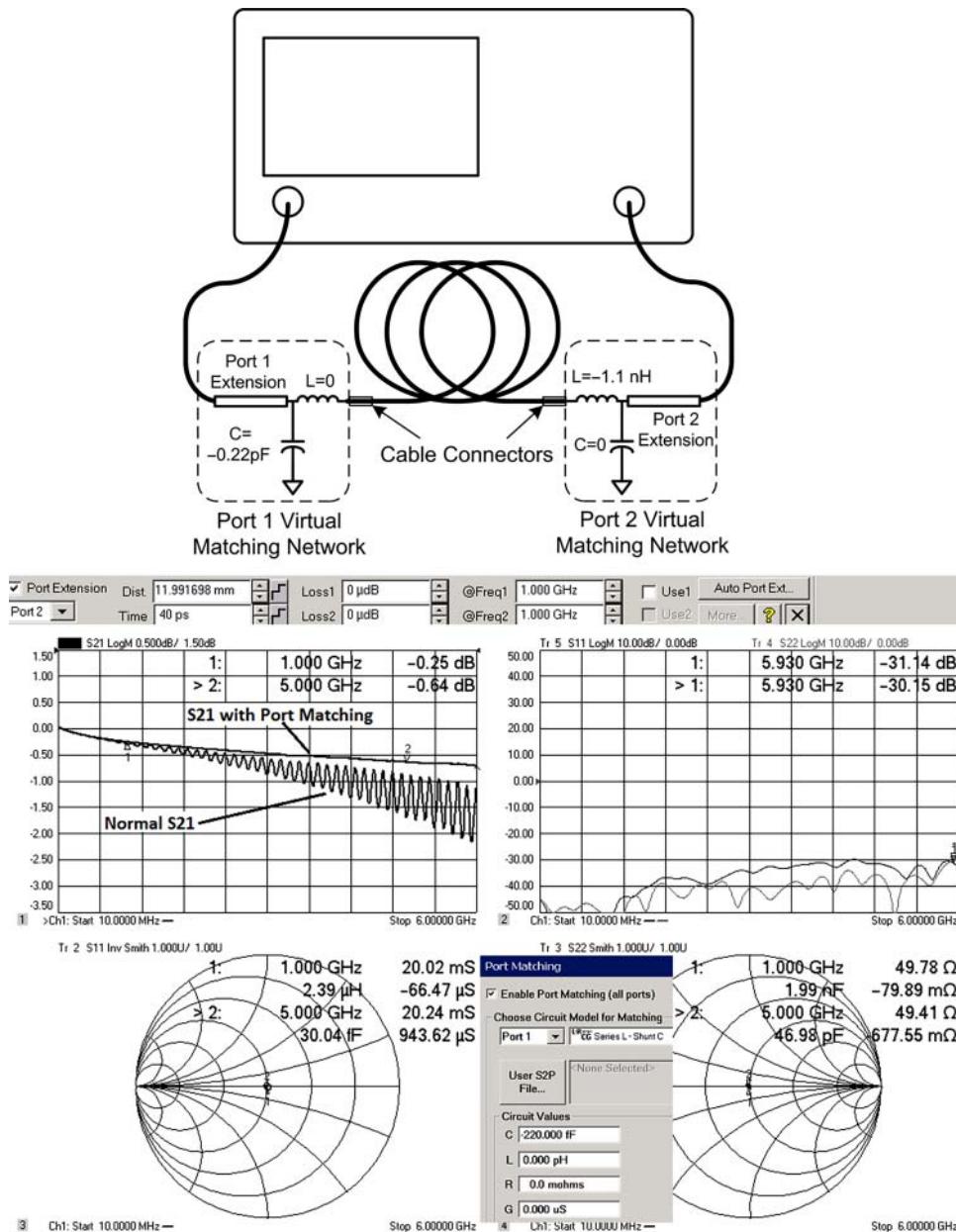


Figure 5.10 Port matching adds the negative of the reactive element for cable compensation.

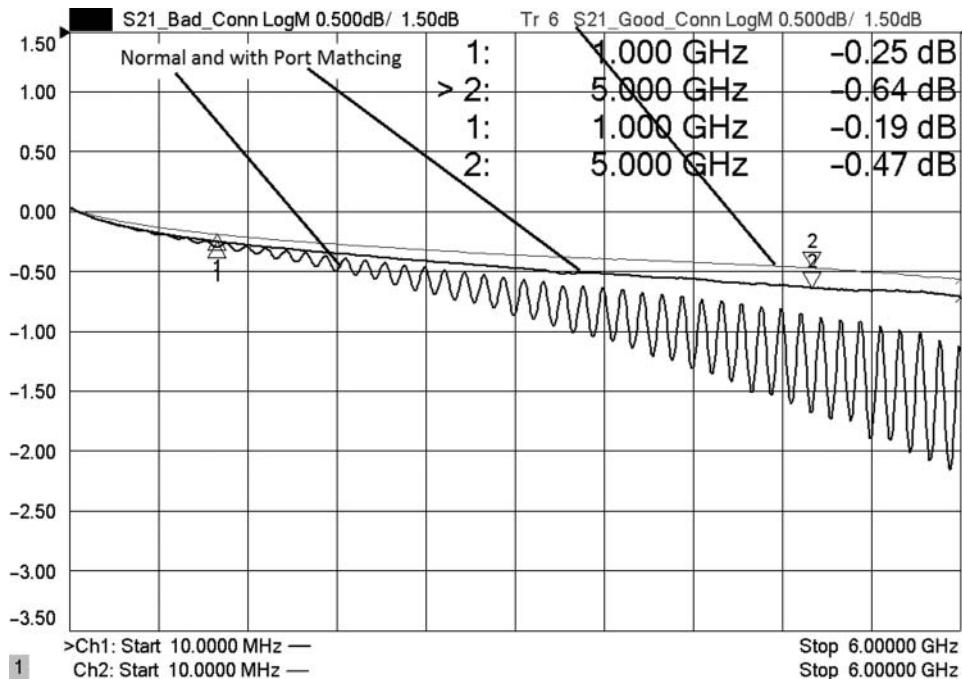


Figure 5.11 Comparing the cable S_{21} with compensation and with good connectors.

As a final comparison, the same cable is measured using very good connectors so that the input and output return loss are small. The resultant S_{21} trace is slightly less lossy than the measurement with bad connectors and compensation with port matching, as shown in Figure 5.11.

The difference in loss could be due to some slight non-compensation since the capacitive discontinuity was not single-valued (it ranged from -0.188 to -0.251 pF), but more likely the cause is due to resistive or radiation loss in the poor connectors. The fact that there exists a discontinuity that yields a poor return loss also suggests that the connector might have some radiation or other lossy response. However, when comparing the measurements of the uncompensated S_{21} with poor connectors to the compensated value, one can see remarkable improvement, with results nearly as good as using a higher-quality connection.

Connector compensation using port matching is quite robust and does not depend upon any attributes of the DUT; however, it does require that the connector discontinuity be simple if simple port matching is to be used. In the next section, a method that does not require any knowledge of the discontinuity to remove its effect is described.

5.1.3.2 Connector Compensation Using Time-Domain Gating

Time-domain gating involves using the VNA time-domain function to remove the effect of the input and output mismatches from the S_{21} measurement. However, it is common to mistakenly

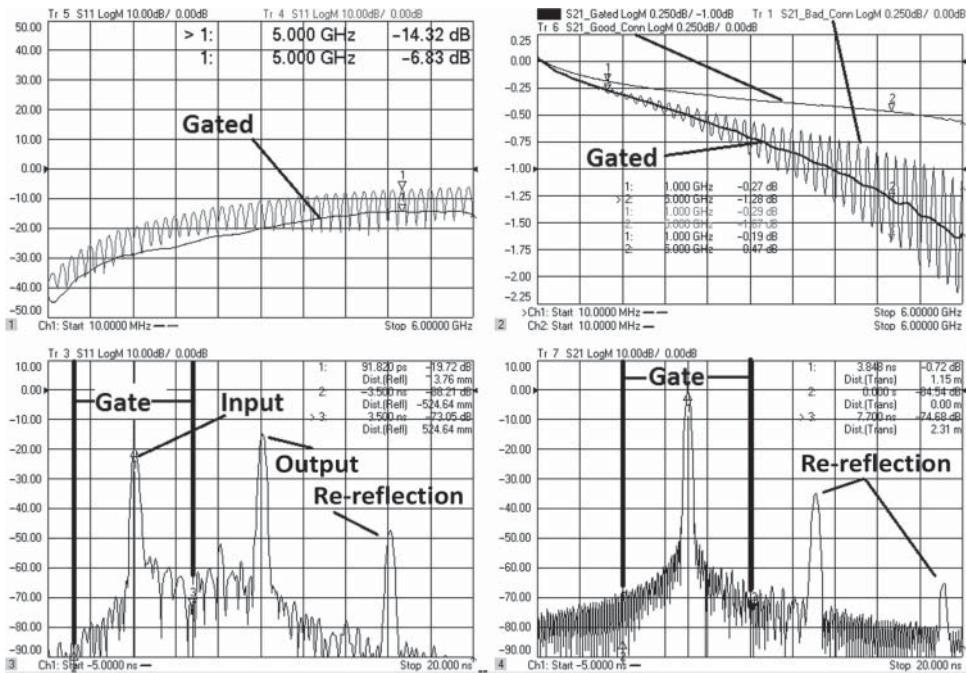


Figure 5.12 A cable with input reflections showing S_{11} (right) and S_{21} (left) with frequency response (upper) and time-domain response (lower).

apply the same gating settings that may have been determined from a reflection (S_{11}) measurement to the transmission (S_{21}), and this can yield completely erroneous results. To be clear, if one uses time domain to remove the effect of reflection (either input or output) in an S_{11} measurement, the same gating settings *cannot* be used with a transmission measurement. To fully understand the time-domain response and reflection effects on a transmission measurement, consider the sample cable measurement shown in Figure 5.12.

In this figure, the same cable and connectors are used as in Figure 5.7; the S_{11} traces are on the left, and S_{21} traces are on the right. The upper portion shows the frequency response, and the lower portion shows the time-domain response. In the S_{11} time domain (lower, left), there are three distinct peaks representing the reflection from the first test connector (marked Input), the second test connector (marked output), and a reflection representing a double reflection from the output, back off the input and off the output again (marked re-refection). The scale is 2.5 ns per division, and the reference (0 time) is at the second division. A time-domain gate is placed around the first reflection, and the gated response is shown in the upper S_{11} frequency response. This represents the energy reflected from the first input test connector. While these connectors are not very good, they are quite typical of some common test port cases, in this case having a reflection coefficient ranging up to -14 dB for the input connector (at high frequency) and -10 dB for the output connector (not shown in the figure). The time-domain response peak represents, in a way, an average reflection over the frequency range of 20 dB

for the input connector and 18 dB for output connector; the output connector appears better in the S11 measurement than it actually is due to the loss of the cable. The gate to separate the input refection from the output reflection is set centered on the input reflection, and the span of the gate is set so that the stop is exactly between the two reflections. Not coincidentally, this time will be equal to the time-domain transmission time, or group delay of the cable, and represents placing the stop gate at exactly halfway into the cable. It is not required to gate so far into the cable, and some experiment can determine when the gate span is sufficiently wide to account for the entirety of the input connector reflection.

The S_{21} time-domain transmission response is shown in the lower-right portion of Figure 5.12. It shows three response peaks, but these represent the transmission response as a function of time. The first, main response shows the time-domain transmission through the cable, and the peak is located at a time that represents the group delay of the cable, in this case, 3.848 ns. The other smaller peaks represent the arrival of the re-reflected signals from the input and output test connectors. Note that if only one of these connectors causes a reflection, there will be no re-reflection in the S_{21} response, and the only effect on S_{21} will be the loss of energy represented by the gated signal of the S_{11} response. However, with a poor input and output connector, the S_{21} response has substantial ripple on it. This ripple is shown in the upper-right portion of Figure 5.12. Also shown is the time-gated response of S_{21} . In this example, it is possible to apply time-domain gating to the S_{21} response, as the re-reflections caused by poor test connectors are easily separated in the time-domain response. But note that the gating factor for S_{21} is not at all related to the S_{11} gate center. In fact, it must be set so that it is centered on the transmission response peak, which is at the group delay of the cable. The gate span must be small enough to eliminate the re-reflection from the output and input connectors. Generally these will occur at three times the delay of the cable, so setting the span at two times the cable delay is a reasonable starting point. In fact, a second higher-order re-reflection can also be seen at five times the delay of the cable, which represents the transit time of the main lobe of the time-domain transmission response, plus two times the transit time for the first reflection and two times more the transit time for the second reflection for the total of five transit times.

The time-gated frequency response is shown in the upper trace, along with the ungated response, which shows the effects of all the re-reflections. It is clear for a low-loss cable that these re-reflections dominate the response. What might not be so clear is that the energy loss from the reflections also dominates the response so that the S_{21} measured response of this cable section, which should not include the mismatch effects of the test connectors, is nearly completely dominated by the test connectors' response. Fortunately, the compensation techniques described in Section 4.8.2 can be applied to this cable measurement with good effect.

The energy lost from the reflection at the input port, and output port, can be described as

$$\begin{aligned} Loss_{Input} &= \sqrt{1 - |S_{11_Gated}|^2} \\ Loss_{Output} &= \sqrt{1 - |S_{22_Gated}|^2} \end{aligned} \quad (5.3)$$

where S_{11_Gated} represents the gated frequency response of S_{11} trace. The gated response of S_{21} is essentially missing this energy but can be compensated by dividing the gated response

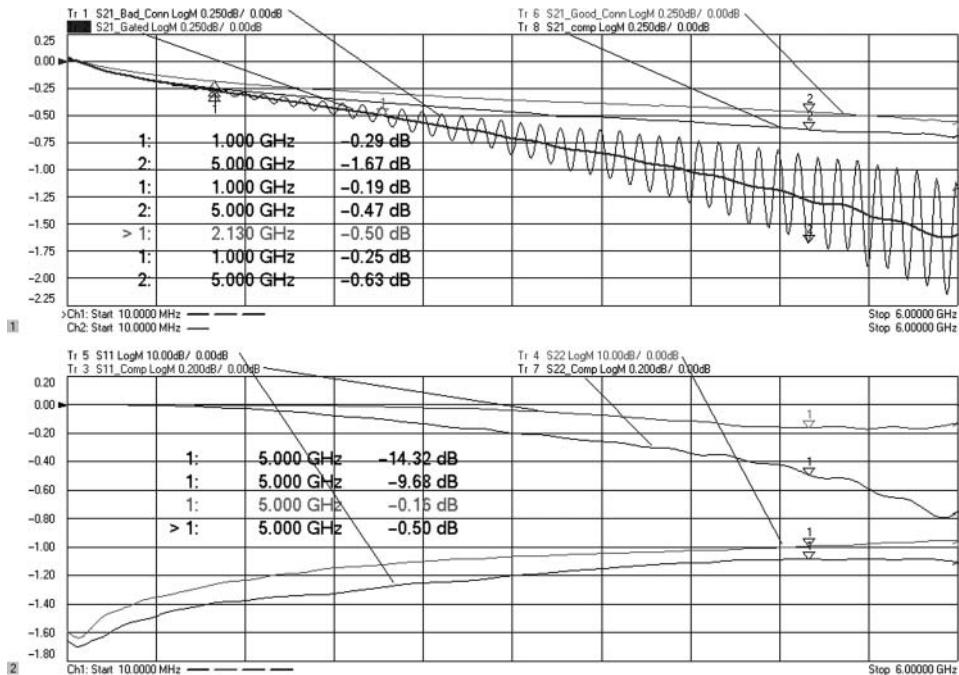


Figure 5.13 Compensating S_{21} for the effect of bad input and output connectors.

by the product of these losses:

$$S_{21_Compensated} = \frac{S_{21_Gated}}{\left(\sqrt{1 - |S_{11_Gated}|^2} \right) \left(\sqrt{1 - |S_{22_Gated}|^2} \right)} \quad (5.4)$$

While not shown in Figure 5.12, the S_{22} gated response is determined in a similar manner as the S_{11} response. These results of these computations are shown in Figure 5.13.

The upper plot shows the S_{21} response with no gating or compensation showing the ripple (marked $S_{21\text{Bad_Conn}}$) and with gating applied (marked $S_{21\text{Gated}}$, and cutting through the middle for the ungated response). Also shown is a trace with the gating and compensation applied (marked $S_{21\text{Comp}}$) and the ideal response for this cable (measured with good connectors). Clearly, the gated-compensated S_{21} is a much closer match to the expected result than the original S_{21} measurement or the gated-only S_{21} measurement (no compensation) from the upper-right plot of Figure 5.12.

The lower portion of Figure 5.13 shows the gated S_{11} and S_{22} measurements, on a scale of 10 dB per division, and the computed compensation for each from Eq. (5.3), marked $S_{11\text{Comp}}$ and $S_{22\text{Comp}}$ in the figure, and scaled at 0.05 dB per division. The equation editor feature of the VNA was used to compute the compensation traces. It was also used to apply these individual compensations to the S_{21} trace utilizing Eq. (5.4).

A small note on practical application of the equation editor: for this VNA, the equation editor math can be applied to either S-parameter results or trace results. In this case, since Eq. (5.3) must be applied to the gated response of S_{11} and S_{22} , a pair of traces displaying

the gated results must first be created (Tr5 and Tr4 in the figure). Next, the equation editor function is applied to these traces to generate Tr3 and Tr7. Finally, the S_{21} trace (the gated S_{21} trace from Figure 5.12, Tr8) is compensated in the equation editor using the equation

$$S_{21_Comp} = \frac{S_{21}}{(Tr3 \cdot Tr7)} \quad (5.5)$$

This trace also has the S_{21} transmission time-domain gate applied.

Since time domain is a linear function, it would be equally valid to create a gated S_{21} trace first and then another equation editor trace taking the gated response of the S_{21} divided by the compensation for S_{11} and S_{22} . Care must be used when applying time gating to equations to ensure that the gate is applied in the proper operational order. In the case of Eq. (5.5), the values of Tr3 and Tr7 are simple scalar numbers due to the absolute value function in Eq. (5.3) and do not affect the phase response of the result and thus can be included either inside or outside the Fourier transform, as

$$\mathbf{F}\left(\frac{S_{21_Gated}}{Tr3 \cdot Tr7}\right) = \frac{1}{Tr3 \cdot Tr7} \mathbf{F}(S_{21_Gated}) \quad (5.6)$$

Note, in this case we refer to the forward Fourier transform as we are in a sense converting back from the gated time-response to the frequency domain.

Finally, here again as in Figure 5.11, the compensated S_{21} trace is slightly lower than the S_{21} trace of the cable with good connectors. This is likely due to the uncompensated resistive insertion loss and radiation of the poor connectors, just as in the example of Figure 5.11. It is reassuring that in both methods of connector compensation, the resulting S_{21} traces agree with 0.01 dB!

This second method of connector compensation does not rely on obtaining any models or de-embedding of the test connectors. Some care should be taken if applying this technique to more complex transmission structures, as a prerequisite for this type of compensation is that the entire S_{21} time-domain transmission response must be able to be identified in the S_{21} time-domain transform, separately from the input and output reflections. In the case of more complex structures, re-reflections in the DUT itself might appear in the time-domain response near the test connector re-reflections and render the use of gating impractical. In such cases, using port matching may yield a better result.

5.1.3.3 Attenuation Measurements on Very Long Cables

One problem that is particular to cable measurements is the occasional need to measure very long cables in situ. This might be an example of a cable between two buildings or a cable in some other structure such as ship or airframe or up to the top of a radio tower. There are two use cases for long cable measurements; one is when both ends of the cable are accessible, although perhaps with some difficulty, and the other is that only one end of the cable is accessible. Options for the first case are described here; options for the second case are described in Section 5.1.3.5. Another use for long cables is in the measurement of a DUT that is not itself large but is situated far from the VNA. This might occur in a satellite test scenario where the satellite is in a thermal-vacuum (TVAC) chamber. A long test port cable may be required to connect through the chamber to the satellite.

The difficulty with measuring long cables is that the input and output must be widely separated. One method that can be employed is to use a remote power sensor and a source to measure the cable loss. This is a scalar technique that can provide reasonable results as long as the loss to the power sensor is not too great. Power sensors are broadband detectors and have a dynamic range limited to about 50 dB for the thermistor type sensors and to 90 dB for the diode sensors. Many power meters can be controlled and triggered by LAN or USB control, so it is possible to remotely locate the head. Some systems provide source and power meter control directly, such as the Keysight PNA, which can use a power-meter-as-receiver (PMAR) mode to directly display a synchronized trace of the power meter reading at triggered source frequencies. Such a measurement is shown in Figure 5.14. In the figure, the dark trace is the power meter response, and the light trace is the traditional S_{21} , for a 10 m cable. The reference for the traces is intentionally offset by one division for clarity, but from the marker readings, one can see that they have nearly identical values.

The noise floor of the power meter does affect the reading at higher frequencies, as well as the lack of match correction, so the worst-case error using a power meter in this way exceeds 0.4 dB. The power-meter based trace in the figure shows more ripple at high frequency than a measurement of the same cable using a full 2-port calibration.

However, in many cases, the phase or delay response of the cable is required, which means that a full vector measurement must be made. The most common method is to position the

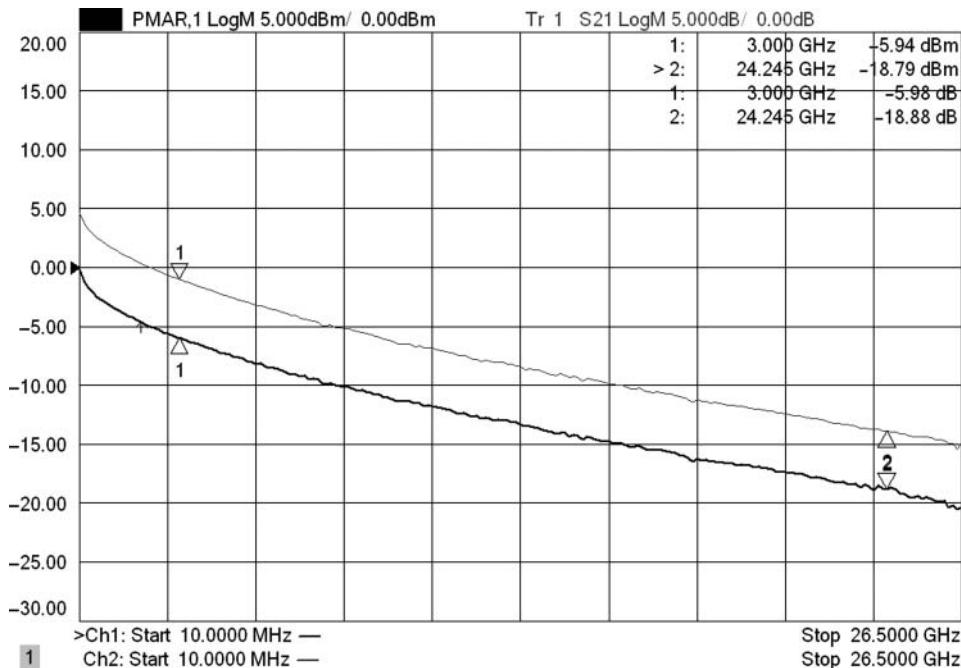


Figure 5.14 Using a power meter as a receiver at the end of a long cable, compared to S_{21} , offset one division.

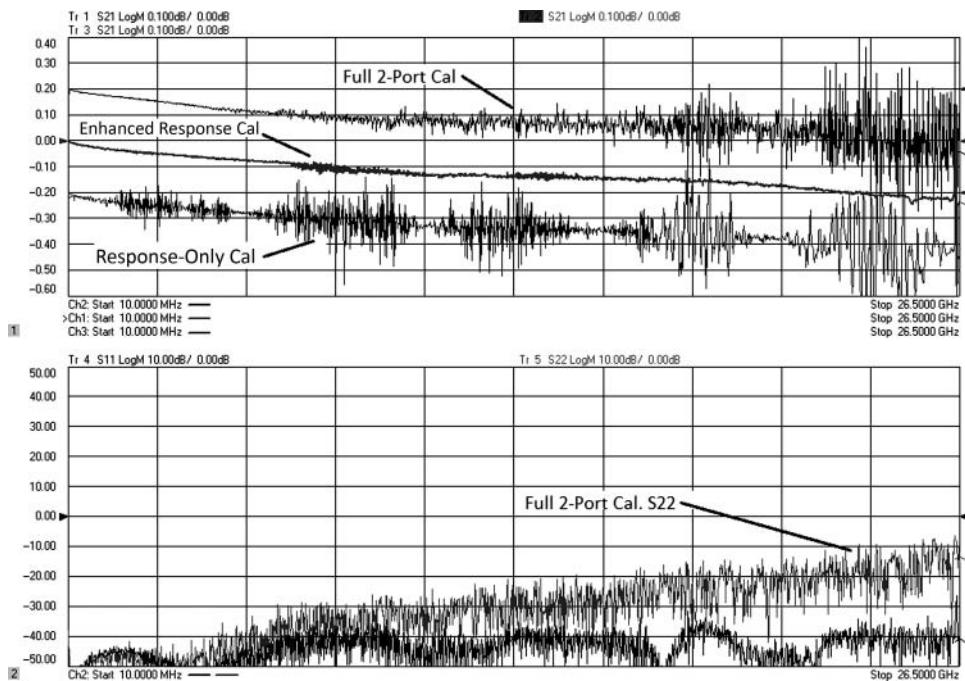


Figure 5.15 Measurement at the end of a long test port cable comparing full 2-port calibration with enhanced response calibration, after movement of the test port cable.

VNA at one end of the cable run and then use a long test port cable to connect to the other end. This presents a particular problem in that the stability and loss of the test port cable will severely limit the overall response. For long cables, the mismatch can dominate the measurement result of a full 2-port calibration as the output match correction at the end of long test port cable is very poor. One method to remove this output match correction error is to use response-only (S_{21} normalization) or enhanced response calibration (ERC) that removes the input mismatch from the S_{21} measurement. Figure 5.15 shows a comparison of a measurement of a short, low-loss line using a 10 m test port cable after flexing the test port cable following calibration. Shown are the S_{21} traces using a full 2-port correction, a response correction, and an enhanced response correction, each one separated by two grid divisions for clarity. In the lower window, the S_{11} and S_{22} traces are shown from the full 2-port calibration. The full 2-port correction shows errors due to the poor measurement quality of the S_{22} trace. The response-only correction shows errors due to mismatch at the input and output. The enhanced response correction shows the best overall measurement of the cable under test, as the input match is corrected for, but the output mismatch correction is not performed. Since the actual value of the DUT S_{22} is essentially hidden from the VNA by the loss and poor match stability of the test port cable, performing the output match correction actually yields poorer results.

5.1.3.4 In-Situ Calibration and CalPods

Recently, a new system/method has been developed that addresses the stability issue. The cause of the stability issue is the fact that the test port cable can have loss and delay errors due to flexing and changes in loss or mismatch. This new system makes use of an in-situ calibration method that can reduce or eliminate the instability of a long test port cable. The heart of the system is an electronically controlled module that provides three discrete reflection standards, nominally an open, short, and load, as well as a thru state, connected directly in-line with the DUT. An example of the system is shown in Figure 5.16.

To use the modules for recorrection, a calibration is performed at the output end of the recorrection module. Immediately after calibration, the standards of the recorrection module are measured and recorded. This is known as the *initialization* process, and it is critical that the initialization measurements are performed before any drift in the test port cable occurs. It is not even necessary to do the initialization using the long test port cable; in fact, a short cable or no cable at all can be used for the initial calibration and initialization. Next, the output end of the in-situ calibration module is connected to the DUT. This can be using the same cable as the initial calibration, or a different, longer cable. The three standards are remeasured, and a recorrection array is computed that represents the difference between the initialization and the recorrection measurement. Any drift, or change in loss or match of the test port cable, is captured in this difference array. The mathematics involved is described in Chapter 11, as a difference array.

The quality of recorrection is shown in Figure 5.17 where the light traces show the poor S_{21} and S_{11} measurements after drift in the test port cable. In this case, the test port cable is a 10 m high-quality cable typically used in satellite test systems. The dark traces show the result after an in-situ recorrection. In fact, this method of recorrection is so beneficial for removing cable drift that it is useful in very low-loss, low-reflection measurements where slight cable drift can degrade performance. In most cases, it is actually more accurate than a calibration



Figure 5.16 Example of a recorrection system for removing test cable drift. Source: Picture Courtesy of Keysight Technologies.

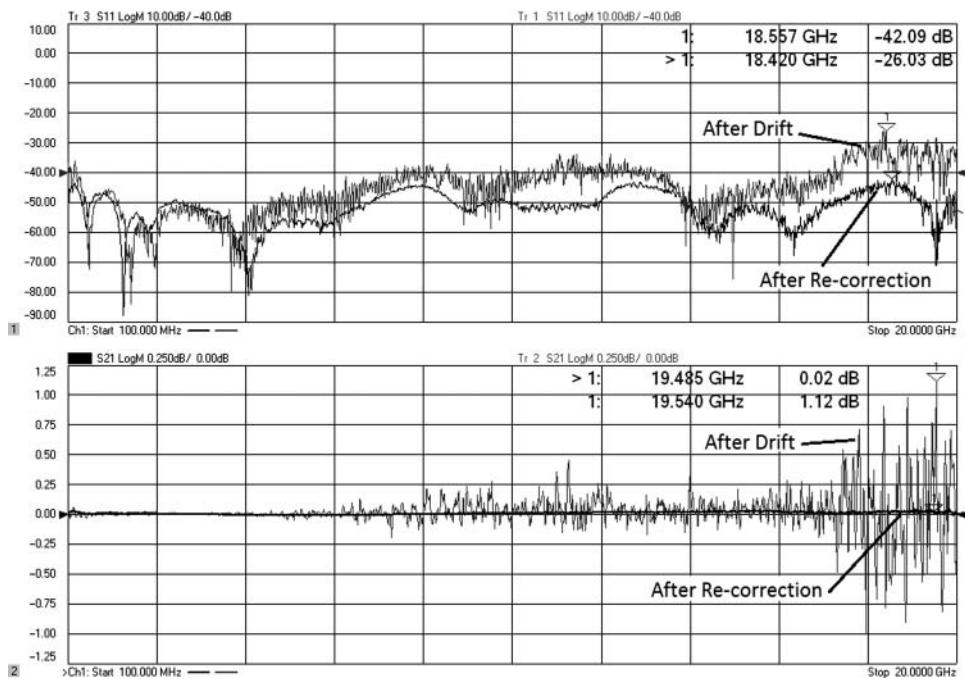


Figure 5.17 Light traces: drift in S_{21} and S_{11} due to a long cable; dark traces: result after in-situ reconnection (utilizing Keysight CalPod).

with a normal Ecal or mechanical standards, as simply the act of disconnecting the Ecal and connecting the DUT can cause a change in the test port cable response that is great enough to benefit from reconnection.

The only limitation is the assumption that the change in loss is reciprocal, and for the proper computation of the change in phase, there must be less than 180° phase change between the initialization and the reconnection between each measurement point. For normal differences due to an unstable cable, this is a minimal restriction. However, if the initial calibration is made with a short cable and the measurements are made with long cables, then the phase change for the difference can be quite large; in such a case, a large number of points is required to ensure the phase response is captured without any aliasing. It is often desirable to do the initial calibration with shorter, lower-loss cables to reduce the error and noise in the initial calibration. The reconnection quality is limited by the noise of the initialization data and the repeatability of the reflection states of the in-situ calibration module. These errors after reconnection can typically be kept below 0.02 dB insertion loss repeatability and -50 dB return loss repeatability for in-situ calibration modules.

The greater loss between the test port coupler and the reconnection module, the more the re-connected performance is degraded. With commercially available modules, the degradation is minimal with less than 16 dB loss between the module and the VNA test port coupler. At lower frequencies, this might extend as far as 20 dB. Beyond 26 dB loss, reconnection results are typically poor. Figure 5.18 shows a plot of the degradation in residual directivity and

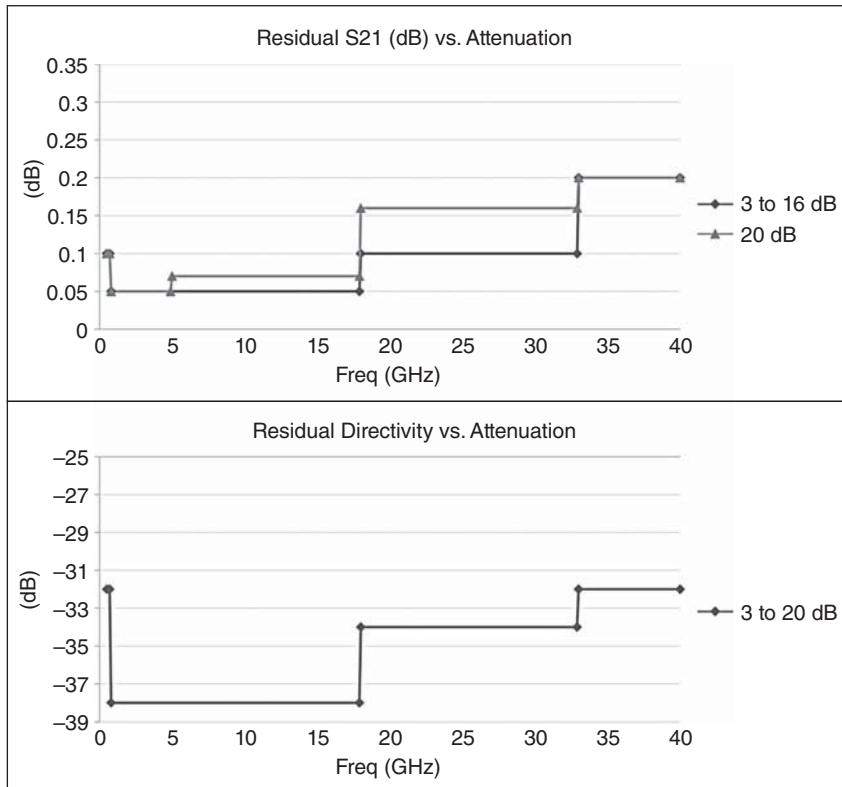


Figure 5.18 Residual directivity and insertion loss as a function of loss between the VNA test port and the recorrection module.

insertion loss tracking as external loss (such as a long cable) is added between the VNA and the in-situ calibration module (in this example, using Keysight 85540B CalPod modules).

While rather new, the technique of using in-situ calibration provides dramatic improvement in making measurements at the ends of long or lossy test port cables and can also correct for a variety of other instabilities including switch repeatability in the case that a multiport switching network is placed between the VNA test port and the DUT.

5.1.3.5 Time-Domain Responses and One-Way Measurements

Occasionally, a cable is located in such a way that only one end is accessible to test equipment and it is still desired to measure the frequency response (magnitude and phase). A classic example is a cable in an airplane wing connected to an antenna element. In such a case, one may still be able to characterize the cable using time-domain reflection techniques.

To use reflection techniques, it is necessary to provide a reflection at the far end of the cable. This can be accomplished by disconnecting it from its terminating element or, in the case of a cable to an antenna, putting a reflector close to the antenna to provide a total reflection into

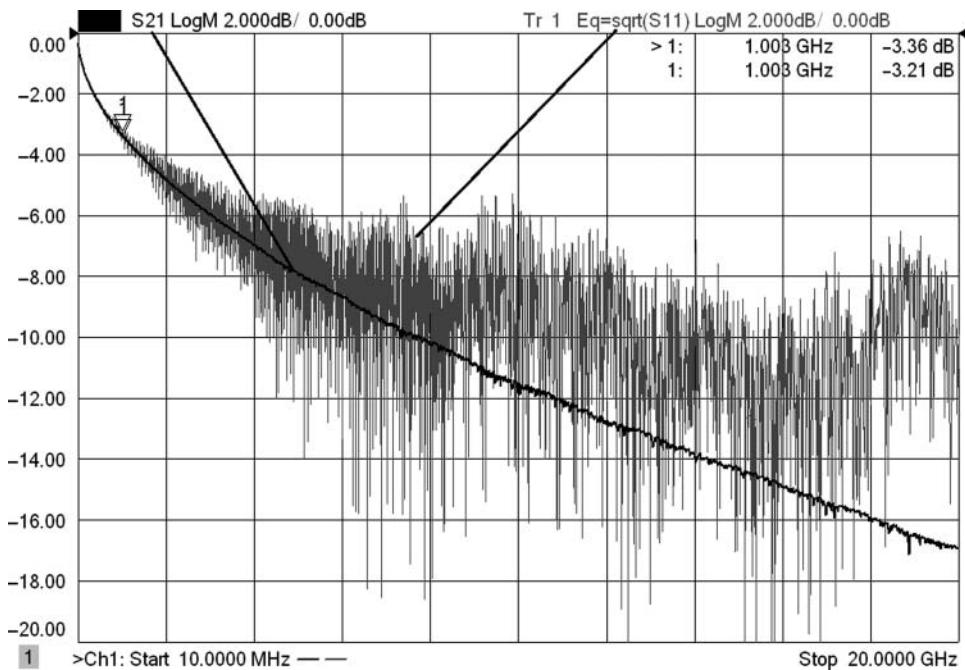


Figure 5.19 Dark trace: the transmission of a cable; light trace: square root of S_{11} of the same cable terminated in a short.

a cable. As an example, consider a planar antenna embedding in the structure of a wing; this could be wrapped in a conductor (aluminum foil) to provide a total reflection.

A simple way to approach this is to simply look at S_{11} of the cable and determine that the one-way loss is simply the square root of S_{11} (or one-half the dB value). However, if the cable has much loss and any reflections, this method will yield poor results. Typically, such a cable will have a substantial reflection at the input connector.

In Figure 5.19, the dark trace shows the S_{21} of a cable, and the lighter trace shows the square root of S_{11} . At lower frequencies, where the cable loss is small, this gives a good estimate of the transmission loss. However, at higher frequencies, where the loss is greater and the reflection from the input connector is also greater, the loss determined by the square root of S_{11} deviates substantially from the transmission response. This is because of the input connector's reflection is adding and subtracting from the two-way loss of the cable and effectively masking the cable's loss.

One can use the techniques similar to those of Section 5.1.3.2 to improve this one-way measurement of a cable's attenuation by taking advantage of time-domain gating (Dunsmore 2007). Figure 5.20 shows in the upper window both the normal S_{21} measurement of the cable and the gated time-domain transform of the S_{11} of the shorted cable.

In the lower window is the time-domain response with the transmission response shown with gates around it. With even limited knowledge of the cable's length, it is easy to determine the location of the shorted end of the cable. The peak on the time domain is lower

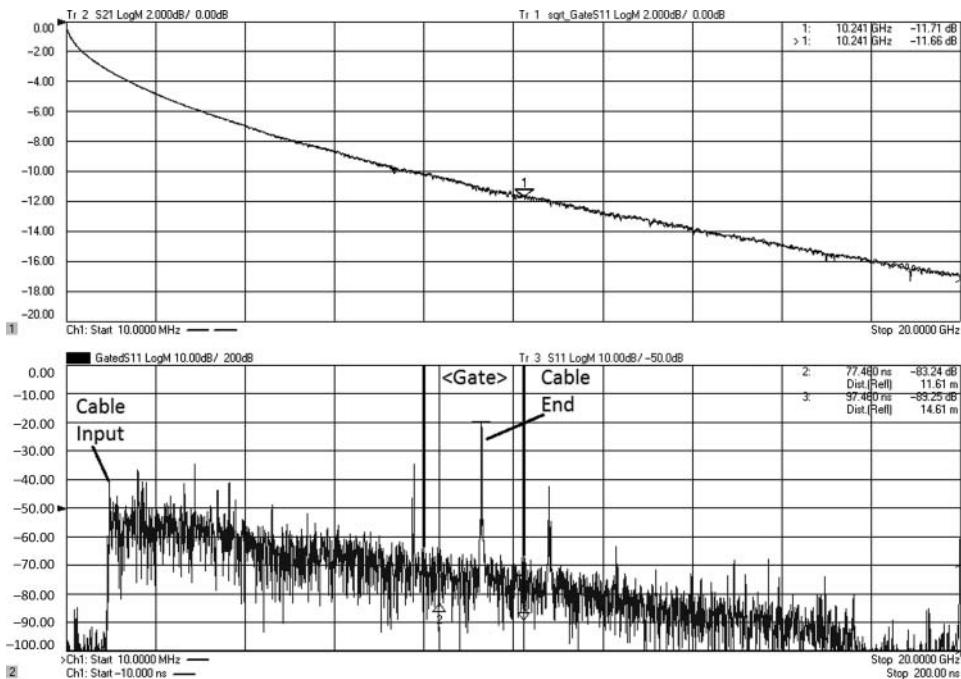


Figure 5.20 Lower window: time-domain response of the shorted cable, with gate set about the short; upper window: frequency response of the cable showing S_{21} , and time-gated S_{11} .

than expected for a short due to the average (across frequency) two-way loss of the cable; but also, a close inspection of the impulse-like response associated with this transmission peak shows some spreading (or dispersion) at the base of the time-domain response for this transmission impulse. This widening of the transmission response is due to an effect similar to windowing and implies that there is a nonconstant frequency response. Thus, it contains information on the magnitude and phase response of the two-way transmission. By placing a gate around this reflection and gating out the other reflections (principally from the input connector, and a discrete reflection in the cable), the loss of the cable is isolated from the other reflections.

The traces in the upper window are the square root of the gated frequency-response of S_{11} , along with the previously measured S_{21} ; they are nearly indistinguishable. In comparing it to Figure 5.19, the improvement in the quality of the estimate of S_{21} is remarkable. Even at this small scale of 2 dB per division, this is nearly identical to the S_{21} measured using standard techniques and demonstrates the usefulness of this method. Further, if one or a few large reflections are identified in the time-domain response, their masking effect can be compensated for just as described in Eq. (5.3). The lower plot shows the time-domain response with the cable input and cable end identified, along with the gate location. It's also clear that there is a substantial reflection in the cable at about 75 ns, likely from some damage in the cable. The re-reflection from this off the short at the cable end is clearly seen at about 110 ns.

5.1.4 Return Loss Measurements

Return loss measurements on cables and connectors are relatively straightforward for cables with common connectors, measured as an assembly, and for the connectors themselves when both ends of the connector are of a type for which a calibration kit exists. However, return loss measurements on bare-cable, very long cables and on connectors that do not have common calibration kits can be quite problematic.

Return loss measurements on cables and connectors are almost always 2-port measurements, where the cable or connector requires a termination, typically in the VNA port 2. If the DUT has a connector to which a high-quality metrology-grade load can be mated, then often the best return loss measurement will result from a simple 1-port calibration and terminating the DUT with the high-quality load. This is especially true in older, legacy VNAs that lack the newer calibration methods and the DUT is non-insertable, that is, when it does not have a mating male and female connector on each port. For older VNAs, the best calibrations could be performed only with flush (zero-length) mating connectors, that is, male on one port and female on the other. If the DUT had the same sex connector on each port, or connectors of different types on each port, then performing a proper full 2-port cal was not practical, unless a well-matched, and well-known defined thru standard was available. In many cases, the error from using a non-zero-length thru, and ignoring the length and match effect, can be greater than the return loss to be measured.

Later versions of legacy analyzers, such as the HP 8510 and HP 8753, incorporated an adapter removal method that allows mixed connector calibration by performing two 2-port calibrations, one on each side of the single adapter that mated to each port. This calibration process required both calibrations to be performed, and then a new calibration set was extracted from the two calibrations.

For high-quality calibrated measurements on most cables and connectors, some type of non-insertable calibration is really required. A common DUT is a cable with male connectors on each end, and a common error is to perform a normal SOLT calibration with female open/short/load standards, and then use a non-insertable male-to-male adapter as the thru standard. This will cause an error in the load match and insertion loss, as described in Section 3.3.3.3.

For almost all cables and connectors, using the unknown thru calibration will provide the best results. In fact, surprisingly the best calibration is likely to occur if the DUT itself is used as the unknown thru. If the DUT has relatively low loss, for example less than 10 dB, it will provide a sufficiently good unknown thru. If the unknown thru step is the last step of the calibration, then there will be no test port cable flexure after the calibration to add error to the measurement. Most engineers assume that using the small thru out of the calibration kit will provide the best calibration, but for reasonably small loss cable and connectors, the DUT itself will, in practice, be the best choice.

This concept applies to cables or connectors that have the same connectors or mixed-series connectors, provided that a calibration kit or Ecal is available for the each connector type.

As an example, a short length of formed semi-rigid cable is tested by calibrating with an Ecal used as an unknown thru, and another calibration using the DUT cable itself as the unknown thru, with the results shown in Figure 5.21. In this case, a slight flex of the test port cable results in slightly more loss measured using the Ecal thru compared to using the DUT as an unknown thru. In this case, the difference is less than 0.02 dB.

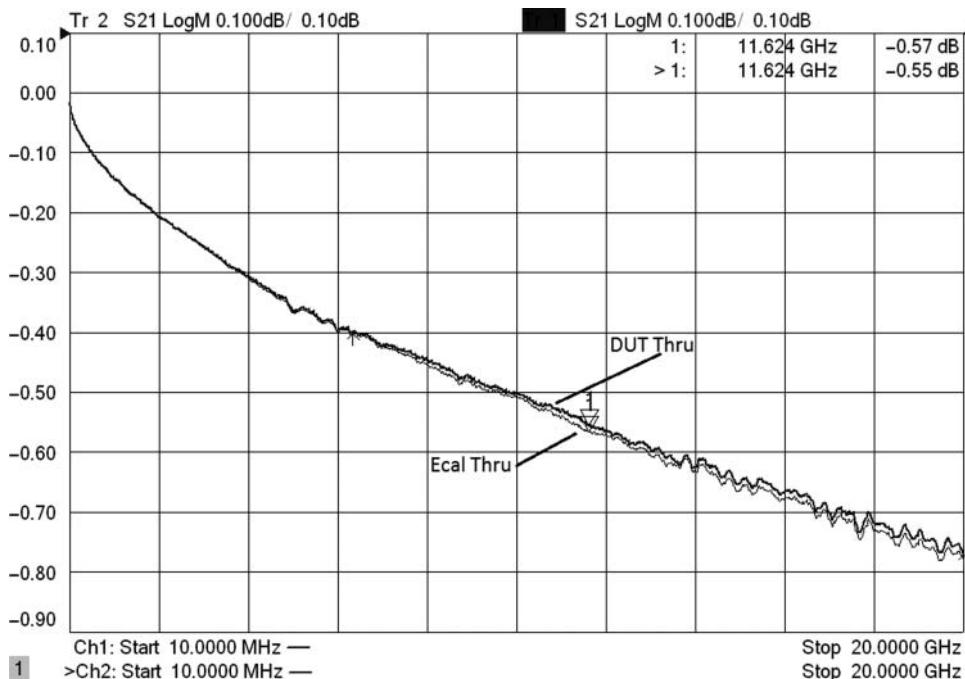


Figure 5.21 Comparing measurements of a short piece of formed semi-rigid cable using an Ecal as unknown thru, and the DUT as an unknown thru.

5.1.4.1 Measuring In-Line Cable Connectors

For some industries such as the cable-TV infrastructure, large hardline cables are used for their very low loss. These cables are connected using an in-line splice or connector. A similar situation occurs for type-F cables used in the home, where the center wire of the cable provides the center pin for a type-F interface, or where the in-line adapter is a type-F female-to-female adapter. In these cases, which can be classified as in-line connectors, the quality of the adapter is difficult to judge in isolation as its effects are apparent only when used between two cables. As such, measurement methods for in-line connectors have been developed that allow them to be measured in-situ.

The difficulty with measuring in-line connectors is that they must be connected to cables, and these cables must then be connected to the VNA through a test connector, which is most likely an adapter as well (transitions from the cable type to the VNA connector type). In many cases, the return loss of these test connectors is worse than the in-line connector to be measured. These in-line connectors are often used as cable splices and have no intrinsic connection type of their own. They are designed to make the splice transitions as clean as possible.

Normal return loss measurements of in-line connectors are almost always dominated by the reflections of the test connectors and other defects in the cable, thus requiring more sophisticated techniques. An in-line test configuration is shown in Figure 5.22. The improved methods turn again to time domain and gating techniques to remove unwanted reflections. Similar

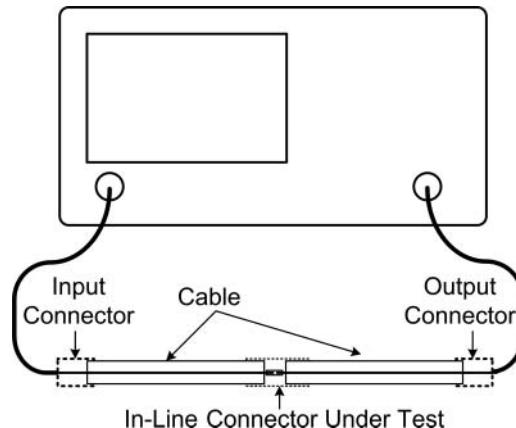


Figure 5.22 Configuration for in-line-connector test.

to the test methods for cables, the in-line connector test method relies on the ability of a time-domain transform to separate the effects of the test connectors from the in-line DUT connector (Society of Cable Telecommunications Engineers, Inc. 2008a).

Figure 5.23 shows a frequency response of an in-line connector, which is inserted in between two approximately 2 m cables; the cable length is chosen to allow good separation

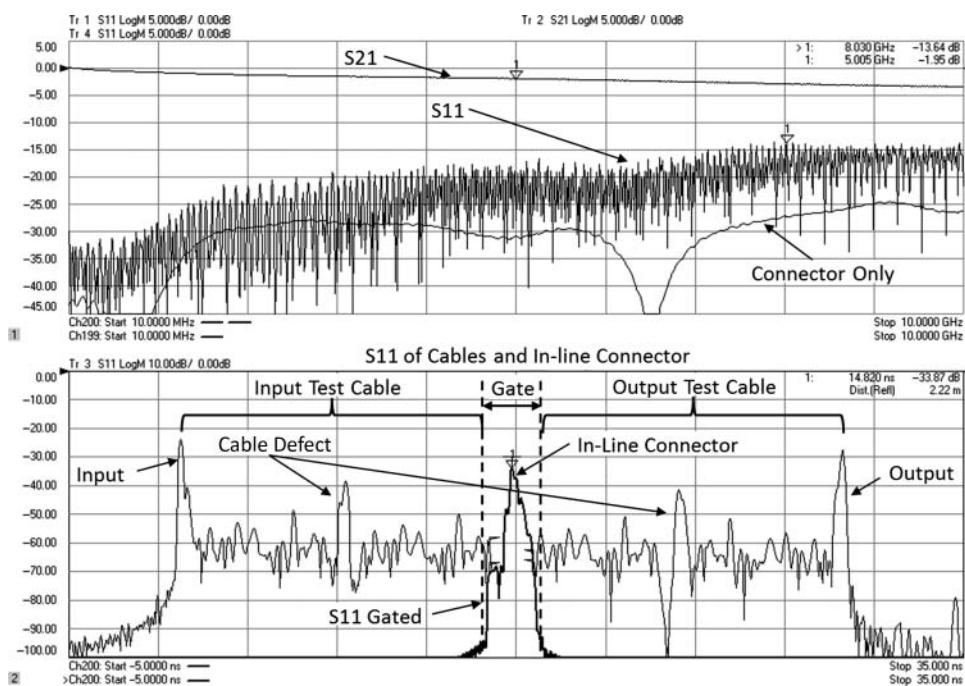


Figure 5.23 Frequency response (upper), time-domain response, for inline connector (lower).

of the responses in the time domain. The upper plot shows the S_{11} and S_{21} response of the overall system of input test connectors, input line, DUT in-line connector, output line, and output connector. Also shown in the upper plot is the trace of just the connector (marked “Connector Only”) derived from a simulation model.

The lower plot shows the overall time-domain response from which the in-line connector response can be discerned in the middle. Knowledge of the approximate position and length (delay) of the in-line connectors helps in setting the gates and in recognizing the in-line connector separately from other defects in the cable (marked as “Cable Defect”). One thing to note from the time-domain response in Figure 5.23, before gating, is that both the input and output connectors of the test line have significantly higher return loss responses than the in-line connector. Further, this device is a semi-flex cable, so it has less uniform response (more return loss peaks) than a typical hardline cable. It is important to ensure that the gates do not include cable reflections from other cable defects (as illustrated in the lower window of Figure 5.23), so care should be taken to ensure that the test cables have reasonable good-quality responses in the region near where the in-line connector is applied. A common practice is to first measure a length of cable equal to twice the length of the test cable to check the quality. This cable is then cut in a region of low reflections, and the in-line connector is applied. The time-domain gated response is also shown in the lower plot.

The gate is applied to determine the in-line connector response, but the response will be lower (showing incorrectly a better return loss than expected), because of the loss of the input cable and connector. The gated frequency response is shown as the “ S_{11} Gated” trace in the lower window of Figure 5.24, while the upper window shows ungated S_{21} and S_{11} responses. Also shown in Figure 5.24 is the response of just the in-line connector from Figure 5.23 (still marked “Connector Only”).

Finally, a third trace (marked “ S_{11} Gated and Compensated”) shows the response of the gated S_{11} is compensated for the loss of the input cable. In this case, since the DUT connector is placed halfway in the test cable, the masking due to the cable loss in reflection at the halfway point is nearly equal to the S_{21} transmission loss, since an S_{11} measurement will travel forth and back through the one-half input cable. Therefore, the S_{21} trace of the upper window is used in the equation editor to compensate for the input loss. Gating may also be applied to the transmission measurement to remove the extra ripples from the input and output connectors (see Section 5.1.3.2). The “ S_{11} Gated and Compensated” is computed as

$$S_{11_GateComp} = \frac{S_{11_Gate}}{S_{21_Gate}} \quad (5.7)$$

If the DUT is exactly in the middle of the test cable, the gate center-time value for transmission and reflection are the same. This final response is almost identical to the ideal in-line “Connector Only” response, proving the quality of this method. At high frequencies, near the band edge, time-domain gating effects will cause an uptick in the response. As with all applications of gating, it is important to discount the last 5–10% of the frequency response due to gating edge effects, as described in Chapter 4. Therefore, over-sweeping the required frequency range by at least 10% is recommended.

Finally, it is important to note that this method for the most part applies only to return loss measurements. The insertion loss is difficult to ascertain in this method, and it is perhaps better to determine the insertion loss by doing a true insertion test. If the S_{21} of the test cable is measured before the in-line connector is added, then the difference can be determined when

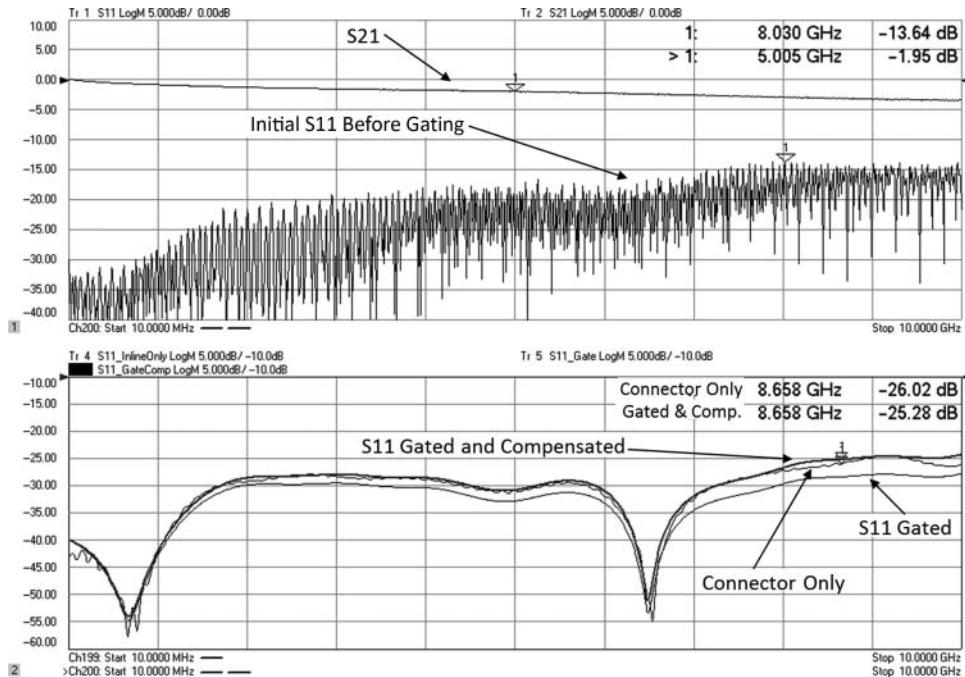


Figure 5.24 Response of the in-line connector, gated and compensated for loss.

the in-line connector is added to look for added loss. In practice, these connectors are nearly lossless, so the in-line S_{21} loss is very nearly

$$S_{21_Inline} = \sqrt{1 - |S_{11_GateComp}|^2} \quad (5.8)$$

This technique for testing in-line connectors and removing the effects of the input and output test connectors can equally well be applied to other embedded component measurements such as surface mount technology (SMT) components on a PC board. In such a case, creating a fixture with long input and output lines will greatly help in separating the effects in the time domain.

5.1.4.2 Structural Return Loss

Structural return loss (SRL) differs from normal return loss in that it measures reflections relative to the average impedance of the cable rather than to some reference impedance (Rowell et al. n.d.). The main reason for making SRL measurements is to look for minute, periodic defects in long reels of very low-loss cable, typically used for cable TV main-line installations, as described in Section 1.8. For many of these systems, there was available some small amount of impedance adjustment in the system, so the absolute value of the cable impedance was not as critical, but variation in impedance or return loss caused substantial problems. As such,

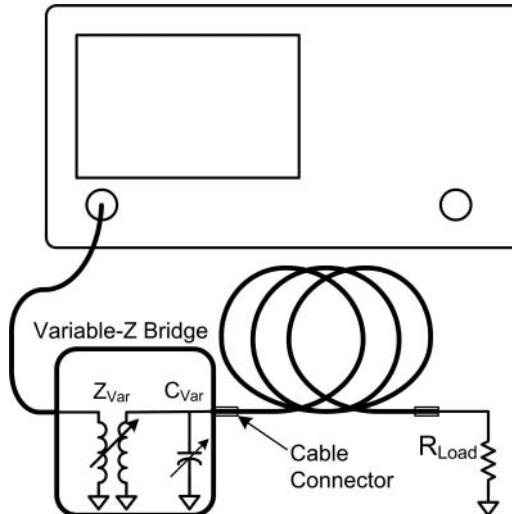


Figure 5.25 Measuring a cable with a variable-impedance bridge.

the SRL was defined relative to the cable average impedance, and the average impedance was specified within some range, typically within plus or minus 1 or 2 Ω of the cable's nominal impedance.

The difficulty in SRL measurements arises from the fact that for long, low-loss cables, small but periodic impedance deviations can add up and create narrow but high return loss spikes. These are often caused by a defect in some part of the manufacturing, often from some rotating spindle that is slightly out of round or has some other defect on it.

Before the use of modern VNAs, the SRL was measured using a variable impedance bridge, illustrated in Figure 5.25. This bridge had an adjustable impedance factor (typically using a variable resistor in one leg) and also included a variable capacitor at the input that canceled another fixed one inside the bridge. Thus, this adjustable capacitor could add or subtract a small capacitance to compensate for the defects in the test connector.

Theoretically, one should also apply a variable impedance load as well to match the far end of the cable to its impedance, but in practice the length and loss of the cable make this unnecessary.

Modern VNAs take a fixed-bridge measurement and use computation to simulate the effects of the variable impedance bridge, using port matching similar to Figure 5.10, with the addition that the impedance of the port can be varied as well as the matching elements.

To investigate how this response appears on a VNA, a circuit created with a 10 mm long, approximately 1 Ω deviation of a cable impedance, every 1 m is simulated, and the data is loaded into a VNA using the built-in S2P file reader. The simulation also modeled a non-ideal input and output connector and made small variations on the size and position of each discontinuity to emulate real-world conditions. A 3200-point sweep and a 201-point sweep were both created in the simulator. The return loss and insertion loss of the 201-point sweep are shown in the upper window of Figure 5.26.

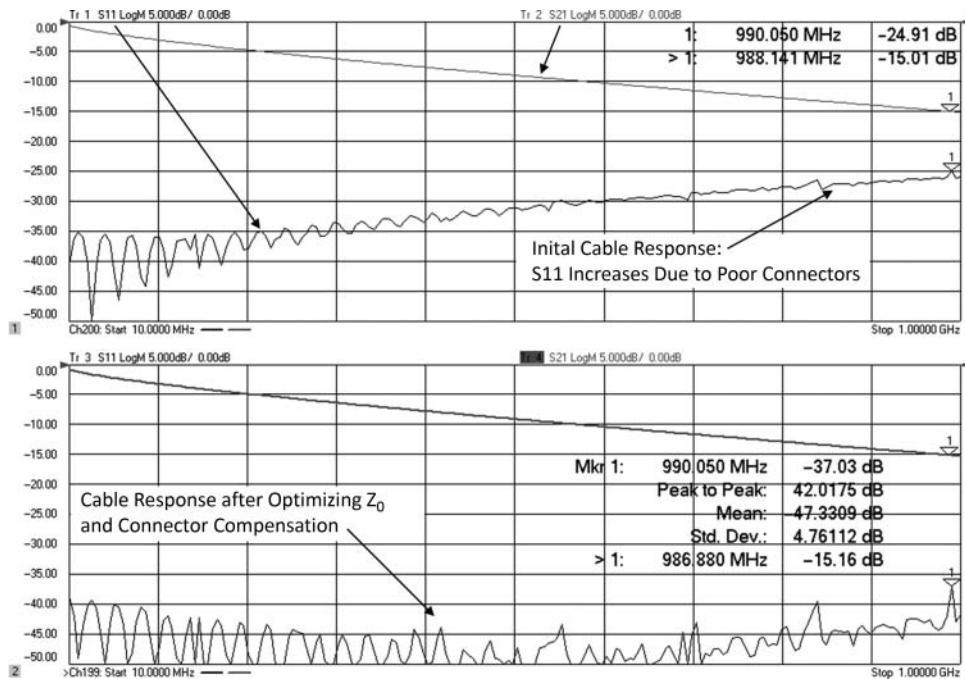


Figure 5.26 Return loss and insertion loss of a long cable; upper is normal return loss and lower is applying a virtual variable impedance bridge.

The lower window of Figure 5.26 shows the resulting SRL after applying the effects of a virtual variable impedance bridge. The first step is to turn on the trace statistics and look at the mean return loss. While monitoring this, the port matching function is used to add either some capacitive or inductive compensation to “lay the trace down” as it is known in the cable test industry. In this case, about -1.1 nH of inductance provides the lowest mean value. Next, some small port extension, the maximum length of which is equivalent to the length of the input connector, is added to further reduce the mean value. Finally, the port impedance transformation is used to both find the average impedance and reduce the mean return loss to its lowest possible value. The impedance at which the return loss is lowest is called the average impedance of the cable. For the cable in Figure 5.26, this value is 76.7Ω .

Finally, the peak of the S_{11} trace is observed to determine the worst-case structural return loss; in this case, it appears to be -37.03 dB . For these types of cables, a typical specification is -32 dB SRL, so it would appear that this cable would pass. However, from a fundamental analysis of the cable, one can determine that the resolution needed to see any and all peaks in the SRL is much smaller than the 201-point sweep above provides.

The cable in question has a length of approximately 500 m and a velocity factor of about 0.9. From this, the frequency at which this represents one-half wavelength can be determined as

$$\frac{\lambda}{2} = \frac{V \cdot c}{2\Delta f} = 500\text{ m} \therefore \Delta f = \frac{0.9c}{2 \cdot 500} = 270\text{ kHz} \quad (5.9)$$

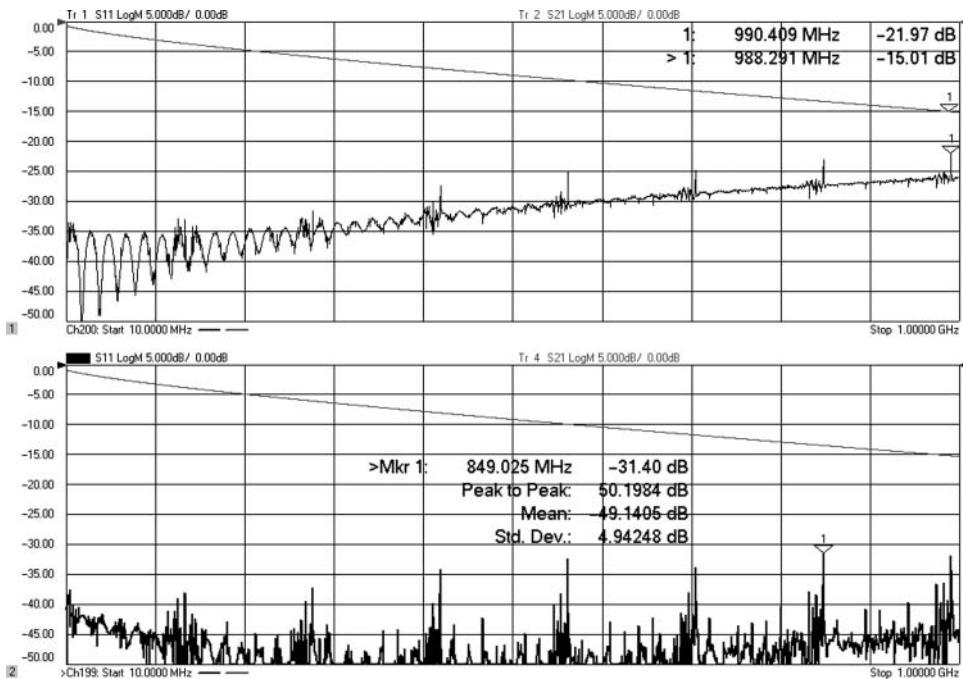


Figure 5.27 SRL measurement at 3201 points; upper trace is before connector compensation.

For the 1 GHz sweep shown, around 3700 points would be required to ensure that a measurement point occurs every half-wavelength. The loss of the cable diminishes the summing of periodic discontinuities, and they will not be spaced precisely periodically, so a somewhat lower number of points may be used. If the measurement is repeated with 3200 points, a different picture of the SRL response appears, as shown in Figure 5.27. Now the very narrow responses usually associated with SRL problems become clear. The adjustments are made to the input connector compensation and the cable impedance with nearly the same results, although the best match for cable impedance was $76.6\ \Omega$. In fact, this matches nearly with the expected value of $76.5\ \Omega$ for this cable from the simulation.

Of course, the most important result is the actual value of SRL, in this case -31.4 dB . This cable would be just out of specification due to the periodic disturbance. The periodic impedance error here was $0.9\ \Omega$, over a 10 mm region, occurring approximately every 1 m. This period corresponds to a frequency of about 140 MHz given the velocity factor of the cable, and this is exactly the repetition frequency where the SRL response is seen in Figure 5.27.

Using a variable bridge with older VNAs, similar results may be achieved with the exception of connector compensation. A consistent problem with the older variable bridge was difficulty in getting the trace to “lay down.” This is because the electrical delay between the variable capacitance in the bridge and the mismatch of the test connector meant that perfect compensation was never possible and resulted in many cases of failed cables or repeated re-measurements with different test connectors. With the advent of modern compensation

techniques, the physical variable impedance bridge has been almost entirely replaced. The first implementation of this technique was found in the HP8711 VNA with its option 100 SRL measurement code. However, these techniques can be applied with nearly any modern VNA.

5.1.4.3 Cable Impedance

While the measurement of the average impedance of a long cable was described in the previous section as part of the SRL discussion, the techniques used are not appropriate for shorter cables where the far-end load will have a more dramatic effect on the result. For shorter cables, the measurement of impedance becomes somewhat problematic as it is seldom realized that impedance of cable is a two-dimensional characteristic. In Chapter 1, the details of transmission lines are discussed, and it is clear that any physical transmission line (one with loss) has an impedance that varies with frequency. Further, any real cable may have perturbations along the length so that the cable impedance varies as a function of distance along the cable as well. So any discussion of cable impedance should really be defined at a particular point in the cable and at a particular frequency. However, almost no one specifies cable impedance in that way, and almost everyone desires to have a simple, single number when referring to a cable.

Thus, cable impedance is often defined as an average over frequency with variation only along the distance of the cable or even as just as an average impedance over both frequency and distance (Society of Cable Telecommunications Engineers, Inc. 2008b). One traditional way of measuring a cable's impedance is using time-domain reflectometry, and a VNA provides a modern version of this method. To illustrate the measurement technique, consider a cable with a stepped impedance every 10 cm, with a less than ideal input connector. An S2P file of a cable simulated with these attributes loaded into a VNA is shown in Figure 5.28. The upper plot shows the normal response, and the lower plot shows the response after applying connector compensation, as described in Section 5.1.3.1.

Note the difference in the value of Marker 1, which represents about a 0.04Ω difference in the apparent value due to the effect of the less than ideal input connector. Also note the frequency response is flattened with the connector compensation in the lower plot. The time-domain response clearly shows the impedance steps in the cable. The values of the Marker 1 are shown in reflection coefficient. Some VNAs allow the marker format to be set differently from the trace format; here Marker 2 is set to show $R + jB$, which provides the impedance in the case of a time-domain trace, which is always pure-real in a low-pass transform. From these plots, the impedance as a function of delay down the cable can be directly determined, by moving the marker along the line.

Using the built-in equation editor feature, one can also compute the impedance as a function of delay down the cable directly by converting the reflection coefficient to impedance. The conversion is simply

$$Z = Z_0 \frac{1 + S_{11}}{1 - S_{11}} \quad (5.10)$$

and is applied to the upper time-domain trace, as shown in Figure 5.29. One anomaly of using such an equation editor function is the time-domain transform occurs on the raw data before the equation editor math is applied. Since converting to impedance is not a linear function, the conversion must be performed after the transform is performed. Thus, the transform is

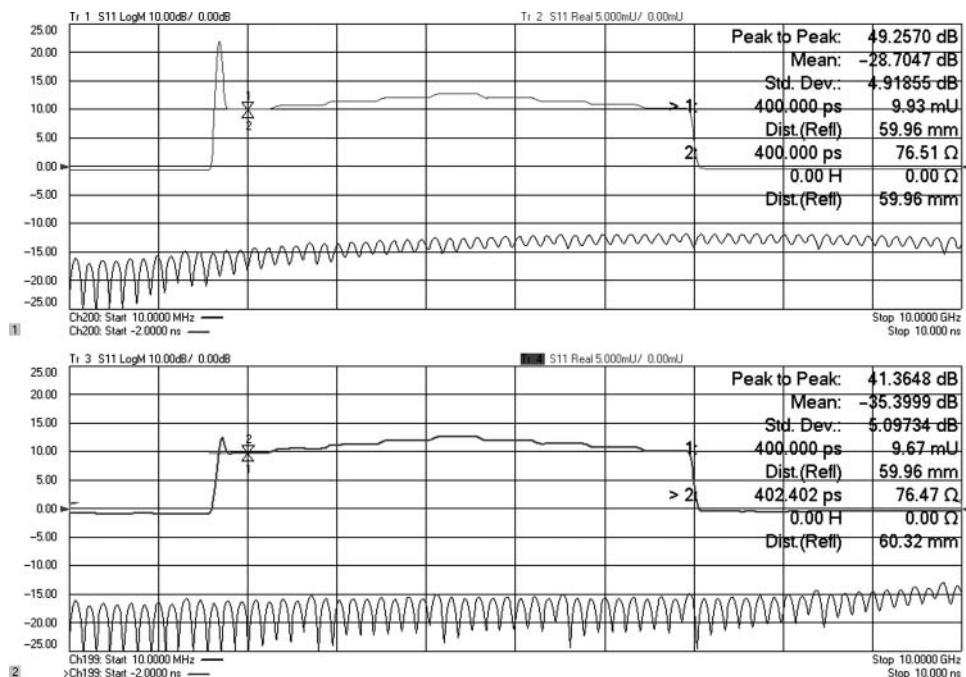


Figure 5.28 Frequency and time domain of a cable with stepped impedances. Upper plot is normal and lower plot is with connector compensation.

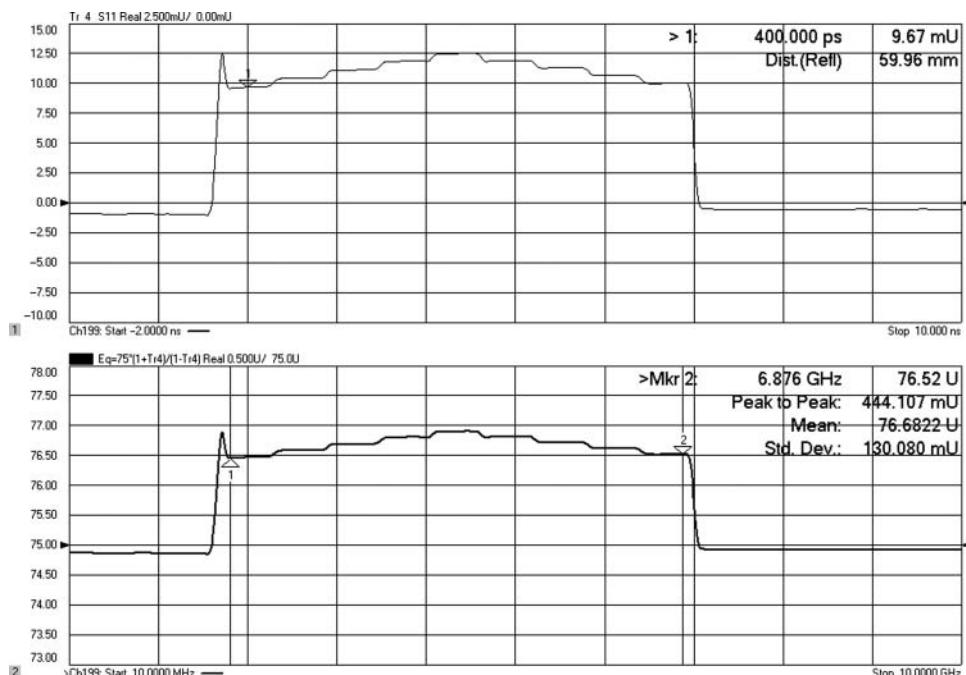


Figure 5.29 Cable impedance as a function of delay down the cable. Upper window shows linear reflection coefficient; the lower window has is impedance in Ω as the Y-scale.

performed in the upper window (trace 4), and then the equation is applied in the lower window. An unfortunate consequence is the x-axis retains its frequency label rather than having the time label. Trace statistics can be applied to a segment of this converted measurement to produce a mean value. In the case of the lower window, the mean is computed between the region between Marker 1 and Marker 2, showing a mean value of 76.68Ω . This is close to the value expected from the design used in the cable simulation.

One caution, however, with using the time-domain transform is that because the zero frequency (DC) value is not directly measured in the VNA, it must be extrapolated. The DC value represents the value of the transform at some distant time and should be equivalent to the termination value plus the DC loss of the cable. This value is set as described in Chapter 4. However, both alias effects (which occurs if the time-domain response does not return to zero inside the alias free range) and extrapolation error effects can cause the apparent impedance at the time just before zero to be different from the expected value, which is usually the system reference impedance. In such a case, one should measure the apparent reference impedance on the time-domain trace and accommodate for any difference in the measurement results.

5.1.5 Cable Length and Delay

A common question when measuring a cable is what is its length or delay. Usually this entails measuring the group delay of the cable, although time domain can also be used for a similar measurement. For long cables, a key issue is under-sampling the phase response, which means that the frequency spacing is such that more than a 180° phase change occurs between measurements points. Figure 5.30 shows the (an admittedly complicated looking) result of measuring a 10 m long cable with four different choices of number of points over the same frequency span. Since the length of the cable is already known, it is straightforward to compute the frequency spacing needed using the formula from Eq. (5.9), which yields about 10.5 MHz for a cable with a 70% velocity factor. In the first case shown in the upper-left window, an 8 GHz span is measured with the default 201 points for a frequency spacing of about 40 MHz, and it is clear that the response is under-sampled. In each window, the phase response and group delay are shown, and in the upper left the phase looks abnormal, and the group delay is negative. The upper-right window shows the result when the number of points is increased to 301 points. In this window, the phase looks somewhat normal and the delay looks positive, but the delay number does not correspond to the expected delay of a 10 m cable, which should be around 47 ns, with a presumed velocity factor of 0.7. The 9.26 ns delay must be wrong, but the under-sampling is such that the phase slope is negative and the delay is positive. The lower-left window shows the response for 401 points. Here the delay is again negative. Finally, the lower-right window shows the response for 801 points. Here the delay is positive again, and the delay number is near the expected value of 47 ns. The frequency spacing here is just under 10 MHz per point, and so we would expect this result to not be aliased.

The plots on the left illustrates caution in increasing the number of points: if the number of points is simply doubled and the phase response is under-sampled sufficiently, doubling the points can result in the same incorrect delay indication. Therefore, it is best to increase the number of points in non-uniform increments to illuminate the phase response with various spacings.

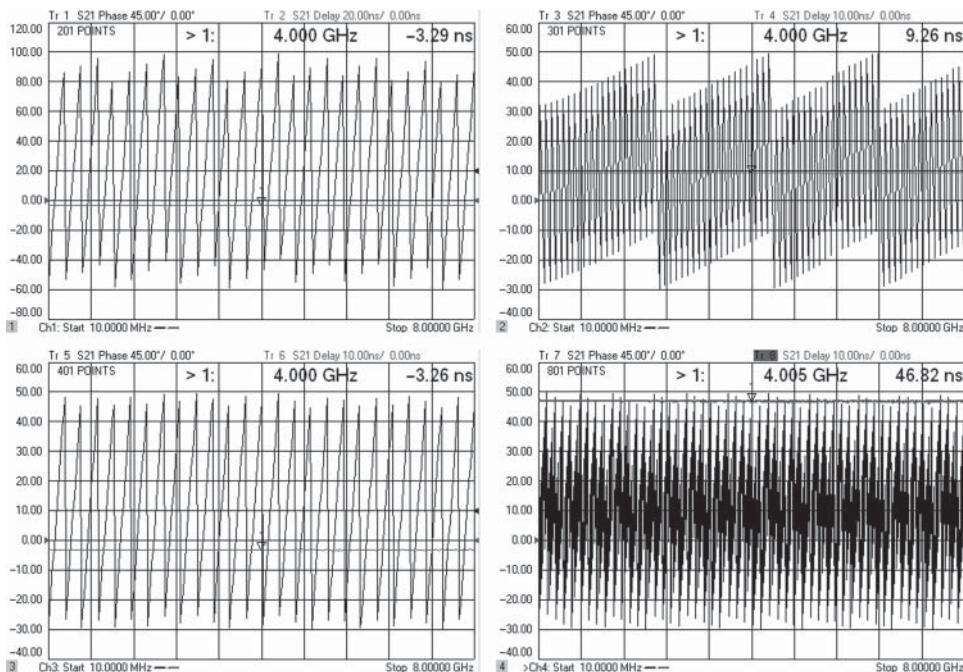


Figure 5.30 Examples of changing the phase sampling for a long cable; only the lower-right plot is correct.

5.2 Filters and Filter Measurements

Of all the devices that absorb measurement test time, filters most likely lead the field. High-performance filters, those used in cellular-phone base stations and satellite multiplexers, must be tuned to achieve the desired performance. The product tolerances in manufacturing filters do not come close to being controlled enough to create these filters without substantial tuning. The tuning process on a complex base-station filter might take up to an hour, and the tuning on a satellite multiplexer might take several days or weeks. During all this time, the careful measurement of the filter's S-parameters must be maintained. The very low loss of these filters means that any residual error in the VNA calibration will accentuate the ripple and mismatch of the filter. Often, these same filters have extremely high stopband attenuation requirements; this means that very high dynamic range is required to accurately assess the isolation of the filter. During the tuning process, which typically involves an operator manually turning a tuning screw while watching a result on a VNA screen, the measurement speed must be fast enough to provide for real-time tuning, usually considered 10 updates per second or faster. Specific aspects of filter measurements are discussed in the following sections.

5.2.1 Filter Classes and Difficulties

The wide range of filter applications yields many different filter types that can be roughly classified as to their performance and application, and from these classifications, common

Table 5.1 Filter classification by application type

Filter classes/ applications	Technology	Loss (dB)	Match (dB)	Complexity/ tuning	Other attributes
Base station transmitter (Tx)	Silver-plated, larger air-coupled resonators	0.1–1	20–26	Complex/tune resonators, couplings	Orders from 6 to 20, deep notches near pass band
Base station receiver (Rx)	Silver-plated, larger air-coupled resonators	0.1–1	20–26	Complex/tune resonators, couplings	May include LNA integrated into filter, very high Tx isolation
Base station duplexer	Silver-plated, larger air-coupled resonators	0.1–1	20–26	Complex/tune resonators, couplings	3-port combination of Tx and Rx
Satellite multiplexer	Highest quality silver plating, tuned resonators and couplers	0.1–1	20–26	Very complex/tune resonators, coupling, multiplexer	High port count (up to 20 or more), multi-stage adjustment
Handset	Ceramic coupled resonator	1–3	10–15	Low/one-way	Low cost
RF subsystem	Microstrip coupled line on PCB	0.5–3	10–20	Simple/no tuning	Often integrated into RF PC boards
RF subsystem	LTCC on ceramic	0.5–3	10–20	Simple/no tuning	Often part of RF system on chip
Channel	SAW	1–10	5–15	Moderate/no tuning	Very narrow, very high order

needs, difficulties, and measurement methods can be described. A sorting of the filter classes is shown in Table 5.1. These can be roughly sorted into three main classes: fully tunable filters, trimmable (one way adjustable) filters, and filters fixed by design.

Tunable filters often have the most strenuous requirements of loss, isolation and match, which is why they must be made tunable to achieve the difficult specifications. These are used in high-power applications where any loss must be minimized, and in low-noise applications where a receiver's noise figure is directly degraded by filter loss. These also have high isolation requirements to ensure than in a transceiver application, the high transmitter (Tx) power is isolated from the low-noise receiver (Rx) path. These filters are often found in a combined form as a diplexer.

5.2.2 Duplexer and Diplexers

A common question is “What’s the difference between a duplexer and a diplexer?” While there is no official answer, one can understand the difference in nomenclature by appreciating the difference in application. A diplexer is a filter that combines signals of two different frequency ranges at two different ports, into a single combined port, and isolates each port from

the other. A multiplexer is a diplexer that combines more than two ports into a single port. Most commonly, a diplexer is used to channelize a receiver or a transmitter.

A duplexer allows a transmitter and receiver to operate at the same time using the same antenna. So, a diplexer is used as a duplexer when it combines a transmitter and receiver such as in a base station.

5.2.3 Measuring Tunable High-Performance Filters

Making measurements on tunable filters means setting up a VNA in a way that allows high-speed updates while maintaining the desired performance. Typically, this is a passband only measurement, so high isolation is not generally needed during the tuning process. In such a case, wide IF bandwidths can be used. However, many VNAs change the sweep type (from stepped to swept source) depending upon the IF bandwidth, with wide bandwidths using swept source by default. The fast sweep times needed for filter tuning are accomplished in some VNAs using continuous swept frequency (as contrasted with stepped frequency sweeps), and this can in some cases cause IF delay problems, as described in Section 5.1.2.1. For the most part, these effects occur only with narrow filters that by nature must have long delays. Typically this is seen only in crystal filters or SAW filters. Thus, the IF delay errors seldom affect tunable filters.

In measuring tunable filters, there are many choices to be made in the configuration of the test and calibration that can affect both the quality and speed of the measurement; some of the common attributes are described here:

IF bandwidth: This is the single most important setting affecting the trade-off between speed of measurement and the noise in the results. For most tuning measurements, the IF bandwidth can be set quite high, but with some limits. At wide bandwidths, other overhead such as computing and displaying the trace results, and band switch time and retrace time will be a larger percentage of the overall sweep cycle time. At some wide bandwidths, the data-taking time is swamped by the other overhead times and increasing the IF bandwidth further does not result in any substantial increase in overall cycle speed. A typical value for tuning is between 10 and 100 kHz IF BW. Finally, the frequency resolution is limited by the IF BW, in that the frequency response of a filter is “smeared” across the IF BW. One cannot use a wide IF BW on a narrowband filter. A narrow IF bandwidth might be required to accurately assess the corner frequency of a filter.

Number of points: Like IF BW, the number of points for a sweep directly impacts the cycle time unless the number of points is very small, where other overhead will swamp out any improvement from further lowering the number of points. For many applications, the number of points is set by the required resolution at the band edges, but the interpolation of the trace response is quite good in modern VNAs, and marker or limit line values are computed off the interpolated result. Usually, this interpolation is good enough to allow reducing the number of points to a reasonable value over the bandwidth. A typical value for most tuning applications is 201 or 401 points.

Sweep mode: As discussed in the section on IF delay, changing from swept to stepped mode may have a substantial effect on the sweep rate. However, modern VNAs have fast synthesizers so that for IF BWs below about 10 kHz, the effect is quite small and stepped mode may be used without a dramatic increase in cycle time. Some VNAs provide two forms of step sweep: standard and fast step. In fast step, some of the settling times associated with the source automatic level control (ALC) loop and the individual receivers used in a non-ratio mode are decreased or eliminated. For almost all ratioed measurements, such as S-parameters or gain, the errors from settling are removed by the ratio process so there is no ill effect in reducing or eliminating these settling wait times. But for cases where absolute power control is important, such as amplifier test, the normal stepped mode should be used to avoid power settling issues.

Calibration type: For most high-performance tunable filters, a full 2-port calibration is necessary to remove the effects of the test system load match from the return loss measurement. Older legacy VNAs, such as the 8753, provided a special mode that would update the reverse (S_{22}) sweep only occasionally at a user-specified interval, such as every 10 forward sweeps. However, the improved dynamic range of modern VNAs allow low-noise measurements at wide IF bandwidths so that even full 2-port calibrations can proceed with near real-time speed.

5.2.3.1 Filters with Very Low Loss and Well Matched In-Band

Measuring filters with very low loss that are well matched for the in-band portion, requires the same techniques discussed in Section 5.1.1. Careful attention to using good calibration techniques and good cables are required. Because filters have a long delay for their physical size, they can have a lot of phase change in the pass band, which can cause extra ripples in the response if the VNA system source and load match are not characterized properly; in particular, since most filters are non-insertable (they have either different connector families on each port or the same type and sex on each port), it is important to use either the unknown thru calibration method or, for older VNAs, an adapter removal calibration method. Using the traditional SOLT calibration and ignoring the thru delay is a common cause of poor calibration that can make filter tuning difficult due to poor load match correction.

5.2.3.2 Measuring Filter Return Loss

The return loss of a filter is almost always used as the principle way to tune a filter, even though the key specification of a filter is S_{21} insertion loss and isolation. However, the return loss is more sensitive to tuning variations, and obtaining a good return loss provides better system performance and nearly always guarantees a good insertion loss result.

When testing for return loss, it is common to use a limit line to set a visual pass/fail criteria. Another convenient technique is to place a marker at the passband edges to see the actual values while tuning. A third, less well known but very convenient technique is to use a feature

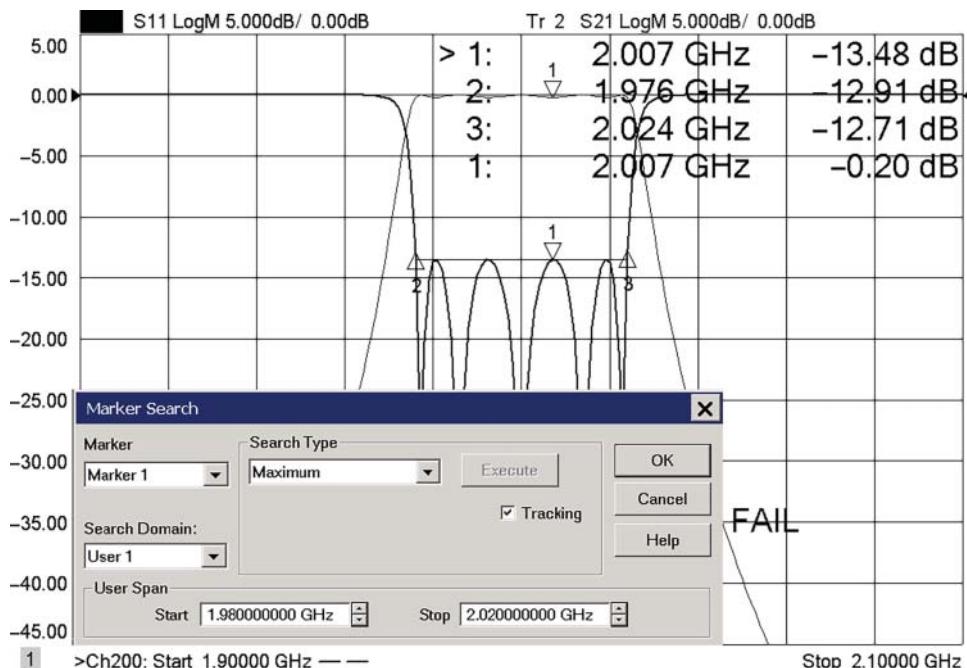


Figure 5.31 Testing S_{21} and S_{11} of a filter, using marker tracking to find the worst-case S_{11} .

found in some VNAs to have a marker track the worst-case (maximum) return loss value in the passband. Often, a technician is instructed to tune for “best possible match” even after the filter passes the minimum specification. Using a technique of tracking, the worst-case point makes this quite convenient to look for the best case. This test scenario is set up in Figure 5.31. Marker 1 is set to track the maximum value of return loss within a narrow region of the passband. Markers 2 and 3 track the edges of the filter.

In trace 2, Marker 1 is set to track the minimum value of S_{21} over the same region as in the return loss trace; it’s common to show both passband insertion loss and return loss at the same time when tuning filters. While tuning this filter, these marker values will update on every sweep and provide a convenient way of tracking the filter’s performance. The limit-line indicator will change from fail to pass when the filter is tuned below the limit-line; usually failed regions of the test trace are highlighted in a red color.

5.2.4 Measuring Transmission Response

The transmission response of filters represents a measurement of the fundamental purpose of a filter. The response has two key attributes: the passband insertion loss and the stopband isolation. Band-pass filters have both an upper and lower stop band, though typically the stopband requirements are much more stringent on one side of the filter than the other, especially in filters used for duplexers in communications systems.

5.2.4.1 Passband Measurement

Measuring a filter in the pass band is similar to measuring low-loss cables and connectors, and many of the details described in Section 5.1.1 should be followed for low-loss filters. One attribute of filters that distinguishes them from other devices is that the match in the passband is not, ideally, zero, but rather the reflection is determined by the number of sections, the ripple in the passband, and the passband isolation. In most cases, the filter response is a trade-off between accepting poorer return loss in exchange for sharper cutoff frequency. Thus, while most filters are tuned for a good return loss, even an ideal filter (one that operates precisely as designed) will not have zero reflection across the passband. In contrast, a good cable or connector will have a nearly zero reflection coefficient.

If a filter is nearly loss-less, then the ripple in the transmission is directly related to the peaks in the return loss through the well-known equation

$$|S_{21}|^2 \leq (1 - |S_{11}|^2) \quad (5.11)$$

On most VNAs, the bandwidth of the filter can be automatically determined using marker search functions. Most such functions allow the bandwidth level, typically -3 dB, to be specified. For an equal-ripple filter, the bandwidth should be defined as the ripple level. The search function finds the maximum value and computes the bandwidth from that level.

After finding the bandwidth, markers are placed on the upper and lower corner frequencies (Markers 2 and 3 in Figure 5.32), at the center (Marker 4) where the loss of the filter is reported, and at the position of the maximum value (Marker 1).

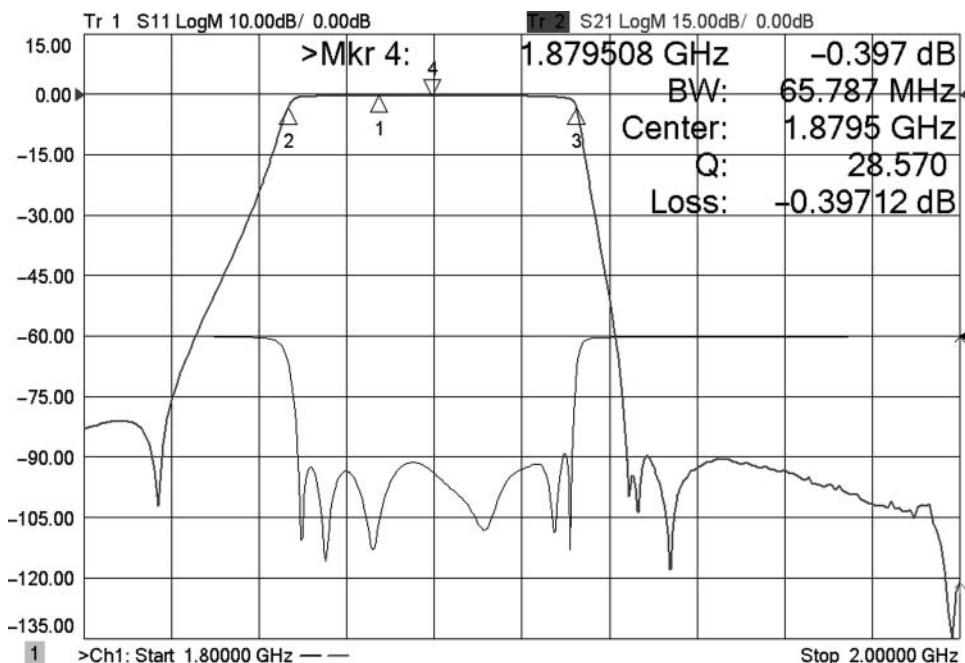


Figure 5.32 Using the marker search function to find a filter bandwidth.

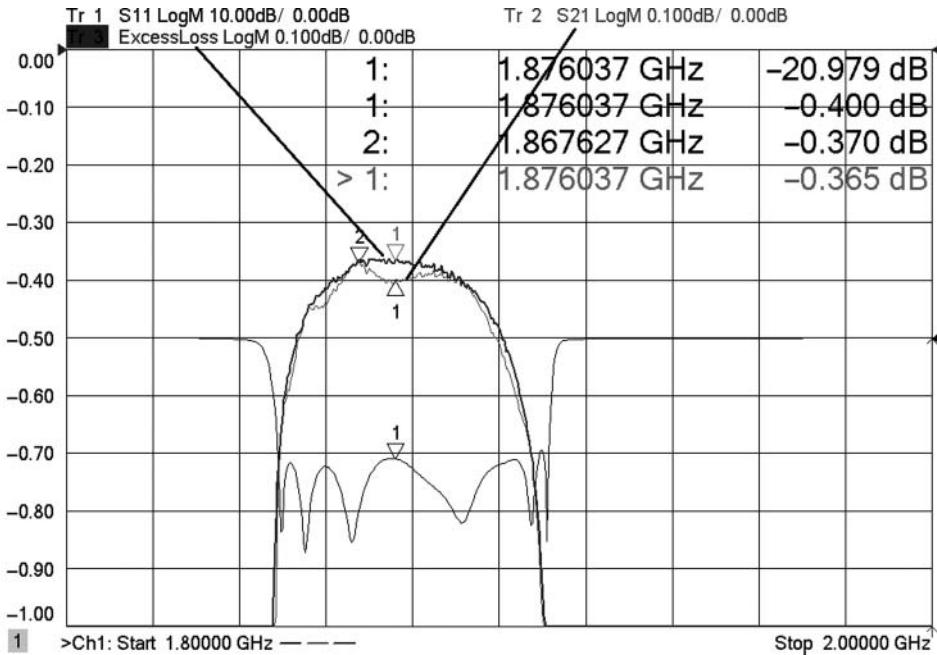


Figure 5.33 S_{11} , S_{21} , and excess loss of a filter.

5.2.4.2 Excess Loss

Though not a typical specification, a useful figure of merit is the excess loss of a filter, defined as

$$L_{\text{Excess_dB}} = 20 \log_{10} \frac{S_{21}}{\sqrt{1 - |S_{11}|^2}} \quad (5.12)$$

This represents the energy absorbed by the filter (or any other passive device), in excess of the mismatch loss. Figure 5.33 shows the S_{21} and S_{11} of a filter and, using the equation editor, the excess loss.

This loss is mostly independent of filter tuning and represents the best-case insertion loss of an ideally tuned filter. It is sometimes useful to see the excess loss when tuning a filter to understand if the desired insertion loss is achievable with tuning. In some cases, degradation of manufacturing processes, such as poor plating or bad soldering, can lead to an unexpected increase in excess loss, creating a filter that cannot be tuned for low-pass insertion loss, even if the return loss is properly tuned. In such a case, it is useful to find this out before a substantial amount of time has been wasted in some fine-tuning process.

Because even ideal filters have reflections in the passband, the transmission response is sensitive to the VNA port mismatch. This can lead to larger mismatch error in the filter response as the additional uncertainty is S_{11} times residual source-match, and S_{22} times residual load match. Thus, for filters with tight tolerances, it is important to use good calibration techniques.

5.2.4.3 Limit Testing for Transmission

Transmission responses are usually tested against a minimum limit line value, but in some VNAs, the limit testing is performed only on discrete data points, and the actual limits might lie in between data points. For example, if a filter is tested across a 60 MHz span, with 201 points, the point spacing will be every 300 kHz. If a limit line centered on the filter center frequency and set to have a span of 50 MHz, then the limit edge points will lie exactly 25 MHz above and below the filter, but the discrete data points of the filter will be at a span of 49.8 and 50.4 MHz, just inside and just outside the limit. If the point just inside the limit passes and the point just outside fails, the limit will pass, even though the displayed traces may cross the limit line. In such a case, the number of points must be increased or segmented sweeps used to ensure that the limit test occurs exactly on the measurement point. Figure 5.34 shows an example where the S_{11} measurement points do not match the edges of the limits, thereby causing a pass indication when the filter actually fails the true limit criteria.

In particular, Markers 1, 2, and 4 are positioned on discrete measurement points, where Marker 3 is interpolated between Markers 1 and 2. It is clear that the trace breaks the limit line, but the limit test passes as there are no measured points outside the limit. Doubling the number of points so that Marker 3 lies on a measured point will result in the limit test failing.

5.2.4.4 Evaluating Ripple Using Statistics

One key figure of merit in filters used for communications systems is the amplitude deviation or ripple in the passband; this is sometimes called *filter flatness*. This ripple is usually

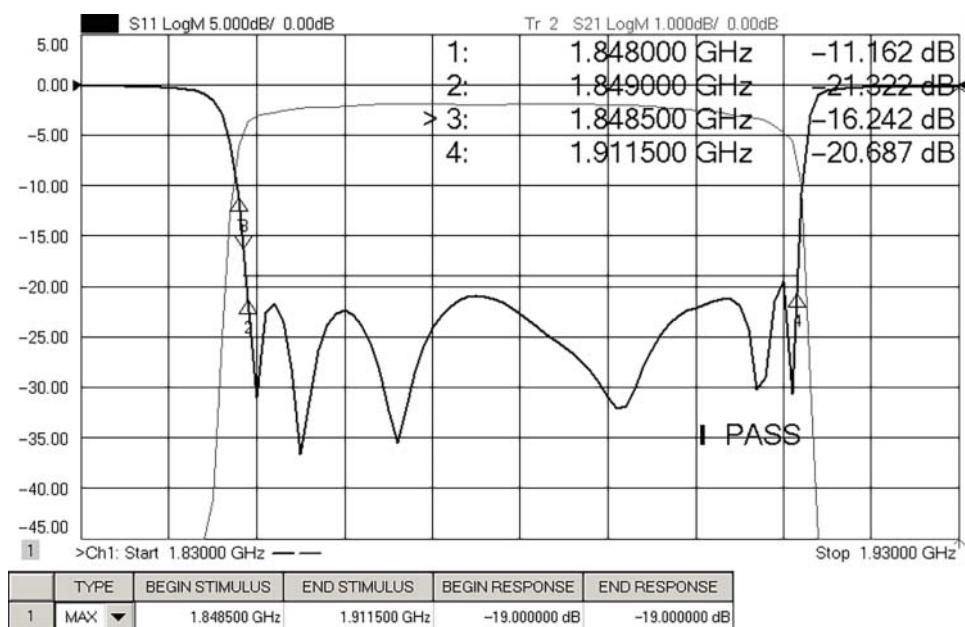


Figure 5.34 Limit testing when the measurement point does not equal the limit edges.

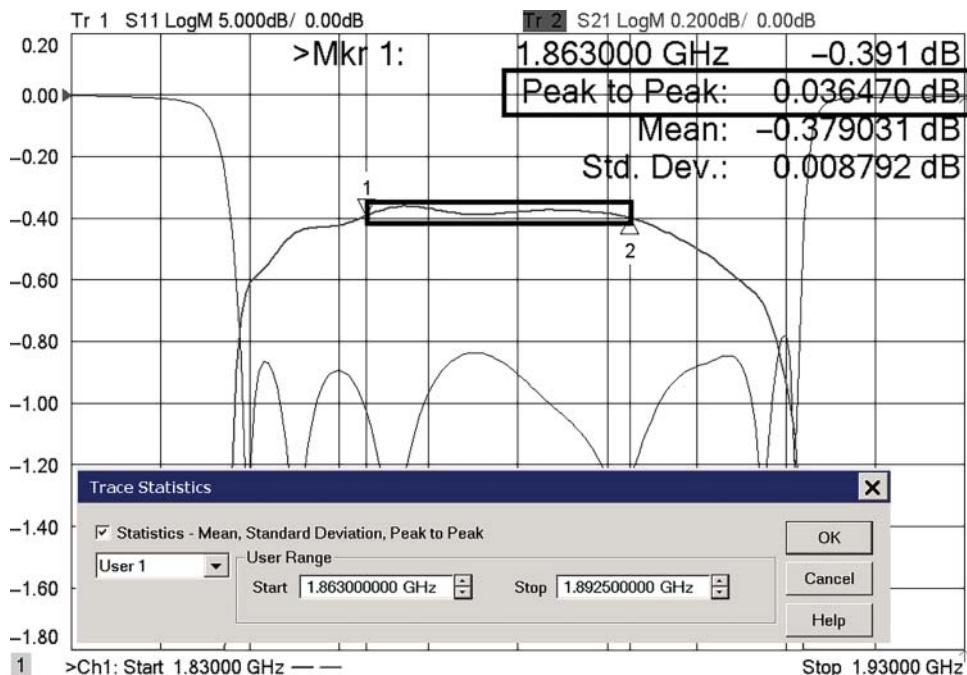


Figure 5.35 Using trace statistics to report the peak-to-peak ripple in the passband.

caused by the reflection of the filter and is one of the design parameters. In many cases, the ripple of the passband will be acceptable if the return loss is properly tuned, but in some cases the return loss specification is less important than the filter flatness. The passband ripple is easy to characterize using the trace statistics function and the statistics user range feature of some VNAs. The peak-to-peak value of the trace statistics directly shows the ripple in the passband, as demonstrated in Figure 5.35. Other VNAs allow using a ripple limit line, which floats at the mean value of the trace and shows whether the ripple is in or out of specification.

In some systems, the frequency flatness response can be compensated for using equalization, and so slope in the loss of a filter (or cable) can be removed, but amplitude deviation cannot. In this case, the flatness cannot be simply characterized as the peak-to-peak value over a region but is defined as deviation from a linear response. In such a case, the slope of the amplitude response should be removed. There are many methods for fitting a straight line to an amplitude response, and perhaps the most common is a least-squares fit. In some modern VNAs, least-squares approximations are provided as a post-processing function (as an imported function of the equation editor, for example), and both the flatness (deviation) and the slope (sometimes called *tilt*) of the response can be directly displayed. Figure 5.36 shows the flatness (around a 0 dB reference) and slope (called the *tilt parameter*) of a filter in the passband, as well as the original S₂₁ and the best-fit line. The tilt parameter is a single value that is equal to the slope of the best-fit line and, in the case of the figure, is the difference between Marker 1 and Marker 2 on the best-fit trace.

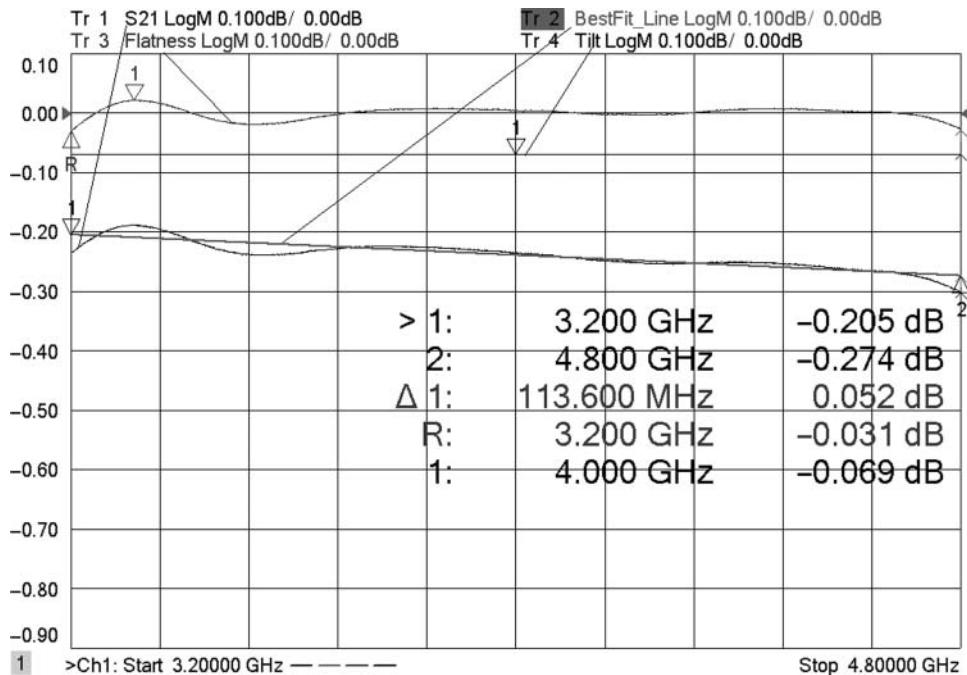


Figure 5.36 The flatness and slope of a filter are displayed, along with the S_{21} response.

In some cases, the flatness computation can be restricted to a small portion of the underlying trace, allowing a single trace or channel to provide both in-band (flatness) and out-of-band (isolation) responses. These functions are commonly needed for amplifiers used in cable-TV systems where the slope or tilt is set to compensate for cable loss, and the amplitude ripple after the slope compensation must be constrained to a small value to ensure a high quality of service.

5.2.5 High Speed vs. Dynamic Range

In the transmission measurement of filters, there is a direct trade-off between the speed of measurement and the dynamic range or noise floor of the measurement. When tuning filters for passband response, the dynamic range is not usually of a concern, as the tuning is primarily performed on the passband response. But some high-performance filters have transmission zeros (or notches) in the S_{21} response, and in some occasions, they are required to be tuned. This typically occurs in coupled-resonator filters with an adjustable cross-coupling element. The strength of the cross-coupling sets the position and depth of the S_{21} transmission zero or null. In such a case, real-time measurements are required for both the passband and a portion of the stop band of the filter to facilitate tuning.

The conflict occurs between measurement cycle time and dynamic range due to the effect of the measurement IF bandwidth. Wider bandwidths provide for high-speed measurements

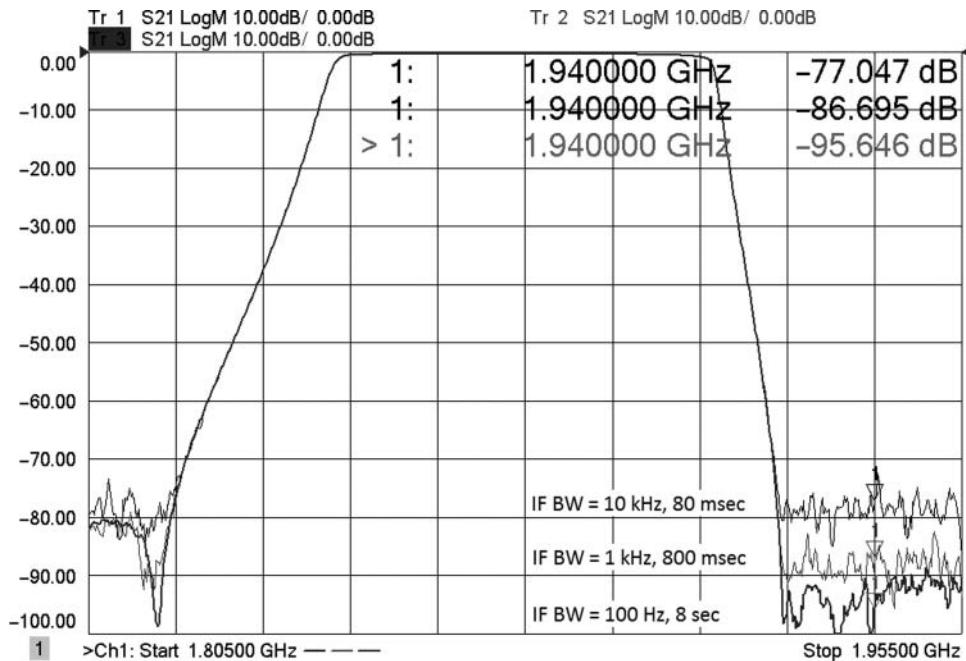


Figure 5.37 Transmission response with three different IF bandwidths and three different measurement speeds.

but have a higher noise floor, simply due to the wider bandwidth allowing more noise into the measuring receiver. Usually a transmission measurement has three regions of interest: passband, stopband high isolation (from the transmission zeros), and other stopband regions. The insertion loss in the pass band is quite low, so the signal-to-noise is very high and a wide IF bandwidth may be used. Over much of the stop band of the filter, only moderate dynamic range is needed, and only a moderately narrow IF bandwidth is required to obtain an adequate noise floor. However, over some of the regions of the transmission zero, very high dynamic range is needed, which requires narrow IF bandwidths. Figure 5.37 shows the result of testing the transmission of a filter with three different IF bandwidths, 10 kHz, 1 kHz, and 100 Hz.

The figure clearly shows the improvement in noise floor using the narrower IF bandwidths, but what is not normally apparent in the plot is the change in sweep time that occurs. For the trace shown and using a full 2-port calibration, the measurement cycle time is 80 ms for 10 kHz, 800 ms for 1 kHz, and 8 s for the 100 Hz IF bandwidth. The noise floor goes down 10 dB for each 10 \times decrease in IF BW, while the measurement cycle time goes up 10 times. For a linear sweep covering the full transmission range, using an IF bandwidth narrow enough to see the transmission zero will cause the sweep to slow down so much that it is completely unusable for filter tuning.

The VNA noise floor may also be reduced by increasing the source power. Modern VNAs have very linear receivers, but the receivers of many older VNAs would compress at maximum source power (up to 0.5 dB compression at max power or more) causing the passband insertion loss readings to be in error. Such compression, while not very consequential in the stop band,

may cause many filters to fail their passband specifications, so such a high power level could not be used on a trace measuring both the passband and the transmission zero region.

5.2.5.1 Segmented Sweeps

Most VNAs provide a convenient solution to the problems described previously, which can be succinctly stated as follows: in the passband region, lower power and wider IF bandwidths can be used, and high point density is needed to ensure the band edges are properly measured, but in the stop band region, very narrow bandwidths and high source power levels are required to obtain a low enough noise floor. What is needed is a way to provide both, and such a method is found in the segmented sweep feature.

The segmented sweep feature originally provided a way to have different point densities over different regions of the measurement sweep. For a filter, the passband could have a high point density, and the stop band regions could have lower point density, with points placed right at the band edges to ensure proper evaluation of the limit tests. Starting with the HP 8753 (and at the insistence of a large Swedish cellular-base-station company), the segment options included the ability to assign different power levels and IF bandwidths to individual segments. Now a sweep could be defined that had high power in the stop band regions, to give good dynamic range, lower power but wide bandwidths in the passband regions, and a few points with very narrow bandwidths where the transmission zeros were to be tuned. An example of a segmented sweep is shown in Figure 5.38 for the same filter from the previous figure.

In this case, the same point density is maintained in the pass band with 200 points and a 10 kHz wide IF bandwidth, but the lower and upper stop bands have only 40 points each with a higher power and narrower 1 kHz IF bandwidth, and the transmission zero region has 32 points so that the transmission zeros are clearly seen, also with high power but with a 100 Hz IF bandwidth. The overall measurement cycle time for this segmented sweep is 1.1 seconds, which is not quite the desired real time but much better than the eight-second alternative shown earlier.

A further optimization of the measurement setup can be utilized to achieve the desired cycle time, as discussed in the next section.

5.2.6 Extremely High Dynamic Range Measurements

For filters with very high to extremely high dynamic range requirements, there are other modifications to the test system that can provide the needed speed and dynamic range. The modifications are available only on VNAs that provide an option that is commonly referred to as a configurable test set. In this option, the cables that interface to the test port coupler are routed to the front panel so that the user has access to the coupled port and input port (or coupler thru) of the coupler, as well as the test port. With this access, it is possible to reroute the port 2 coupler so that it provides a lower noise path directly to the port 2 receiver, at the cost of lower source power available from port 2. This is commonly referred to as *reversing* the coupler and is illustrated on port 2 in Figure 5.39.

In this configuration, the VNA receiver is connected on the thru arm of the coupler to the test port, so it has substantially more sensitivity, typically about 14 dB more. This will

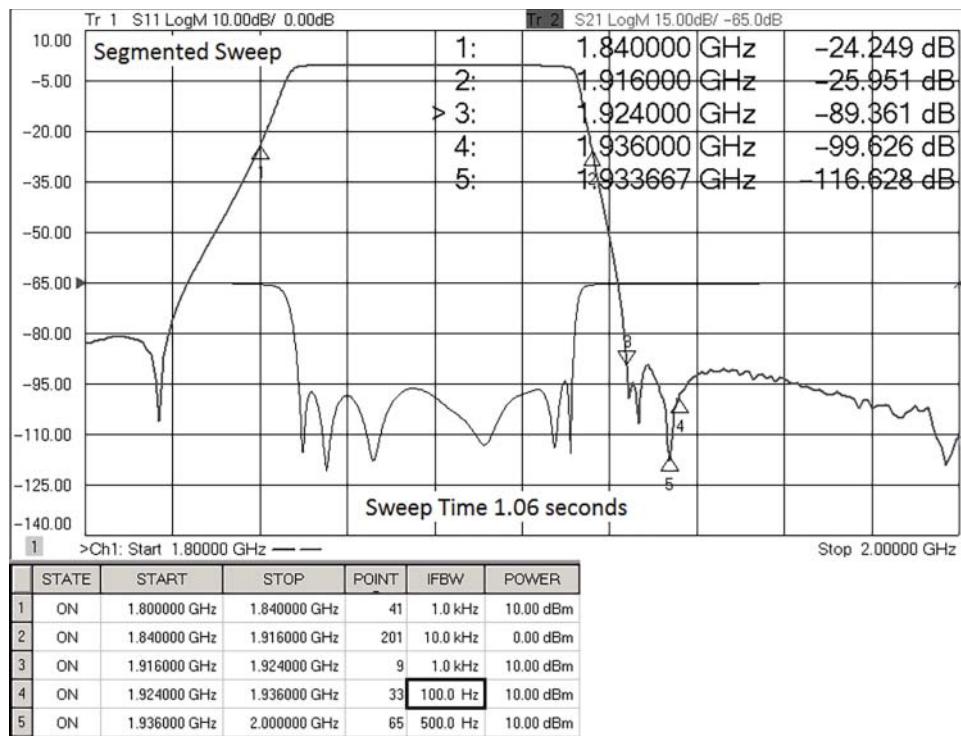


Figure 5.38 Segmented sweeps allow optimized measurements of filter transmission response.

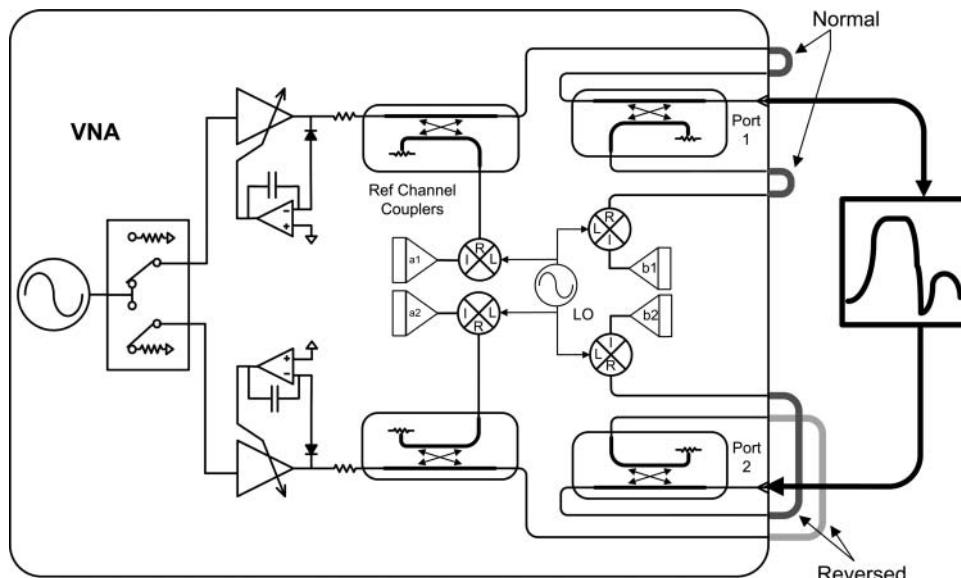


Figure 5.39 Block diagram of a VNA with configurable test set and reversed port 2 coupler.

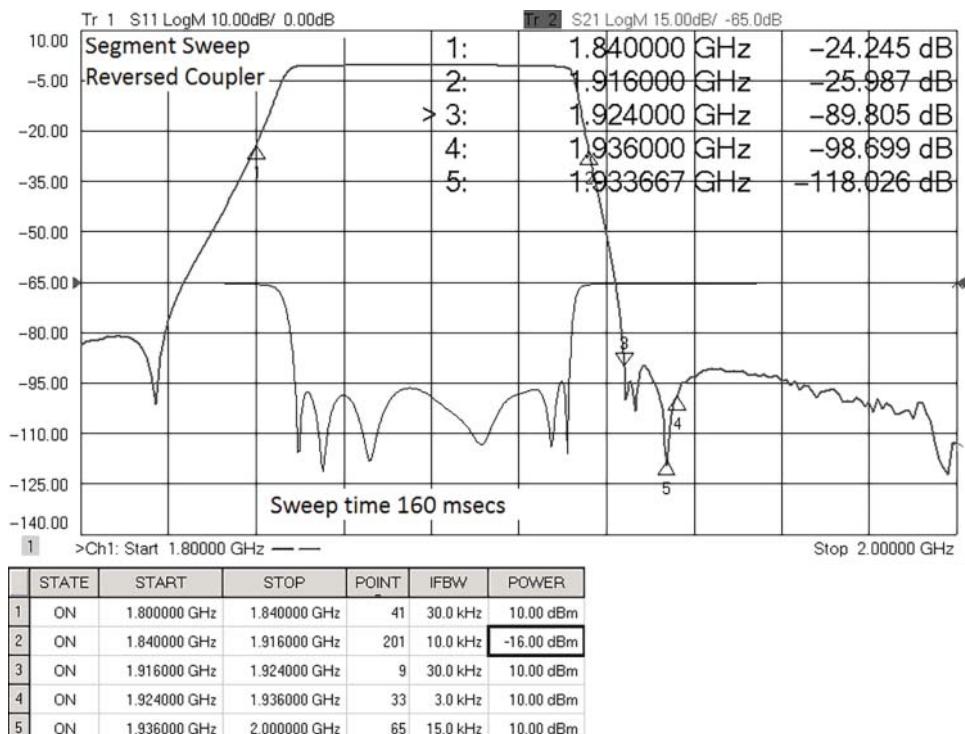


Figure 5.40 Increased dynamic range and speed using a reversed coupler.

improve the noise floor sufficiently to allow the IF bandwidth to be increased about 30 times, making the sweep rates much faster for the same dynamic range. From the previous example, the sweep rates can be as fast as about 160 ms. The passband sweep rate cannot be increased much because the power level in the passband must be reduced during the sweep to maintain a sufficiently low power in the receiver so that it does not compress the receiver. This power level is essentially equivalent to the receiver power level of the passband in the normal configuration and so the IF bandwidth must remain the same to give the same trace noise. However, in any portion of the stop band with sufficiently high loss, the power level can be set to maximum, and the full 14 dB improvement in noise will become apparent. Figure 5.40 shows the measurement result on the filter with a 160 ms sweep time, which fits the requirement for near-real-time adjustment and provides for the ability to see the transmission zeros down more than -115 dBc.

When reversing the coupler, it is critical to use segmented sweeps with adjustable power levels for each sweep to avoid overdriving the test port receiver in the passband. Also, one must be aware that the dynamic range and noise floor on the reverse (S_{12}) sweep will be degraded by the same 14 dB. Since filters are almost always linear and bilateral, $S_{21} = S_{12}$, and there is no need to measure the reverse transmission. On the reverse sweep, the source power is dropped by 14 dB hitting the DUT port 2, but the lower source power is offset by 14 dB more sensitivity on the port 2 receiver, so the signal-to-noise of the S_{22} measurement

is not changed by reversing the port 2 coupler. The higher noise on the S_{12} measurement for high-loss region has almost no effect on the corrected value of the other parameters; the contribution of S_{12} to the error correction math is always as a product of $S21 \cdot S12$. Since both are very small values, the product is small indeed. Such is not the case when testing high-gain amplifiers where the product can be greater than 1 if the noise of reverse path limits the S_{12} measurement; this is discussed at length in Chapter 7.

Other test parameters are important besides the insertion loss and isolation measurements of the filter. In many cases, the phase response and other derived measurements form a critical part of the filter specification. Some examples of these are given next.

5.2.6.1 Group Delay Measurements

A common figure of merit for many filters is group delay, in particular group delay ripple in some passband and, on occasion, the absolute group delay. Group delay measurements on filters by default are only an approximation of the actual group delay because the analog value must be estimated from a set of discrete measurements. The definition of group delay is classically

$$\tau_G \triangleq -\frac{d\phi_{Rad}}{d\omega} = \frac{-1}{360} \frac{d\phi_{Deg}}{df} \quad (5.13)$$

However, phase measurements of a filter are by nature discrete, and the analytic slope cannot be computed through differentiation but must be computed through a discrete finite-difference computation.

$$\tau_{G_meas} = -\frac{\Delta\phi_{Rad}}{\Delta\omega} = \frac{-1}{360} \frac{\Delta\phi_{Deg}}{\Delta f} \quad (5.14)$$

This discrete differentiation inevitably leads to confusion as the span of the Δf , called the aperture or delay aperture, has a strong influence on the overall measured group delay response.

The first issue of concern for measuring filters is that the phase shift must be less than 180° per point, and for some high-order filters, it may require a large number of points to avoid aliasing issues, as described in Section 5.1.5. But generally, most filters require sufficiently close frequency spacing so that under-sampling is not an issue. On the contrary, often a major issue in making group delay measurements on a filter is that the data points are so closely spaced that the group delay becomes inordinately noisy due to a very small Δf .

In most legacy VNAs, the group delay was computed by simply taking the change in phase at each point divided by the frequency step. This causes two issues: first, it produces a slight skew in the data as the actual delay on each point is offset by one-half of the span, if only two points are used. Second, because there are $N - 1$ frequency segments, there are $N - 1$ group delay computed points for an N -point sweep. In most legacy analyzers, this is handled by making one of the delay points repeated, usually the first point. One can avoid the skew issue by always using an odd number of points for the delay or smoothing aperture.

If the frequency step was very small, even very slight trace noise on the phase trace would cause very large trace noise in the group delay trace. Figure 5.41 shows the group delay response of a filter with two different frequency spans in the upper window, and for various number of points in the lower window, where the reference position is offset by 1 division for

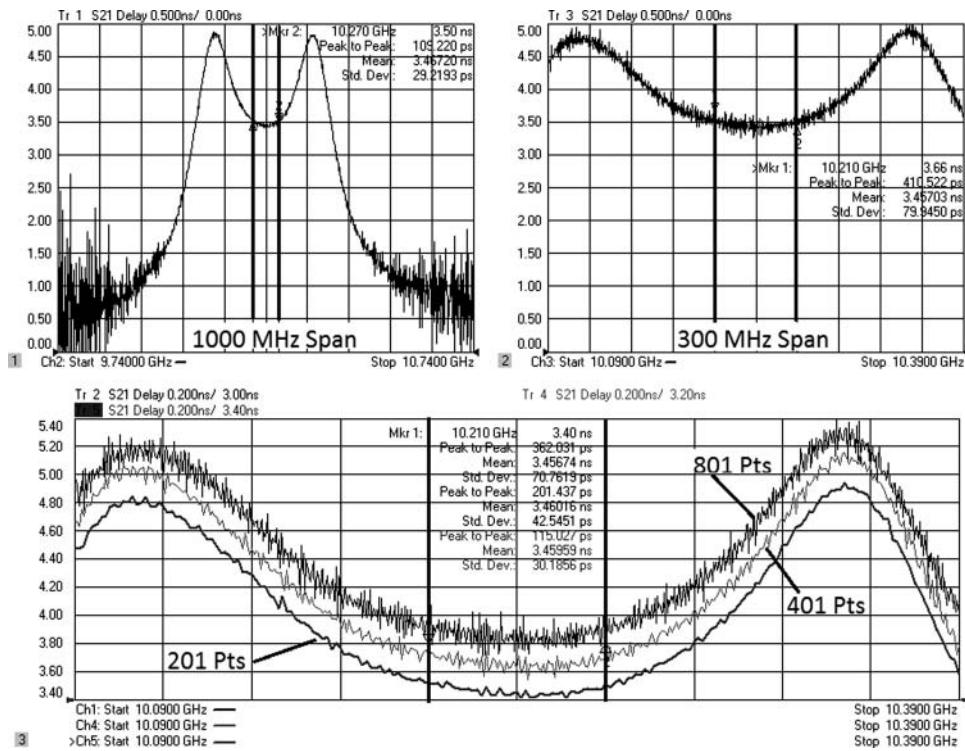


Figure 5.41 Group delay on a filter with various number of points and various spans.

each setting to clarify the resulting traces. For each measurement, a trace statistics function is shown for the middle 60 MHz of the group delay response, and the relative value of trace noise can be seen in the Std. Dev. result.

A wide span trace has less noise than a narrow span trace, roughly in proportion to the span. Traces with fewer points show less delay noise than greater point density traces roughly in proportion to the number of points. In all cases, the trace noise on the phase trace is the same, but the divisor in Eq. (5.14) changes. All of these results are attributes of the group delay aperture and the fact that the aperture changes with span and number points.

While the trace noise in the phase responses is essentially a constant level, the changing number of points changes the aperture with a corresponding increase in the trace noise of the group delay trace. One way to avoid this issue is to use fewer points on the group delay trace, but in many instances a large number of points is required to ensure that there are no issues in the amplitude response, and it would be convenient to have the group delay computed on the same trace. Most legacy VNAs incorporate a smoothing function that applies a moving-average smoothing on a trace, and this can be used to set a wider effective aperture on a group delay trace. The smoothing function operates on a moving average window

$$Y_{n_Smooth} = \frac{[Y(n-m) + Y(n-m+1) + \dots + Y(n) + \dots + Y(n+m-1) + Y(n+m)]}{2m+1} \quad (5.15)$$

where $2m + 1$ is the smoothing aperture, in points. Smoothing is often given in percent of span, and this is converted to smoothing points by

$$m = \text{int} \left[\frac{N \cdot (\text{smoothing percent})}{2} \right] \quad (5.16)$$

where N is the number of points.

When this is applied to most traces, each point forms an average of the surrounding points and is similar to the video bandwidth function of a spectrum analyzer. In *most* cases, smoothing is an invalid way to reduce noise, as it can also remove important structure in the response. In the case of an amplitude response, smoothing can be used to eliminate mismatch ripple due to a poor calibration, but it can also hide the fact that a response has some excess ripple due to the DUT response.

In group delay traces, however, a quirk of mathematics makes using smoothing *identically equal* to reducing the number of points around the target point, and thus any intermediate values have no effect on the delay; a simple example of applying the definition of group delay from Eq. (5.14) to the definition of smoothing from Eq. (5.16) for a five-point smoothing ($m = 2$) illustrates this point.

$$D(n)_{\text{Smo}} = \frac{\left[\frac{\varphi_{n-1} - \varphi_{n-2}}{\Delta f} + \frac{\varphi_n - \varphi_{n-1}}{\Delta f} + \frac{\varphi_{n+1} - \varphi_n}{\Delta f} + \frac{\varphi_{n+2} - \varphi_{n+1}}{\Delta f} + \frac{\varphi_{n+3} - \varphi_{n+2}}{\Delta f} \right]}{360 \cdot 5}$$

$$D(n)_{\text{Smo}} = \frac{\varphi_{n+3} - \varphi_{n-2}}{360 \cdot 5 \Delta f} \quad (5.17)$$

From this, it is clear that none of the intermediate points contributes anything to the computation of the smoothed delay. Unlike smoothing in other traces, the values of the measurements of intermediate points have no consequence on the result, such that the value after smoothing is identical to the value that one would see if only the phase of the endpoints of the smoothing aperture were measured.

In more modern VNAs, the group delay aperture can be sent independently from the trace smoothing, so it is no longer necessary to turn smoothing on and off when changing the trace format from delay to other formats such as log magnitude. In some VNAs, the aperture can be set as either a point aperture, a percent of span, or a fixed delta-frequency aperture. In these cases, it is necessary for the delay to be computed from interpolated phase points if the fixed delta frequency or percent of span does not lie exactly on measurement points. Using a fixed delta frequency is a convenient way to specify aperture as the resulting group delay trace, and noise on the trace does not change with changes to frequency span or the number of points. Figure 5.42 shows the resulting response when a fixed frequency group delay aperture is applied to the measurements of Figure 5.41. Here the aperture of 3 MHz was used, which matches approximately the default aperture for the 201-point trace above. Looking at the Std. Dev. numbers, one sees that they are nearly identical for all traces. Thus, specifying the aperture in terms of delta frequency yields the most consistent results.

In most filters, the absolute value of the group delay is not important; only the group delay ripple matters to the communications channel. Group delay ripple can cause distortion in modulated measurements. However, in some classes of filters, the absolute group delay is critically important, and these are called, not surprisingly, *delay filters*. These types of filters are often found in feed-forward amplifiers to provide a fixed delay to a signal applied to an

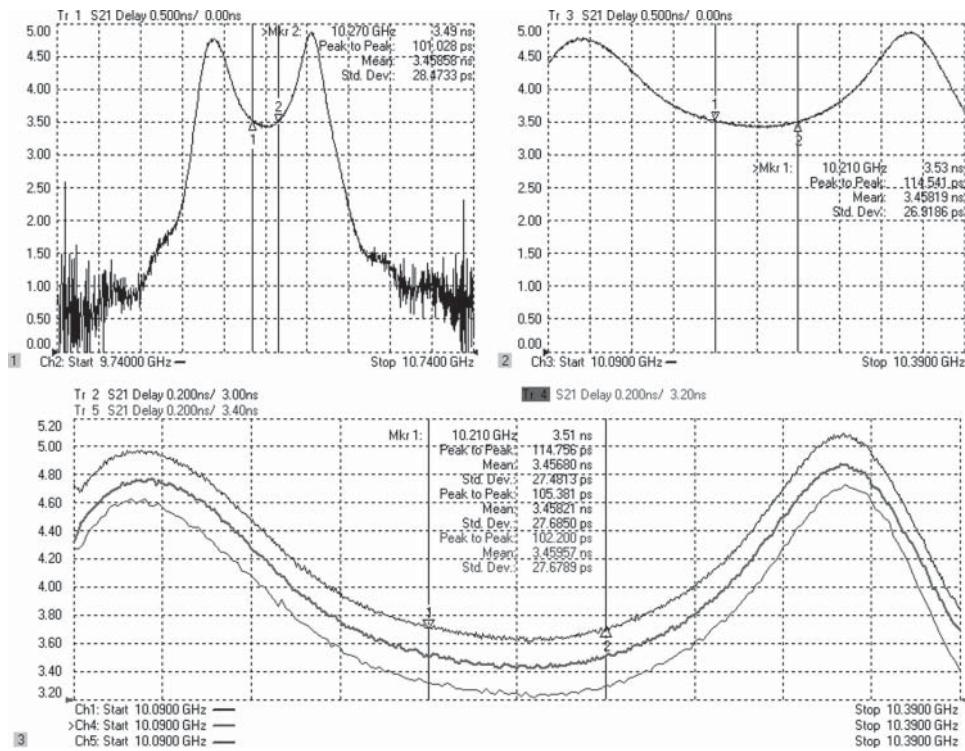


Figure 5.42 Group delay results from applying a fixed-delay aperture to the previous figure's response.

error amplifier that is eventually combined at the output and used to cancel the distortion of the main power amplifier. These filters must be tuned for a precise delay, as well as other characteristics.

5.2.6.2 Long Delay (SAW) Filters

Some filters, particularly SAW filters, have long delays. The delay of a filter can be approximated as one over the bandwidth times the number of resonators. SAW filters use a transducer that can have many effective resonators, high Q , and narrow bandwidth thus producing the long delays. SAW filters are often tested in an on-wafer form, before being packaged, and the isolation response may be limited by the RF leakage that occurs directly from probe to probe. In the packaging process, the physical ends of the SAW material are treated to prevent the acoustic waves from reflecting, but before they are packaged, or if they are poorly packaged, a re-reflecting acoustic wave, called *triple travel*, will degrade the isolation and increase the ripple in the passband. In some instances, such as on-wafer test, it is desired to see the response of the filter in the absence of the triple travel signal. This effect can be eliminated by using time-domain techniques (Keysight Technologies, Inc., 2002).

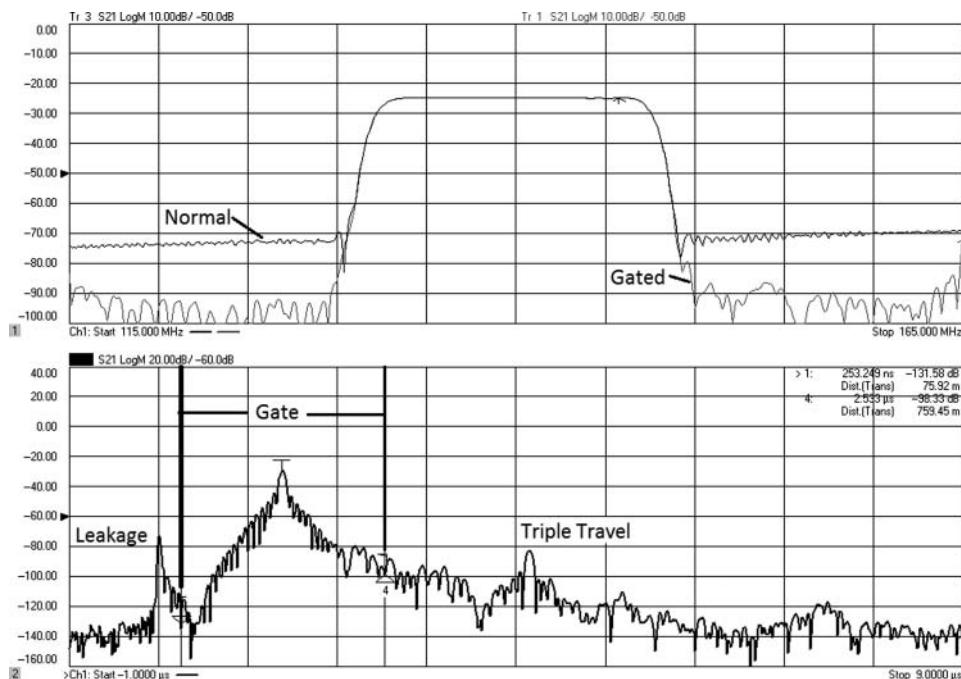


Figure 5.43 Saw filter response frequency (upper), time domain (lower).

Since the group delay of SAW filters is so long, the response from the leakage can be easily distinguished from the main transmission response. Similarly, the triple travel response will occur at a time sufficiently long that it is also easily distinguished from the main response. Thus, the response of the main lobe in the time domain may be gated to exclude the leakage and the triple travel and yield a response of the filter that depends only upon the design of the transducers. Figure 5.43 shows the response of a SAW filter.

The upper window is the gated and ungated frequency response of the filter. It is clear that the gated response has better isolation, indicating that the transducer is well designed to provide the isolation. The ungated response shows the effect on the isolation of the RF leakage and triple travel. The lower window shows the ungated time-domain response along with the region where the gate will be applied. The RF leakage and triple travel are easily identified in the time-domain response.

5.2.6.3 Deviation from Linear Phase

An important transmission response related to the group delay is the deviation from linear phase. This is a more direct statement of the phase error or phase flatness and is sometimes specified instead of group delay ripple. In fact, group delay ripple is probably a better measure of the phase flatness, as a sharp change in phase over a small change in frequency might still pass a peak-to-peak phase ripple specification but would cause a large peak in the group delay ripple measurements. Nonetheless, many systems specify the phase deviation as a figure of

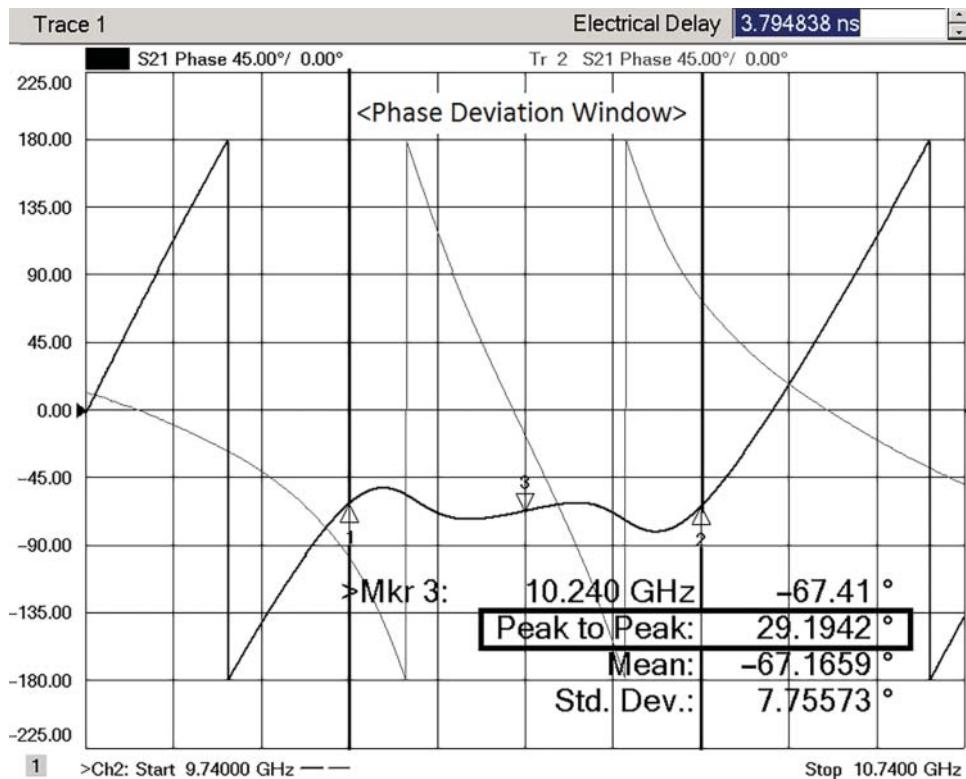


Figure 5.44 Phase response of a filter before and after setting electrical delay.

merit, and for phase deviation measurements, one must normally remove the linear slope portion of the phase response to show just the non-linear deviation.

Most VNAs provide a convenient feature, sometimes called Marker → Delay, which computes the electrical delay of the DUT within $\pm 10\%$ of the marker position and removes that amount of linear delay from the phase response. This is usually accomplished through a scaling function called *electrical delay*; it provides a mathematical normalization of the phase response to the phase slope associated with an ideal transmission line that has the entered delay. The phase response of the filter from Figure 5.41 is shown in Figure 5.44, before and after applying Marker → Delay, with the electrical delay function value shown.

The Marker → Delay feature is convenient for providing a quick offset, but it is clear from the trace that some adjustment to the electrical delay will provide a flatter response with less peak-to-peak phase deviation, due to the downward tilt of the phase trace in the phase deviation window. Typically, this fine adjustment is performed by manually adjusting the electrical delay while monitoring the peak-to-peak phase ripple using trace statistics. The trace statistics have been limited to the narrower user range as shown by the phase deviation window, as in Figure 5.35.

In some modern VNAs, equation editor functions are available that compute a variety of phase deviations and offsets directly. The simplest form is similar to the flatness feature of

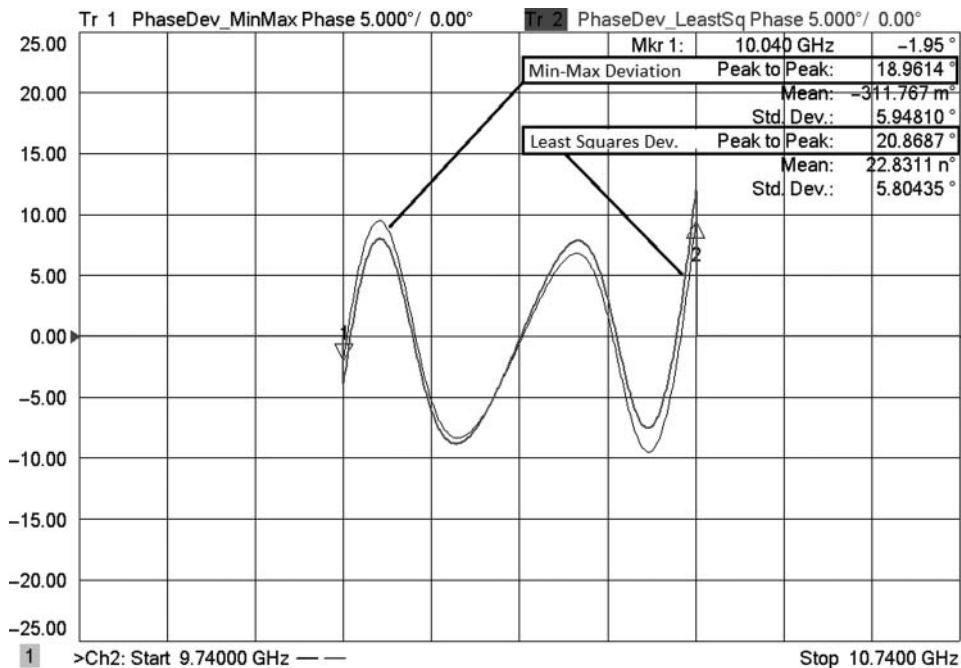


Figure 5.45 Least-squares and min-max fit of a phase deviation.

amplitude response, where a least-squares fit is computed. This is easy to compute numerically and provides a quite good answer for the best-fit line through a phase ripple trace, but not the optimum answer. It is possible to compute mathematically the best-fit line minimizing the maximum peak-to-peak error (min-max), and some VNAs provide this as an additional equation editor function. The results of both the least squares fit and the min-max fit are shown in Figure 5.45.

The min-max fit is compute-intensive and will create very slow updates if the number of points is large. The least squares fit provides better computational speed, at the trade-off of slightly less than optimum results. In this case, Min-Max provides 2° lower deviation. Each one provides about 10° improvement on the basic Marker → Delay function. For both of these functions, only the portion of the phase trace inside a user-specified window is displayed.

5.2.7 Calibration Considerations

In measuring filters that have both high isolation and low loss, the stimulus settings to use during calibration may be different from those that are used during measurement. One of the common misconceptions with regard to calibration is that it is presumed to be a requirement to exactly match the same stimulus conditions during calibration as during measurement. This misconception most likely came about because of an unfortunate choice of annunciators used in some of the earliest network analyzers. After calibration, it was common to put a “C” or “Cor” annunciator on the screen to indicate that the measurements had error correction

applied. If *any* stimulus settings were changed, the annunciator would change to “C?,” leading many users to presume this indicated a questionable calibration. To avoid this unpleasant label, many users would select settings for calibration that gave poorer results, after being applied, than would be the case if the settings were changed.

An excellent example of this is apparent in the segmented sweep settings for Figure 5.40. It is not possible to calibrate at the maximum power used in the isolation segments, because during the thru portion of the calibration, the VNA receiver would be in hard compression. An unwitting user would select a lower power to calibrate the isolation portion, just low enough to avoid compression (about -6 dBm), and then leave that power for the measurements. However, this would result in at least 10 times the noise in the isolation case, perhaps many dB of noise degradation. The other alternative, calibrating at a lower level and then changing the power after calibration, would result in a few tenths of a dB error at most in the isolation measurement but would produce the regrettable “C?”.

Since the days of the legacy VNAs, the calibration annunciators have become more informative; it is common to see C^* to indicate that the error correction array has been interpolated, and ΔC to indicate that some setting has been changed, but the underlying calibration is still valid. The “C?” annunciator is no longer used.

Another common error in calibration for filters is to set the power very high (usually in a non-segmented sweep) to ensure that isolation will have low noise without changing the power level after calibration. This can cause one of two distinct, and differing, errors in the transmission pass band.

If the calibration uses mechanical standards, the thru portion might put the VNA receiver into compression, causing the receiver response to read lower than it should be for the calibration trace. If the filter to be tested has some substantial loss, then the receiver will *not* be in compression during the measurement and will read a lower insertion loss value than the filter actually has, giving an optimistic view of the filter loss. While this will allow filters to appear to pass specification more easily, it is not valid.

If the calibration uses electronic calibration modules, the loss of the calibration module may take the VNA receiver out of compression during the calibration portion, but a very low loss filter will put the receiver into compression during the measurement. This will lead to a reading during the measurement that presents a higher insertion loss than it should, causing some filters to fail their limits when their performance is actually satisfactory.

Thus, in many cases, it is best to use a moderate power for calibration and change the power after to provide the optimum trade-off between dynamic accuracy and trace noise in the measurements. Some of the most modern VNAs have receiver linearity so good, that there is no need to be concerned at all about varying the power after calibration so the best power for a low noise calibration should be chosen.

5.3 Multiport Devices

Many linear devices have more than two ports, but most discussion of measurements focus on the 2-port case for convenience. In the past, the measurement of multiport linear devices proved difficult when only 2-port VNAs were available. The termination effect on the ports not connected to the VNA could have substantial effect on the measurements, and the process of changing cables to access all the combination of ports was tedious and error prone.

In the late 1990s the first VNAs with more than two ports were introduced. These systems were 4-port systems that added an additional 2-port test set to the standard VNA. Soon after, the integrated 4-port VNA became a standard offering, especially in the RF frequency range where differential devices were being developed. Now 4-port measurement systems are available up to THz frequencies, and extension test sets are available that provide full calibration capability up to more than 32 ports.

So, while the problem of multiport measurement systems has largely been solved, there are some important considerations when making measurements on multiport devices that will be discussed next.

5.3.1 *Differential Cables and Lines*

Differential devices have become common in the RF and even microwave frequency ranges. Details about active devices will be discussed in later chapters, but a common passive component is a differential transmission line or differential cable-pair used to connect between differential devices. Often, these are used only for test purposes, but their careful characterization is necessary to provide accurate measurements of devices when these test lines or cables are used.

Differential lines are by their nature coupled, but the definition of a mixed-mode measurement system presumes in its definition that a 4-port measurement system has essentially uncoupled ports. A simple mental test to decide whether a pair of ports is uncoupled is to consider what happens if power is applied to just one of the ports that comprise the pair. If power is transmitted to the other port, then they are *not* uncoupled. A differential probe, for example, will likely have some leakage from one signal line to the other, and so is not completely uncoupled. A 4-port VNA will likely have extremely high isolation between ports, and so is completely uncoupled. When ports are uncoupled, the two modes of transmission from the port pair, common-mode and differential-mode, are easily defined, and this is the basis for most differential measurements. The details of differential measurements are discussed in Chapter 10.

5.3.2 *Couplers*

Directional-couplers are 3- or 4-port devices that are used for monitoring and signal separation. In the past, measurements of directional-couplers were typically performed using two ports at a time, while loading the other ports with good terminations. This works quite well, but some special cases occur, and using a 3- or 4-port VNA can provide substantially more convenience. One case is when the test port of the coupler cannot be well terminated. The isolation of a directional-coupler depends upon a good match at the test port, when a 2-port VNA is used. However, if a 3-port or 4-port VNA is used, multiport error correction can be applied that provides a full correction for the mismatch applied at all ports. This allows a good measurement of isolation even if a high-quality load is not available for that port. Some cases that this applies to are testing PC board or integrated circuit couplers that must use fixtures or probes for testing.

Another example where using a 3-port VNA for coupler testing is useful is testing and tuning high-performance couplers. Because a coupler's directivity depends upon all three of the

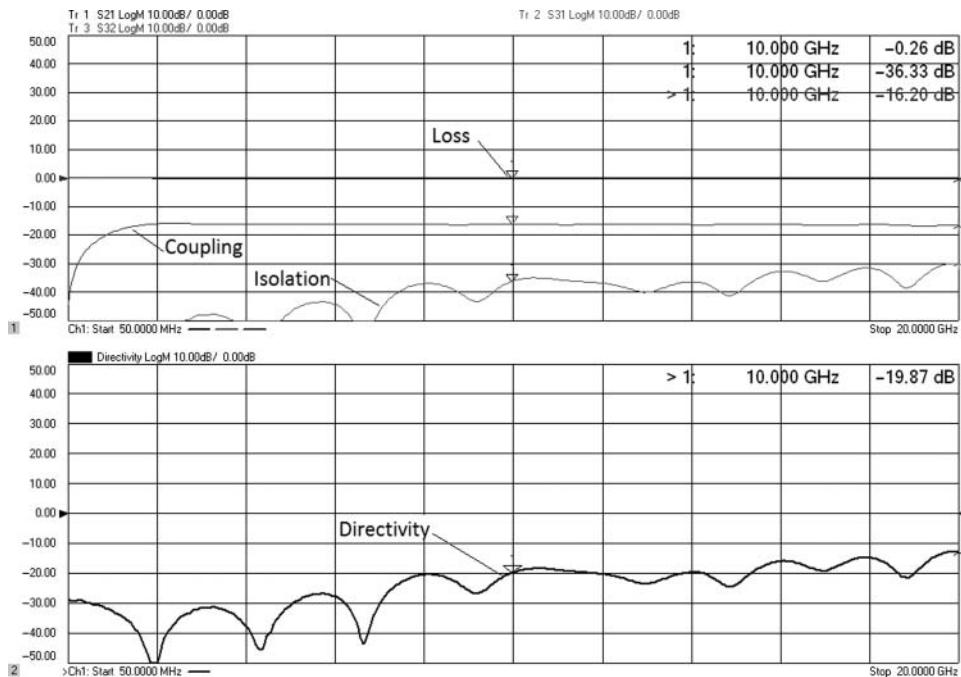


Figure 5.46 Measuring couplers: upper plot is the three main terms, lower plot is the directivity computed using an equation editor function.

path measurements, main-arm loss, coupling, and isolation, all three must be measured and accounted for when measuring directivity. Often, the overall isolation measured comes from a combination of the intrinsic isolation of the coupling structure and mismatch at the coupler output or test port. Normally mismatch acts to degrade the isolation, but in some cases, it can work to cancel some intrinsic leakage, and thus slightly mismatching the coupler output can improve the overall isolation. For high-performance couplers, where the directivity approaches 40 dB, some tuning is typically required. In such a case, making all three measurements on a 3- or 4-port VNA allows direct computation of directivity so that it can be adjusted in real time. Figure 5.46 shows an example measurement of a coupler.

In the figure, the directivity is computed in the equation editor using the formula in Eq. 1.90. In the case where the test port is at port 1, with forward coupling to port 3, and the main thru arm at port 2, then

$$\text{Directivity} = \frac{\text{Isolation}}{\text{Coupling} \cdot \text{Loss}} = \frac{S_{31}}{S_{21}S_{23}} \quad (5.18)$$

An example where a 4-port VNA is useful is in the case of a 4-port coupler that uses an external fixed load instead of an integrated load. Often the load is attached in a semi-permanent way (typically painted over), and the coupler is treated as a 3-port (see Figure 1.33). However, during the manufacturing process, or if the attached load is damaged (perhaps due to high power) then the coupler's intrinsic characteristics can be determined by

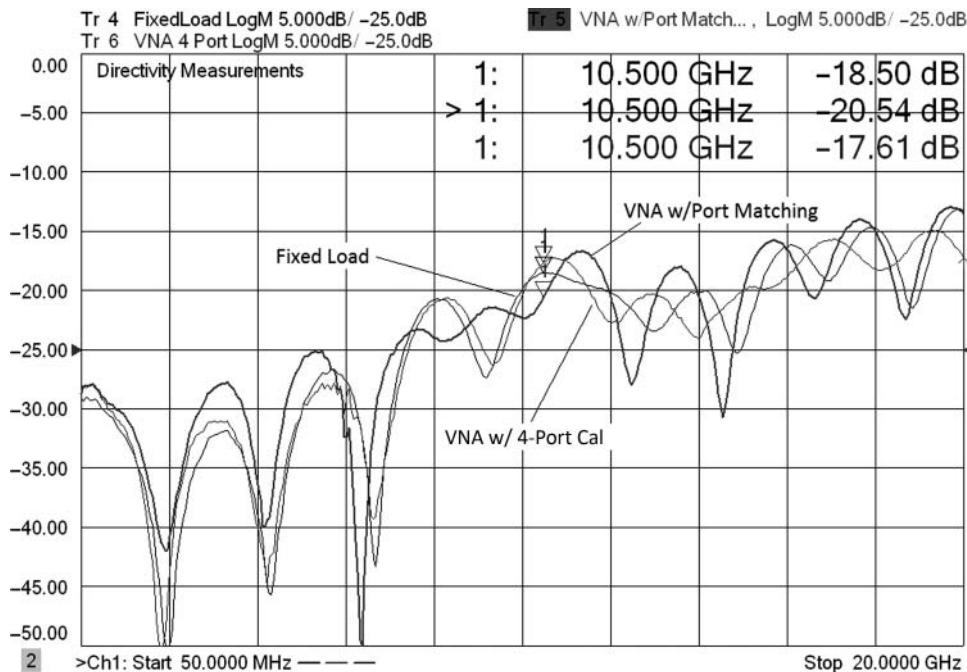


Figure 5.47 4-port coupler using a fixed external load, a 4-port VNA, and port matching.

a 4-port measurement without using a load. Equation (5.18) still applies, but the load for the fourth port is supplied from the VNA, and the VNA test port impedance error corrected to the residual load match of the calibration.

Interestingly, as a 4-port measurement, the effective port matching of the load port can be modified using the built-in VNA fixturing function such as port matching. An example is shown in Figure 5.47, which has a measurement of the coupler directivity with the attached load, again with a 4-port VNA measurement. Also shown is the result of applying some port-matching elements to improve the directivity.

The port matching was performed by adjusting a virtual circuit of a shunt C, series L, and series R to show the effects of termination impedance on the coupler directivity response. These were adjusted to provide the best directivity up to 10 GHz. With this port matching, one can determine that a 3 dB improvement was possible if an optimized load is used.

Other important measurements on couplers are the mismatch at the input and output ports. Normally, the port match of the coupled arm is not very important, but if there is substantial mismatch, it can cause ripples in the coupled signal if the termination impedance of the device connected to the coupled arm is also not well matched. In the case where a coupler is used to monitor a signal level, this will cause a ripple error in the apparent coupled signal. Thus, if the coupled arm is poorly matched, it will still have a flat response when measured into a good load, but the flat response will not be maintained if the circuit into which the coupler is eventually terminated does not provide a good match. In contrast, if the coupled arm has a good match, the match of the receiving circuit connected to the coupled arm will not have a strong effect on the ripple seen in the coupled signal.

Ripple in a wideband coupler is to be expected as they are often designed to have some equal ripple response if the coupling bandwidth is wider than one octave. Thus, a common specification on a coupler is peak-to-peak ripple in the coupled signal.

5.3.3 Hybrids, Splitters, and Dividers

Hybrids are a common name for a circuit that provides a nearly loss-less split between two signal paths, and sometimes also called 3-dB hybrids, splitters, dividers, or baluns. They can be considered a special kind of coupler, where the coupling factor and loss are equal and around -3 dB. Loss in a hybrid is often specified relative to the ideal loss, so it is not uncommon to see a specification for a hybrid of -1 dB maximum loss, and this is to be interpreted as being less than 4 dB loss from the input to the output.

Hybrids come in three distinct types: 0° , 90° , and 180° . Many hybrids are 4-port devices that have two inputs in addition to having two outputs. One input provides an equal power, 0° phase split to the two output ports, and the other input provides each output with equal power and a 180° phase split. The 90° hybrids have similar versions with two inputs, one of which provides 0° and 90° to the two outputs, the other providing 0° and -90° . The 90° hybrids are often used to provide signals for I/Q mixers or image-reject mixers. Example illustrations of both types of hybrids are shown in Figure 5.48.

Common measurements on hybrids include input and output match, isolation between the two output ports, and, most importantly, loss, and magnitude or phase imbalance of the two output ports. For the imbalance measurements, the equation editor function of VNAs can be used to directly compute the balance of the splitter. Using the equation

$$Bal = \frac{S_{21}}{S_{31}} \quad (5.19)$$

one can plot on separate traces both the magnitude balance and the phase balance. For all hybrids, the ideal balance is 0 dB (equal power). For splitters, the ideal phase is 0° . For 180° hybrids, it is obviously 180° , and of course 90° for 90° hybrids.

For 4-port hybrids, the balanced can be computed for both the inverting and non-inverting ports, or for the $\pm 90^\circ$ ports. Two examples of measurements for hybrids are shown in the following figures.

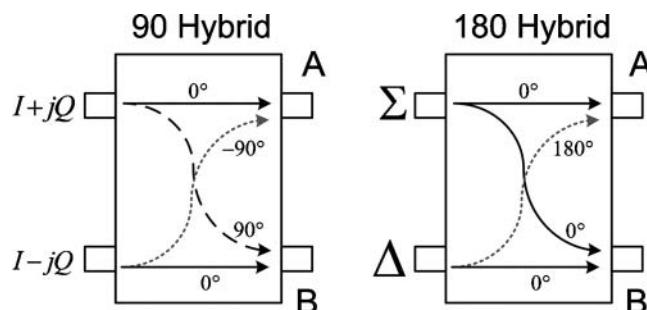


Figure 5.48 A 4-port 90° hybrid and a 4-port 180° hybrid.

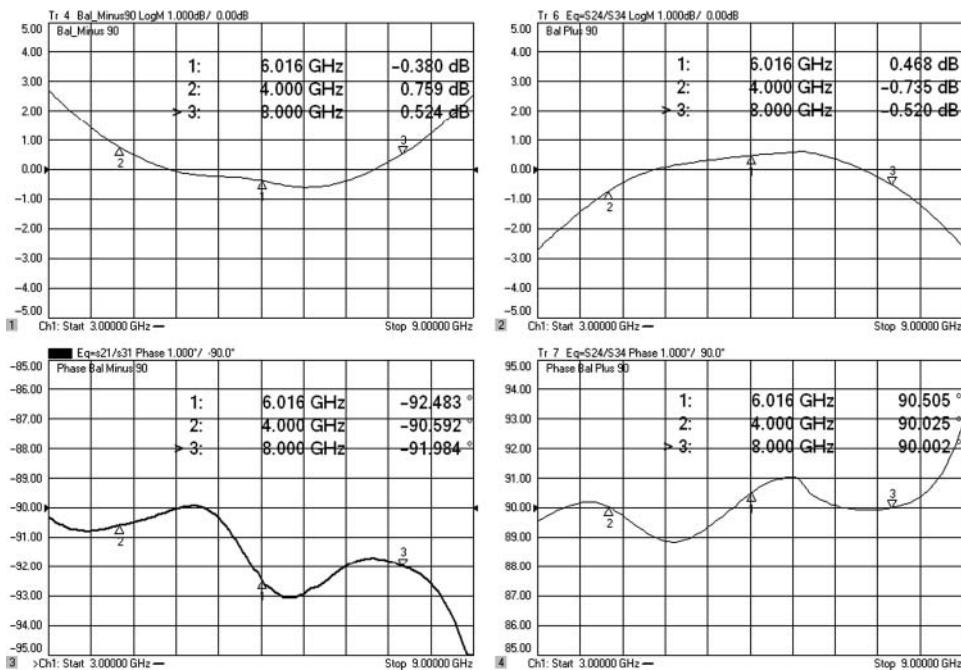


Figure 5.49 Response of a 90° 4-port hybrid.

5.3.3.1 90° Hybrids

Figure 5.49 shows the response of a 4-port –90° hybrid. The upper window shows the individual magnitude balance, and the lower windows show the phase balance. The plus and minus 90° responses are generated by driving different ports on the input of the hybrid (port 1 for –90, port 4 for +90), so different equations are used based on the source port to show the relative phase. In this case, as is most common, the hybrid has frequency limitations on the order of one octave, ranging from 4 to 8 GHz.

5.3.3.2 Balanced Hybrids

Often, balanced hybrids are used to drive differential devices. In such cases, their measurement is typically reported as mixed mode S-parameters. An ideal balance-to-unbalanced hybrid will have two key parameters: common mode gain from the summing or sigma port (S_{CS21}) and differential mode gain from the differential or delta port (S_{DS21}). Even though it is a 4-port device, when used as a balun, it is presumed to be used to drive either a differential or common mode signal, so it is modeled as one of two different 3-port devices. Hybrids are used for this purpose as they allow matching of both the common mode and the differential mode simultaneously. Transformer type baluns do provide for a differential drive, but the common mode impedance at the balanced port is usually either infinite (for a 4-port simple transformer) or zero (for a center-tapped transformer). For more details on differential and mixed mode behavior, see Chapter 10.

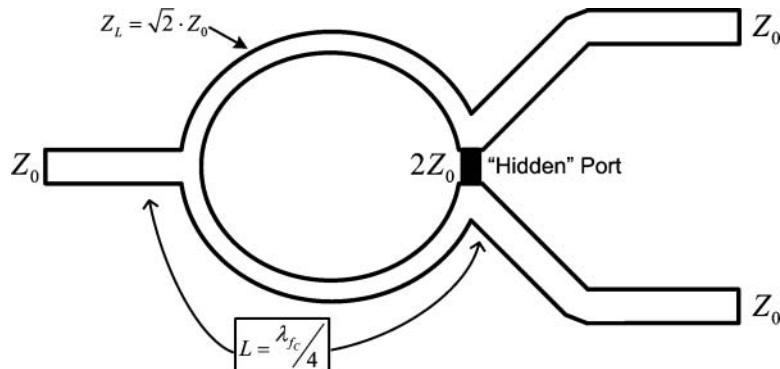


Figure 5.50 Typical form of a Wilkinson power splitter.

5.3.3.3 Splitters and Dividers

Splitters and power dividers come in lossy or lossless versions. Lossy versions split the signal while maintaining good match on all ports by use of resistors. The two main lossy splitters are three-resistor splitters (sometimes called *power dividers*), used when the power is simply divided between to go to two different loads, and two-resistor splitters, used when the ratio of the two outputs paths are used, as in the reference arm of a network analyzer.

Lossless splitters use many different structures to divide the power between two paths. One type of loss-less splitter is a 3 dB coupler, which in fact has a fourth port that is the internal load of the coupler.

Figure 5.50 represents a layout of a different kind of splitter; this particular splitter uses a Wilkinson structure (Wilkinson 1960) and has a “hidden” fourth port (a resistor of value $2Z_0$) that provides for a matched condition and absorbs power only when the two output lines are not equally terminated, or as a power combiner, when lines are not driven in a balanced way.

Figure 5.51 shows the frequency response of this splitter. The line length sets the center frequency of the splitter. While the isolation and match are degraded at low frequencies, the loss is still relatively flat. An equation editor function incorporates Eq. (5.19) for each trace in the lower window, with the format set to LogMag on one trace and phase on the other.

If the device is driven in the forward direction to isolated test ports, the balance is directly discerned. The definition of balance in this case is

$$Bal = \frac{S_{21}}{S_{31}} \quad (5.20)$$

However, in the case where the splitter has an unequal drive into the split ports, some signal will be absorbed inside the internal load of the splitter. Thus, this style of hybrid is not ideally lossless, and in fact it is to be expected that nonmatched impedances on the split ports will result in reflections into the DUT that will be absorbed by internal load. It can be shown that for any 3-port network, it is not possible to have the network be lossless and matched at each port. This type of splitter/combiner demonstrates that with certain drives or loads, it does absorb power in its internal load and is thus not lossless. But in normal operation, it is nearly lossless. Also, the amplitude and phase balance is good, with less than 0.1 dB of amplitude imbalance and less than 1° of phase imbalance.

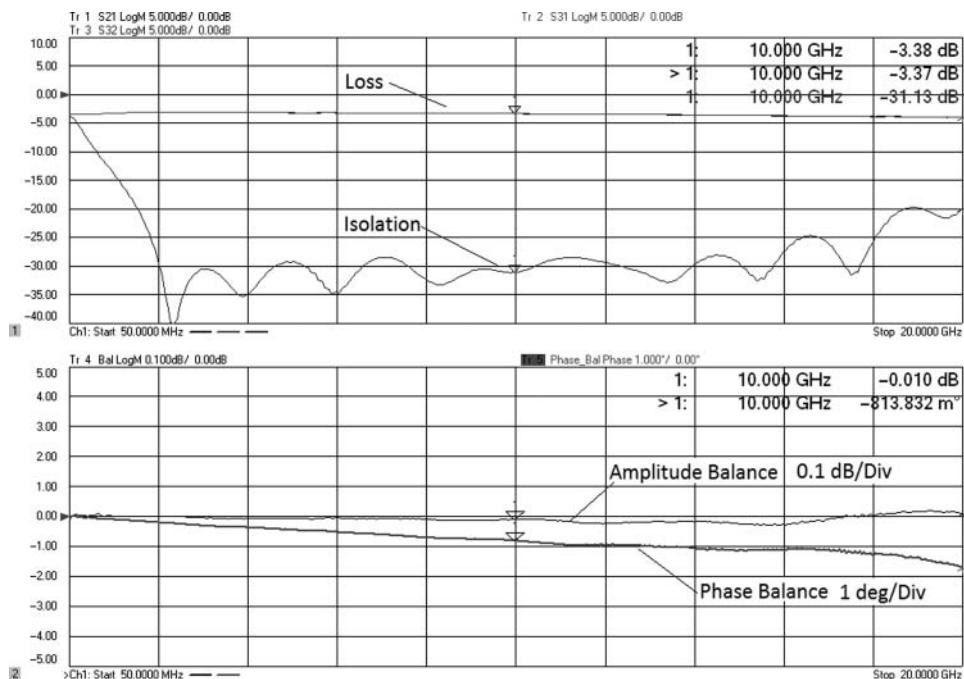


Figure 5.51 Response of a splitter.

5.3.4 Circulators and Isolators

Circulators have much in common with directional-couplers, in that the transmission response from one port to another depends strongly upon the load or return loss on a third port. Circulators have a unique role in microwaves and RF as they are perhaps the only component that is at the same time linear and *not* bilateral. The circulation function is a result of a non-uniform response of a magnetic element that provides for a differing phase delay when a signal is circulating in a clockwise or counterclockwise direction.

Most circulators cover approximately one octave and provide for high isolation in one direction of signal flow and low loss in the opposite direction. A key figure of merit is the loss in the low-loss direction and the isolation in the reverse direction. However, using a 2-port network analyzer presents similar problems to measuring the isolation of a coupler, in that the quality of the load on the third port, that is, its return loss, will directly affect the isolation measurement. That is because any reflection from the third port will continue to circulate in the low-loss direction and cause a leakage-type signal to appear at the isolated port, as illustrated in Figure 5.52.

The measurements of an example isolator are shown in Figure 5.53. The upper window shows the isolation measurements, and the lower window shows the insertion loss measurements. For comparison, the isolation measurement of S_{21} is also shown in the presence of a non-ideal load applied to port 3, in the left upper window. In this case, the non-ideal load presents approximately a -32 dB return loss. This is a narrowband high-isolation circulator,

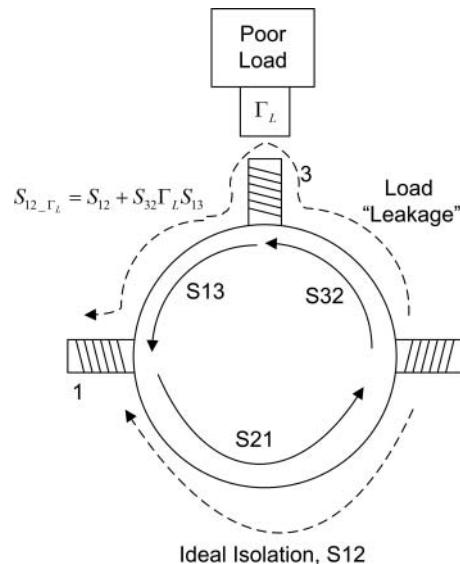


Figure 5.52 Isolator behavior in the presence of a non-ideal load.

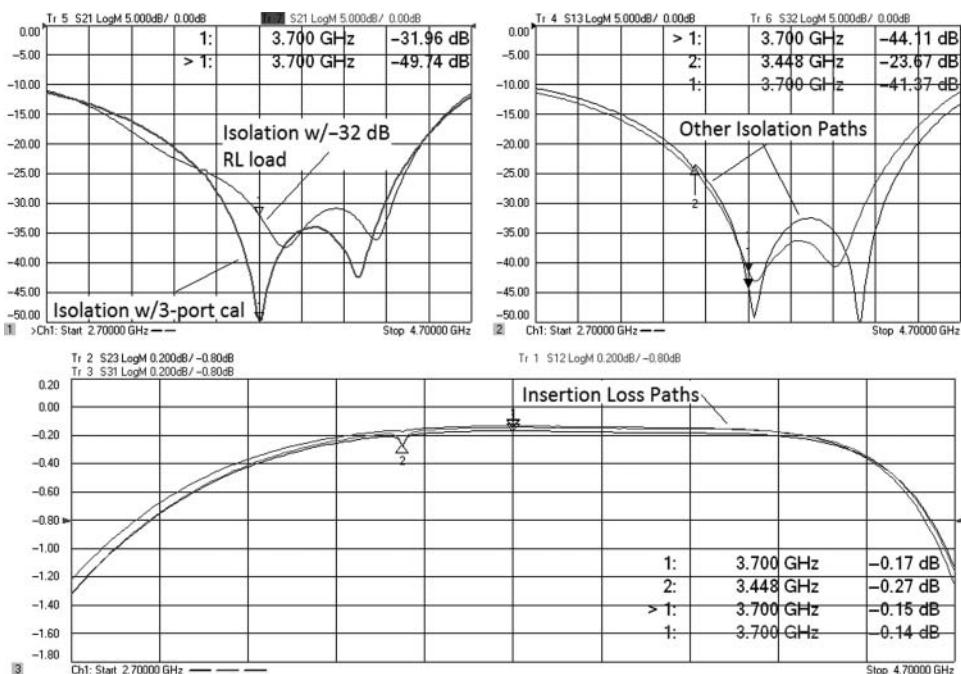


Figure 5.53 Measurements of an isolator.

and with the full 3-port calibration, the true isolation of the S_{21} path is shown to be well over 40 dB at the center frequency.

For a high isolation circulator, a poor load can cause a substantial degradation of the isolation measurement. If the isolation approaches the return loss of the load, the worst signal can be as much as 6 dB higher than the actual isolation, as shown in the trace marked "Isolation w/-32 dB RL load." As a reference, when combining a large signal with a small error signal, if error signal is 19 dB below the other, it will cause a 1 dB error in the response measured for the larger signal. These errors become important when looking at the isolation response of circulators.

Finally, it is apparent that this isolator has a small mode (on the order of 0.1 dB) around 3.448 GHz. There is evidence of this mode also in the isolation paths.

5.4 Resonators

5.4.1 Resonator Responses on a Smith Chart

One key measurement made with Smith charts is the measurement of resonators to find the Q and center frequency. For 1-port resonators, the procedure to find the Q and center frequency is relatively simple. Consider the circuit of Figure 5.54. The key figures of merit for the resonator are f_0 , the center frequency, and Q_0 , the unloaded Q . The center frequency and Q can be found from

$$f_0 = \frac{1}{2\pi\sqrt{LC}}, \quad Q_0 = 2\pi f_0 R_0 C \quad (5.21)$$

The common measurement of Q is to look for the 3 dB loss points of a transmission (S_{21}) response, relative to the center frequency, where the loaded Q is defined as

$$Q_L = \frac{f_0}{f_2 - f_1} \quad (5.22)$$

And f_2 and f_1 are the frequencies of the lower and upper 3 dB down points. Loaded Q means that the circuit is loaded by the external resistance of the VNA source, typically 50Ω . The usual figure of merit for a resonator is unloaded Q , or Q_0 .

In a 1-port measurement of Q , the measurement is S_{11} , so the normal definition doesn't apply. When a high Q resonator is measured with a 1-port S_{11} measurement as in Figure 5.54,

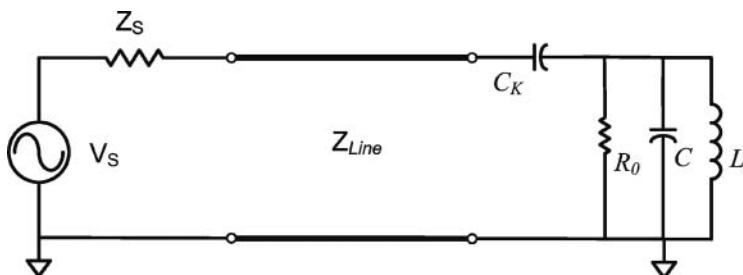


Figure 5.54 Schematic of a 1-port resonator with coupling capacitance.

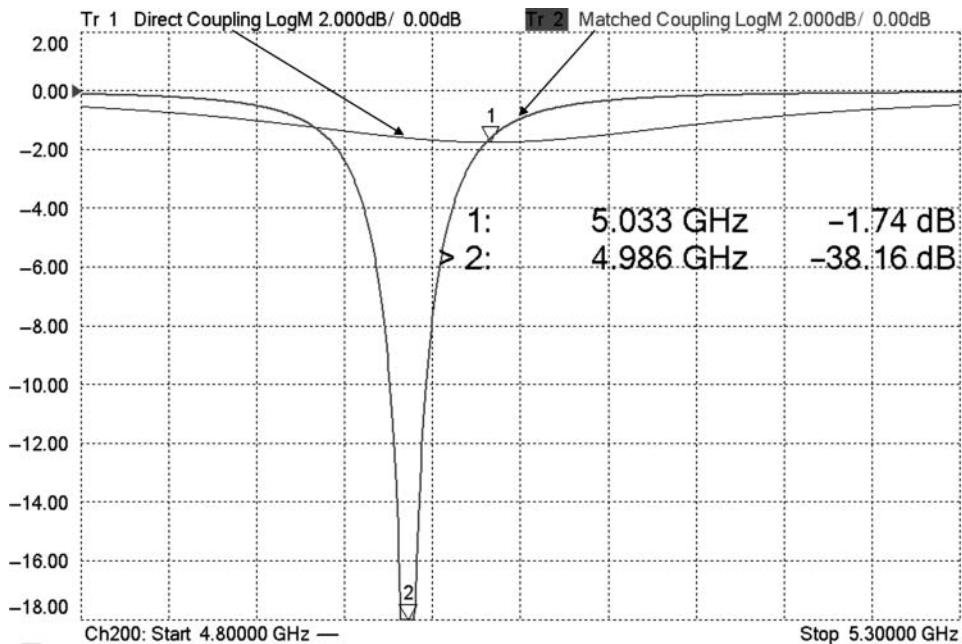


Figure 5.55 Return loss plot of a resonator with direct coupling and with matched coupling.

there may not be any point below 3 dB, so the concept of looking for a magnitude response fails to find a Q value. When viewed on a log magnitude return loss plot, the reflection from a high Q circuit is hard to see, as the return loss is very nearly 1 for all frequencies. Therefore, it is common to add a coupling structure, often a small capacitance, C_K , which transforms the impedance of the lossy elements to match the impedance of the test system. Figure 5.55 shows the return loss of a relatively low Q resonator with a direct connection and with a coupling capacitance added to match the circuit to Z_0 . If the resonator was of a higher Q , the direct connection trace would show almost no change in return loss and appear as a flat line of S_{11} nearly equal to 0 dB return loss.

When plotted on a Smith chart with a direct connection, the resonator forms a perfect circle and crosses the real axis at f_0 . But in real-world measurements, there is almost always some external transmission line that shifts the response of the resonator as shown in Figure 5.56. Also shown is the same resonator with the same value of coupling capacitance as used in Figure 5.55. From these Smith chart plots, the Q factor can be directly computed (Kajfez 1994).

The center frequency is found by looking for the point where the trace crosses over itself at the outside of the Smith chart. Drawing a line from this point through the center of the Smith chart marks the position on the trace trajectory that represents the resonant frequency, f_0 . From this line, a position is marked at $\pm 45^\circ$ angles and a line drawn from the crossing point, at these angles, until it crosses the Smith chart trace at frequencies f_1 and f_2 . One can use these frequencies to measure the loaded Q of the circuit as in Eq. (5.22). The diameter, d , of the circle formed on the Smith chart is a measure of the difference between loaded and

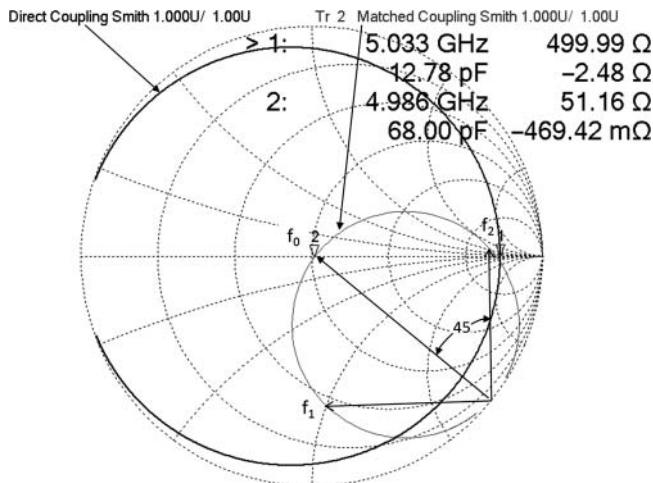


Figure 5.56 Smith chart plot of a directly connected resonator and one matched to Z_0 using a coupling capacitance. The Q is determined from f_1 and f_2 , taken as 45° from the diagonal chord to f_0 .

unloaded Q , and the unloaded Q can be computed from

$$Q_0 = Q_L \left[1 + \left(\frac{d}{2d-1} \right) \right] \quad (5.23)$$

Thus, measurements of unloaded Q can be made using the Smith chart trajectory without knowing anything about the value of the coupling capacitance.

5.5 Antenna Measurements

The topic of antenna measurements could encompass an entire book, from the area of near field and far field patterns to phased array measurements to antenna efficiency. However, for many RF uses, the key aspect in measuring an antenna is determining the correct frequency of operation of the antenna, and the principal way of doing this is through a return loss measurement.

One key aspect of antenna measurement is that the antenna to be measured must be sufficiently removed from any conducting surfaces that might reflect energy back into the antenna and cause the apparent impedance to change. The ground conductor of the test cable can also act to change the effective radiation of the antenna, and thus care must be used to ensure that an effective ground plane is used that emulates the mounting of the antenna in use.

The measurement of an antenna return loss is rather straightforward, requiring only a 1-port calibration. Figure 5.57 shows the measurement of a whip antenna at the end of a coaxial cable. This one is suitable for use at about 300 MHz. Also shown is a transmission measurement where port 2 is just a small pickup coil loop some distance away. It is interesting to note that the S_{21} peak is just off frequency from the antenna's apparent center frequency.

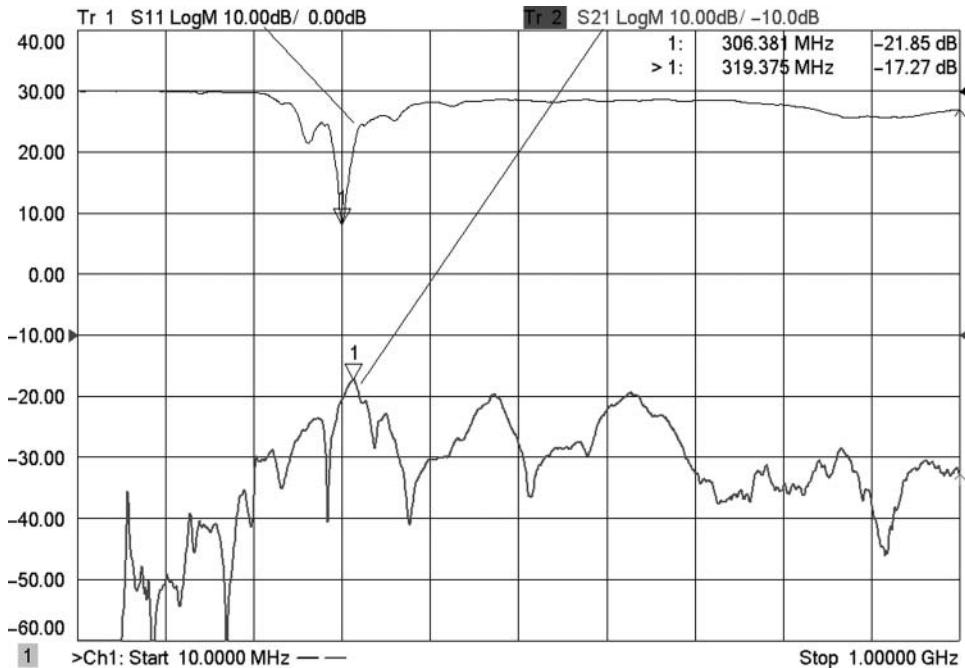


Figure 5.57 Measurement of an antenna return loss.

One source of uncertainty on the antenna frequency is caused by the directivity of the directional-coupler. While normally a directivity error is assumed to cause the return loss to be larger than actual, in fact it can add or subtract from the value of the return loss. In the case of an antenna measurement, the return loss causes a characteristic null at the tuned frequency of the antenna, but this happens only because that is the point at which the antenna's impedance matches that of the VNA's reference impedance. If there is a directivity error or if the antenna's effective impedance is different than Z_0 , the effective impedance of the VNA will be either lower or higher than the antenna as the frequency sweeps, and the null from the antenna will subsequently move lower or higher in impedance, as the impedance of an antenna is not fixed but varies as the frequency varies. Thus, the apparent frequency of an antenna depends entirely on the directivity, or effective impedance, of the VNA. Since the effective impedance of a VNA is set by the calibration, errors in calibration will lead to errors in the apparent tuned frequency of an antenna, as illustrated in Figure 5.58. Also, if the characteristic impedance of the antenna is not $50\ \Omega$, it will cause an apparent shift in the S_{11} null of the antenna.

In the figure, the lower plot shows the real and imaginary parts of the antenna impedance. If one chooses the reference impedance to equal the point where the imaginary part crosses zero, then the real part has a value of $72.7\ \Omega$. The upper plots show the antenna response for the both $Z_0 = 50\ \Omega$ and $Z_0 = 72.7\ \Omega$. Interestingly, the shift in the S_{11} null corresponds better to the peak in the S_{21} response seen in the upper-left window.

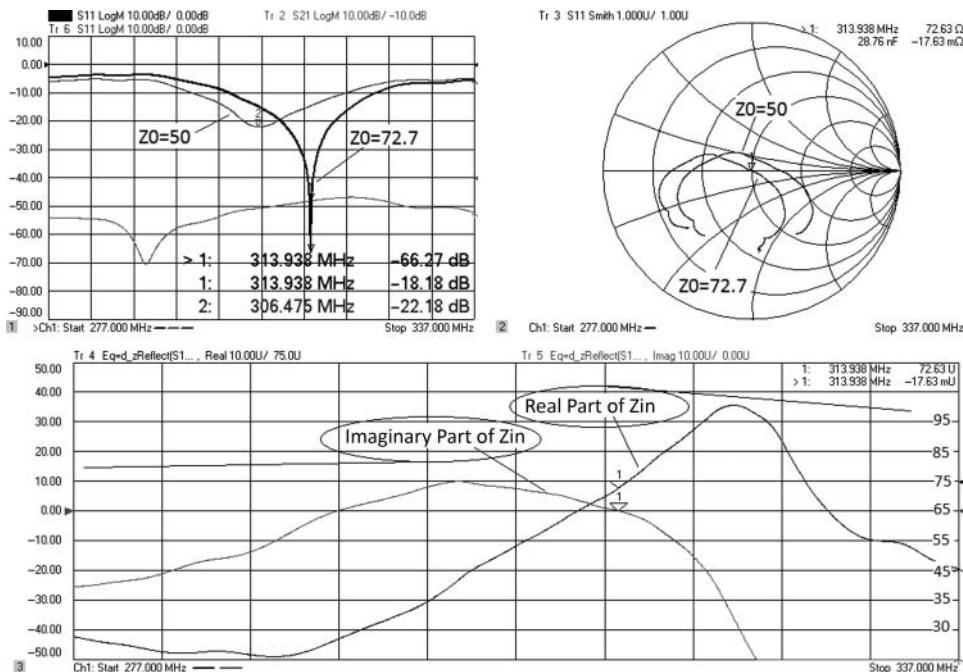


Figure 5.58 Change in apparent tuned frequency due to directivity errors or change of reference impedance.

5.6 Conclusions

While many aspects of testing linear devices have been discussed here, no one chapter can possibly cover every aspect of every linear device. However, attending to some key guidelines as outlined in many of these sections should allow extending the specific techniques described to almost any other situation.

A practice important enough to be re-emphasized here is the practice of pre-measuring a device before starting any calibration or serious measurements. Almost all modern VNAs have a built-in factory calibration that provides approximately correct results for most measurements. In pre-measuring a DUT, countless hours of wasted calibrations and measurements will be saved. During the pre-measurement, cable connections can be tested for stability, the insertion loss measurements can be evaluated for trace noise and appropriate averaging or IF bandwidths can be set, delay measurement aperture effects can be evaluated, and even unnecessary measurements on damaged or incorrect parts can be avoided.

The basic principle of measuring linear devices is to understand the system stability and noise, understand the best calibration to be applied, and understand the interactions between the measurement system and the DUT. This chapter provides the foundation for such an understanding.

References

- Keysight Technologies, Inc. (June 2002) User's Guide: Keysight Technologies 8753ET and 8753ES Network Analyzers [Online], Available at <http://literature.cdn.keysight.com/litweb/pdf/08753-90472.pdf>
- Keysight Technologies, Inc., 2002. Keysight Time Domain Analysis Using a Network Analyzer Application Note 1287-12 [Online], Available at <https://www.keysight.com/zz/en/assets/7018-01451/application-notes/5989-5723.pdf>
- Joel P. Dunsmore, Jan. 30, 2007, Transmission Response Measurement System and Method of Using Time Gating, Agilent Technologies assignee, Patent No. 7170297
- Kajfez, D. (1994). *Q Factor*. Oxford, MS: Vector Forum.
- Rowell, J., Dunsmore, J., Brabetz, L, n.d. Cable Impedance and Structural Return Loss Measurement Methodologies, Hewlett-Packard white paper, Available at: <http://na.tm.agilent.com/8720/applicat/srlpaper.pdf>
- Society of Cable Telecommunications Engineers, Inc. 2008a ANSI/SCTE 05 2008, Test Method for "F" Connector Return Loss In-Line Pair, available at http://www.scte.org/documents/pdf/Standards/ANSI_SCTE%2005%202008.pdf
- Society of Cable Telecommunications Engineers, Inc. 2008b ANSI/SCTE 66 2008: Test Method For Coaxial Cable Impedance, available at http://www.scte.org/documents/pdf/standards/ANSI_SCTE%2066%202008.pdf
- Wilkinson, E.J. (1960). An N-way hybrid power divider. *IRE Transactions on Microwave Theory and Techniques* 8 (1): 116–118.

6

Measuring Amplifiers

When considering active device measurements, most engineers naturally think of testing amplifiers. These devices form the heart of any communications, radar, or satellite transponder system, and are required in almost every other radio frequency (RF) or microwave system. Amplifiers are often considered in the system design as a unilateral gain block used to boost signals, and for the inexperienced system designer, that is often as far as the amplifier characteristics are considered. Unfortunately, RF and microwave amplifiers have complicated and subtle behaviors that challenge such a simple view. Amplifiers can be roughly split into two categories: low-noise amplifiers (LNAs) used in receiver applications, and power amplifiers used in transmitter applications. While amplifiers for each application might share many attributes and require many of the same measurements, the attributes of the tests and the relative importance of various factors of the test system vary greatly depending upon the particular requirements for each amplifier and application. A power amplifier used to drive the probe of a nuclear magnetic resonance (NMR) imager has very different requirement from an ultra-low-noise cryogenically cooled amplifier used in a space-based radio telescope. However, both have many similar issues including such things as stability, compression, power consumption, and distortion. This chapter will explore most of the attributes of amplifier measurements important to an RF or microwave designer or test engineer, with the exception of noise figure, which is discussed for both amplifiers and frequency converters in Chapter 9.

6.1 Amplifiers as a Linear Devices

The most basic attributes of an amplifier are often considered as purely linear; thus, the fundamental description of most amplifiers is the S-parameter matrix of the amplifier. However, amplifiers differ from passive linear circuits in that the power scattered from an amplifier will always be greater than the power incident on the amplifier. The basic definition to call a circuit an amplifier is that the RF power out of the device is greater than the RF power incident on the device, or

$$\sum |b_n|^2 > \sum |a_n|^2 \quad (6.1)$$

When applied to the forward signal of a 2-port device, this can be interpreted as

$$|S_{21}|^2 > 1 - |S_{11}|^2 \quad (6.2)$$

For amplifiers operating in the linear mode of operation, the incident and scattered waves are defined exactly the same as for passive devices, described by Eq. (1.17) and repeated here.

$$\begin{aligned} b_1 &= S_{11}a_1 + S_{12}a_2 \\ b_2 &= S_{21}a_1 + S_{22}a_2 \end{aligned} \quad (6.3)$$

The principal measurement performed on an amplifier is the small signal gain measurement, followed by the input match, output match, and isolation measurements. These are traditionally made in the same manner as linear passive devices, stimulating the device-under-test (DUT) one port at a time, measuring the resulting waves, and computing the S-parameter error-corrected results. But this method relies on the DUT being a linear device, so one might ask how to define the linear mode of operation.

6.1.1 Pretesting an Amplifier

For most simple devices, the linear mode for amplifiers as applied to Eq. (6.3) occurs at approximately 20 dB below the 1 dB compression point: there is approximately 0.1 dB compression at 10 dB below the 1 dB compression-point, and 0.01 dB compression at 20 dB below. At such a level, other practical limitations such as drift, of both the amplifier and the test system, render any distinction between linear and non-linear meaningless as far as computation of the S-parameters is concerned.

Figure 6.1 shows the relative gain of two different amplifiers for a power sweep at a fixed frequency, with compression markers shown at 1 dB (Marker 1), 0.1 dB (Marker 2), and 0.01 dB (Marker 3). One amplifier is a normal class A amplifier operating at a lower power, which follows the rule that compression drops roughly 10 times (in dB) for every 10 dB back-off from the 1 dB compression point. From this, it is clear that 20 dB back-off is a good estimate of the linear range in an amplifier.

The other amplifier is a high-linearity amplifier with about a 10 dBm higher compression point, designed so that the bias changes at higher powers to increase the gain, thus creating a more linear transfer curve. As is common in these types of compensated amplifiers, there is some expansion before compression occurs. Even here, the clearly linear region is about 20 dB below the 1 dB compression point. The x-axis of the plot is different for each amplifier; both are scaled at 4 dB/division, and the stop power is set just above the 1 dB compression point. By placing both on the same grid, it is easy to see the difference in compression curvature between the two designs of amplifier.

Before measuring the S-parameters of an amplifier in earnest, it is a good practice to try to ascertain some of characteristics such as its linear operating point first. This may be simply done by starting at a low power and measuring raw (uncorrected) S_{21} . If the amplifier is to be tested over some narrow frequency range, it is wise to extend this range to ensure that there are no frequencies at which the amplifier is unstable under the source and load conditions of the vector network analyzer (VNA) being used. Unstable operation will show itself as a large gain peak at some frequency, or possibly even a large spike in the normal gain plot if

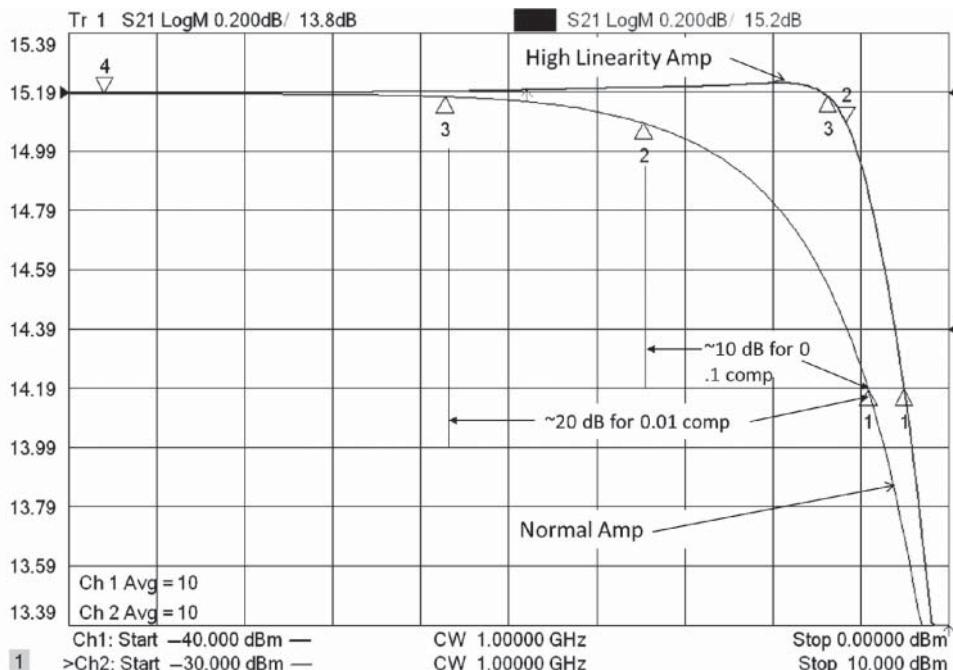


Figure 6.1 Relative gain of an amplifier versus compression level.

the device happens to be oscillating. These investigations can be done with wide bandwidths, no calibration, and a large number of points to quickly detect problems.

Once it is confirmed that the device is not actively oscillating when being measured, then the linear operating region may be determined. For this pretest, the VNA source power is set to a low level where it is expected that the amplifier must be linear, and the S_{21} trace is placed into memory, and the data and memory traces are displayed. The VNA source power is then increased, while a marker is used to track the minimum point on the trace. It is also useful to have a second trace of output power (typically using the $b2$ receiver) with a marker tracking the maximum output power level, monitoring to ensure that the power into the VNA receiver does not drive it out of its operating range. A third trace of input power may be added and provides a convenient trace of drive power that can be displayed using the trace markers. Continue raising the source power until a clear offset in the S_{21} trace occurs. The most common figure of merit is 1 dB compression, but in some cases it will not be possible to achieve 1 dB compression from the level of source drive available from the VNA. Figure 6.2 shows, in the upper plot, a trace of the wideband measurement of the S_{21} of an amplifier, at low power (for linear response, marked S_{21_Lin}) and at high power (for compressed response, marked S_{21_Comp}). In the lower plot is shown the results of looking at the input and output power, as well as the compression. The compression trace is the high power gain divided by the memory trace, which was saved at a linear power. All these measurements are before a correction and so rely on the VNA factory cal.

One object of particular interest in this figure is the gain peaking in the linear trace (upper plot) at the very low frequency range, around Marker 3. This shows that the gain peaks 4 dB

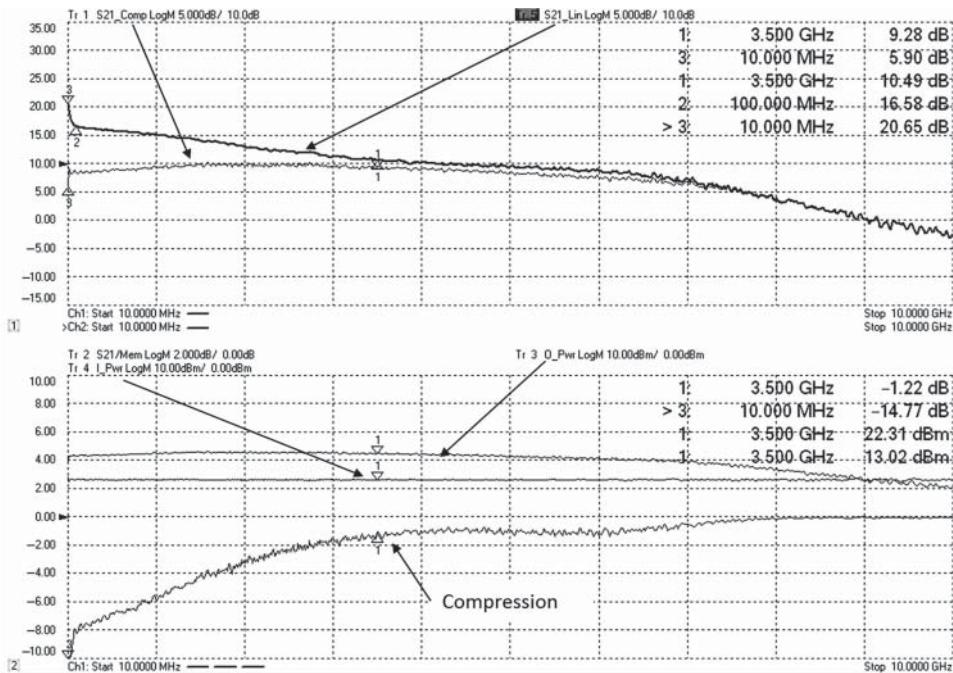


Figure 6.2 Amplifier pretest: a wide band sweep looks for instability, and changing the drive power looks for compression.

in only 90 MHz, likely indicating some feedback mechanism that is limited in low frequency response, probably by an AC capacitor. This wideband plot shows that the low frequency gain may be an issue if the low frequency matching is not properly considered.

6.1.2 Optimizing VNA Settings for Calibration

If investigations into the compression point of the amplifier are anticipated, now is the time to adjust source and receiver attenuators in the VNA to ensure sufficient drive power to drive the amplifier fully into compression (more than 1 dB compressed at all the operating frequencies) and to ensure that the VNA receiver is not compressed by monitoring the VNA output receiver's measured maximum value and ensuring that the receiver attenuation is set sufficiently to keep the power into the receiver below the top 5–10 dB of the VNA maximum operating power. In some VNAs, the receiver has automatic over-range detection and will post a message if the VNA receiver is overloaded. For high-power amplifier testing, where the overload might exceed the damage level of the VNA, the overload detection may be set to shut off the source power if overload conditions are detected. In the previous figure, the maximum power from the amplifier is about +23 dBm, and with an expected gain of about 10 (from the raw S_{21}), the maximum source power is about +13 dBm, meaning that the 0 dB source attenuation state must be used. The maximum operating power for the VNA receiver is about +13 dBm, so to ensure that the test power is at least 5 dB below this value, 15 dB of

receiver attenuation is required. For the amplifier shown, the source power for linear operation is about -10 dBm , approximately 20 dB below the compression point power.

While monitoring the S_{21} trace at the linear drive power, the IF bandwidth should be adjusted to ensure that the trace noise over the desired range of the amplifier is sufficiently small. Because the source power control of the VNA occurs before the reference coupler, the power in the reference channel might be at the low end of the receiver's normal range when the amplifier is in the linear operating region. At higher source powers, the trace noise reduces at approximately the rate of 3 times for each 10 dB increase in power. So near the compression point, the trace noise will be about 10 times smaller than at the linear operating region.

6.1.3 Calibration for Amplifier Measurements

Once the amplifier has been pretested and the VNA settings optimized, the calibration methods and acquisition can be performed to produce a calibrated instrument state providing the highest quality of measurements for the amplifier under test.

For most moderate gain amplifiers (gain below 20 dB), the calibration does not require any special care. Before calibration begins, it is wise to look at the S_{21} response of a thru, or the thru state of the electronic calibration (Ecal) if it is to be used, and determine if IF BW setting is such that the trace noise on the thru state is acceptable for the measurement results. The trace noise of the thru state will be embedded in the response calibration, and measurement results of the amplifier will never be smoother than this trace. For this reason, it is common to reduce the IF BW even more than needed for good amplifier measurements to obtain good calibration results (one should always use stepped-sweep mode if changing the IF BW, as discussed in Chapter 5). Perhaps a better alternative is to add averaging during the calibration. With averaging applied, the trace noise is reduced, and also some other, subtler, dynamic effects may be reduced. Trace deviation (drift of the S_{21} trace) can occur for many reasons such as non-coherent spurs in the source or receiver, external noise, environmental drift, or dynamic effects in the Ecal. Using trace averaging to reduce these effects during calibration may be more effective than reducing IF BW as the effects may come and go at a rate sufficiently slow that the sweep-to-sweep averaging will have an effect where the one-shot IF BW measurement will not.

6.1.3.1 Source Power Calibrations

If power measurements of the amplifier are desired, it is best to do the power calibration before the S-parameter calibration. Almost all VNAs support some form of source power calibration. The details of source power calibration are described in Section 3.6. For older legacy VNAs, the source power calibration is an independent function, which creates a source output offset table that would be applied to the factory source power setting to provide the correct power at the power calibration plane. A power meter is used to measure the power for each frequency, and the source power is normally iterated until the target power is reached within some tolerance. This source power setting is valid for power changes around the calibration region, but if the power is changed too much (more than perhaps 10 dB), source linearity errors can cause significant errors in the drive power. For this reason, it is usually wise to perform a reference receiver calibration immediately after the source power calibration, as the reference receiver

can be used to continuously monitor, and even level, the drive power during measurements. This will be discussed in detail in Section 6.1.4.1.

6.1.3.2 Receiver Power Calibrations

The reference receiver calibration can be simply a response calibration that is acquired on the reference receiver, with the source power calibration applied. In older VNAs, the response calibration presumed 0 dB for the reference point, so it was required that the source power calibration be performed at 0 dBm. Most modern VNAs have explicit receiver power calibrations that allow any power for the reference power and automatically use the last source power calibration value. For the reference receiver cal, it is best to leave the power meter in place to ensure the same match condition for the receiver calibration as the source calibration.

The VNA output receiver may be calibrated for power at this time, by providing a thru connection between port 1 and port 2 and acquiring a receiver response cal for the output receiver, typically the *b2* receiver. The details of receiver power calibration are discussed in Section 3.7. A simple response calibration method does not remove mismatch effects from the receiver response calibration so that if the VNA measurement system does not have good source or load match, the response calibration can have significant error. For example, consider a system with 15 dB source and load match at some frequencies (not untypical for a VNA with test port cables), the receiver response calibration would be in error by approximately 0.27 dB. If the DUT also had a 15 dB return loss, the receiver power measurement error could be as much as 0.52 dB!

6.1.3.3 Advanced Techniques and Enhanced Power Calibration

Sections 3.6 and 3.7 discuss some advanced power calibration techniques that greatly improve the source power calibration and the receiver response calibration. The basic idea is to make return loss measurements during the power meter acquisition so that mismatch effects can be removed from the source power settings. The reference receiver or *a1* power calibration uses the power meter mismatch and knowledge of the source-match to remove the mismatch effects from the reference receiver calibration. The step of calibrating the output or *b2* receiver is not performed explicitly. Instead, the *b2* receiver response calibration is computed by combining the *a1* calibration with the S_{21} tracking from the full 2-port calibration, yielding a match-corrected *b2* receiver response calibration. Given the VNA scenario described earlier, with 15 dB source and load match, a typical receiver response error would be less than 0.03 dB using full match correction.

These advanced techniques can be used on legacy VNAs by using the details of Section 3.7 to apply the mathematical corrections to error terms acquired during the power calibration and S-parameter calibration using remote programming commands, and downloading the corrected receiver response arrays to the VNA calibration memory.

The chief difficulty is in situations where the power meter cannot be connected directly to the VNA, such as in the case of an on-wafer or in-fixture measurement.

Fortunately, some of the most modern VNAs now support a guided power calibration as part of the normal calibration engine. As part of these advanced calibrations, even the situations where the power meter cannot be directly connected to the port are nicely handled

by adding an additional calibration step. If the power meter cannot be connected to the measurement port, such as an on-wafer probe, or in the case where only a coaxial power meter is available for calibrating a waveguide system, the user is prompted to add (or remove) an adapter to allow the power meter to be connected. In the case of an on-wafer measurement, the coax connection from the VNA to the probe might be removed and the power meter connected directly to the VNA. In the waveguide case, a waveguide-to-coax adapter might be added to the test port and the power meter connected to the adapter. After the source power calibration, the user is prompted to perform a 1-port calibration at the power meter measurement plane using the coax standards appropriate for that power meter. From the information of this 1-port cal and a later 2-port cal at the measurement plane, the adapter's full characteristics are determined and removed from both the source power calibration and the receiver calibration.

It should be noted that it does not matter at all if an adapter is added or an adapter is removed; the mathematics work exactly the same. In the case of an added adapter, the de-embedding network computed has loss, as would be expected. In the case of removing the adapter, the de-embedding network has gain. In fact, an adapter may be both removed and another one added; for example, in an on-wafer probing situation, an RF switch may be added to the cable path from the VNA to the wafer prober, with the switched port set to apply the power meter and corresponding 1-port calibration. The adapter from the probe to the switch is effectively removed and any cable from the switch to the power meter is added, but the de-embedding network computed for transferring the power cal becomes the difference (in a full S-parameter matrix manner, including mismatch) between the two paths.

With these recent improvements in power calibration techniques, the enhanced power calibration should be considered the new gold standard in power measurements. In fact, one can demonstrate that except for the case of a DUT having both a nearly ideal output impedance and an output power near 0 dBm, the VNA measured power using these calibration methods is more accurate than a power meter, in measuring output power. And if the DUT has a high harmonic content, the error in power meter measurements versus VNA measurements may even be greater.

6.1.3.4 S-Parameter Calibrations for Amplifiers

After the power calibration, or as part of a guided power calibration, the S-parameter calibration is performed. It is important in legacy VNAs to perform the power calibration first to avoid a C? or C Δ indicator as the source power calibration level is one of the stimulus attributes that signal the indicator change. In fact, for some VNAs, if the S-parameter calibration is performed first, then the source power cal and then the receiver calibration, the stimulus difference between the S-parameter cal and the receiver cal will cause one of them to be marked invalid in the cal arrays.

If the amplifier is to be measured only in its linear region, there is no real need to perform a source power cal as most VNAs have built-in factory calibrations on the source that give a reasonably flat and correct power. Some older VNA systems, such as the HP 8510 or Wiltron 360, used external sources so that the power at the test port was substantially different than the source power settings. Care is needed in such a case to ensure the proper power is set at the test port.

S-parameter calibration for amplifier test is relatively straightforward. Any of the techniques such as Ecal, SOLT, thru-reflect-l (TRL), or unknown thru may be used. In the case of Ecal, care should be taken that the drive levels to the Ecal don't exceed its recommended levels, which might occur in the case of a high-power amplifier, where the drive power is set quite high. This is even a concern for the SOLT calibration where high power might cause heating in the load element, and more details on calibration issues for high power cases are discussed in Section 6.4.

One method to reduce noise while calibrating is to increase the source power during calibration to the level near that which the amplifier would produce at the test receiver. It was the common wisdom of users of older VNAs that the source power should not be changed after calibration, or a questionable cal (C? indicator) would appear. Common, but wrong. The C? indicator showed that a change in power was made, but in most cases, the measurements that resulted would be more accurate than ones made at lower power levels.

To understand the trade-offs, consider the two competing effects that change measurement accuracy with changing power levels. First is dynamic accuracy, which is another way of stating the ability of a receiver to measure correctly a change in power level. If the calibration is performed at a higher power level and the source is then set to a lower level, the power in the reference receiver has changed and so it will have a dynamic accuracy error. However, in the case of an amplifier measurement, the *b2* or output receiver will have the same power as calibration, and so it will have no dynamic accuracy error. On the other hand, if the calibration is performed at the lower level, the *b2* receiver will have a change in power during the measurement due to the amplifier gain and pick up a dynamic accuracy error. Further, most VNAs have more attenuation in the reference receivers than the test receivers, so the test receivers typically show more dynamic accuracy errors than do the reference receivers. Thus, the first trade-off is no trade-off at all; the measurements will always be more accurate on an amplifier if the source power is set near the expected output power of the amplifier, provided it does not exceed the power handling capabilities of the calibration standards.

The second trade-off is also no trade-off; higher source power during calibration will yield lower noise during calibration, period.

Unfortunately, most VNAs still report at least a $C\Delta$ indication, which concerns many users if they see it after calibration. On older VNAs this can be overcome by starting the calibration in with the low source power settings, which will be saved as the stimulus value, and then raising the power level or adding averaging and resuming the calibration. One caution with calibrating at a higher power: after the power is lowered for measurement, if the calibration is re-applied (turned off and back on again), it can sometimes reset to the original calibration power. To avoid overdrive in high-power situations, it is often wise to turn off the RF source power prior to making any changes and turn it on again only when one has confirmed the proper settings have been made.

Finally, on some VNAs, there is a remote command that allows one to simply copy the calibration files to a new calset that has the current stimulus settings applied. This allows the new calset to be used with whatever stimulus changes the user desires (source power, IF BW, averaging, even interpolation, fixturing, or de-embedding) as though there were no change. This is sometimes called "flattening the calibration" as the effects of fixturing and de-embedding are notionally stacked on top of the original calibration, and the fixture math "pushes" them down into the calibration array. Creating a new calibration array with these attributes embedded has, in a sense, flattened the calibration and fixturing state. In fact, at least

one VNA vendor even uses the word *FLATTEN* in the remote command. This has the added benefit of embedding the new stimulus in the calset and avoids the unfortunate situation of turning on a calibration and having the high power used for calibration destroy one's amplifier.

Calibrations should be done at the densest point spacing expected for any future measurements. If measuring a wideband amplifier and one anticipates some need to zoom in to inspect a small region, the calibration point spacing should be sufficient to allow good measurements in any region. Interpolation of the calibration can be used if the point spacing is sufficiently dense to allow an accurate acquisition of the VNA raw response. As a rule of thumb, the point density should be greater than 200 points per GHz per meter of test port cable with a minimum of 200 points per GHz span. For example, a 6 GHz span with normal (less than 1 m) test port cables should have 1200 points in order to use interpolation. At this point spacing there is about 12° phase shift between points so that circular interpolation generally works well and the error from interpolation is on the same order as the uncertainty from the calibration standards. One caution about interpolation is that some VNAs change characteristics over different bands, and the change at the band edge in the raw response is abrupt. Many VNAs have factory calibrations to remove most of this effect, but it can still be seen when using interpolation, as small discrete errors.

6.1.4 Amplifier Measurements

Once the calibration is complete, the basic S-parameter measurements may be directly made. For simple testing, this is all that is required. More complicated testing might include power measurements, compression measurements, and voltage and current measurements to compute efficiency, as well as other changes in test conditions. Further analysis may be performed on a combination of these measurements, which is discussed more fully in Section 6.1.5. Some particular details of measurements are discussed next.

6.1.4.1 S-Parameters, Gain, and Return Loss, Input Power and Output Power

The S-parameters are the most common measurement of amplifiers with S_{21} or gain being the principal defining feature of an amplifier. For the most part, all four S-parameters are important for further analysis, and on most VNAs, all four are measured and saved as part of the full correction. It's convenient to create multiple windows to group various amplifier measurements. For example, S_{11} and S_{22} are often displayed on the Smith chart, while S_{21} and S_{12} are displayed principally in a LogMag format. It may be convenient to show the gain and isolation on separate grids as the gain is often viewed at small scale (e.g. 1 dB/div) to look for ripple, but the isolation is typically viewed at 10 dB/div or more.

When measuring power, it is good practice to show both the measured input power and output power. The source power setting should not be relied upon for giving the input power as it is not typically corrected for mismatch effects. The output power should ideally be the input power times the gain, and so should have the same ripple as the gain, but error in the input power setting will cause additional ripple in the output power. Figure 6.3 shows a typical measurement configuration with the S_{11} and S_{22} plotted on a Smith chart, isolation on a large Log-scale plot, and gain, input power, and output power on a separate LogMag plot with a smaller scale.

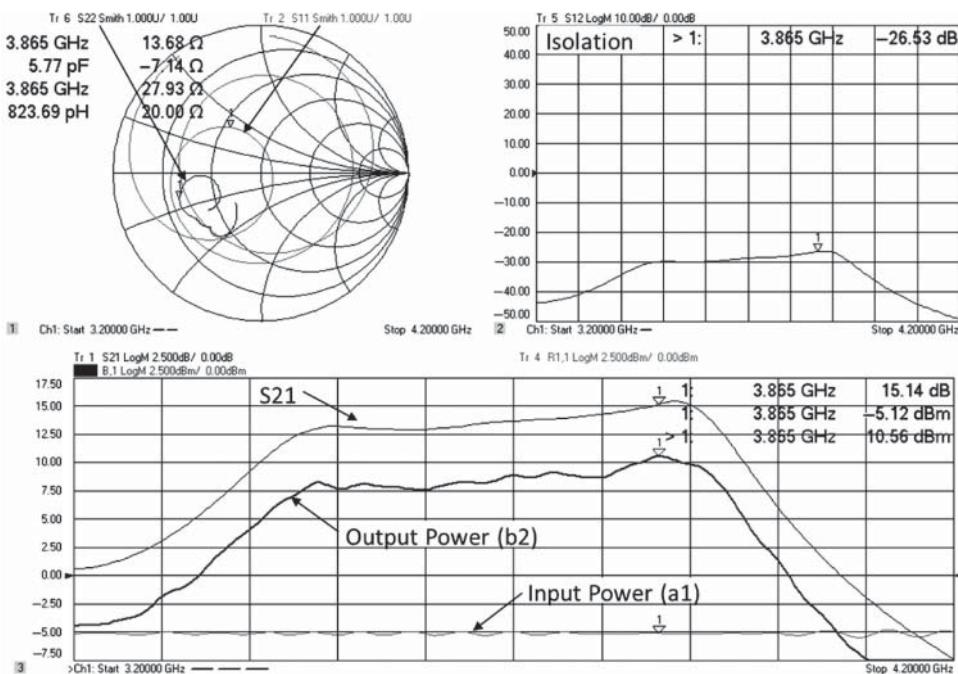


Figure 6.3 Typical plot showing S-parameters, gain, isolation, input, and output powers.

This measurement illustrates the results using a simple source and receiver power calibration; the measurement is performed at nominally -5 dBm input power. The input power and output power show substantially more ripple than the S_{21} trace. The S_{21} trace, of course, is fully corrected for mismatch errors, but neither the input nor output power is corrected for mismatch.

Enhanced power calibration provides for input and output match correction of the power measurements, but the results from VNAs without such capabilities can be manipulated to provide the same corrections provided that the source and load match terms are available. In modern VNAs, there often exists the ability to display one or more error term results in the VNA display. If so, the built-in equation editor function can be used to apply match correction to the result using the formulations in Eqs. (3.65) and (3.70).

For measurements of input and output power, the equations cited can be applied in the equation editor or in offline analysis. For more modern VNAs, the match-corrected power measurements are available as a part of a Guided Power Calibration and are displayed along with the corrected S-parameters. An example of such a Guided Power Calibration is applied in Figure 6.4 to the amplifier from Figure 6.3 to give an improved measurement of input power (IPwr) and output power (OPwr). In the upper window, three traces are shown. One trace shows the gain computed by taking the output power divided by the input power (labeled OPwr/IPwr) is very nearly the same as the match corrected gain of S_{21} . The slight difference is due to the fact that the “loop” term of $S_{21} * S_{12} * \text{ELF} * \text{ESF}$ is not fully corrected for, although the use of 1-port corrected Γ_1 and Γ_2 compensates for most of the effect of this loop term;

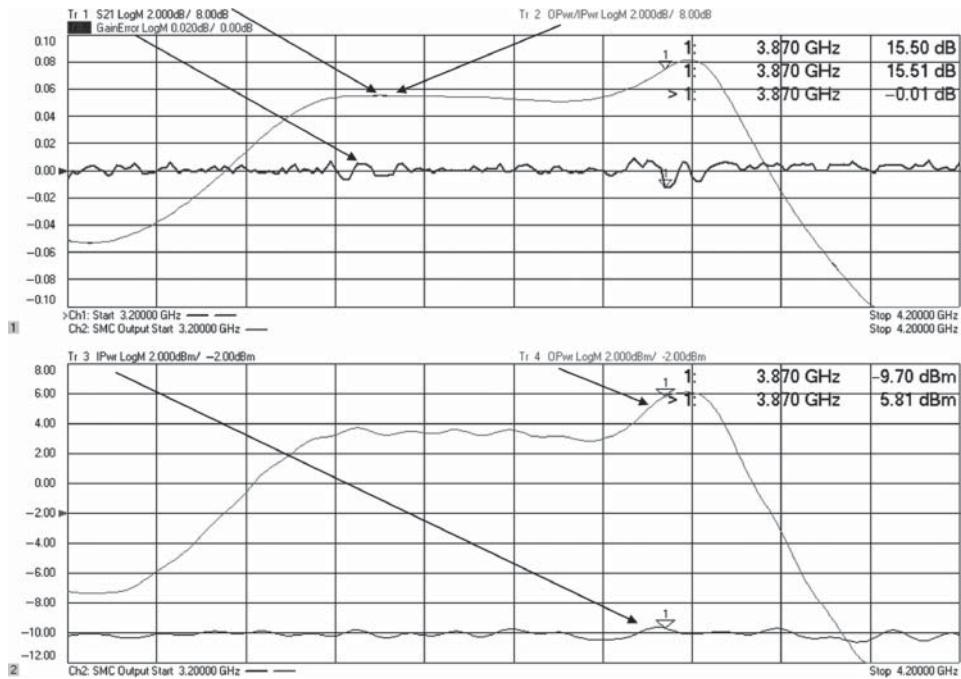


Figure 6.4 Match-corrected power measurements of an amplifier.

with S_{12} being small, the error is small indeed. This error is shown as “GainError” in the plot and is less than 0.01 dB. At the scale of 2 dB/division, the comparison of S_{21} and the gain computed from OPwr/IPwr are indistinguishable.

The output power (OPwr in the lower window) trace still shows some excess ripple when compared to the S_{21} trace, and this is directly due to the input power ripple, seen in the lower plot of Figure 6.4. This ripple in the input power occurs even though the source power calibration was performed at 0 dBm, with virtually no error. However, when the amplifier is connected to port 1 during the measurement, even at the same source-power setting, the input power shows ripple because a different match is applied to port 1: the effective input match of the amplifier terminated in port 2 of the VNA, or Γ_1 .

From Eq. (3.65), we see that the input power correction properly compensates for this error, and the input power displays the actual incident power, but that power is not exactly 0 dBm. Instead, it has ripple associated with the amplifier input match and the source power match (see Section 2.2.2). The source ALC does not respond to the change in match – either from the power meter used in calibration or to the DUT input match – in a way that completely compensates for the mismatch, so a ripple in the actual incident power occurs. This ripple in the input power causes an equal ripple in the output power which is measured at the port 2 receiver. Both of these power measurements are exactly correct, as the incident power is really changing and the output power is really changing, so the gain computed as the ratio of the two is almost exactly the same as the fully corrected S_{21} . However, it is desirable that the incident power exactly match the source power setting, and for this an additional

measurement adjustment must be made in the case of devices with even moderate reflections at the input port.

The cause of the ripple is that while the reference receiver does properly detect the change in incident power due to DUT mismatch, the source ALC loop does not. This can be corrected for by using the *a1* receiver, with its match correction, as the reference for the source ALC loop rather than the internal source detector. A function, sometimes called *receiver leveling* or *Rx-leveling*, provides such a capability. This can be performed either in hardware, by adding a detector to the *a1* receiver path, or in software by iterating the source power until the desired level is achieved. In either case, the detected signal is only the raw reading, and some correction must be applied to ensure the *a1* leveling provides the proper output level. In software Rx-leveling, match correction to the *a1* receiver reading can be easily added so that the source power is adjusted to provide a true incident power. And because the *a1* receiver is used as the source level detector, the power source-match and the ratio source-match are identical, meaning the incident power level is exactly corrected.

Figure 6.5 shows an example where the Rx-leveling method is applied to the amplifier from the previous figure. The input power is now nearly exactly -10 dBm (within the tolerance setting of the software Rx-leveling function), and the output power has almost the same shape as the gain or S_{21} , offset by 10 dB. Of course, the output power will differ from the gain since the source power is not 0 dBm. The gain computed by taking output power divided by input power (marked OPwr/IPwr in the figure) matches the S_{21} gain almost exactly.

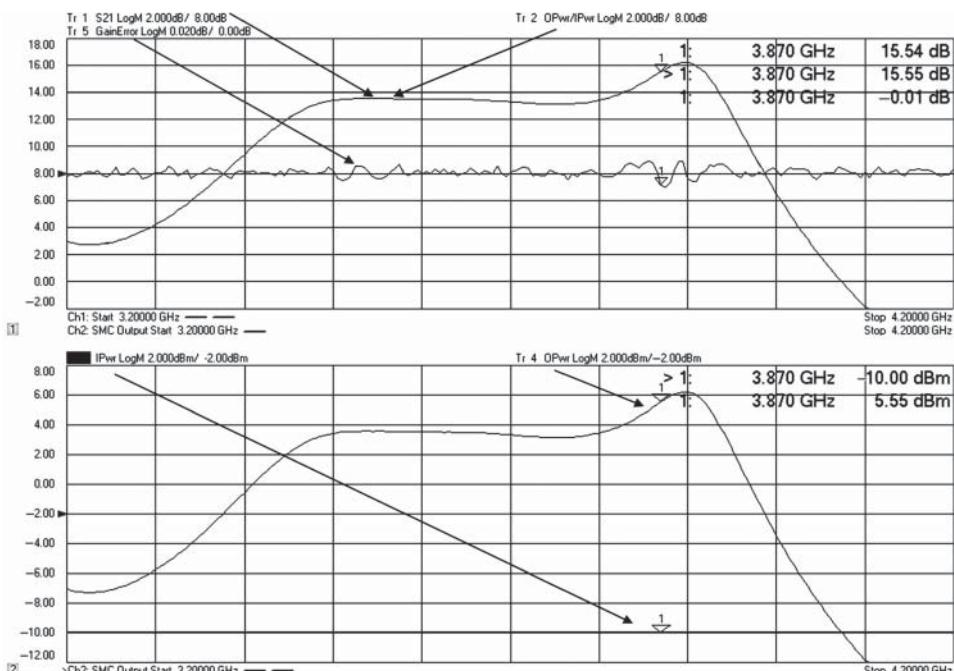


Figure 6.5 Match-corrected powers with software Rx-leveling.

This final figure demonstrates the highest level and performance of calibrated amplifier measurements with input power, output power, and gain all fully corrected for test system mismatch, and source power set exactly to the desired incident power. Note that the gain error, while very small (less than 0.01 dB), remains the same as in the previous figure. This gain error represents the fact that while the output power is corrected for load match of the VNA and the incident power is corrected to provide the same power as a perfect $50\ \Omega$ source, there is a re-reflected term that is not fully accounted for, but it is very small.

6.1.4.2 DC Measurements

A final set of measurements on amplifiers are often required to evaluate the efficiency of the amplifier. These are measurements of the direct current (DC) voltage and DC current into the bias pins of the amplifier under test. While it is common to just measure the DC at a single frequency or power using a DC power analyzer or source measurement unit (SMU), the actual DC power consumption of an amplifier can change over its frequency range or power range, if it is in a non-linear mode of operation. Almost by definition, the DC operating point cannot change if the amplifier is operating in a linear region, as the definition of linear implies that the signal level is so small as to not affect the operating point of the amplifier.

However, many amplifier test scenarios are with the amplifier in or near compression, so in such cases the DC operating point must be measured at each RF frequency point and power point.

Some modern VNAs provide integrated DC measurements as built-in functions. The DC readings of voltage and current can then be displayed simultaneously with the RF input and output power measurements. Figure 6.6 shows a typical connection scheme for measuring the DC power consumption of an amplifier. In this case, the VNA has two built-in analog-to-digital-converters (ADCs) that can measure two analog input signals (AI1 and AI2).

Some voltage scaling resistors are needed if the DC meters have limitations on maximum voltage. The current sense resistor should be scaled to give a reading sufficiently large to provide a difference voltage that is easily detected by the DC meters. The AI2 signal is the

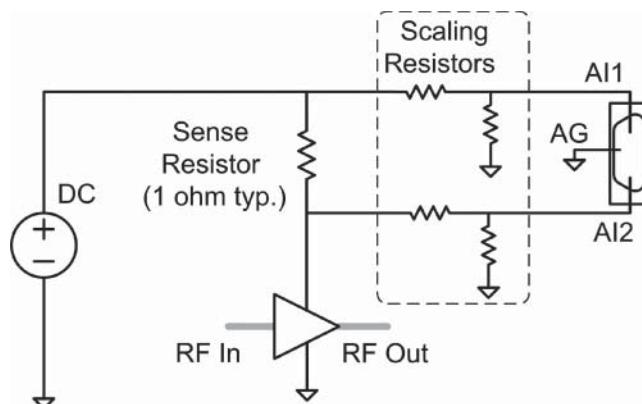


Figure 6.6 Typical configuration for measuring DC power consumption.

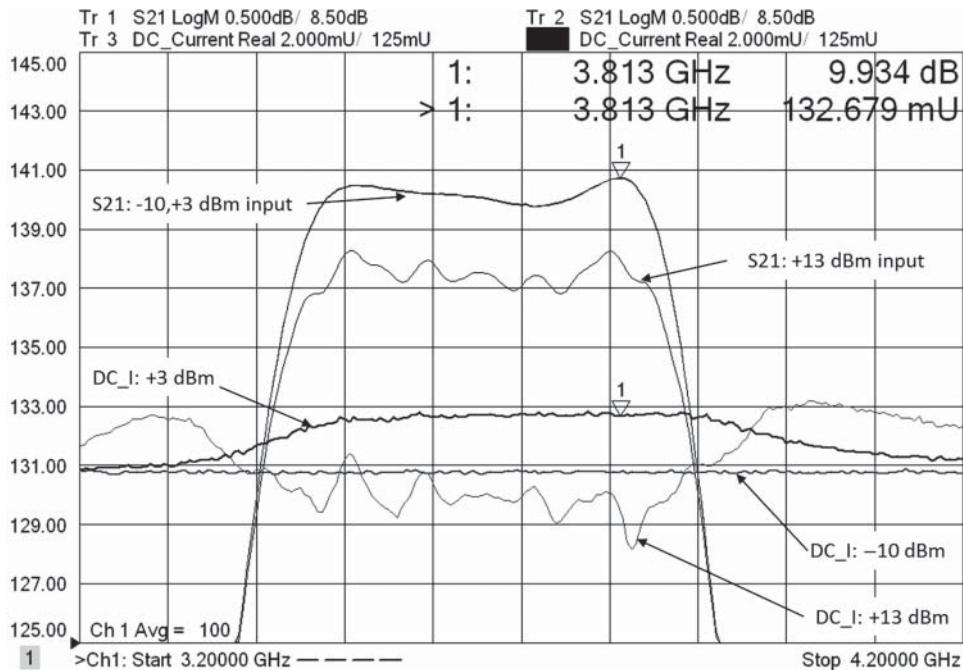


Figure 6.7 Measurement output power and DC current, for three different input power levels.

voltage drive to the amplifier. The DC current is computed using the equation editor function by taking the difference in AI1 and AI2 and dividing by the current sense resistor. If the current requirement of the amplifier is approximately known, the offset from the current sense resistor can be programmed into the DC power supply. If not, then the AI2 voltage can be measured and DC power supply adjusted to provide the correct DC supply to the DUT.

An example measurement is shown in Figure 6.7 displaying the gain (S_{21}) and DC current of an amplifier for three different RF input power levels.

In this case, the current consumption changes with output power so that the efficiency also changes. It is interesting to note that while the S_{21} trace doesn't perceptibly change between a drive level of -10 and $+3$ dBm, the DC does change indicating that the power of $+3$ dBm is sufficient to increase the DC operating point. At higher drive power, near the 1 dB compression drive, the DC current actually drops below the level of a small signal bias point.

These DC measurements may also need some calibration to account for offset or gain errors in the DC meter or for offsets from the measurement setup. For example, if a DC-resistive divider is needed to limit the voltage to the DC meter, then the current consumption of output divider needs to be subtracted from the current reading (the current draw of the input divider does not affect the voltage across the current sense resistor). Further, if there is DC resistance, from the voltage sense at the output-side current sense resistor to the amplifier bias input, then the voltage drop associated with the current passing through that resistance also needs to be compensated for. This compensation would normally need to be iterative, if a particular output voltage is required at the DUT bias pin.

6.1.4.3 Conclusions on Amplifier Measurements

The principal measurements of amplifiers are output power, input power, gain, isolation, input and output match, and DC power consumption. All of the RF measurements normally require some form of correction for systematic error, and the most modern of VNAs provide full match correction for every RF measurement, as well as integrated DC measurements. From this set of basic measurements, many other derived parameters and results can be computed as part of a data analysis. Some examples of post-measurements analysis are discussed in the next section.

6.1.5 Analysis of Amplifier Measurements

While gain, or S_{21} , is the principal measurement of an amplifier, many other parameters have very practical application in the design process and in understanding the amplifier's effect on system performance. Many useful results can be obtained by applying additional analysis to the basic measurement parameters acquired in the previous section. These are mathematical reformulations of the parameters to obtain insight into the device. Some select analysis functions are described here with examples for many of them.

6.1.5.1 Stability Factors

For engineers having only a casual familiarity with RF circuits, gain and S_{21} are used interchangeably. This imprecise usage works because many systems utilize matched gain blocks that have S_{11} and S_{22} very near zero. In fact, many hours of design are absorbed with the tasks of producing both the desired gain and power from an amplifier while at the same time providing a matched condition. However, many more hours of troubleshooting occur when the true nature of the amplifier is not understood and both in-band and out-of-band matching conditions provide issues with the overall design.

When placed in a circuit, the behavior of an amplifier can be rewritten as in terms of various gain attributes such as maximum available gain when conjugately matched as

$$G_{MA} = \frac{|S_{21}|}{|S_{12}|} \left(K \pm \sqrt{K^2 - 1} \right) \quad (6.4)$$

where the sign is negative if $1 - |S_{22}|^2 + |S_{11}|^2 - |S_{11}S_{22} - S_{21}S_{12}|^2 > 0$ and K is the stability factor (Vendelin 1982). For many designs, especially at high microwave and mm-wave frequencies, the goal is to achieve this maximum gain, and since the design is often defined for matched source and load conditions, the design task is principally creating a matching network, or transformer, which changes the (normally) 50Ω reference impedance into the conjugate match of S_{11} and S_{22} . However, this matching condition is true only if the amplifier is unconditionally stable, that is, if $K > 1$. Thus, measuring the stability of an amplifier is one of the key linear measurements.

In many modern VNAs, the stability can be computed and displayed on the screen directly either using a built-in function or using equation editor math on the S-parameters. The stability function most commonly used is the K-factor, which is defined as

$$K = \frac{1 - |S_{11}|^2 - |S_{22}|^2 + |S_{11}S_{22} - S_{21}S_{12}|^2}{2|S_{21}S_{12}|} \quad (6.5)$$

An amplifier is unconditionally stable if $K > 1$, and

$$\begin{aligned} |S_{12} \cdot S_{21}| &< 1 - |S_{11}|^2 \text{ and} \\ |S_{12} \cdot S_{21}| &< 1 - |S_{22}| \end{aligned} \quad (6.6)$$

the latter two conditions are met if $|S_{11}S_{22} - S_{21}S_{12}| < 1$. These last conditions, often written as

$$|\Delta| = |S_{11}S_{22} - S_{21}S_{12}| < 1 \quad (6.7)$$

is sometimes forgotten but is an important criteria and becomes very significant in out-of-band amplifier cases where S_{11} and S_{22} can become quite close to one, due to tuning conditions.

Both conditions must apply; if so, the amplifier is called *unconditionally stable*, and no combination of source and load impedances will cause the amplifier to oscillate.

If the amplifier does not meet the stability factor conditions, it is often erroneously called *unstable*. More correctly, such an amplifier should be called *conditionally stable*; that is, the amplifier will not oscillate under some conditions of source and load impedances.

A conditionally stable amplifier is measured in Figure 6.8. The upper-left window shows the S_{11} and S_{22} on a Smith chart. It is clear that S_{11} is poorly matched. The upper-right window shows the gain and isolation, S_{21} and S_{12} . The lower-left window shows both the K-factor (dark trace) and Δ term (called *delta*, light trace), with limit tests for values. The K-factor fails the stability criteria; thus, this amplifier is conditionally stable. In the lower plot, the

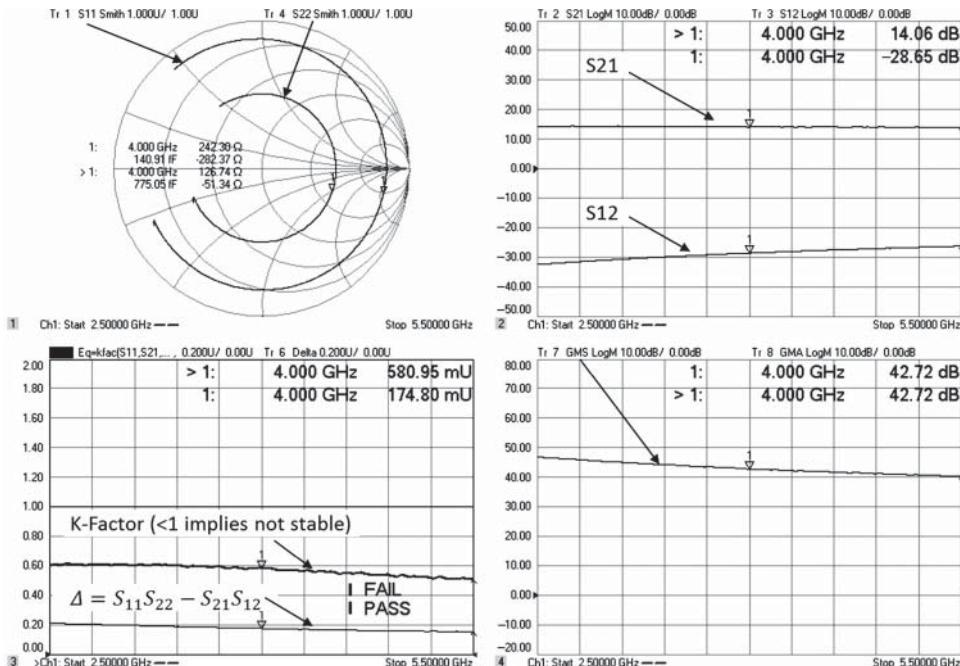


Figure 6.8 S-parameters, K-factor, and max stable gain.

GMS is computed as $|S_{21}|/|S_{12}|$ (called *maximum stable gain*) and is displayed along with G_{MA} based on Eq. (6.5); but here G_{MA} is set equal to GMS as $K < 1$.

A common situation is to measure the four S-parameters and save the S2P file for further investigation using some off-line simulation tool. However, using the built-in features of many modern VNAs, a great deal of investigation of other circuit topologies and matching scenarios can be applied in real time.

From the Smith chart trace of S_{11} in Figure 6.8, it is clear that this might be a parameter that affects K to a great extent, since it is so large. One aspect of a conditionally stable design is that it generally cannot be made stable by reactive matching alone; typically some resistive matching is needed.

In the case of the circuit from Figure 6.8, a small amount of series and shunt resistance was added to the input port using the Fixture Simulator function called *port matching* (see Figure 5.10). Port matching gives several circuit topology choices plus the option to import an S2P file to create a matching network between the test port and the DUT. In this case, adding about 10Ω of series resistance and about 330Ω of shunt resistance improved the K factor dramatically. The resistor values were chosen to keep the K factor above 1.5 everywhere. The result of this fixture simulation is shown in Figure 6.9.

This does have the effect of lowering the maximum gain (shown along with the computed GMS value which is identical to the previous figure), and the input and output matches are still not very good.

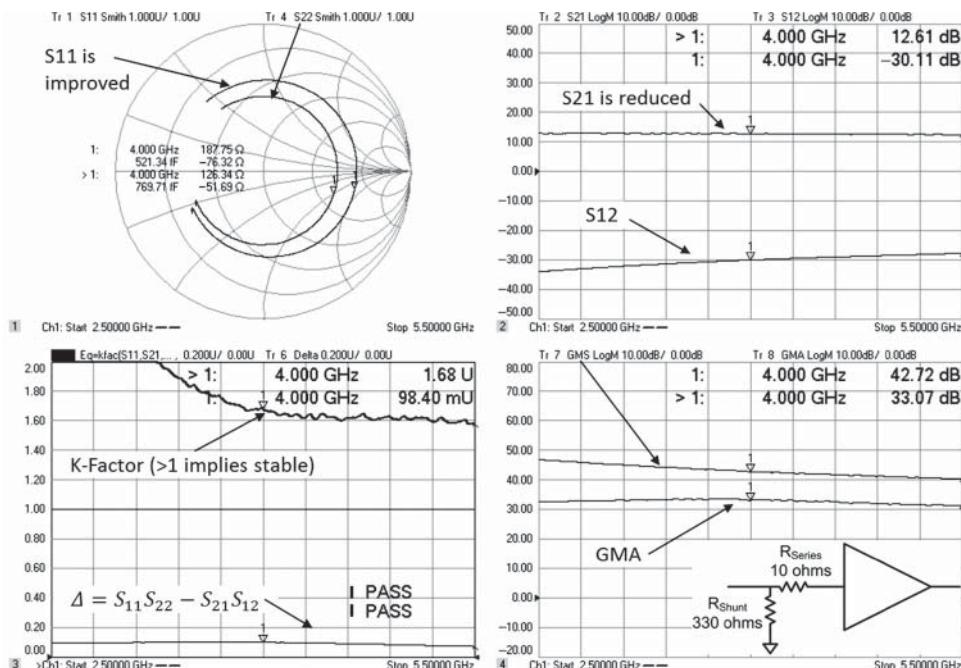


Figure 6.9 Resistance is added to the input network to improve the stability.

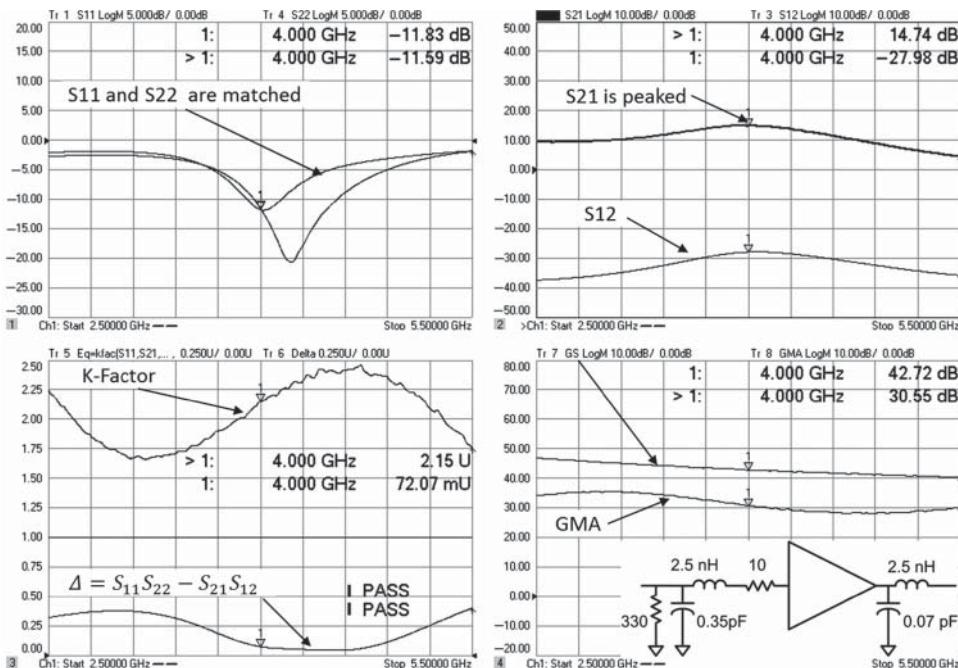


Figure 6.10 Circuit response after matching.

The reactive elements in the fixture simulator can be utilized to improve the reactive matching. From the G_{MA} value, which is much larger than the S_{21} , there is sufficient room to make matching choices that would allow an increase in the gain. Here, the matching is changed to recover the gain lost from the stabilization of adding the resistive elements. After some trial and error, final values were obtained, and the overall response is plotted in Figure 6.10. The matching that was employed has the effect of peaking the gain at the marker value above the nominal 50Ω gain obtained in Figure 6.8. The stability factors are well away from any limits, and the input and output match are tuned to a reasonable value.

While this is not an amplifier design text, it's useful to note that with modern VNAs, the normal workflow of measurements and modeling can be combined into real-time analysis. This also facilitates other investigations such as K-factor versus bias.

6.1.5.2 Stability Circles

In many cases, it is not desirable to use resistive matching to improve the stability. Resistive matching can degrade the noise figure at the input of an amplifier and degrade the maximum output power at the output of an amplifier. To avoid these degradations, designers often choose to use a conditionally stable device but must find the proper impedance transformations to ensure they are stable in operation. Since most amplifiers are used in a matched environment, the stability criteria is to add matching elements so that Z_0 is a stable impedance for both input and output ports.

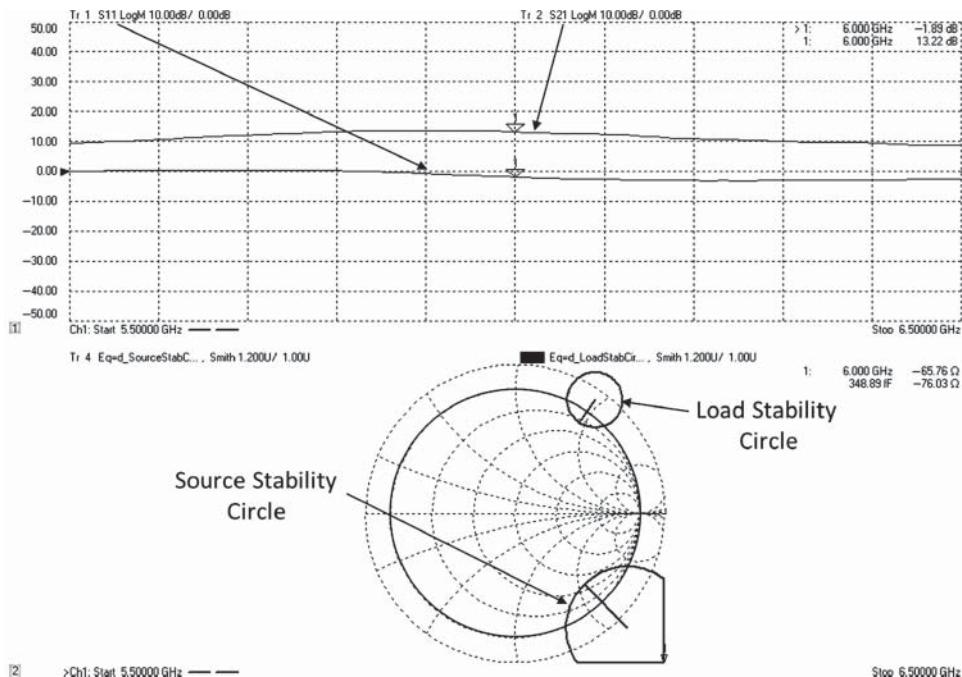


Figure 6.11 Stability circles at the center frequency.

Stability circles are often used to assist in determining the proper matching. These circles lie on the locus of impedance points where the amplifier circuit is just on the brink of instability. On one side of the circle the amplifier is stable; on the other side it will oscillate if the both ports of the amplifier are terminated in unstable areas. These circles can be drawn for both the input and output ports. Circuits that are designed to be oscillators have a stability circle that covers nearly the entire Smith chart, with Z_0 inside the unstable region. Circuits designed to be amplifiers strive to have the stable region as large as possible, and of course encompassing the Z_0 point, or center of the Smith chart.

In the past, creating stability circles was quite tedious, and simulation tools were used to create sets of stability circles for frequencies across the simulation bandwidth. In modern VNAs, one can once again turn to the equation editor function to create stability circles directly on the response traces of the VNA (Agilent Technologies n.d.), as shown in Figure 6.11, with the upper window showing the S-parameters (S_{11} and S_{21}) and the lower window showing the stability at the center frequency.

For the stability circles, the line to the center indicates an unstable region. This amplifier is conditionally stable, and a high reflection on either the input or output port, with the correct phase, could cause it to oscillate.

6.1.5.3 Mu Factors

An alternative stability factor has been derived that is finding favor with many designers based on a geometric analysis of stability circles. The mu factor is essentially a measure of

the distance from an unstable region to the center of the Smith chart. If the distance is greater than 1, the amplifier must be stable, as the nearest unstable region is outside the Smith chart.

There are two mu stability factors, originally called μ and μ' , but more recently called μ_1 and μ_2 or μ_1 and μ_2 . These are the input and output mu factors, and they are related to the maximum distance for stability of the input or the output load. The factors are related, and if either one is greater than 1, the device is unconditionally stable. If they are less than 1 but greater than zero, the device is conditionally stable including the matched Z_0 point. If μ_1 and μ_2 are negative, then the Z_0 point is not stable. The magnitude of mu indicates the distance, on a Smith chart, from the origin to the closest unstable impedance. From this definition, it is clear that a magnitude greater than 1 means the device is stable for any points inside the unit circle of the Smith chart.

The formulas for μ_1 and μ_2 are

$$\begin{aligned}\mu_1 &\equiv \frac{1 - |S_{22}|^2}{|S_{11} - S_{22}^* \Delta| + |S_{21}S_{12}|} \\ \mu_2 &\equiv \frac{1 - |S_{11}|^2}{|S_{22} - S_{22}^* \Delta| + |S_{21}S_{12}|}\end{aligned}\quad (6.8)$$

where Δ is the same as defined in (6.7). These formulas could be created directly in the equation editor of a VNA, but some modern VNAs provide this as a built-in function. Figure 6.12 (lower) shows the result of plotting μ_1 and μ_2 for the amplifier of Figure 6.8, along with the S-parameters (upper left), and K-factor (upper right).

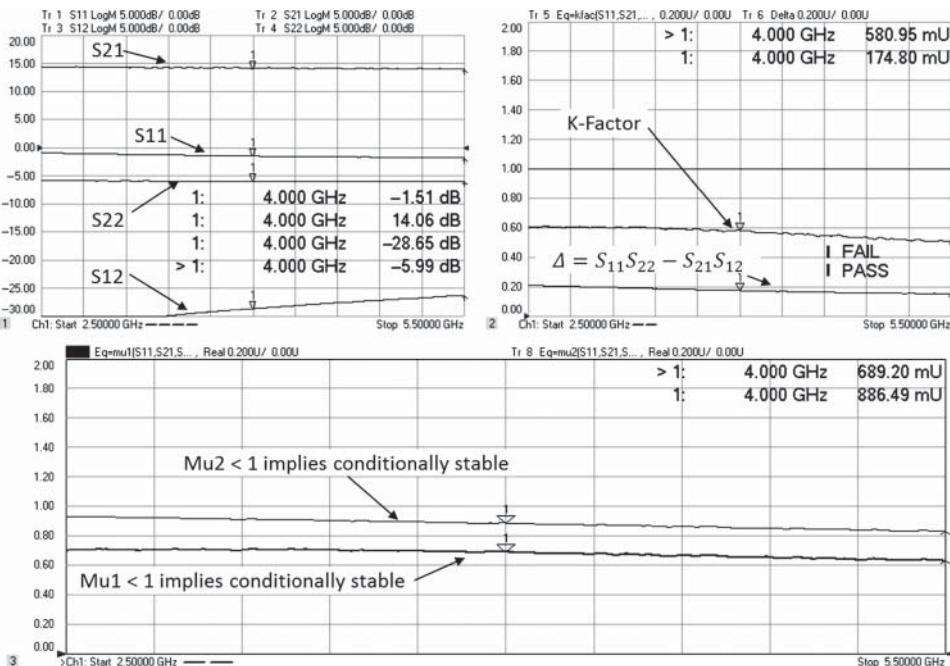


Figure 6.12 Mu1 and Mu2 for a conditionally stable amplifier.

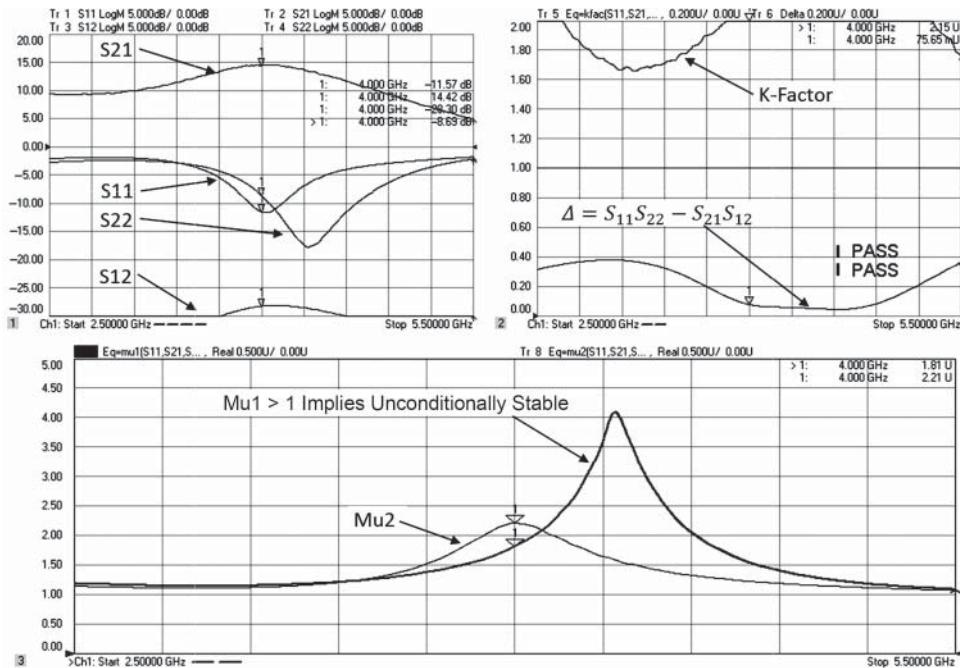


Figure 6.13 Mu1 and Mu2 for an amplifier after port matching to make it unconditionally stable.

Just as with the K-factor, the stability of an amplifier with respect to the mu factors can be evaluated using the port matching functions of the VNA. Using the same techniques as described earlier, the conditionally stable amplifier may be evaluated with the port matching applied, as shown in Figure 6.13. Here the values of the loss elements are adjusted until μ_1 is just greater than 1, and one can note that μ_2 also becomes greater than 1, and the same reactive matching as in Figure 6.9 is used to peak the gain.

6.1.5.4 Gain Factors

Besides S-parameter gain, which is defined as the gain from a matched source into a matched load, there are many other gain factors that provide useful insight into how a part may be used. Maximum stable gain and maximum gain were discussed as part of stability, but some other important factors are available gain and transducer gain.

The transducer gain gives the ratio of power available from the source to power delivered to the load, so both the source and load match values of the system must be known.

Available gain is the gain that an amplifier can provide to a conjugately matched load from a source or generator of a given impedance, and it is computed with the formula

$$G_A = \frac{(1 - |\Gamma_S|^2)|S_{21}|^2}{|1 - \Gamma_S S_{11}|^2(1 - |\Gamma_2|^2)} \quad (6.9)$$

where $\Gamma_2 = \left(S_{22} + \frac{S_{21}S_{12}\Gamma_S}{1-S_{11}\Gamma_S} \right)$

Since this analysis depends upon the particular source impedance of the previous stage, this additional information must be provided and can be obtained by measuring the output match of the previous stage and saving it in a different trace, or, on some of the more modern VNAs, loading the output match into a trace using and S1P or S2P file. This response can come from a measurement or from a simulation of the matching at the input or the output. Often, the preceding stage may be from a filter or other element whose match varies with frequency, such that the available gain also varies with frequency even if the amplifier under test has a flat frequency response.

Figure 6.14 shows the computed available gain (label GA) compared with the S_{21} gain, and the source impedance (label G_S), utilizing the equation editor function of the VNA.

In this case an intermediate result of the Γ_2 of the amplifier is computed and shown as well, called out as the trace G_2.

Note that where there are peaks in the G_S of the amplifier, there are also peaks in the available gain, and everywhere the available gain is greater than S_{21} .

The transducer gain gives the ratio of power available from the source to power delivered to the load, so both the source and load match values of the system must be known. Just as in available gain, both the source and load impedance of the system into which the amplifier will be placed can be loaded into the VNA trace data, and equation editor functions can be used to directly display the transducer gain.

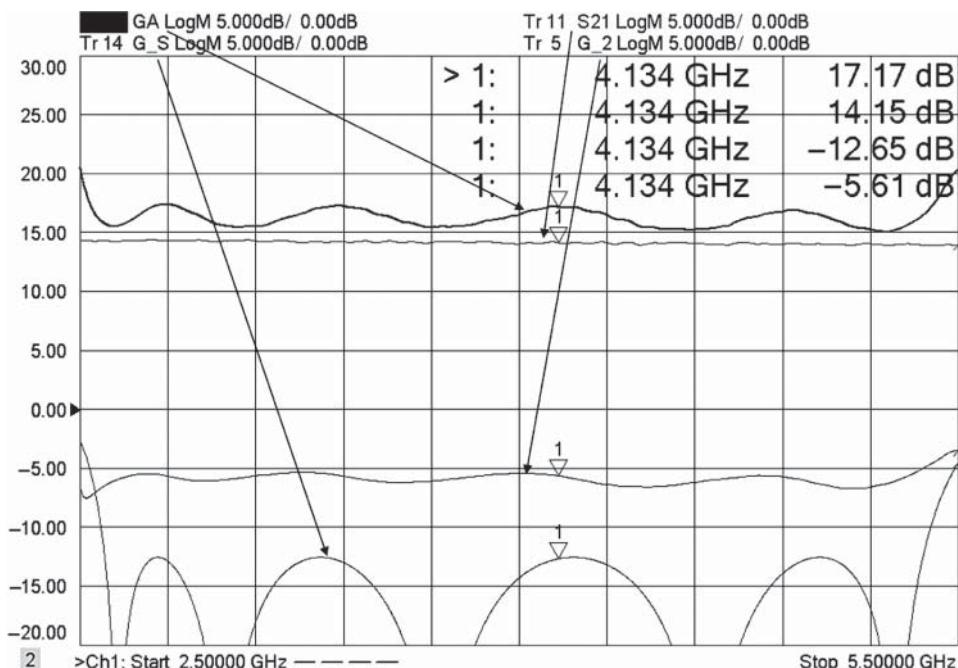


Figure 6.14 Available gain of an amplifier computed from the output match (G_S trace) of a filter and measurements of the DUT S-parameters.

Transducer gain is computed as

$$G_T = \frac{(1 - |\Gamma_S|^2)|S_{21}|^2(1 - |\Gamma_L|^2)}{|1 - \Gamma_S S_{11}|^2 \cdot |1 - \Gamma_2 \Gamma_L|^2} \quad (6.10)$$

An example of transducer gain (G_T) computation for an amplifier placed in a system with an identical filter on either end is shown in Figure 6.15. For reference, both the S_{21} of the amplifier and the available gain (GA) of the amplifier are shown.

Available gain and transducer gain were developed before the widespread use of simulation technology and so created a way to evaluate the effect of source and load matching of system elements on an amplifier's gain response. However, for the most part, these design techniques have been almost entirely replaced by simulation where the complete S-parameter matrix of system components can be embedded in a single measurement result.

Consider the amplifier from earlier, with a known source and load impedance from a pair of filters before and after the amplifier. If the filter characteristics are known, then the full effect of the S-parameter matrix of the filters at the input and output of the amplifier can be embedded in the response using the T-matrix math provided in the port matching function of the VNA (see Eq. (3.5)), as shown in Figure 6.16 (labeled S21_Simulated), along with an equation showing simply the product of S_{21} of each of the elements (labeled S21_Product), which would be the supposed result if one ignored the mismatch effects. Using port matching is a good way to simulate the overall gain effect and can essentially replace other forms

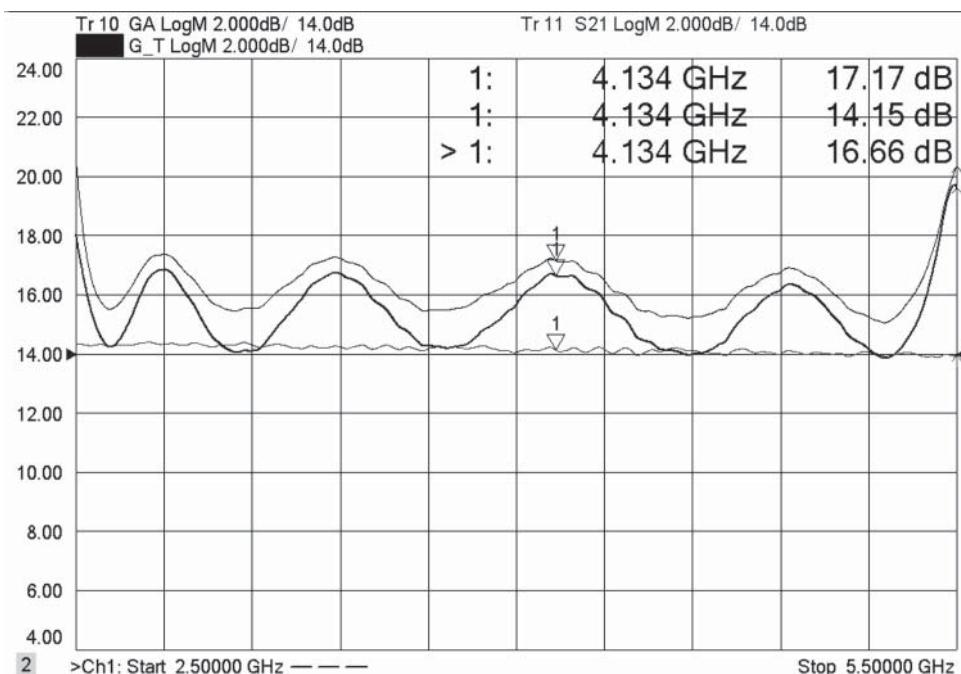


Figure 6.15 Transduce gain for an amplifier between two filters (dark trace).

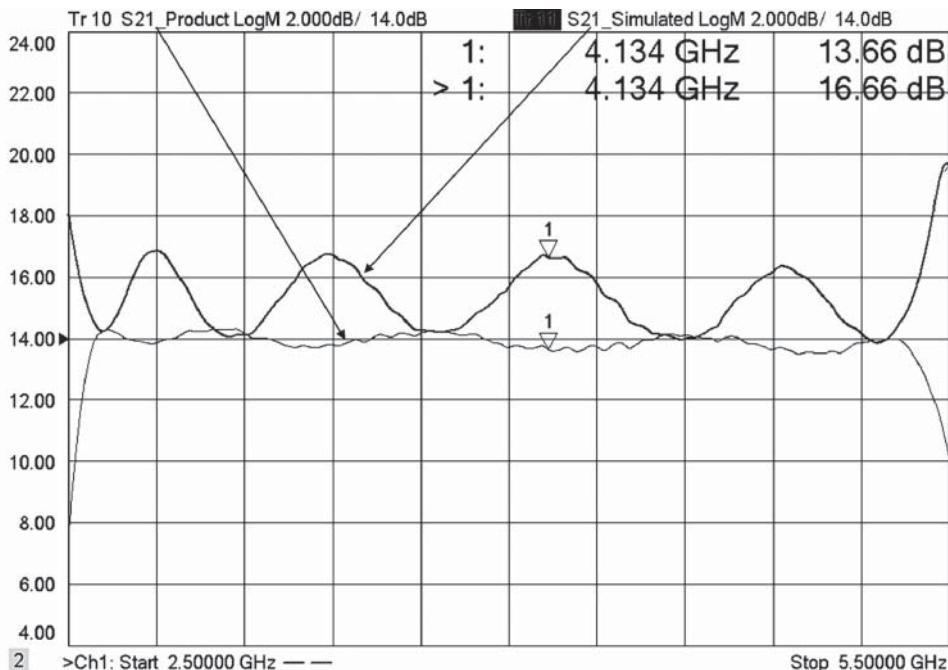


Figure 6.16 The overall gain from embedding the filter response using port matching, compared to the product of S_{21} .

of analysis of source and load influences on amplifier gain; since the filter loss is small, the simulated concatenation of the filters, the amplifier's gain ($S_{21_Simulated}$), is nearly identical to the transducer gain in the previous slide.

The port-matching gain does not give the same answer as taking the product of S_{21} of each of the components whenever the match of the filters and amplifier become large, especially if the amplifier is conditionally stable. This difference is a source of many errors in system design when mismatch effects are ignored.

6.1.5.5 Analysis Conclusions

Sophisticated real-time analysis of DUT performance can be created and displayed on modern VNAs using a variety of built-in functions. Fixturing applications allow application of port matching or embedding of the response of other components on the measurement results of the amplifier under test.

The equation editor function is another method of providing virtually unlimited types of analysis on the underlying data, including RF or DC or any other measured parameter. The resulting analysis traces can update in real time, allowing for tuning of the response of the amplifier on more than just the S-parameter results. An example would be tuning for improved stability or improved maximum gain. Other important analysis examples are presented in the sections to follow on specific applications of amplifiers including high power, high gain, and low noise.

6.1.6 Saving Amplifier Measurement Results

A final step is often needed in making any measurements, and that is saving the measured data in a form that can be used at a later time or in another program. For many legacy analyzers, saving data consisted of printing or plotting the results. To obtain the data in a format compatible with other processing, one often had to resort to reading the data programmatically, using some quite involved methods to recover the complex data, by ensuring that the data was expressed in the proper format, with the proper correction and the proper byte size and ordering. Once legacy analyzers could support floppy disk drives, several direct data dump formats were provided that could be used for such data archiving. Some of the important and more widely used formats are described next.

6.1.6.1 CITI File

Perhaps the first commonly used file format was the CITI file format, introduced for VNAs in the HP 8510. CITI file stands for *Common Instrumentation Transfer and Interchange* file format. This format had a linear structure, with a single column of data that required no in-line parsing. CITI file format has the nice advantage in that it is almost completely flexible in the kind of data that it can represent. The data is formed into packages, based on the stimulus. The structure of each package is a header with information about the instrument model and firmware revision, date and time, the names of the data traces, the independent variable (or stimulus) type, and the format of the data. The body of the file has variable lists with a “begin” and “end” keyword. Multiple channel files are all sorted into such packages according to channel number. If a channel contains two different stimulus domains, all the traces from each domain are combined into a package, and a separate package is created for each stimulus type. A sample CITI file is shown in Table 6.1, for a two-channel measurement, where the stimulus is different between channel 1 and channel 2, and channel 2 has one trace with time domain. This creates three packages. The table has sorted the file into three columns, for clarity and comparison, but in reality all three packages follow one after another in a single column in the actual file.

The example shown is an output from an Agilent PNA-X analyzer. As with some modern analyzers, this provides the user with several options on how the CITI file is formatted and what data is saved. The choices include whether to save all the displayed traces or just a single trace. An Auto choice saves all the corrected data for the selected trace; for example, if an S_{11} trace is selected that has a 2-port calibration applied, the Auto setting will also save S_{21} , S_{12} , and S_{22} .

The Save function also allows the format of the saved data to be different than the format of the displayed trace, by forcing all traces to either LogMag/Phase, LinMag/Phase, or Real/Imag. An Auto format choice saves all the data in the format of the currently selected trace.

While CITI file is a flexible format and was the format used in the first VNAs, several newer formats have come about that have found common use.

6.1.6.2 S2P or Touchstone® Files

Originated by EEsof before it was acquired by Hewlett-Packard, the Touchstone format, or S2P format as it is most commonly called, was created as a compact form to encapsulate

Table 6.1 CITI file output format; the file output is broken into three columns only for comparison

CITIFILE A.01.01 !Agilent Technologies !Agilent N5242A: A.09.42.08 !Format: LogMag/Angle !Date: Sunday, November 13, 2011 05:26:09 NAME CH1_DATA VAR Freq MAG 5 DATA S[2,1] DBANGLE DATA S[1,1] DBANGLE VAR_LIST_BEGIN 1000000000 1250000000 1500000000 1750000000 2000000000 VAR_LIST_END BEGIN -0.18913588,107.97729 -0.2556681,-134.61461 -0.29677463,-17.619053 -0.25021815,99.132004 -0.38517338,-143.36929 END BEGIN -23.06007,-128.98009 -35.266006,-0.99580592 -28.660841,151.64325 -41.02335,-23.198109 -32.739304,-28.0033107 END	CITIFILE A.01.01 !Agilent Technologies !Agilent N5242A: A.09.42.08 !Format: LogMag/Angle !Date: Sunday, November 13, 2011 05:26:09 NAME CH2_DATA VAR Freq MAG 5 DATA S[2,1] DEANGLE DATA S[1,1] DEANGLE VAR_LIST_BEGIN 2500000000 2750000000 3000000000 3250000000 3500000000 VAR_LIST_END BEGIN -0.38341591,91.595734 -0.19906346,-152.15067 -0.60013449,-34.987034 -0.52427602,81.872543 -0.43853623,-161.16911 END	CITIFILE A.01.01 !Agilent Technologies !Agilent N5242A: A.09.42.08 !Format: LogMag/Angle !Date: Sunday, November 13, 2011 05:26:09 NAME CH2_2_DATA VAR Time MAG 5 DATA S[1,1] DEANGLE VAR_LIST_BEGIN -4e-009 -2e-009 0 2e-009 4e-009 VAR_LIST_END BEGIN -35.606487,-87.11322 -26.621368,-126.83741 -35.606487,-87.11322 -26.621368,-126.83741 -35.606487,-87.11322 END
---	--	--

S-parameter data of a 2-port device. There are also higher port count versions, and as a class these are often referred to as SnP files, where n denotes the number of points. The structure of the S2P file is shown in Table 6.2. The Touchstone format has been officially accepted by several organizations including the Input Output Buffer Information Specification (IBIS) open forum group.

The version 1 Touchstone format was never formally adopted, partly because it has one major deficit in that the reference impedance for the S-parameter file must have only one value. This does not present any theoretical limitation, because if a network were measured with a different reference impedance on each port, it is a simple transformation to change one of them so that they are both the same. However, this shortcoming likely prevented formal adoption of the version 1 standard, which is found on many websites marked “draft,” and has been at the draft state for 20 years. In 2009, the IBIS Open Forum officially adopted version 2 of the Touchstone format, which includes modifications to support different port impedances on each port and includes a format specification for mixed-mode S-parameters. At this time, very few instrument or electronic design automation (EDA) companies have implemented the new format.

For a higher port count, the SnP file is used, and the format of the file differs from the previous discussion in that the data appears as a first line of five-to-a-line data (frequency plus the first four data points), and the remainder of the data is presented four to a line, all in the normal matrix order of S_{xy}, with x being the row number and y being the column number.

One curious attribute of the S2P file is that the standardized parameter order is a bit odd. It does not match the normal row/column description of the S-parameter matrix. Rather, it matches the order of parameters that is the default for many legacy VNAs: S₁₁, S₂₁, S₁₂, S₂₂. This has the unfortunate effect that the S2P file parser must be different than the SnP file parser.

One remaining deficit of the SnP file is that only S-parameters can be represented. If a VNA has other traces representing power, time domain, or any other parameter, the SnP file cannot be used to save the data.

6.1.6.3 CSV Files and Exporting Data to Excel

One new format to appear in some modern VNAs is the .csv format, which stands for comma-separated-values. This is a very generic format and in some ways matches the CITI file format in that it sorts the data into similarly grouped packages. An example of a .csv file is shown in Table 6.3.

The format is completely flexible just like CITI file but has the advantage that for a single channel, if the stimulus data is consistent, all the parameter data is on a single line for each stimulus value. The file can store any trace, including power traces, or the results of any equation editor function. Of course, a channel may contain traces with more than one stimulus; for example, channel 2 in the previous table contains a frequency and time-domain trace. In such a case a package for each stimulus is created. The .csv data is convenient because it may be directly read into an Excel spreadsheet. The .csv file output function can give several choices such as outputting only a single trace or all the displayed traces. The data may be formatted automatically, similarly to CITI file, or the format can be set to a specific format

Table 6.2 S2P data format

```
!Agilent Technologies,N5242A,US47210094,A.09.42.08
!Agilent N5242A: A.09.42.08
!Date: Sunday, November 13, 2011 06:01:33
!Correction: S11(Full 2 Port(1,2))
!S21(Full 2 Port(1,2))
!S12(Full 2 Port(1,2))
!S22(Full 2 Port(1,2))
!S2P File: Measurements: S11, S21, S12, S22:
# Hz S dB R 50
1800000000 -25.337969 -132.64902 -14.879494 -46.526134 15.25239 12.145417 -37.272678 -64.085381
1850000000 -26.51457 -74.607704 -14.982295 -7.8928113 15.358265 12.418506 -39.203392 -43.029419
1900000000 -31.968689 15.852159 -15.061513 31.229309 15.434734 12.457278 -33.634251 -77.491875
1950000000 -24.418478 -107.31781 -15.066053 70.428352 15.442039 12.643203 -30.444292 -44.296356
2000000000 -22.84832 -27.07604 -15.031354 109.00504 15.421839 12.837379 -30.570721 -29.915812
```

Table 6.3 CSV file example

```
!CSV A.01.01
!Agilent Technologies,N5242A,US47210094,A.09.42.08
!Agilent N5242A: A.09.42.08
!Date: Sunday, November 13, 2011 06:33:10
!Source: Standard

BEGIN CH1_DATA
Freq(Hz),S21 Log Mag(dB),S11 Log Mag(dB)
1800000000,-14.877501,-25.325617
1850000000,-14.982286,-26.499651
1900000000,-15.063152,-31.963058
1950000000,-15.069975,-24.410412
2000000000,-15.031799,-22.855028
END

BEGIN CH2_DATA
Freq(Hz),S21 Log Mag(dB)
2500000000,-0.37367642
2750000000,-0.19583039
3000000000,-0.59756804
3250000000,-0.52162892
3500000000,-0.43824977
END

BEGIN CH2_2_DATA
Time(s),S11 Log Mag(dB)
-4e-009,-35.611397
-2e-009,-26.620441
0,-35.611397
2e-009,-26.620441
4e-009,-35.611397
END
```

such as LogMag/Phase, LinMag/Phase, or Real/Imag. For the most flexibility, one can also choose the currently displayed format for each trace.

With these more flexible and comprehensive data save formats, it becomes much less necessary to read data from a modern VNA using the programmatic interface. Instead, a simple “store data as .csv” manual front-panel function can provide a file on that has all the desired data formatted in the desired way. And, of course, the same command can be sent programmatically. This removes the need to create data buffers, select traces, and read data programmatically over the bus, saving substantial programming time for a test engineer.

6.2 Gain Compression Measurements

The previous section described a wide variety of linear measurements that are commonly made on amplifiers, but amplifiers are often used in a non-linear portion of their operating range, and there are several important non-linear descriptions that are found in common use. Perhaps the most fundamental measure of non-linear performance is gain compression.

6.2.1 *Compression Definitions*

Just as the name implies, gain compression is a measure of the degradation of an amplifier's gain as a result of increasing input power. While amplifier non-linear gain is almost always referred to as compression, it is not uncommon to find some amplifiers that have a slight expansion, or increase, in their gain before the compression goes into effect.

Figure 6.1 showed an example of trying to identify the compression point of two amplifiers, before optimizing the VNA settings to ensure that driving the amplifier into compression will not overdrive the VNA. One amplifier shows only compression, while the other shows some slight expansion before the onset of compression occurs.

For the most basic compression measurement, the same technique can be used to find the onset of compression. After calibration, and starting from a power level sufficiently small to ensure linear amplifier operation, the S_{21} response is put into memory and the ratio of S_{21}/Memory is displayed. The power level is stepped up while monitoring this normalized response until the onset of compression is clearly found. This onset can be easily tracked by adding a marker and setting the marker search function to continuously search for the minimum value of the trace. In Chapter 5, the response of the VNA to increasing the source power by 20 dB was shown to reduce the trace noise by 10 times. Thus, if one wants to optimize trace noise and measurement speed, one should increase the averaging or reduce the IF BW for the initial linear response by a factor of 100. Since compression comes on slowly, the first power step can be 10 dB higher than the initial power, and the IF BW increased or averaged reduced by a factor of 10. If the onset of compression is not apparent, the power may be increased in increments until the marker shows some compression on the order of 0.5–1 dB. An example of this progression of measurements is shown Figure 6.17. The trace is normalized at -20 dBm , and the trace noise is evident in the measurement. The progression of power at -10 and 0 dBm input is shown, and the compression versus frequency is clearly seen. Since this is a filtered amplifier, the out-of-band area shows no compression. The peaking in the compression is likely related to peaking in the gain. The fact that the compression changes dramatically where the gain is peaked could also indicate that the value of S_{22} is affected by compression as well, and the gain peaking is reduced (thus more compression) the most at higher powers for frequencies near Marker 1 as compared with other frequencies.

6.2.1.1 *Compression from Linear Gain*

Once the onset of compression is found, the marker's frequency value can be entered into the CW frequency for a power sweep, and a power sweep over the linear to non-linear power range can be acquired. Such a compression sweep is shown in Figure 6.18. Also shown in the plot is a trace of input power and a trace of output power. The compression of the amplifier's

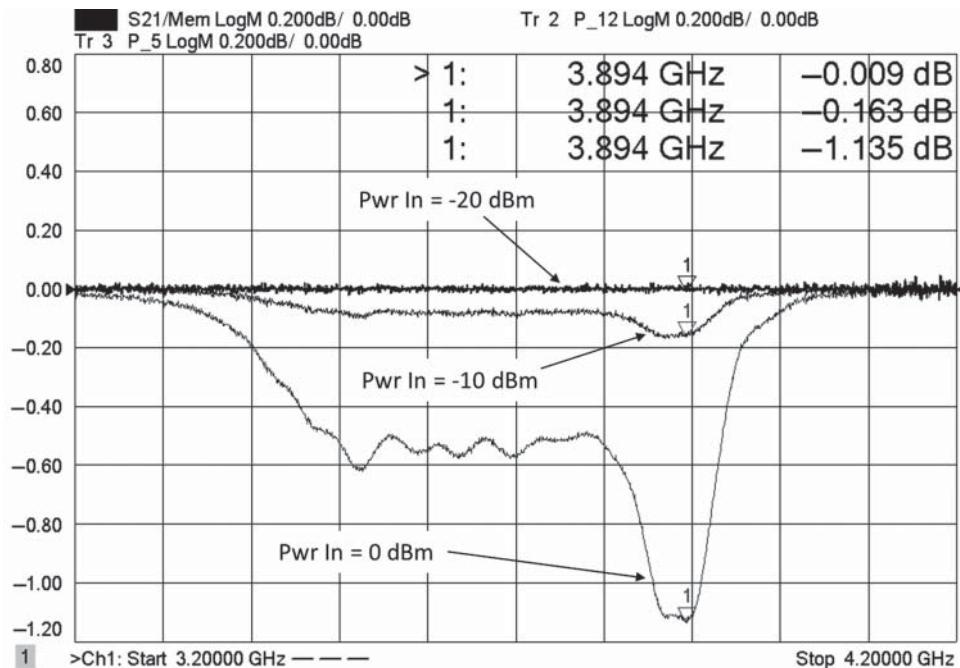


Figure 6.17 Detecting the onset of compression.

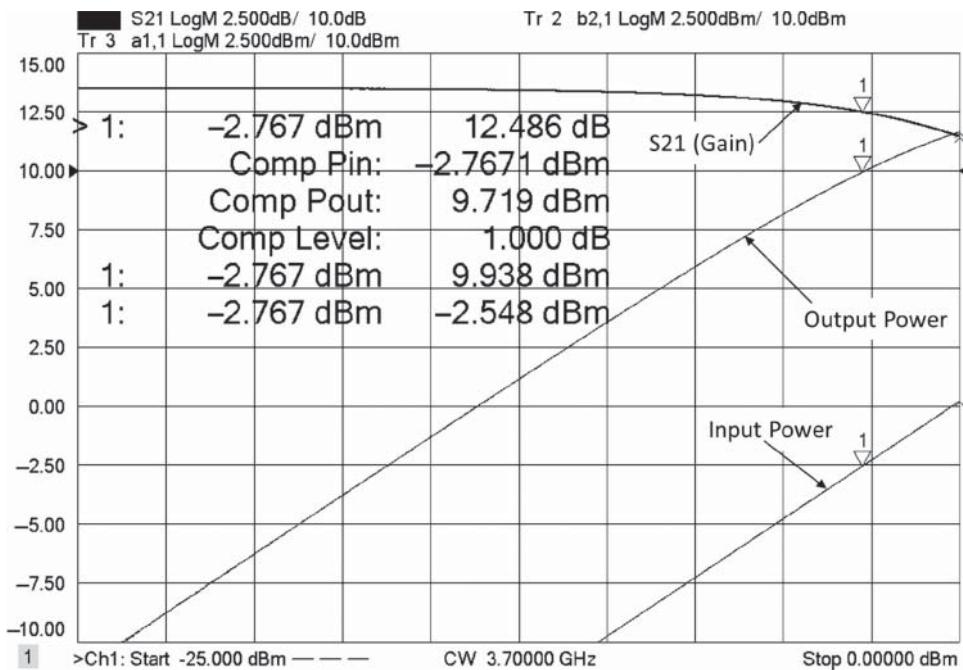


Figure 6.18 CW power sweep to find compression.

gain is clearly shown in the S_{21} trace. For a power sweep, there are three key powers: the linear power or power at which the reference gain is measured, the maximum power or stop power for the power sweep, and of course the power at which the gain is compressed by 1 dB.

A marker can be used to manually determine the 1 dB compression point (the most common definition of compression) or, in many modern VNAs, the compression point can be found automatically using a gain-compression marker function. In this method, the definition of compression is the point on the S_{21} trace that is 1 dB below the S_{21} value at the linear power measurement, usually defined as the first point of the power sweep. The compression marker will search for the 1 dB down point and then report the gain at that point, as well as the input power (the x-axis stimulus value), and the output power computed from the input power and the gain at the marker. This is in keeping with the definition of a trace marker in that it only operates on the data from the measured trace itself, rather than from other data, such as input power or output power. This is a very quick and quite good measurement but has some small drawbacks.

The main issue is that the x-axis values or marker stimulus values represents the source setting, but not necessarily the input power. However, if the input power and output power traces are also shown, as in Figure 6.18, then the markers can be coupled together, and the compression marker search will place the markers on the input and output power at exactly the right point to show the exact input and output power at compression. In this case, the output power at compression is about 0.2 dB higher than the marker search value reported on the S_{21} trace. This is because, from the input power trace, one can see that the actual measured input power is about 0.2 dB above the stimulus setting for that point (-2.5 vs. -2.7 dB). Alternatively, receiver leveling can be turned on so that the source setting exactly equals the input power reading.

While the definition of compression given is the most common, there are several other definitions that are also found in various industries, as listed next.

6.2.1.2 Compression from Max Gain

Some amplifiers experience an expansion or increase in gain just before the onset of compression, often due to a subtle re-biasing of the amplifier that increases slightly the gain. In fact, some amplifiers are designed particularly for this effect as a means to extend the linear operating range. In these amplifiers, the gain peaks versus drive power and then compression quickly ensues (refer to Figure 6.1). For these types of devices, the gain compression is sometimes specified as compression from maximum gain, rather than compression from linear (or low power) gain. This definition is somewhat more conservative than the prior one, in that it will report a lower output power for 1 dB compression than will compression from linear gain. Note that for an amplifier following a normal compression curve, where maximum gain is at the linear power, the definitions are identical.

6.2.1.3 Compression from Back-off or X-Y Compression

Another compression measurement that harkens to the first days of measuring non-linear behavior is compression from back-off, or the so-called *X-Y compression*. These are different implementations of the same method, and they essentially define compression as a defined

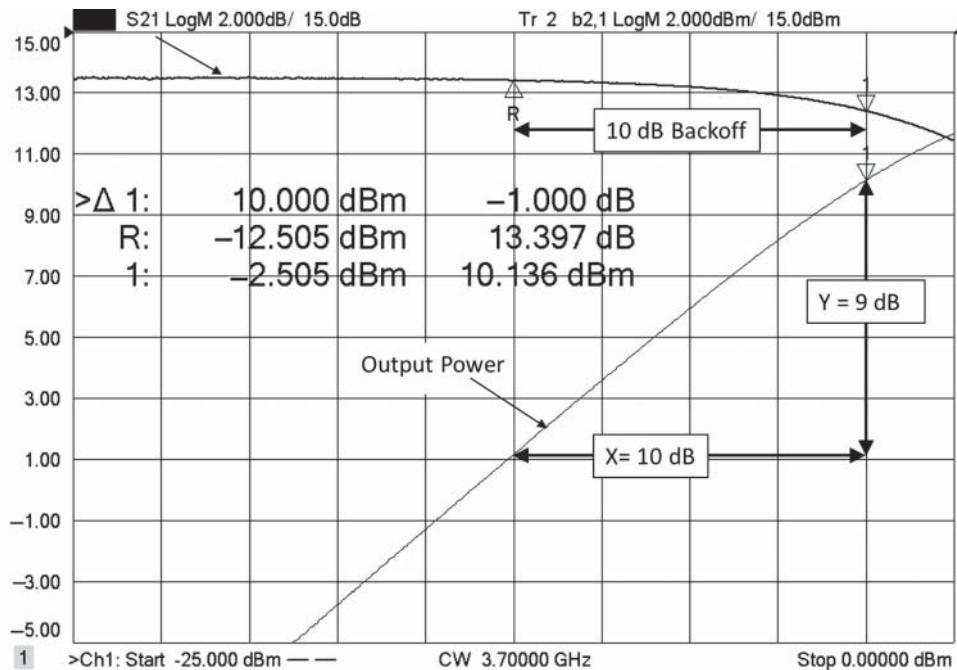


Figure 6.19 Back-off and X-Y methods of finding compression.

change in gain (or the y-axis of the S_{21} plot) over a specified change in input power (or x-axis). Compression from back-off looks at the S_{21} versus power curve and looks for a point where the gain drops a specified value (usually 1 dB) over a specified change in drive power (usually 10 dB). In the past, compression was measured by moving a 10 dB pad from the input to the output of an amplifier, while monitoring the output power. The input power was increased for each iteration until a power was found where moving the pad changed the output power by 1 dB, giving the same result as the back-off method. The X-Y method, which is functionally equivalent, looks for a specified change in output power (usually 9 dB) over a specified change in input power (usually 10 dB). These compression methods are illustrated in Figure 6.19.

In some ways, this is the best method of finding compression because it is insensitive to noise at the linear power range and it incorporates a concept of looking for a change in gain over a nominal change in input power. Compare this with the max gain or compression from linear methods: max gain requires the maximum gain be determined, which might require a very dense point spacing in the power sweep to ensure the maximum power is correctly measured; compression from linear means that trace noise at the very low power of the linear measurements will directly affect the compression computation at high power. Further, for some amplifiers, the gain drops very slowly with increasing power so that the power sweep range must be very large to ensure that the linear power is achieved. However, the X-Y or back-off compression is always found within the X dB (e.g. 10 dB) range of the compression point. For many modulated signals, which maintain some average power level and have some specified peak-to-average ratio, the compression from back-off provides a more real-world use case for a compression measurement. If the modulated signal provides an average power,

it is probably more reasonable to use that power as reference for the measurement of compression at the peak than it does to use a very low linear power, which some modulations formats will not ever impart. Notice that X-Y or back-off always gives a larger compression value than compression from linear; perhaps this is one reason why it is favored by many amplifier manufacturers.

6.2.1.4 Compression from Saturation

While somewhat of a misnomer, compression from saturation is a method that is applied to amplifiers that are normally used at or near their saturation point. For some amplifiers, such as traveling-wave-tube (TWT) amplifiers, a very clear saturation point occurs in the input-power output-power curves. The amplifier is operated backed-off, just below this saturation point, with a specified back-off level. Sometimes this level is quite low, such as 0.03 dB below saturation (or max power) typically found in TWT amplifiers. This level is very close to the maximum output power level, but one key operating point metric is the input power for maximum output power, sometimes called the *normal operating point* (NOP). When the saturation curve is very flat, even the slightest noise in the trace can cause large swings in the input power measurement at saturation, so backing off just slight amount, such as 0.03 dB, provides a much more stable number for the NOP.

Sometimes the back-off is more substantial such as 8 dB often found in solid-state power amplifiers (SSPAs) that are used in some satellites replacing TWT amplifiers. The method to find this back-off level from saturation is identical to the TWT method, with the only change being the value of the back-off. Note that while the same term is used, the value of back-off here is unrelated to the term used in a similar sense in Section 6.2.1.3.

6.2.2 AM-to-PM or Phase Compression

When viewing the effects of gain compression on a complex modulated signal, a common figure of merit is error from the intended magnitude and phase. The total error in this signal is the vector difference between the desired value and the actual value, and this includes both amplitude errors and phase errors; the size of this error is called the *error vector magnitude* (EVM). In some cases, the change in phase at the compression point can cause a larger vector error than the change in magnitude. To quantify these effects, a common measure of compression's effect on phase is expressed as AM-to-PM, although two distinct definitions have emerged.

One definition is the change in phase from a reference value at the defined amplitude compression point. Thus, displaying the phase versus drive along with the magnitude response will provide a direct measurement of AM-to-PM. If the marker compression search is used, the phase marker can be coupled to the compression marker, and the phase at 1 dB compression can be simply read from the marker display, as shown in the first marker of Figure 6.20. Here the X-Y, or 10 dB back-off, method is used to find the 1 dB compression point. It is clear from the phase trace that trace noise at the lowest power will add significant error to the estimate of phase change at the 1 dB compression point. This is because the reference would have almost 1° of trace noise. Using the 10 dB back-off as a reference eliminates much of the noise issue.

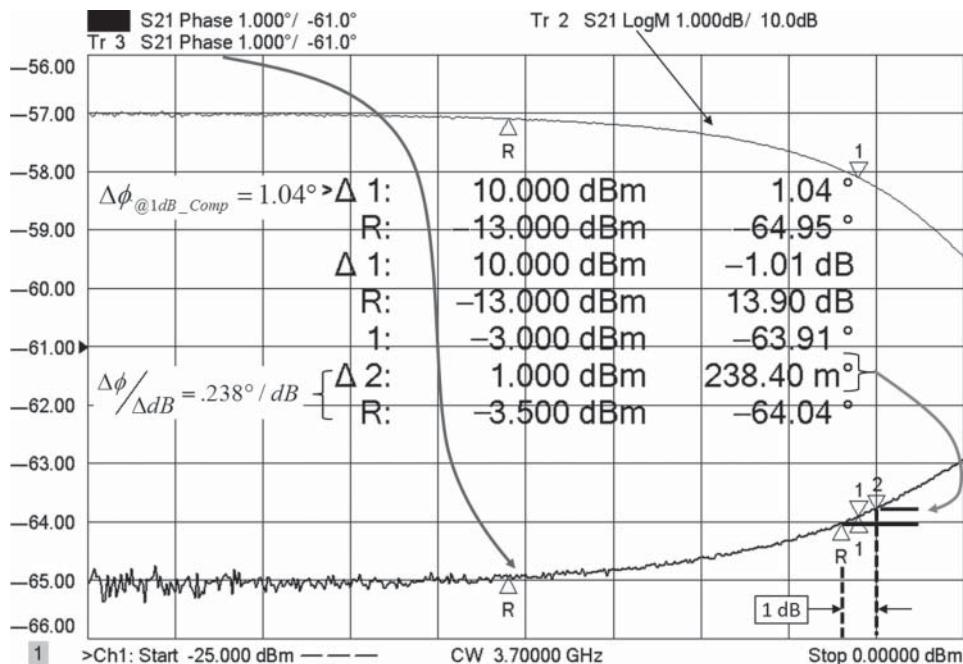


Figure 6.20 Phase vs. drive and AM-to-PM.

An alternative definition of AM-to-PM is the phase slope in degrees/dB, at the 1 dB compression point. This can also be directly displayed, although in a roundabout way. An S_{21} phase versus drive trace is used with one marker set to 0.5 dB below the input power of the compression point, and another at 0.5 dB above the power. One of the markers can be set to be a reference marker, and the other will read directly the change in phase for a 1 dB change in drive power. This measure of AM-to-PM is shown in the bottom marker set of Figure 6.20.

6.2.3 Swept Frequency Gain and Phase Compression

The compression definitions used so far are single-frequency measurements where the power is swept and the compression point is determined. However, in many cases the compression of an amplifier, particularly a narrowband tuned amplifier, can change over the frequency band of the amplifier. In this case, it is desirable to measure the gain compression across the entire frequency band.

Traditionally, the swept frequency gain compression measurement was performed by repeating the CW power-sweep method over each frequency in the frequency range, using an external computer to control the process. The 1 dB compression points thus found were collected and displayed as the 1 dB compression versus frequency. The collected data essential creates a two-dimensional measurement over a frequency and power range of the amplifier.

While this is a very direct approach, there are much better algorithms for finding the compression point over a span of frequencies. The biggest drawback from the swept-power

stepped-frequency method is that a fixed power sweep range, as is most common, can overdrive an amplifier severely across high-gain regions of its frequency response, and underdrive it across lower-gain regions. If the amplifier under test is overdriven in its high gain region, it may go into deep compression and change the operating point or temperature of the amplifier; if so, the measurement for the next frequency, when the power sweep is reset to a lower power, gives an inaccurate gain reading because the amplifier has not recovered from the overdrive condition.

Rather than sweep power at each frequency, a better approach is to sweep the frequency at each power step. The first sweep is made at a linear power, and each subsequent frequency sweep is made at a higher stepped power. The gain and compression, as well as input and output power, are recorded for each frequency sweep across each power step. When the power has been stepped through its entire defined range, the compression is computed for each frequency from the 2-D frequency and power recorded. In this way, the linear measurements are all made at the same operating point, and the compression levels are likewise made at similar operating points, provided the compression is relatively constant across frequency.

Recently, some VNA vendors have implemented variations of automated swept frequency 1 dB compression algorithms. One method goes a step further in speed, accuracy, and safety, the details of which are described next.

6.2.4 Gain Compression Application, Smart Sweep, and Safe-Sweep Mode

For a swept-frequency gain-compression measurement, the data acquisition can be performed in two ways: sweep power and then step frequency, or sweep frequency and then step power. The former method is most commonly applied but has several serious drawbacks, the main one being that the DUT is exposed to maximum power at the end of one power sweep just before the measurement changes to the new frequency and minimum power. This can often cause a change in DUT behavior that makes subsequent linear power acquisitions incorrect. On the other hand, sweeping frequency starting at the minimum or linear power and then stepping up the power for a new frequency sweep avoids any issue with overdrive.

For the very fastest measurements of compression, it is not necessary to measure every power level between the linear power level and the compression level. An initial measurement of the linear power is first acquired. Then a first guess at the power compression level can be made. For example, choose a power halfway between linear and maximum power, and measure the compression. A second guess can be made at a higher power. From the measurement of these two powers, a third guess can be computed for the level at which 1 dB compression will occur. A new frequency sweep is performed with the new power settings, and another compression acquisition is obtained. This continues until the desired compression level is achieved within a specified tolerance. One VNA vendor refers to this as SMART sweep in gain compression, because the sweep attributes are adaptive on a point-by-point basis to the DUT behavior.

At this point, the input power table to which the source is adjusted is exactly the input power for 1 dB compression, sometimes called CompIn or CompIn21 to identify the path of the compression measurement. The gain at this compression level is called the CompGain21, and the output power is called the CompOut21. These represent the fundamental compression measurements. An additional parameter is often useful to view, the DeltaGain, which is the

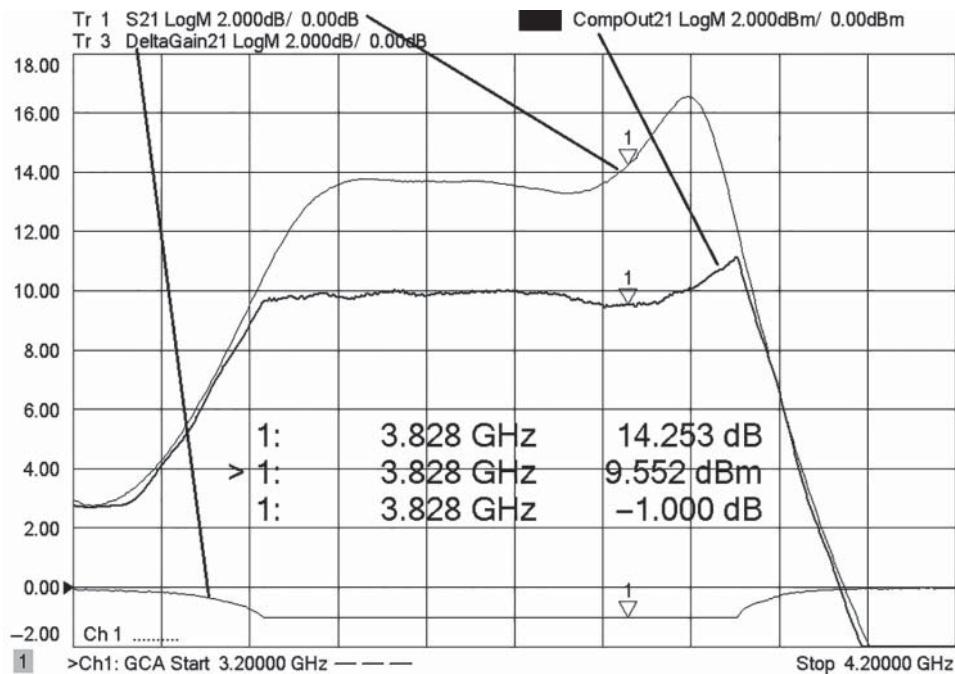


Figure 6.21 Swept frequency 1 dB compression measurements.

measured value of compression. Ideally, this should be exactly -1 dB, but if the amplifier is not driven into compression or the tolerance is large, it can vary from that value.

An example of a swept frequency 1 dB compression measurement is shown in Figure 6.21.

From the DeltaGain21 trace, it is clear that the amplifier is not in compression at the band edges, due to the filtering function at the input of the amplifier not allowing sufficient drive power to compress the amplifier. The CompOut21 shows an interesting result in that the compressed output power is not constant but is higher where the gain is peaked. A marker on the CompOut21 trace reads the value of output power at 1 dB compression at any frequency along the trace.

6.2.4.1 Safe Modes of Measuring Compression

While the iterative method of finding the 1 dB compression point can be very fast, it does have a drawback in that it can also overdrive a DUT if the DUT has a lot of gain variation and the power settings for the initial gain readings are high enough to overdrive the amplifier. Also, if the amplifier does not follow a normal compression curve, it is possible that the predicted power for the next iteration will overdrive the amplifier. Particularly for very high-power amplifiers, an overdrive condition must be avoided both to protect the DUT and to protect the test equipment connected to the DUT.

In such a case the iterative method may be modified to provide a safe mode of operation. Such a safe mode would need to have some defined limits to the power setting and the output

power; in particular, there should be limits set so that the input power is not increased in the next iteration if the output power has exceeded a pre-determined limit, even if the amplifier is not compressed. Also, the step size of the power change should be limited to avoid jumping from an underdrive to an overdrive condition in one step.

One method is to specify a maximum power step size, say 1 dB. Then during the iterations, the input power could not increase by more than the specified step, ensuring that no overdrive could occur by more than this step. However, if the linear power is far from the compression point, a great deal of iterations will need to occur before the compression level is found. A smarter method is to define two step sizes: a course step (perhaps 5 dB) and a fine step (perhaps 1 dB). In addition, a compression threshold should be defined such that when an amplifier exceeds some safe compression threshold (perhaps 0.5 dB), the step size automatically switches from course step to fine step. Such a scheme ensures the amplifier will never exceed a specified output power (thus protecting the external equipment) and never be overdriven by more than the fine step size (thus protecting the amplifier).

6.2.4.2 Full 2-D Gain and Compression Characterization

For some cases it is desirable to have a full description of the amplifier's performance, over a predefined set of power points and at every frequency point, as a two-dimensional array of gain versus power and frequency. The previous adaptive sweep methods do not give a deterministic number of power points as the power steps depend upon the amplifier's response. Thus, it would be convenient to modify the gain compression acquisition to make a full two-dimensional array of frequency and power. The safe mode would support this if the fine step and course step are set to be the same size, but the safe mode stops increasing the source power after the compression point has been passed, and subsequent iterations simply zero in on the compression point, so the data is not regularly formed in the data set.

Instead, a separate mode of operation should be used to sweep the frequency and step the power for each one of a defined power sweep range, even if it overdrives the amplifier. In this way a regular array of Pin, Pout, and Gain data can be created. Some VNAs already provide 2-D sweeps, and these also allow the choice of sweeping power and stepping frequency, or sweeping frequency and stepping power. Except in a rare circumstance, the frequency should always be swept as the power is stepped from low to high to avoid overdrive-related issues with measurements of the linear power.

If a data set with this two-dimension array of gain versus input power and frequency is exported, a surface plot may be simply created using a variety of plotting tools such as MATLAB to create a three-dimension surface of compression versus power and frequency, as shown in Figure 6.22.

Also shown in the figure is a selection of points that represent the points closest to the 1 dB compression value.

6.2.4.3 Calibration in Compression Measurements

One important point to note is that for gain compression measurements, the error correction math of Eq. (3.3) does not strictly apply. The concern about this math function is that the input reflection errors are not properly accounted for when an amplifier is in compression. Take the

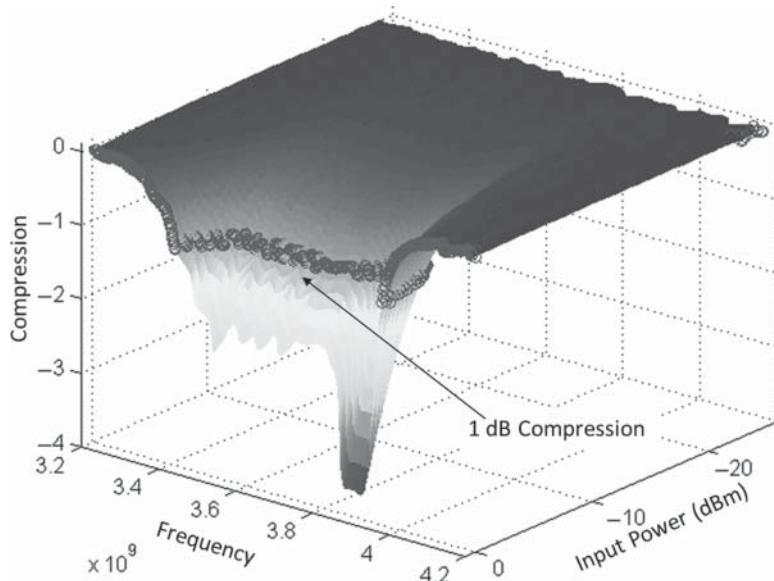


Figure 6.22 A 3-D surface of compression versus frequency and input power.

extreme case of an amplifier in saturation. The output power is limited at a fixed level by the saturation. If there is mismatch ripple at the input that causes the incident source power to be greater than that of a matched source, there will be no corresponding increase in the output power. However the error correction function of Eq. (3.3) supposes that an increase would occur (S_{21} being linear in the assumption) and so the corrected gain and output power (which is sometimes computed as input power times gain) are reduced by the input mismatch error. In fact, the input mismatch error causes *no* change in the output power (the amplifier being saturated), and thus the gain will show ripple that it does not really have. Thus, one should not use normal error correction for amplifiers tested in their non-linear mode of operation.

Instead, an alternative computation of gain and output power can be performed using the definition of gain as the match-corrected output power over the match-corrected input power. In such a definition, there is no assumption of linear behavior for S_{21} , so any error caused by mismatch at the input will not be reflected in output power. The gain of amplifier will be properly adjusted for the input mismatch in such a case. This type of correction does have one problematic assumption: that the output match of the amplifier does not change with drive power. In fact, this is not likely the case, but for a simple compression setup, this assumption cannot be avoided. More advanced non-linear VNA methods can address such situations, as discussed in Section 6.8 on X-parameters. In the case where an amplifier's output impedance varies greatly with drive level, the best approach to getting quality measurements may be to add some small attenuator at the end of the port 2 cable to improve the raw match.

Another issue that is more prevalent when an amplifier is just starting to go into compression is the error in the input power due to input mismatch. Because of this effect, the DUT will have a higher than expected power at some frequencies and a lower than expected power at other frequencies. This is true even if the source power of the VNA is calibrated, because

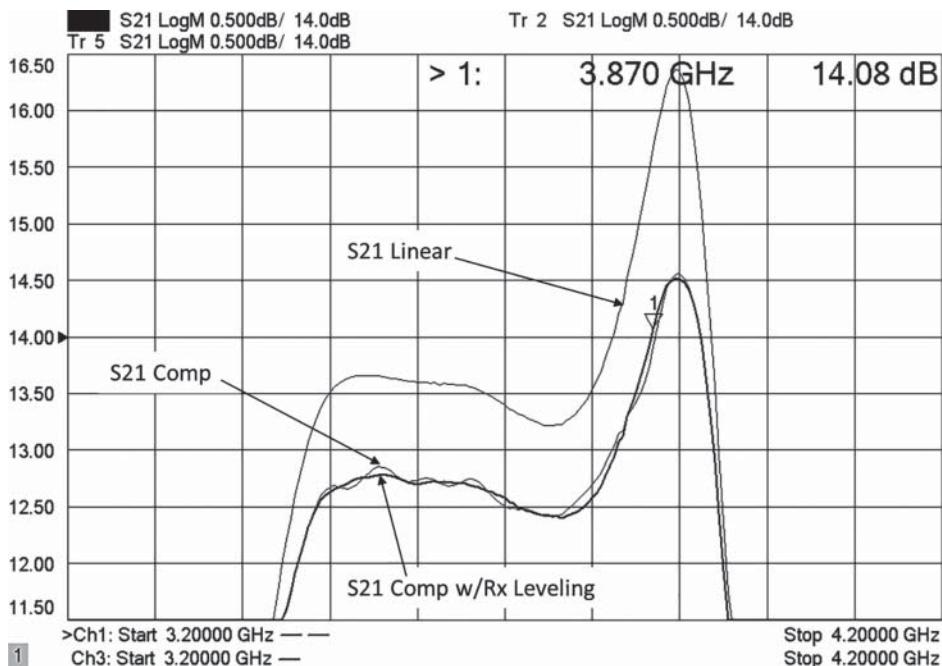


Figure 6.23 S₂₁ gain of an amplifier in compression, normal and with match corrected Rx leveling applied.

the calibration was performed into a well-matched power meter but the DUT may present a poor match.

An example of the S₂₁ at linear power and near the 1 dB compressed power is shown in Figure 6.23. For each case, the normal full 2-port correction is applied, but in the case of the compressed power, the S₂₁ gain (labeled S₂₁ Comp) shows excessive ripple. Also in the figure is a trace of S₂₁ with receiver leveling applied, near compression. In this case, the input power is controlled to provide a flat, leveled power even in the case of mismatch, so the compressed gain shows a smooth response as one might expect. A reference trace of S₂₁ for linear drive powers is shown as well.

From this figure, one can see that the apparent ripple in the S₂₁_Comp trace is not truly representative of the gain at that power, due to the non-linear change in S₂₁ responding to the mismatch induced ripple of the input power.

6.2.4.4 DC Power Analysis

For many amplifiers, the efficiency of the amplifier in terms of creating RF power from the DC power consumed is a critical performance parameter, with the most common form being the power-added efficiency (PAE), which is defined as

$$PAE = \frac{Output_Pwr - Input_Pwr}{DC_Pwr} = \frac{(S21 - 1)}{DC_Pwr} Input_Pwr \quad (6.11)$$

where the power is expressed in watts or milliwatts. Since the DC power of a linear amplifier is essentially constant (one definition of linear being that the RF signal level is so low, the DC operating point of the amplifier is not changed by it), the PAE will increase with increasing input power until the amplifier becomes non-linear and gain compression occurs. For many amplifiers, the measurement of PAE in the non-linear region is a key design parameter, and this measurement depends greatly upon other aspects of the design such as the termination impedance of the amplifier at the fundamental, second and third harmonics. Often, PAE is measured as part of a load-pull measurement to find the optimum load for power efficiency. Once this load is determined, a matching network is designed, and the overall matched amplifier response must then be remeasured into a matched impedance load (normally $50\ \Omega$).

As described in the measurements in Section 6.1.4.2, some VNAs provide a means to measure the voltage and current of an amplifier synchronously with the frequency or power sweep. In such a case, the equation editor function can be used to implement Eq. (6.11) directly, or even some built-in functions have the definition of PAE already specified. An example of a swept frequency PAE measurement and gain is shown in the upper-left trace of Figure 6.24, along with voltage and current in the left lower.

A similar measurement is shown in right windows of Figure 6.24 for a swept-power PAE measurement. For this device, as is common in most amplifiers, the efficiency improves as the device approaches compression.

When a full two-dimensional gain compression measurement is performed, it is convenient to also acquire the DC voltage and current for each frequency and power point. Some VNAs

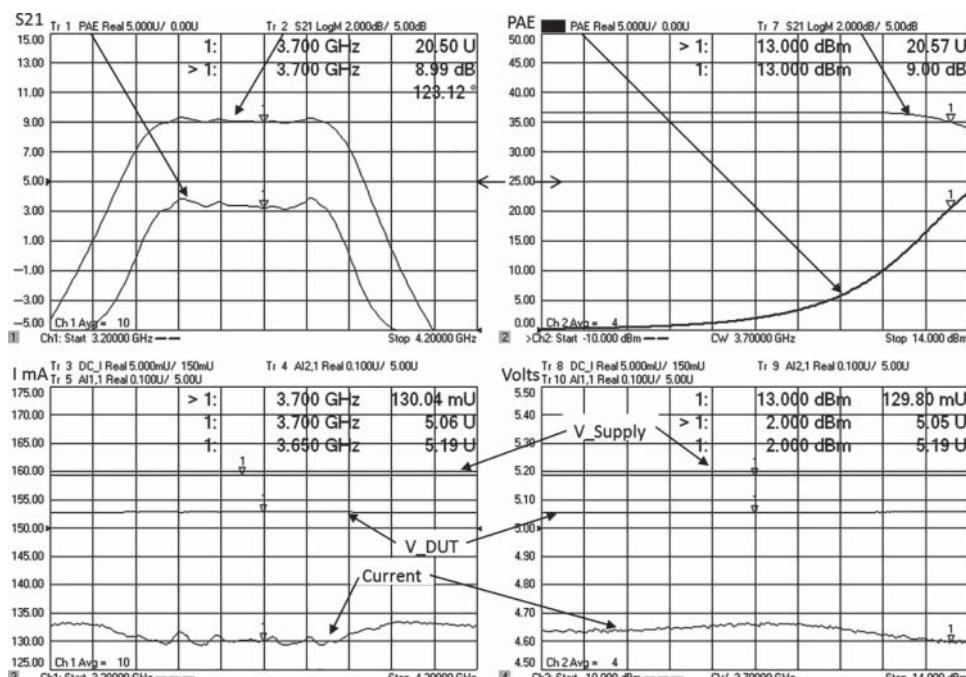


Figure 6.24 Swept power PAE (upper); swept frequency PAE (lower).

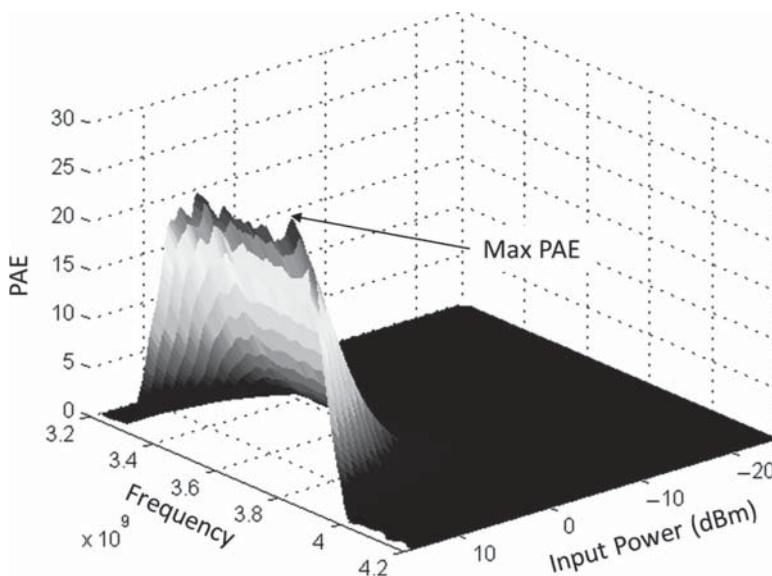


Figure 6.25 PAE versus power and frequency on a 3-D surface.

provide this as part of the built-in gain compression data acquisition, and the saved data can be processed to create PAE surfaces that show the PAE as a 3-D surface function of input power and input frequency, an example of which is shown in Figure 6.25.

6.2.4.5 Error Correction and Compression

The normal error correction methods rely on the DUT being having a linear response, but compression is, by definition, a non-linear behavior. While traditional gain compression measurements are made as described earlier, recently new methods have been developed that perform non-linear characterization in a manner that does not depend upon the assumption of linear correction to remove the effects of test system mismatch. These methods, called *hot S-parameters*, are described in Section 6.8.3.

6.3 Measuring High-Gain Amplifiers

Most amplifiers are low to moderate gain, in the region of 10–30 dB of gain. For these amplifiers, no special techniques are needed in setup, calibration, or measurement as the normal setup and source ALC range will allow a reasonable calibration at the maximum of the source power, and then a reasonable measurement at a lower source power. At these gains, the input and output match are quite reasonable to measure as well as the reverse isolation. However, as the gain of the amplifier becomes high, the input power must be greatly reduced, and the raw measurements of the other parameters can become very noisy. This in turn affects the quality of the error correction resulting in poor measurement results for all S-parameters.

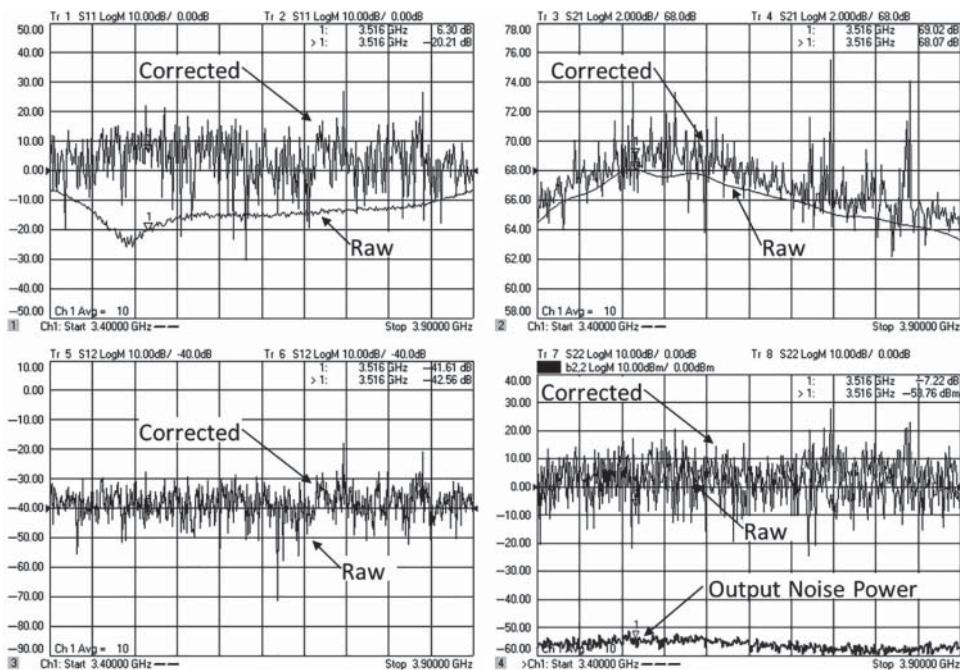


Figure 6.26 Error-corrected measurements on a high-gain amplifier.

Consider an amplifier with greater than 60 dB gain and +10 dBm compression level, as might be commonly found in a communications system LNA. The input drive level for linear operation of this amplifier is found by subtracting 20 dB from the compression point, and 60 dB due to the gain, to yield -70 dBm input drive level. From a preset condition, if the power level is simply changed to -70 dBm very poor error-corrected measurements will result. An example of such a measurement is shown in Figure 6.26, with both raw measurements (uncorrected) and corrected measurements. The default setup for most VNAs is to couple the port powers, so setting the port 1 power to -70 dBm also sets the power of port 2 to -70 dBm. Notice that the raw S_{21} measurement is much less noisy than the corrected measurement, and the same is true of the S_{11} measurement. The S_{12} measurement is essentially just noise, in both raw and corrected. The most interesting result is the S_{22} measurement: while the raw S_{11} looks OK (with some noise), the S_{22} is completely full of noise and is in fact showing a measurement greater than 0 dB. To understand how this can be, consider also the $b_{2,2}$ trace, which is a measure of the b_2 power received during the S_{22} sweep. Since the port 2 test port power is -70 dBm, one would expect the b_2 power to be below -70 dBm; instead, it is around -58 dBm. The reason for this is the high gain of the amplifier produces high noise at its output, even with no signal, and so this high noise swamps the measurement of S_{22} , and both the raw and corrected results show just noise. Methods for removing these effects and improving these terrible measurements are described in the following sections.

The question remains: why are the S_{11} and S_{21} corrected measurements so bad, since their raw measurements are relatively noise free? The root cause of this is that for the low drive

power, the S_{11} measurement has some noise, the S_{22} measurement has terrible noise, and the S_{12} measurement is substantially all noise. This is because the low input drive power from port 2 does not allow enough signal to capture the behavior of the amplifier. Consider S_{12} : the S_{12} of the amplifier is actually lower than -110 dB, but with the drive level of port 2 being just -70 dBm and the noise floor of $b1$ receiver (used for S_{12} measurements) being -110 dB, the apparent S_{12} is only -40 dB ($-110 - (-70)$) due to the noise floor limitations. Thus, the “loop-gain” term for error correction from Eq. (3.3) becomes very large as the product $S_{21} \cdot S_{12}$ is nearly 30 (~ 30 dB), much greater than one (0 dB) that is required for a stable amplifier. In such a case the error correction essentially adds the noise of S_{12} onto the S_{11} and S_{21} traces.

In addition to the noise from S_{12} , noise in the raw S_{11} and S_{22} will also be translated into noise on the S_{21} trace through the error correction mathematics.

6.3.1 Setup for High-Gain Amplifiers

Avoiding these noise issues is quite simple if the setup of the measurement is slightly modified. For any amplifiers, but especially high-gain amplifiers, the test port powers for each port must be uncoupled. The power for port 1 of course must be set to the linear power level, but the power for port 2 does not need to be set to the same level. In fact, during normal operation, the power level of port 2 should be set at the linear input power, plus the gain, less about 10 dB. This ensures that the power is sufficiently large to avoid noise issues but is always small enough to ensure linear operation of the amplifier, as it is much below the normal output power of the DUT. Raising the port 2 power has the benefit of reducing the noise in both the S_{12} and S_{22} traces. In this case, the noise is reduced by the equivalent of 50 dB, or about 300 times reduction in trace noise.

In some cases, raising the port 2 power still leaves a noisy S_{11} trace; the noise on the S_{11} trace is imparted on the S_{21} corrected results yielding a noisy S_{21} trace. An additional modification of the test setup can help reduce this noisy effect. If the VNA being used has a configurable test set, that is, if the connections between the VNA source and receiver to the directional coupler are exposed as loops externally, then the port 1 test coupler can be “reversed” in much the same manner as described in Section 5.2.6, shown in Figure 6.27. With the source routed to pass through the coupled port, a higher-reference channel power will occur for the same test port power. In addition, the $b1$ reflection receiver is connected to the through arm of the test port coupler lowering the loss to the $b1$ receiver. In this way, the signal-to-noise ratio of the S_{11} measurements will improve by about 14 dB, resulting in around a 25 times improvement in S_{11} trace noise. This same change also improves the signal-to-noise ratio of the S_{12} trace, yielding further benefits for noise reduction in the overall response.

6.3.2 Calibration Considerations

When testing high-gain amplifiers, the source power must be reduced, but this leads to low signal levels during calibration and excess noise in the calibration error terms, which are imparted on the corrected trace appearing as a stationary noise-like error or ripple.

For modern VNAs, the receiver linearity is so good that there is no degradation raising the source power for calibration and then lowering it for measurement. In setting the measurement

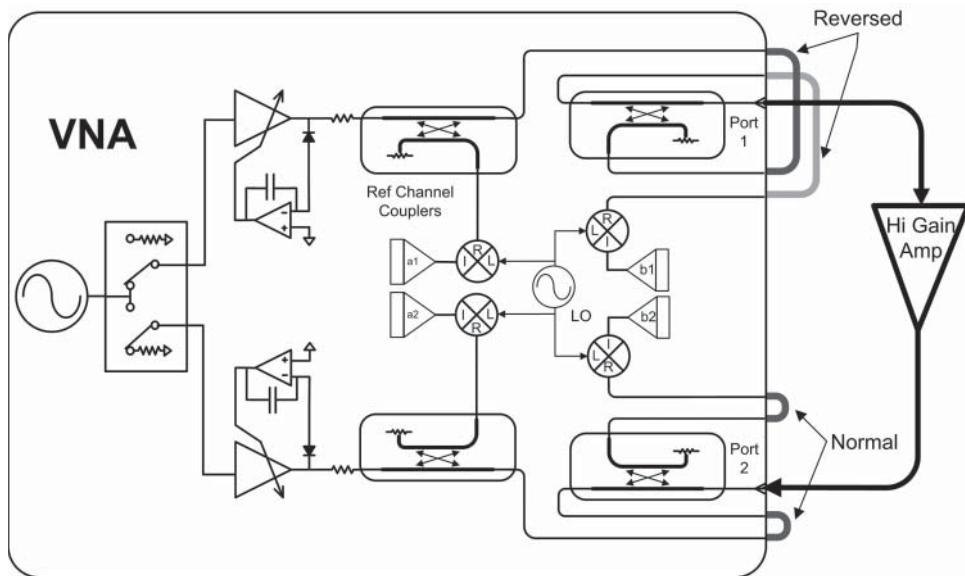


Figure 6.27 VNA block diagram with port 1 coupler reversed.

for optimum results, the source attenuator should be set so that the linear power level should be at the very bottom of the ALC range for that attenuator setting. Presumably, there is no reason to measure the amplifier below the linear power. Then, for calibration, the source power level is raised to nearly the top of the ALC range. For very best results, the top 5 dB should be avoided in some older VNAs due to some compression of the reference receiver, but if the linear power is very low, even the top 5 dB of the ALC range should be used as the compression may be much lower than the trace noise.

While the idea of calibrating at a high power and measuring at a low power is disturbing to those who were brought up on the idea that changing power after calibration invalidated the calibration, it is very easy to demonstrate that in almost every case, the error from raising the power for calibration is orders of magnitudes less than the error due to noise in the calibration.

Further, if the gain of the amplifier is greater than the change in source power, additional averaging or IF bandwidth reduction should be enabled during calibration at the rate of 10 times change for each 10 dB of DUT gain above the power offset. For example, a 60 dB gain amplifier should have the power set to maximum during calibration; during measurement the power will be set approximately 40 dB lower. To ensure the noise during calibration does not limit the quality of the measurements, an additional 20 dB worth of noise reduction should be used, meaning 100 times IF BW reduction or 100 times averaging. Of course, this supposes that during the amplifier pretest (when the power levels and IF BW should be determined for the particular amplifier used) the IF BW was set so that the trace noise during amplifier measurement is just at the desired level. Finally, depending upon the setup, the noise in the reference channel may limit the measurement results when the power is reduced, so some extra averaging or IF bandwidth reduction may be necessary during the measurement as well.

Figure 6.28 shows several measurements of S_{21} of a high-gain amplifier for different test conditions. Traces marked A show the result of simply lowering the port powers (both port 1

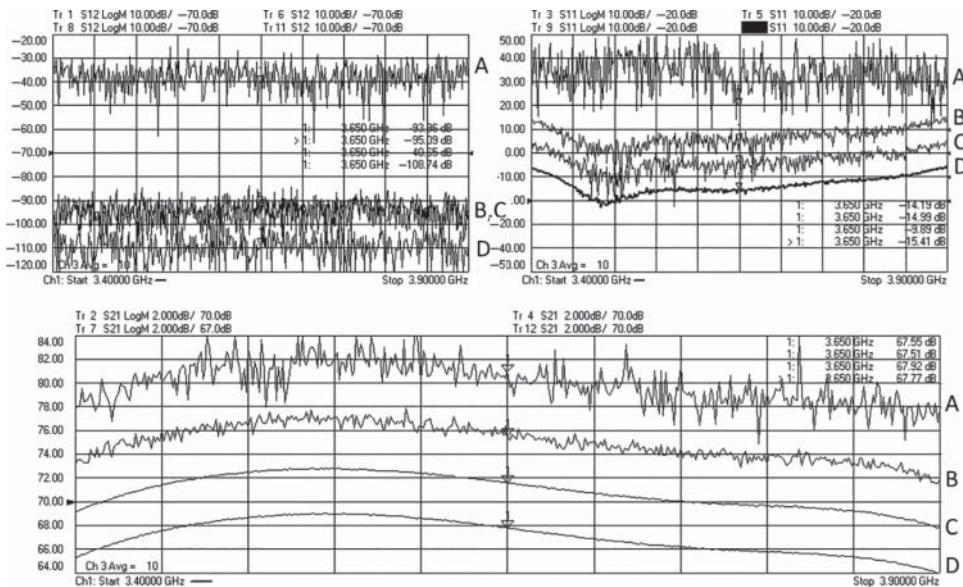


Figure 6.28 S21 noise on a high gain amplifier with various settings; S11, S22, and S21 reference offset 1 division for clarity.

and port 2) to -70 dBm, after preset, and calibrating at the lower power (this is the same result as Figure 6.26). The traces in the S_{11} and S_{21} windows have identical scale/div, but each offset by one division. The traces in the S_{12} window (upper left) have identical scales with no offset. The lower window shows the S_{21} trace.

For this initial calibration state, while the S_{21} trace has about the correct amplitude, it is extremely noisy with about 5 dB of peak-to-peak noise, even after 10 averages.

The S_{12} trace is completely noise at -40 dB, and the S_{11} trace shows only noise centered above 0 dB! Clearly, this measurement is almost entirely invalid, with the error correction itself causing substantial noise, as shown in Figure 6.26, when compared to raw (uncorrected) measurements. This is primarily due to the low test port power of port 2 causing completely invalid readings of S_{12} and S_{22} . Several stepwise changes are shown that each improve the corrected result.

Traces marked B show the effect of uncoupling the port powers and setting the port 2 power to a higher level, at 0 dBm (which is still well below the expected saturation of the amplifier). In the B traces, each of the S-parameters is now more valid, but there is still substantial noise on the S_{21} trace of about 2 dB peak to peak (here, unlike trace A, no averaging is used). Here, the residual noise in S_{21} does not change sweep to sweep, indicating that the measurement itself is not noisy, but rather the noise is embedded in the calibration trace due to low source power used during calibration.

Trace C shows the result of performing the calibration at a higher power while measuring at a lower power. Now the S_{21} trace is clean (the trace noise on the thru-tracking term during calibration has been reduced), but this has no effect on the S_{11} trace, nor on the S_{12} trace, and they are unchanged relative to the B traces.

Finally, trace D shows the result of all the above settings, plus reversing the port 1 coupler. The port 1 attenuator and power setting are changed to increase nominal power by 13 dB to provide the same -70 dBm incident power to the DUT. With this configuration, the S_{12} trace drops by approximately 13 dB, indicating improved dynamic range in the reverse direction, and the S_{11} trace has much lower trace noise, indicating improved S_{11} sensitivity in the reversed coupler case, for the same -70 dBm signal applied to the DUT. There is even an improvement in the S_{21} trace noise as the effect of a noisy S_{11} trace is eliminated. Thus, the configuration of Figure 6.27 shows the clear benefit of proper setup and settings for calibration and measurement of high-gain amplifiers.

6.4 Measuring High-Power Amplifiers

High-power amplifiers are widely used in radar and communications systems. For purposes of this book, amplifiers are considered high power if they cannot be measured using the normal configurations of the VNA and require either external-booster amplifiers, external couplers and attenuators, or both.

High-power amplifiers can be segmented into some classes that dictate the changes required to the setup for the VNA. Amplifiers with high drive requirements will require a booster amplifier in the source path. Moderate power amplifiers, below 1 watt ($+30$ dBm) output, can generally be tested directly at the test port of many VNAs, sometimes with some small amount of simple padding at the port.

Medium power amplifiers, between 1 watt and 20 watts ($+30$ to $+43$ dBm), can be tested with many VNAs directly using the built-in test port couplers but require some reconfiguration behind the test port coupler to add isolators or pads to reduce the signal level to components behind the test port coupler.

High-power amplifiers, above 20 watts, generally need to have external couplers and external high-power isolators and attenuators connected to the VNA source and receiver, essentially bypassing the internal VNA test set entirely.

6.4.1 Configurations for Generating High Drive Power

6.4.1.1 Moderate Drive Levels (Less Than $+30$ dBm)

There are two basic configurations for creating high-power drive signals from a VNA. The first simple configuration is available with some modern VNAs that provide rear-panel loops, which allow direct access between the VNA source and the reference channel coupler. Adding high-drive power for this configuration is as simple as adding a booster amplifier in the loop. Drive levels up to approximately $+30$ dBm are possible with this simple scheme. Some other modifications might be required such as adding an attenuator in the reference path to reduce the signal to the $a1$ receiver, as shown in Figure 6.29. At sufficiently high drive powers, an additional attenuator may be needed in front of the $b1$ receiver; often a VNA has built-in switched receiver attenuators for this purpose.

Typically, there is some setting in the VNA software, called *source offsets*, which allows one to provide an offset to the source power so that the drive levels are approximately correct even without calibration. This is convenient to avoid over drive issues that can occur when the source power settings don't match the test port power.

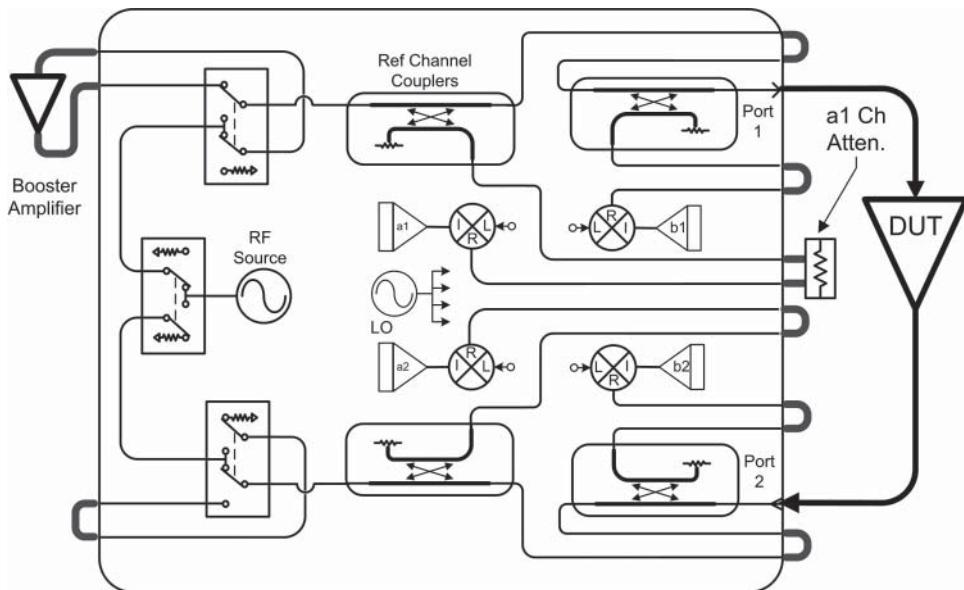


Figure 6.29 Configuration for high power drive using rear panel loops for test port powers to +30 dBm.

6.4.1.2 High Drive Levels (Greater Than +30 dBm)

The second configuration, Figure 6.30, is required if the VNA source path components cannot handle the required drive power. In this version the booster amplifier comes from the source output through a loop between the reference coupler and the test coupler. The output of the booster amplifier is routed through a high-power coupler that provides a signal to the reference receiver (a_1) of the VNA. Some additional attenuation may be needed to avoid over-driving the receiver.

The through path of this high-power reference coupler is then either routed through the test port coupler (if it can handle the drive power), or routed through a second high power coupler, as shown in Figure 6.30. The coupled arm of this reflection coupler is routed to the port 1 reflection receiver, $b1$. The combination of coupling factor and some external attenuators should be added to provide a sufficiently small signal to the VNA receivers for the maximum drive signal from the amplifier. As a safety precaution, one should never set up a configuration that would allow the booster amplifier to overdrive the VNA test receivers. These receivers usually have a guard band of damage level that is 10-15 dB above their maximum operating range. In general, use the smallest power booster amplifier that will support the test needs.

In each of the configurations discussed, the full S-parameter functionality of the VNA is preserved. Of course, if the DUT has gain, then additional steps are needed to allow the VNA to receive high power, as discussed in the next section.

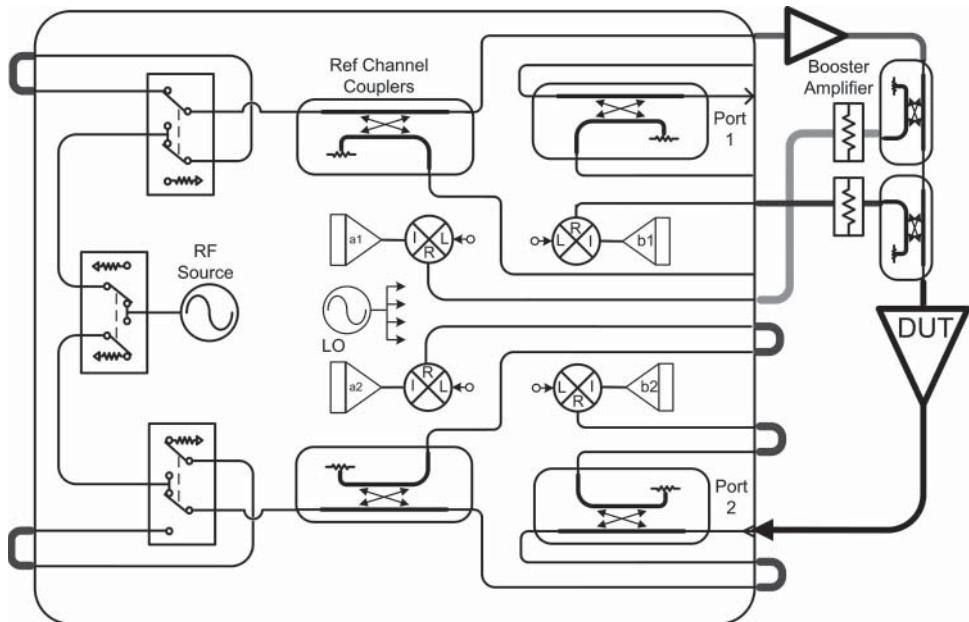


Figure 6.30 High power drive with external couplers.

6.4.2 Configurations for Receiving High-Power

Once a high-power test system has been designed to deliver sufficiently high power to the DUT, the configuration for receiving even higher power from the amplified signal must be considered.

For moderate power devices, less than +30 dBm, the amplifier can be connected directly to port 2, as long as the source and receiver attenuators are set sufficiently high to prevent overload on the internal components such as the source/load switch and the b_2 test receiver. The internal attenuators are typically rated for about +30 dBm. Even for powers up to perhaps +36 dBm, the simplest solution is to add a high-power attenuator to the end of the port 2 cable. In such a case, the S_{22} measurements are generally valid even up to 10 dB loss in front of the test port coupler. For even higher powers, larger attenuation may be used, but in such a case the reverse match, S_{22} , becomes even noisier and less reliable.

In cases where full 2-port calibrations are required, and for powers up to about +43 to +46 dBm, the internal coupler of the VNA may still be used (check manufacturers' specifications for maximum power handling). In this case, external attenuators or isolations should be used to drop the drive power sufficiently to avoid overloading the VNA components such as the port 2 source/load switch. Typically, these components can handle power up to +30 dBm damage level, but operationally they are limited to perhaps +20 dBm. Again, adding a high-power attenuator on the order of 3–6 dB at the port 2 cable can help to dump some power before the test port coupler, as well as improve load match. In trying to obtain

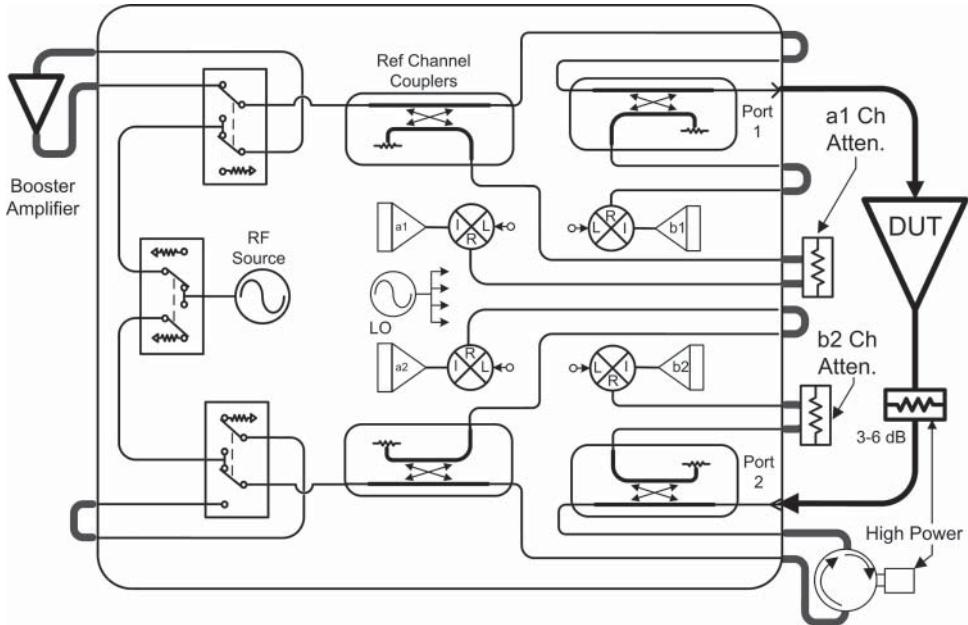


Figure 6.31 Measurement setup for +46 dBm maximum power.

good S_{22} measurements, an isolator or circulator is sometimes used instead of an attenuator behind the port 2 coupler thru arm. This provides for a low loss to the reverse signal while routing power to a load for the forward signal, enabling less noise in the S_{22} measurements. An example block diagram for measurements up to +46 to +49 dBm is shown in Figure 6.31. In some VNAs, the internal bias tee is limited in power, and in such a case it should be removed or an external coupler should be used.

An alternative to adding the isolator is to use an attenuator behind the port 2 coupler, but that can lead to poor S_{22} and S_{12} measurement due to low signal from test port 2. An alternative simple approach is to increase the value of the attenuator on port 2 to some large value and abandon using a full 2-port calibration. Instead, a simple response or enhanced response calibration can be used. Since most attenuators have reasonably good return loss, there will be very little error in using the enhanced response calibration. The only downside to this approach is that the full S-parameter measurements are not possible, so one cannot use the results to do more complex analysis such as K-factor or available gain.

For even higher powers, above +43 dB, if a large port 2 attenuator is not used, an external coupler is required that can handle the high power from the amplifier under test. The block diagram for such a setup is identical to Figure 6.31, but with an external coupler replacing the internal coupler of the VNA.

In some cases, for very high power levels, the attenuator or isolator used for the load may have an issue of changing its impedance due to heating from the high power. If this is the case, then it may be necessary to reconfigure the test system to use three high-power couplers configured as shown in Figure 6.32. With such a configuration, the load impedance may be monitored using the $a2/b2$ trace when the source is coming from $a1$. If the load response is

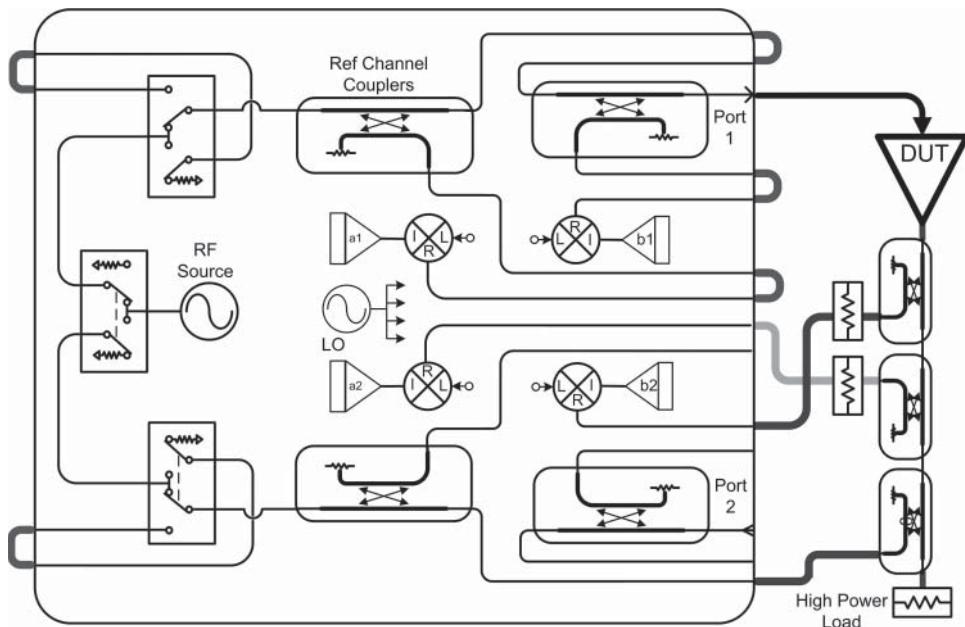


Figure 6.32 Configuration for high power test where the load changes with power level.

not constant when port 1 is driven, the load match term must be monitored and the correction changed to account for drift in the load. These heating effects can be reduced or eliminated if the measurements are made with a pulsed RF signal, as described in the next section.

Accounting for the load in this case may mean rewriting the load match error term after each sweep or applying the computation of S-parameters as described in Eq. (1.21). In this configuration, the reverse power is injected using a third coupler to avoid the need of a high power attenuator and allow the use of high power load instead. The coupling factor of this third coupler should be set so the when the amplifier is driven in the forward direction from port 1, the power will not damage the VNA source at port 2.

6.4.3 Power Calibration and Pre/Post Leveling

The general method of calibration for high-power systems is essentially unchanged, but care must be observed to ensure that the power handling of the calibration standard used is not exceeded if a booster amplifier is used for generating high drive levels. Most particularly, if a power meter or Ecal module is used for calibration, the power for calibration must be controlled to ensure the power does not exceed the damage level of these components.

Because the receivers are very linear in most VNAs, the best solution for calibration is usually to set up the source attenuators such that at the top of the ALC range of the source, the output power is just enough to drive the maximum required output power of the booster amplifier. Then for calibration, the power may be reduced to near the bottom of the ALC range. In many modern VNAs, this can be up to 40 dB lower, or more. Thus, for drive powers

below about 40 dBm, this is no problem for either the Ecal or the power sensor, as the ALC allows the cal power to be as low as 0 dBm. Some VNAs provide a means to bypass or open the ALC loop altogether, allowing power control ranges of more than 70 dB without changing the source attenuator.

The power calibration of the system can be performed independently, or in some VNAs, as part of a Guided Power Calibration method. Great care should be taken during the power calibration portion as the source power is iterated to find the proper drive level, and it is possible that the maximum source power may occur during the iteration process. The difficulty in power calibration is that if a booster amplifier is used, the actual source power out will be higher than the nominal power by the gain of the booster amplifier. Consider the case of a drive amplifier that can produce +35 dBm, with 25 dB gain. The source can be set so that maximum power is +35 dBm and minimum power is -5 dBm. A reasonable power such as 0 dBm might be used for the calibration, and the source power calibration performed to achieve 0 dBm, before the S-parameter calibration proceeds. During the initial setting and reading of the source power, it is common for the VNA software to set the source to the nominal requested power, in this case 0 dBm. But the booster amplifier gain will generate +25 dBm or more, perhaps destroying the power meter. Thus, it is recommended to add some external attenuator to the power meter so that the power meter is safe even at maximum drive. The value of the attenuator can usually be compensated for as a loss factor in the power meter configuration of the VNA.

Some VNAs avoid this problem by providing a source power offset entry, which accounts for external gain (or loss) so that the nominal source power setting includes the effects of the booster amplifier. This is usually a fixed offset so there can still be difficulties if the gain of the booster amplifier has large peaking or is not flat. With this offset, the initial setting of the source is lowered by the gain of the amplifier, and this will avoid overdrive issues. To be safe, the maximum gain of the amplifier should be used for the offset value.

During the S-parameter portion of the calibration, the averaging factor should be increased, or the IF bandwidth should be reduced, so that noise in the calibration does not degrade the measurement results. As a rule of thumb, the averaging should be increased, or IF bandwidth reduced, by a factor of 10 for each 10 dB of power level reduction between calibration and measurement, and for each 10 dB of gain of the DUT, to ensure that the noise contribution of the calibration is on the same order of magnitude as during the measurement.

One final aspect of high power measurements is that for many amplifiers, the performance is specified at a particular output power, rather than some input power. Gain-at-rated-power is one example, where the gain of an amplifier is specified to be at or above some level for a given output power of the amplifier. For these measurements, it is critical that the measurements be made at exactly the output power of the specifications. To achieve this, the receiver leveling function described in Section 6.1.4.1 is modified so that the *b2* receiver is used as the power detector, and the source power is iterated so that the output power is maintained at a constant level.

6.5 Making Pulsed-RF Measurements

For high-power amplifiers or amplifiers operating near compression, the RF dissipation of the DUT can cause heating effects that will change the measurement results. This is particularly

true for on-wafer measurements where it is not possible to adequately heat-sink the die. In these cases, making pulsed measurements with low duty cycles can avoid the problem of device self-heating. Other devices are designed to operate only in a pulsed mode, and so pulsed S-parameter measurements are required.

In older systems, creating and synchronizing pulsed measurements was rather involved. The RF signal from the source must be sent to an external pulse modulator, which in turn must be driven by an external pulse generator, and the pulsed RF signal should be routed to both the reference and test channels to ensure drift in the pulse modulator is accounted for; thus, external couplers were required. Older legacy VNAs had relatively narrow IF bandwidths so that only wide pulses could be used, as the pulse on-time must be long enough for the VNA to capture a data point in the IF, and the triggering of the pulse generator, modulator, and measurement receiver must be all synchronized, which might require special interface circuits. However, these methods have been largely replaced by modern techniques.

6.5.1 Wideband vs. Narrowband Measurements

Wideband pulse measurements refer to the idea of using a wideband IF, with a very fast response time, to measure the RF signal during the pulse on-time. For example, if a $10\ \mu\text{s}$ RF pulse is used, the IF bandwidth would need to be wider than approximately one over the pulse width, or wider than $100\ \text{kHz}$, to capture the energy of the pulse. Typically, an IF BW of 1.5 times is the minimum used to ensure entire IF measurement is made even if there are some timing errors and pulse delays. A $15\ \text{MHz}$ bandwidth, the widest currently available on modern VNAs, can provide for measurements on pulses as narrow as $100\ \text{ns}$. Figure 6.33 illustrates the timing diagram for pulsed measurements, overlaid on an example measurement. Here the pulse is set to $10\ \mu\text{s}$ on-time with $25\ \mu\text{s}$ pulse repetition time. The RF pulse modulator signal is delayed by $5\ \mu\text{s}$ after the start of sweep to see the full pulse rise time. It is common to have some delay due to hardware path differences between the pulse-generator's sync pulse (called Pulse 0) and the start of ADC data taking. Slightly offsetting the delay of Pulse 0 can compensate for these delays. Also, the RF power signal will be delayed by the rise time and

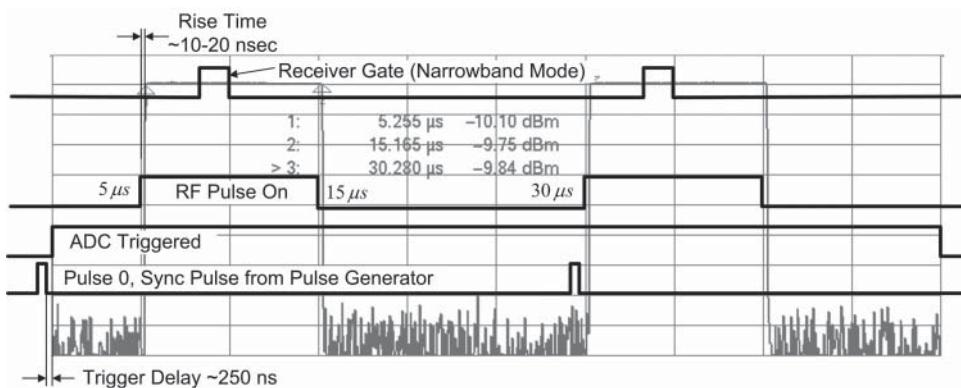


Figure 6.33 Timing diagram for wideband pulsed measurements.

response delay of the pulse modulators in the VNA (or external pulse modulators if the VNA does not have an internal pulse modulator).

For narrower pulses, or for systems with narrower IF bandwidths, an alternative method called the *narrowband approach* can be applied to achieve measurements on very narrow pulses, down to 10 ns of pulse width (Shoulders and Betts 2008). In the narrowband mode, the receiver of the VNA is time-gated to measure just a narrow portion of the pulse signal, as indicated by the upper line of the timing diagram. The narrowband approach relies on the fact that a repetitive RF pulse will have a spectrum of discrete frequencies related to the pulse repetition frequency, as shown in Figure 6.34. If a narrow band IF filter is centered on just one of these discrete frequencies, with filter zeros centered on each of the other frequencies, then the overall RF measurement can be made using very narrow IF filters, by accumulating a signal in the IF. The difficulty in this method is that custom IF filter bandwidths are needed for each different pulse period. In a sense, the narrowband mode averages the results of several pulses in the IF filter and displays the result of this average. The only way this average gives a good response is if all but the central spectral element are removed from the average, through the use of the custom filters. Figure 6.34 illustrates the spectrum of a pulse and time domain measurement of the pulse profile, for a 3.3 μ s pulse with a 50 μ s pulse repetition time (20 kHz pulse frequency). The 20 kHz pulse repetition frequency implies that the spectrum should have 20 kHz spectral components; Marker 1 is set at the spectral peak, and Marker 2 is at the next spectral line, exactly 20 kHz away. A narrowband IF filter would need to pass the center of this pulse and have a zero at 20 kHz intervals. In practice, even narrower

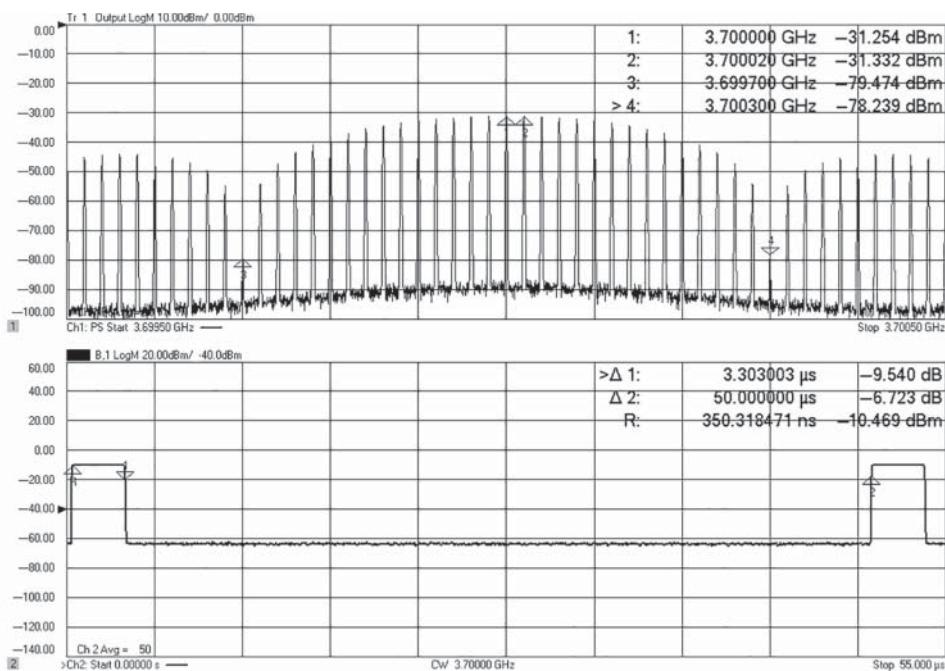


Figure 6.34 Narrowband pulse measurement spectrum and time measurement.

IF filters can be used to lower the noise floor, as long as the IF filter has transmission zeros at each of the other spectral lines. Since the pulse width is $3.3\ \mu\text{s}$, one would expect nulls in the pulse spectrum every $1/(3.3\text{e-}6)$ or $300\ \text{kHz}$. Markers 3 and 4 are positioned plus and minus $300\ \text{kHz}$ from the center of the spectrum and indeed have substantial nulling.

With narrowband pulse methods, the average power displayed is less than the true power by the ratio of the receiver gate time and the pulse repetition time. For power measurements in narrowband mode, the received power should be modified by this ratio to reflect the true power in the pulse, using a magnitude offset function or an equation editor function.

6.5.1.1 Point-in-Pulse Measurements

The most basic pulsed RF measurement is measuring the S-parameters and power of an amplifier in the center of an RF pulse, over a range of frequencies. This is essentially just a standard measurement, but with a pulsed RF stimulus. This is sometimes referred to as *point-in-pulse* measurements and refers to the fact that for each point on the trace for the frequency sweep, one RF pulse is measured, typically in the center of that pulse, as illustrated in the timing diagram in Figure 6.33. Thus, the normal frequency response of the amplifier is measured, and the RF pulse becomes just another stimulus setting.

Most modern VNAs now have a high-speed digital IF, some with sample rates up to $100\ \text{MHz}$. Some high-performance units have built-in pulse modulators and pulse generators, all internally synchronized, making pulsed RF measurements particularly simple. In fact, some application programs require that the user only enter the pulse width and pulse period (or pulse repetition frequency) and every other setting is automatically adjusted to create the proper pulsed measurement.

For many pulsed systems, the pulse modulator is inside the source ALC loop. Since the ALC loop typically has a response slower than the pulse, the ALC loop is automatically disabled and put into an open loop or sampled mode. Since the internal detector is not used, the RF level can have substantial errors in this mode. Once again, the use of the reference channel receiver as a leveling detector can provide a substantial improvement in the accuracy of the drive power. The Rx-leveling function in pulsed mode uses the same method for leveling as the in normal mode, where background sweeps are used to iterate the source power and find the proper level; then a data-taking sweep is performed with the corrected source power. Figure 6.35 shows an example of a pulsed RF measurement on an amplifier, with and without Rx-leveling. The upper window shows the input and output pulsed power (along with S_{21}), in the open-loop ALC mode, with no correction. The lower window shows the same stimulus, but with receiver leveling on the input. The input power and output power flatness are clearly improved by the Rx-leveling function.

Noise reduction for the power measurements are usually achieved by reducing the IF bandwidth; however, the IF bandwidth must remain wide for pulsed measurements. Sweep averaging for power measurements averages the power (rather than the voltage) because the phase is not coherent for power measurements on a sweep-by-sweep basis; this means that noise power is not reduced and adds to the trace value. In contrast, for ratio measurements, the phase is retained on a sweep-by-sweep basis so the noise is averaged away in S_{21} for any averaging mode. However, some VNAs provide for point-mode averaging, which takes many samples of the same point before moving to the next frequency. In the pulsed case,

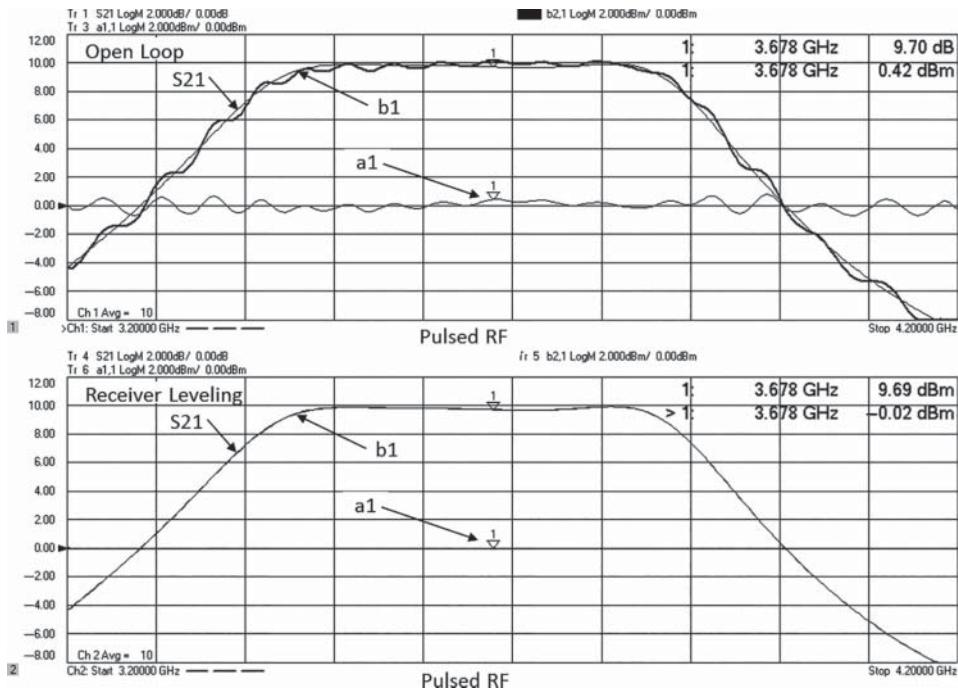


Figure 6.35 Pulsed RF amplifier measurement show $a1$ power before and after Rx leveling.

this means that several pulses are acquired, and the results are averaged. For power measurements, the phase is coherent from pulse to pulse, so vector mode averaging can be used to reduce the noise in pulsed power measurement. Thus, point averaging provides a way to get low-noise power and gain measurements of a pulsed signal, where IF bandwidth reduction is not possible.

In some VNAs, it is possible to use the current source settings of the Rx-leveling to update the existing source power calibration so that the open loop mode can be used without the background sweeps of the Rx-leveling function. This is particularly important when pulse-profile measurements are made on amplifiers, as described in the next section.

6.5.2 Pulse Profile Measurements

For many amplifiers used in pulsed applications, the measurement of the amplifier's response to a pulsed RF input as a function of time is a key concern. These measurements are generally referred to as *pulse-profile measurements* and show the gain, phase, and power response of an amplifier to a pulsed stimulus versus time relative to the pulse. The measurement method differs significantly from the point in pulse measurement in that the entire measurement occurs over a single pulse. The use of a wideband digitizer sets the effective resolution of the pulse profile measurement to essentially the inverse of the widest bandwidth of the digitizer; that is, in wideband mode, the bandwidth of the filter sets the minimum time step size for the pulse profile resolution.

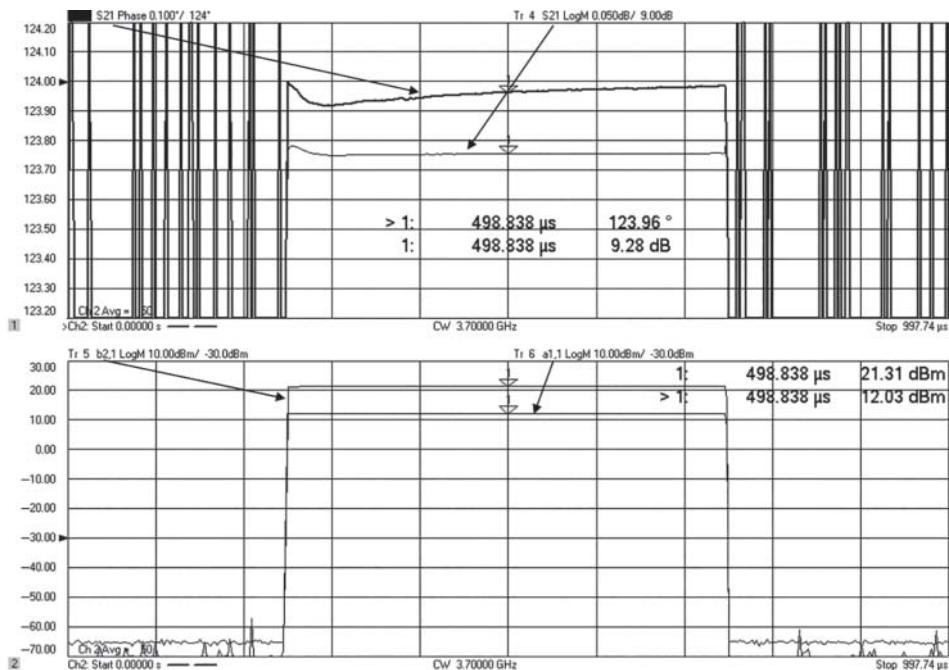


Figure 6.36 Pulse measurement shows gain, phase, and output power of an amplifier during a pulse.

It is possible to set up pulse profile measurements using external pulse generators and modulators, but some modern VNAs have complete built-in pulse-profile functions that make the measurements completely simple. An example of a wideband pulse profile measurement is shown in Figure 6.36. In this case, the noise floor and trace noise of the measurement is limited by the wide IF bandwidth that must be used to capture the pulse. From the timing diagram of Figure 6.33, the pulse profile is measured by moving the receiver gate a small increment. In wideband mode, this increment is the IF BW data taking time.

Trace noise on ratio measurements such as S-parameters can be improved by sweep-to-sweep averaging, but the noise floor for power measurements cannot be lowered in this way. For the figure shown, 100 averages are used. In this case, a 3 MHz IF bandwidth filter is used so the resolution is about 330 ns. The acquisition time for a 1 μs pulse profile is 2 μs including pre- and post-pulse measurements of 25% of the time axis. So, 100 averages take only about 0.2 μs of measurement time and reduce the trace noise on the S₂₁ trace substantially. The total acquisition time for the measurement in Figure 6.36 is about 0.3 μs due to retrace overhead.

Pulse profile measurements are also possible in the narrowband mode. Narrowband measurements require many pulse acquisitions for each measurement result. In a pulse-profile measurement, the effective timing of the measurement is changed by incrementing the pulse signal to the receiver trigger (sometimes called *receiver gating*) so that each successive data point is acquired on a different relative position of the pulse. In narrowband mode, the resolution can be set by the user; narrower gates mean more pulses will need to be measured to average out noise due to a narrower gate causing lower signal level. The gate

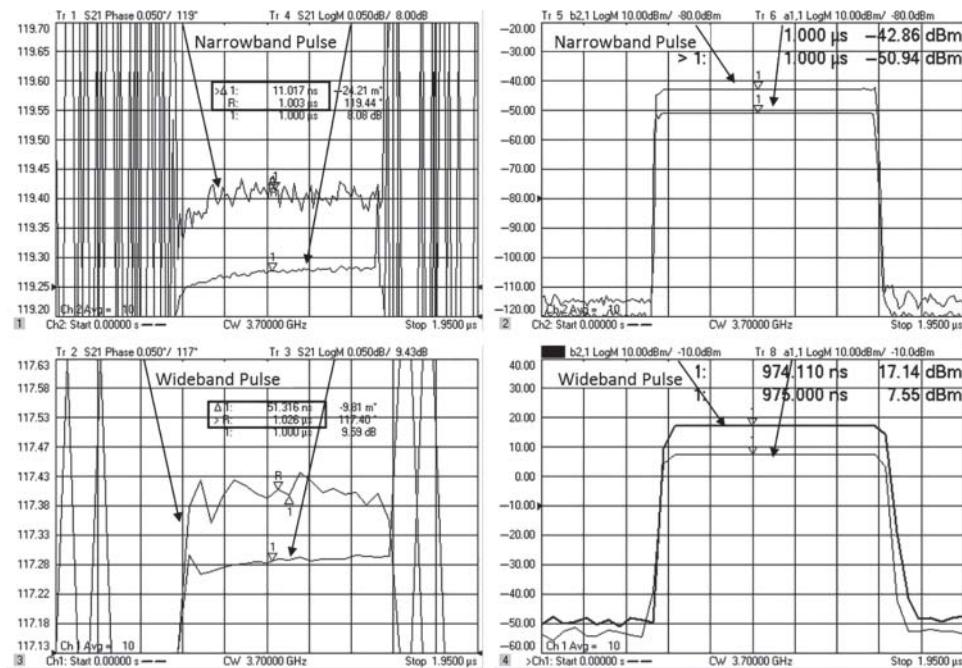


Figure 6.37 A narrowband mode pulse profile on a narrow pulse.

width (pulse resolution) in narrowband mode is set by the pulse-generator minimum pulse width, or the minimum gate response time of the receiver gating circuit.

In this way, the pulse measurements are possible on pulses that are much narrower than the native IF bandwidth of the VNA would allow. For VNAs with built-in pulse generators, this resolution can be as small as 10 ns; or external pulse generators may be used where the limit is now the minimum gating time of the VNA receiver, perhaps as small as 5 ns.

Older, legacy VNAs had narrower IF bandwidths, on the order of 10's of kHz maximum, so the minimum pulse time was limited to 100 microseconds or so. Modern VNAs have high-speed digital IFs with greater than 10 MHz of bandwidth, so much narrower pulses can be measured in wideband mode, reducing the need for narrowband measurements in many cases.

An example of the narrowband mode pulse profile is shown in the upper windows of Figure 6.37. Note that this is really a composite response of many pulses and represents an average of the pulse response over the pulses needed to acquire the IF bandwidth of the narrowband mode. For comparison, a wideband pulse profile measurement is shown in the lower window.

In the previous example, the pulse on-time is 1 μ s and the pulse repetition rate is 10 μ s, a 500 Hz IF bandwidth filter is used to remove the unwanted pulse spectrum. The upper left shows the delta time from one point to the next in a highlighted box; here the step time is 10 ns.

The lower windows show the same pulse measurement in the wideband mode, which has continuous sampling across the pulse. In this case, the VNA has a maximum IF BW of 15 MHz, which yields approximately a 50 ns pulse resolution. In contrast, the narrowband

mode (upper windows) has a resolution that is limited only by the pulse generator resolution, in this case 10 ns, so many more measurement points are available during the pulse period. Note that the pulsed power measurements (right windows) are offset in the narrowband mode (upper plot) by the duty cycle of the receiver gating to pulse period. In this case, this ratio is 10/10000 for about a 60 dB offset. The narrowband method does allow a finer resolution to be seen on the S_{21} trace of the pulse.

6.5.3 Pulse-to-Pulse Measurements

Pulse-to-pulse measurements are used to measure the effect of device performance over time on a pulse train. In particular, on a high-power application such as a radar system, an amplifier may have more gain in response to a first RF pulse than to subsequent RF pulses due to heating or other effects. In a pulse-to-pulse measurement, the triggering is set to gate the receiver to measure only one point, on one pulse, for each measurement pulse trigger. The next pulse triggers a measurement on the same position in the next occurring pulse. If plotted versus time, this pulse-to-pulse measurement shows the drop or gain change from one pulse to another for the same point on the pulse. The idea is graphically represented in Figure 6.38. These measurements are usually performed in a single-sweep mode, and between sweeps the DUT is allowed to return to an ambient level.

6.5.4 DC Measurements for Pulsed RF Stimulus

A final aspect of pulsed amplifier measurements is measuring the DC voltage and current on an amplifier during the pulse to compute the PAE of a pulsed amplifier. This is particularly important when testing on-wafer where poor heat sinking can cause significant amplifier performance variation due to heating in a CW measurement.

There are two basic DC configurations: the first configuration supplies constant DC voltage to the DUT, and the second pulses the DC as well as the RF. For some designs, the current dissipation doesn't occur until the RF signal is turned on. For these devices, constant DC sources may be used. Other devices require that both the RF and the DC be pulsed to avoid heating. For these devices, special DC switching circuits must be used as few commercial DC supplies provide for high-speed pulsed DC outputs. If the DC values can be measured during the pulse on-time, then the PAE of a pulsed amplifier can be determined.

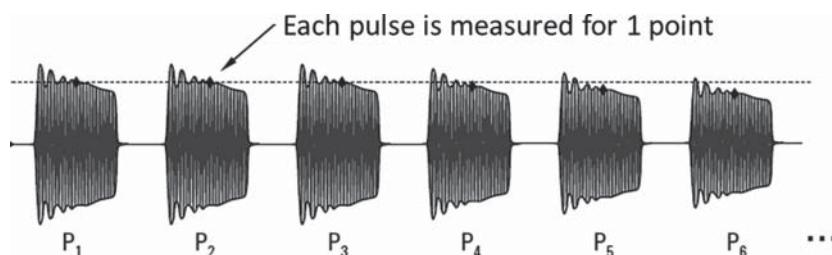


Figure 6.38 Pulse-to-pulse measurements.

6.5.4.1 Pulsed-DC and Pulsed-RF Measurements

Pulsed DC stimulus can be combined with pulsed RF stimulus, provided care is taken to properly sequence the DC stimulus with the RF stimulus and measurement. A multichannel pulse generator is usually needed to provide a wider DC pulse than RF pulse. The DC pulse usually must turn on before the RF pulse is applied. This sequencing can be created by making the DC pulse wider than the RF pulse and then delaying the RF pulse to center it in the DC pulse window.

In either case of pulse or fixed DC stimulus, the pulsed DC measurements are quite similar. As described in Section 6.1.4.2, some VNAs contain DC inputs that can also be gated along with the receiver inputs. Often the bandwidth of these DC inputs is fixed so that the measurement time of the DC input is limited by the bandwidth. For a typical bandwidth of 25 kHz, the DC resolution time will be about 40 μ s. Since these ADCs are not synchronized with the internal DSP, it is difficult to use them for pulse-profile measurements of RF and DC, and their use is typically limited to swept-frequency point-in-pulse measurements.

Some VNAs, however, allow direct IF inputs into the high-speed digitizer used for the digital IF; in this case, the DC pulse profile can be measured in a similar manner to the RF pulse, and higher-speed measurements are possible. To allow this measurement, the digital IF center frequency is changed to 0 Hz, allowing direct measurement of the DC signal. Since the digital IF typically is centered at only one frequency at a time, the DC measurement must be made on one channel while the RF measurement is made on a different channel. This also presumes that the pulse response of the amplifier is consistent on a sweep-to-sweep basis, as at least two sweeps are required for acquiring both DC and RF power.

Figure 6.39 shows an example of a DC measurement on a 1 μ s pulse, using the two-channel approach with the high-speed digitizers measuring both the RF signal (from the VNA test ports) on one channel, and the DC signal (from the direct DC input) on another.

For this measurement, the DC input (V1 above) was set on the DC source, and using a 1 Ω current sense resistance, the voltage on either side of the resistor routed was the IF input of the VNA, marked V1 and V2 in the upper-left window. A 400:1 voltage divider was used to ensure the DC input did not exceed the maximum DC input of the digitizers; the series value of this divider was 10 K Ω , ensuring that very little current was absorbed by the DC measurement. Remote commands were sent to the VNA to set the IF frequency to zero, allowing DC readings with the IF inputs. The DC IF BW is set to be the same as the RF bandwidth, allowing noise reduction on the DC readings as well. The DC offset and scaling, as discussed in Section 6.1.4.2, is applied using the equation editor. When using the digital IF to measure DC in this way, the DC is represented as a complex number, and the proper reading must be computed from $\sqrt{re^2 + im^2}$; this computation is performed by the equation editor in traces 1 (V1) and 2 (V2) in the upper-left window. The current is computed as (V1-V2)/R and is shown as I_DC in the lower right window. The upper-right window shows a close-up scale of the amplifier output power, and the lower-left window shows both the output power (at a large scale) and the PAE, computed from the other traces, again using an equation editor function.

One aspect of interest on this amplifier is the peaking of the output power at the initial portion of the pulse. After the initial turn-on time, the output power decreases somewhat, also seen in the PAE trace. There is also some odd behavior in the DC current consumption when the RF pulse goes off; this may be related to the current draw of the amplifier (as part of an RF dependent self-bias), or it is possible it is related to elements of the bias tee. In this

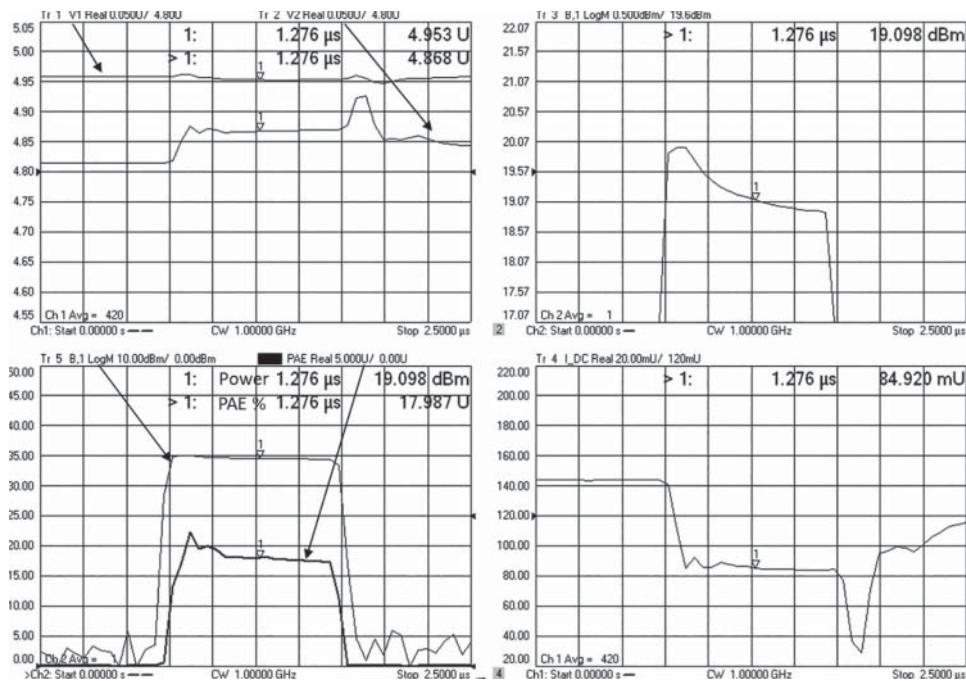


Figure 6.39 Pulse profile showing DC measurements and PAE.

case, the bandwidth of the bias tee is about 2 MHz, which implies it could have some response to very high-speed pulses. It is also interesting to note that the measured value of V1 is not constant, indicating some non-zero source impedance in the DC source.

6.6 Distortion Measurements

For basic amplifier characterization, distortion falls into two categories: harmonic distortion and inter-modulation distortion. Amplifiers operating in the non-linear region create distortion products, which are a direct consequence of the gain compression of the amplifier. Looking at a single-tone CW RF signal in the time domain, compression causes the peak of the signal to flatten, resulting in harmonic distortion. If the compression occurs equally on the positive and negative peaks so that the distortion of the time domain waveform is symmetric, the harmonic distortion will be strictly third order, and that is usually a sign that the bias point of the amplifier has been optimized. However, many amplifiers also generate second-order distortion in the form of second-order harmonics. For narrowband amplifiers, the harmonics created in the active device are often filtered out by the matching networks so that there is essentially no harmonic power at the output of the amplifier. These amplifiers can still cause distortion of modulated signals, and two-tone or multitone signals can be used to ascertain the distortion quality of these devices. The measurement of two-tone intermodulation distortion is discussed in detail in Chapter 8.

6.6.1 Harmonic Measurements on Amplifiers

Traditionally, harmonic measurements have been performed by using a signal source to stimulate an amplifier, measuring the resulting spectrum on a spectrum analyzer. For a direct display of harmonic distortion, the spectrum analyzer must be swept from the fundamental frequency up through the highest harmonic frequency; the other alternative is to make narrower bandwidth (or even zero-span) measurements at precisely the fundamental and harmonic frequencies, but this requires multiple measurements not normally supported on a spectrum analyzer. Some VNAs also provide a spectrum plot of frequency response. For either instrument, the broadband spectrum can take significant time to measure, but the measurement results are very simply interpreted, as shown in Figure 6.40. The markers are used to measure the fundamental and harmonic powers. If a reference marker is placed on the fundamental, the harmonic markers can display their values directly in dB relative to the fundamental.

Most modern VNAs can be used to measure harmonics directly and recently new capabilities have been added that allow direct swept acquisition of multifrequency measurements, such as harmonics, such as option 089 (Different and IQ Measurements Application) in the Keysight PNA series. In this application, a series of related (or unrelated) frequency ranges can be defined, as shown in Figure 6.41. In this case the ranges are coupled with each range (F2–F5) set as a different harmonic multiplier of the first range.

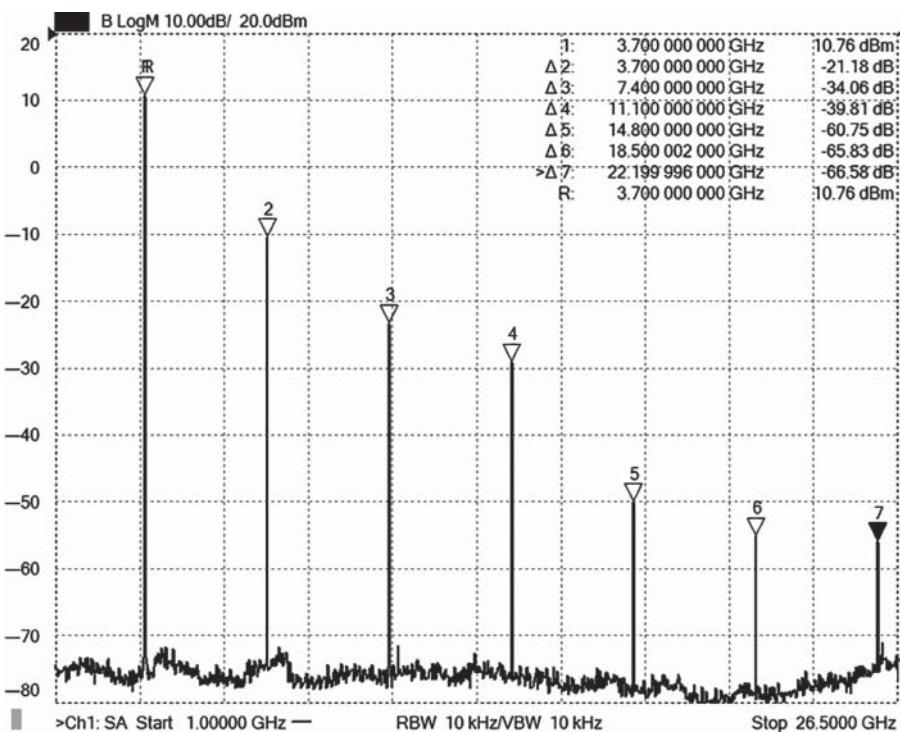


Figure 6.40 Spectrum plot of an amplifier's harmonic response.

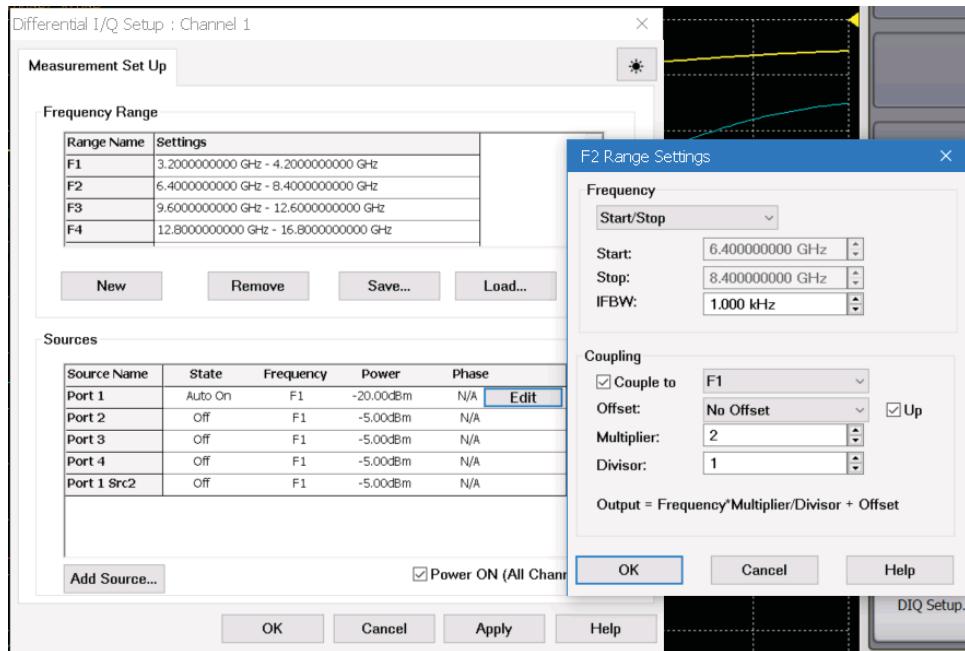


Figure 6.41 Setting up for a harmonic measurement.

These ranges define the frequency measurement points, and any of the ranges can be assigned to the any of the sources. In this case, the first frequency range is assigned to the port 1 source. Each range is divided into the same number of points, and for each point in the sweep, the source is set to the proper frequency and power, and the VNA receiver is moved to each of the different frequency ranges in turn, measuring the value of signal. Thus, a single channel can make quite complex measurements.

The measured (complex) value of the receivers is available at each of the frequencies and at each of the receivers and can be combined in a variety of ways to create quite sophisticated measurements (some are discussed in Chapter 7 with reference to IQ modulators, and in Chapter 10 in relation to measuring the harmonics of a differential amplifier). Here the measurements are quite simple, either the fundamental and harmonic power in dBm, or, as is common, expressing the harmonics as a dBc value in relationship to the power of the fundamental. These are simply defined using the equation function as demonstrated in Figure 6.42. As shown, the second harmonic in dBc (labeled H2dBc) is defined as the power at the port 2 receiver (b_2) at the second-harmonic frequency, divided by the power at the same receiver at the fundamental frequency. Also shown are the direct harmonic powers (labeled H2 through H5) and simply defined as the power at the b_2 receiver at each harmonic frequency.

The dBc value of harmonics parameters can be displayed directly so that the entire measurement is captured in a single-channel measurement and on a single display. An example of this is shown in Figure 6.43. In this way, a harmonic measurement across hundreds of frequency points can be completed in a fraction of a second. Here the fundamental power is labeled “OPwrF1” (for Output Power at frequency F1), and the first through fifth harmonics,

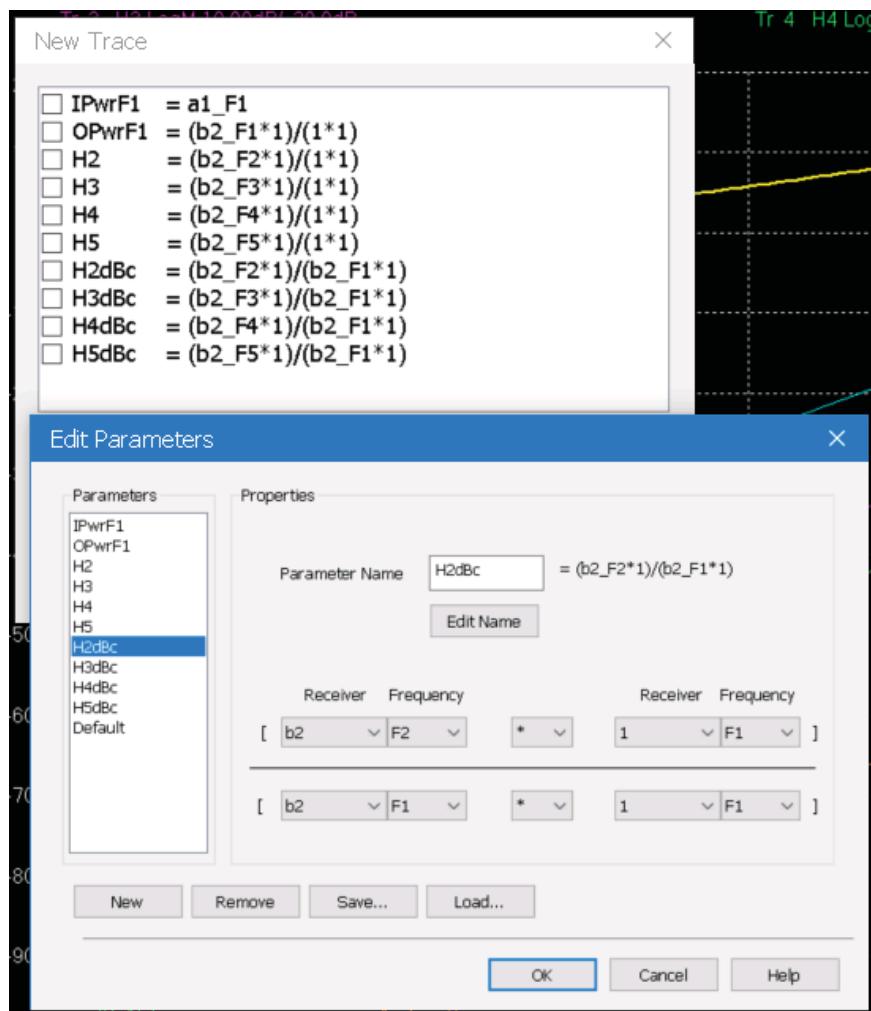


Figure 6.42 Defining harmonic measurement parameters.

in dBc, are labeled “H2dBc” through “H5dBc.” Note that the dBc value of harmonics, at the center frequency, follow almost exactly the single-frequency spectrum measurement shown in Figure 6.40 within 0.1 dB up to the fourth harmonic. By the fifth harmonic, the signal is quite low and trace-to-trace noise will cause variation in the signal. This is a filtered amplifier, such as might be found at the input of a frequency converter, driven with a constant input power. The output power in the amplifier’s main channel is relatively flat, but the swept frequency harmonics show substantial variation over the same channel. Making this measurement using traditional stepped-signal generators and a traditional spectrum analyzer, sweeping once for each step of the signal generator, would be more than 100 times slower than the measurement shown in Figure 6.43.

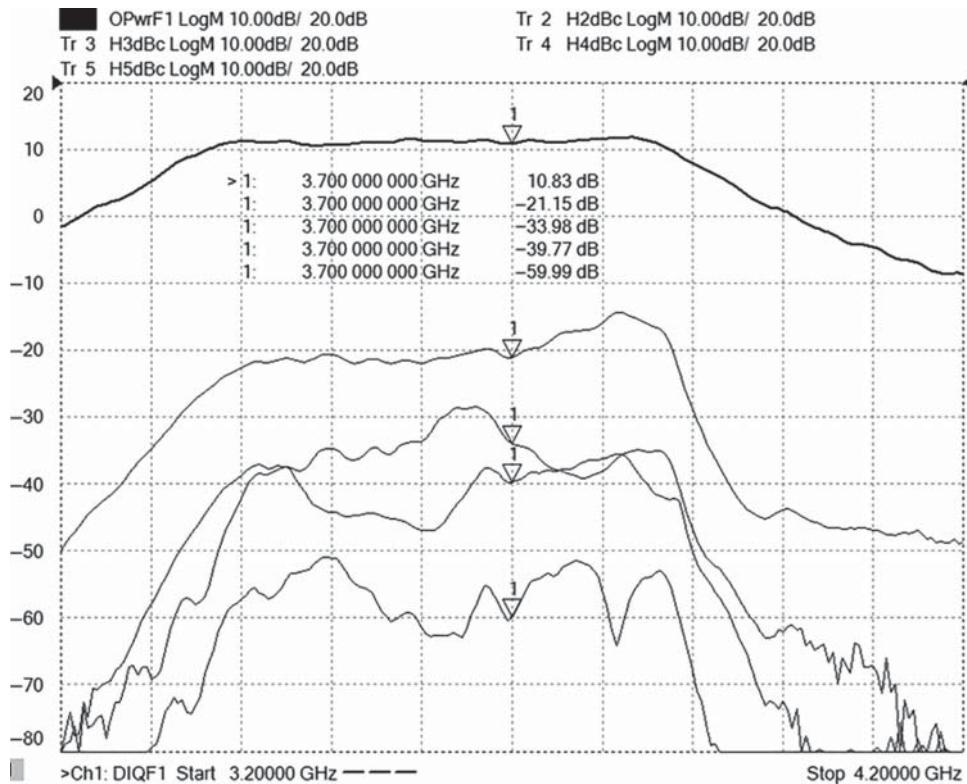


Figure 6.43 Harmonic measurements on a VNA.

For harmonic measurements, one must be sure that the source and receiver of the VNA does not produce harmonics in excess of the DUT harmonics. Most VNAs have rather poor harmonics, so some filtering is required between the source and the DUT. However, some high-performance VNAs include banded switch filtering to produce exceptionally clean harmonic sources for distortion tests. One can evaluate the source and receiver harmonics by making two tests on the VNA with just a thru connection. The first test looks for source harmonics, so a large attenuator is added to the port 2 receiver to ensure that it is operating in a linear mode. The IF bandwidth can be reduced to a very low value (even to 1 Hz) and the source harmonics measured in FOM mode for second-, third-, and fourth-order harmonics. The upper window of Figure 6.44 shows a VNA measurement of harmonics (in dBc) with a 30 dB pad in front of the port 2 receiver (either an external attenuator or an internal receiver-attenuator, if available, can be used). In this case, the source harmonics over the receiver bandwidth are lower than -62 dBc, measured at 10 dBm with a 10 Hz IF bandwidth.

The lower window shows the same measurement, but with the attenuator removed. If the displayed harmonic level is above the source level, then one can assume this is caused by the receiver harmonics. If the source harmonics are too great to evaluate the receiver harmonics, a filter may be used to evaluate a narrow portion of the receiver by removing any source harmonics. In this case, the receiver harmonics are as bad as -37 dBc at $+10$ dBm input.

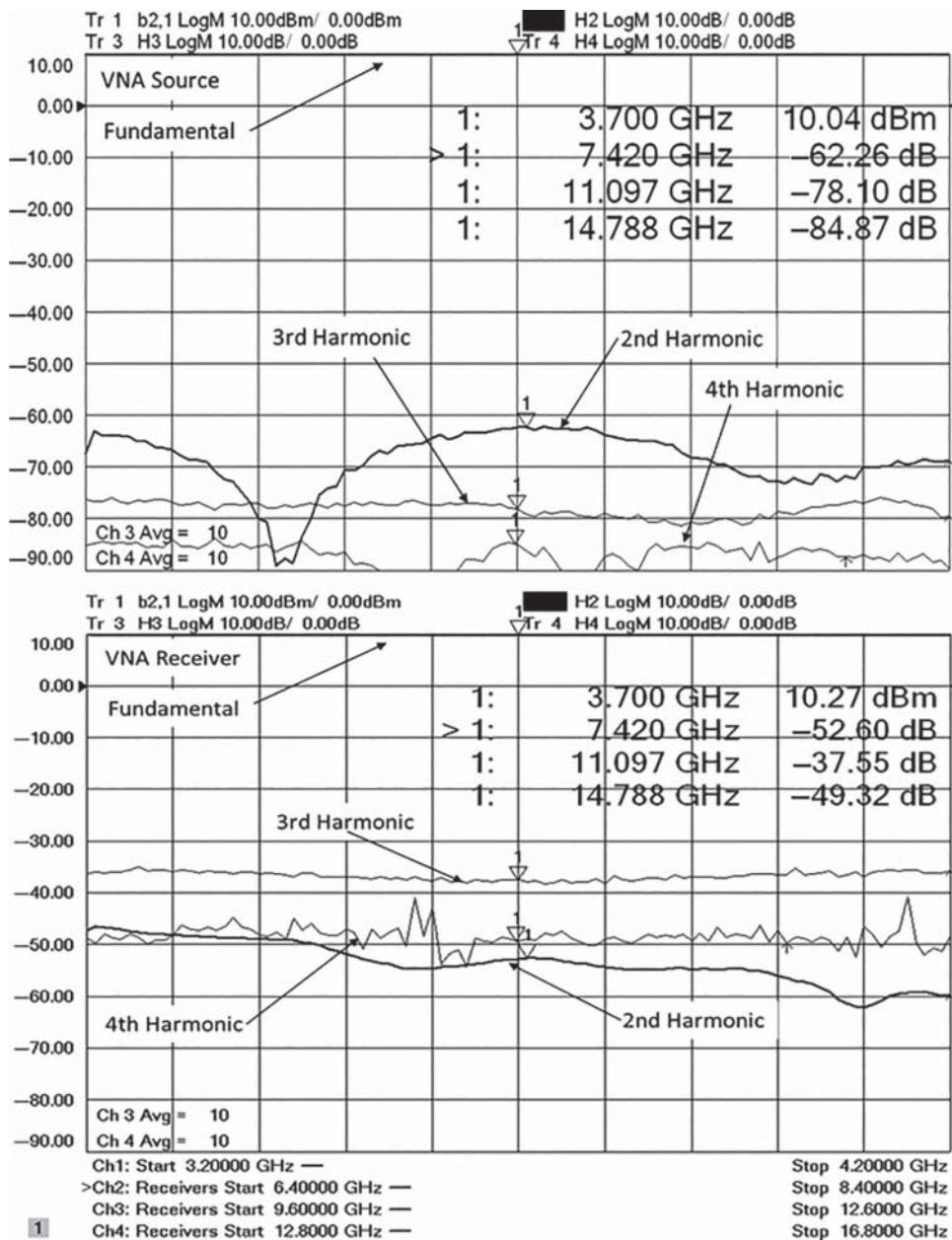


Figure 6.44 Evaluating VNA harmonics: source harmonics (upper); receiver harmonics (lower).

Calibration for harmonics can be derived directly from the receiver power calibration. In the case of the DIQ application, the power calibration is automatically derived from a power meter measurement of the one source, which is transferred to all the other channels using an S-parameter calibration.

In some cases, the VNA receiver harmonics dominate the measurement, and it is not possible to see the device harmonics because of this. This often happens with amplifiers that have built-in matching networks that short out the harmonic response. The normal approach to improving receiver harmonics is to increase the attenuation to the receiver and decrease the IF bandwidth to make up for the loss of sensitivity. However, this method does not work for pulsed RF measurements if the pulse width is narrower than the IF bandwidth acquisition time. In such a case, the minimum IF bandwidth is set by the inverse of the pulse width. This limits the maximum VNA receiver attenuation as the noise floor is set by the minimum IF BW. Unlike a source measurement, one cannot simply filter the output of the amplifier to remove the fundamental, as the harmonic level in dBc requires a measurement of both the fundamental power and the harmonic power.

In these situations, the solution may be to use a harmonic enhancement technique that utilizes a special circuit that provides some substantial but reasonable loss to the fundamental, while providing lower loss to the harmonics of the amplifier. An example of such a circuit can be constructed from power splitters, an attenuator, and a high-pass filter, as shown in Figure 6.45, along with the frequency response of the circuit.

This circuit should be placed between the coupled arm of the port 2 coupler and the port 2 receiver (*b*2) so that any mismatch from the circuit is reduced by twice the test port coupling

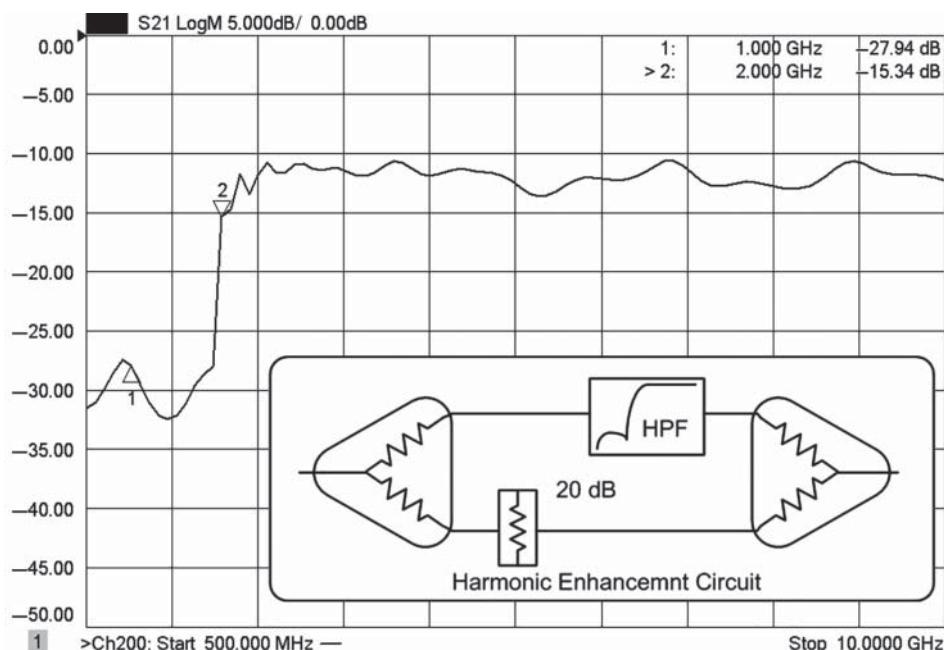


Figure 6.45 A harmonic enhancement circuit with its frequency response.

factor. This technique is particularly useful in amplifiers with harmonic matching networks on the output, which effectively filter the harmonics of the DUT. This circuit gives more than 15 dB of reduction of the fundamental, which gives 15 dB of improvement in the second harmonic measurement, and 30 dB for the third harmonic. This circuit should be in place during the receiver power calibration to remove the effects of its frequency response from the measurement results.

6.7 Measuring Doherty Amplifiers

Doherty amplifiers are a particular configuration that combines two discrete amplifiers into a single gain block, and that has a key advantage of improving PAE at lower drive powers. The configuration includes a carrier amplifier and a peaking amplifier, as shown in Figure 6.46. Doherty amplifiers have the interesting characteristic that they have higher efficiency when the power is backed off from saturation. Typical Class A and Class AB amplifiers have a constant DC current versus input drive and so show maximum efficiency when the power amplifier is driven to saturation (since the DC current is constant, the power consumption is constant so peak efficiency is at saturation, as that produces peak power at the same DC input). For many newer communications formats, such as orthogonal frequency domain multiplexing (OFDM) or high-level quadrature amplitude modulation (QAM), the signal can contain high peak-to-average ratios, which means that the signal has a few, very high peaks occurring occasionally while most of the time the signal is near the average power. Thus, it is desirable to have an amplifier that can handle the peaks without distortion but is most efficient at the average power. Doherty amplifiers provide this characteristic.

The design task is matching the output of the amplifiers in such a way that the impedance is transformed to generate a load that optimizes power. The carrier amplifier is always on and generates power at all times. The peaking amplifier is biased in such a way that it is substantially off for small signals and turns on to provide peaking voltage when the signal is large, such as during the peak of a modulated signal. Because the peaking amplifier is on only a small percentage of the time, it doesn't contribute much to the overall power dissipation and thus improves the PAE while providing for a much higher peak voltage available from the amplifier.

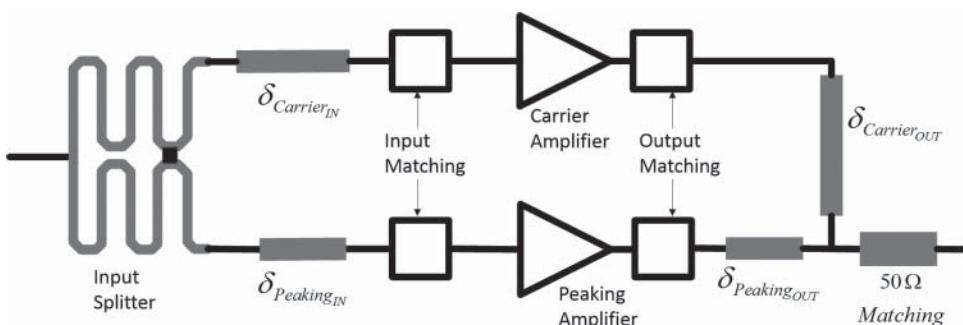


Figure 6.46 Doherty amplifier block diagram.

One of the key design challenges is getting the correct phasing on the input and the output of the amplifier. A typical design utilizes matching networks at the input to get maximum power delivered to the amplifier input, and matching and phase shifting networks on the output to optimally combine the signals from each amplifier. The input is typically driven by splitting circuit that provides the same phase to each amplifier input.

In the past, ensuring the proper phasing of the matching networks and the output phase-combiner has been a difficult challenge, and difficult to verify. Today, using the same differential and IQ application that was used to measure harmonics in Section 6.6.1, the evaluation of proper phasing between the paths has become much easier. During the design phase, if the splitter is removed, a dual-source VNA can be used to drive the input of the Doherty amplifier directly, and the VNA sources phase and amplitude can be precisely controlled to evaluate the performance of the amplifier, as shown in Figure 6.47. A common design technique is to use blocking capacitors or $0\ \Omega$ resistors in the splitter common input, which can be removed and repositioned to isolate the two halves of the splitter into two individual inputs.

A first test is to check the gain and PAE across the frequency range, as shown in Figure 6.48. Here the VNA sources are set to sweep across frequency but maintain a constant phase relationship. A key point here is the PAE measure must include the input power from each port, as well as the DC power. The PAE peaks where the gain (and output power) peak.

To evaluate the behavior of the amplifier with respect to power drive, a fixed-frequency power sweep may be performed, by holding the frequency and phase of the two sources

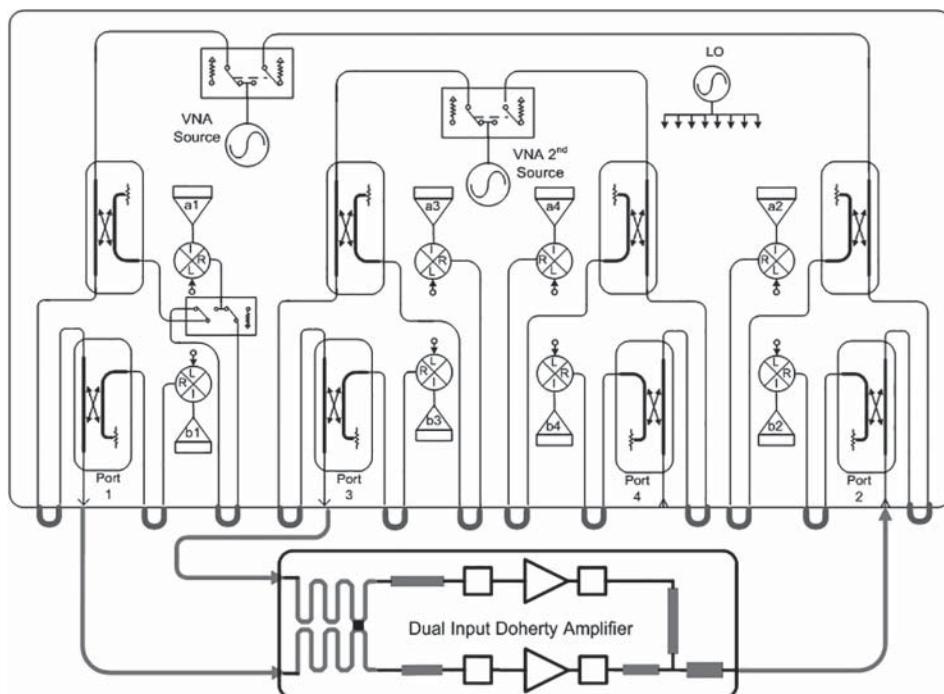


Figure 6.47 Driving an amplifier from a dual source VNA.

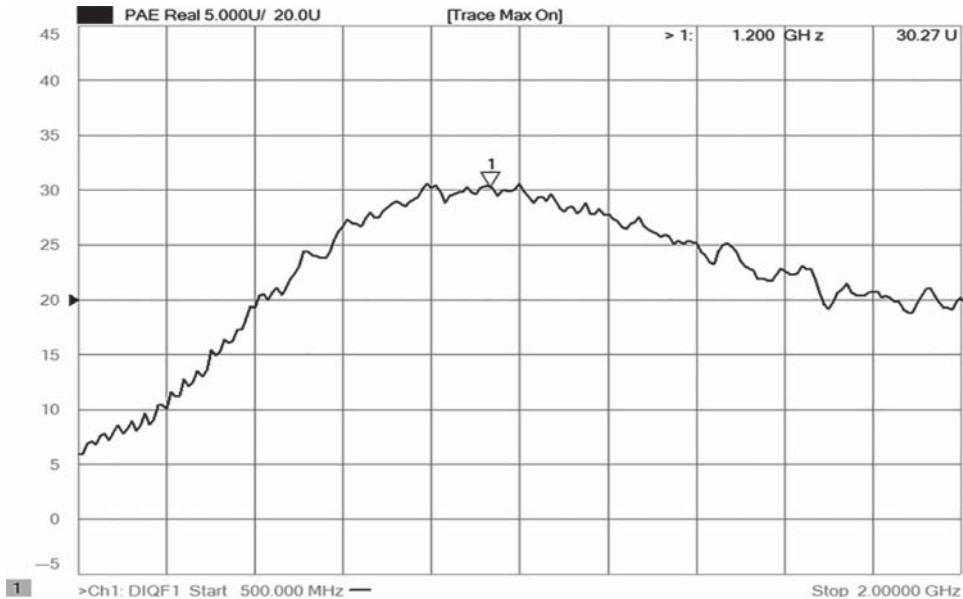


Figure 6.48 Frequency sweep of Doherty amplifier.

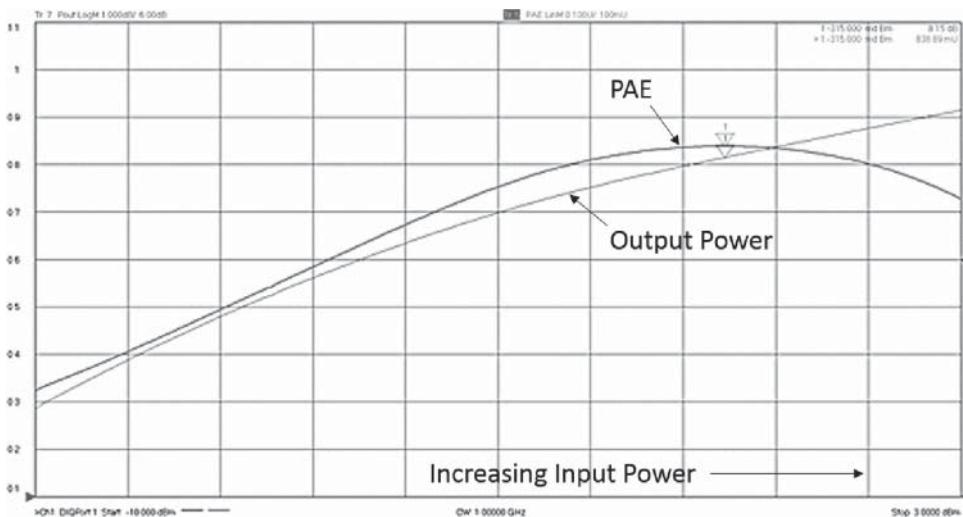


Figure 6.49 Power sweep of a Doherty amplifier.

constant, and sweeping their input power, while monitoring the PAE and output power, as shown in Figure 6.49. This amplifier has a large increase in DC power consumption as the RF drive is increased due to the peaking amplifier turning on. The output power PAE will further increase once the peaking amplifier saturates its DC current.

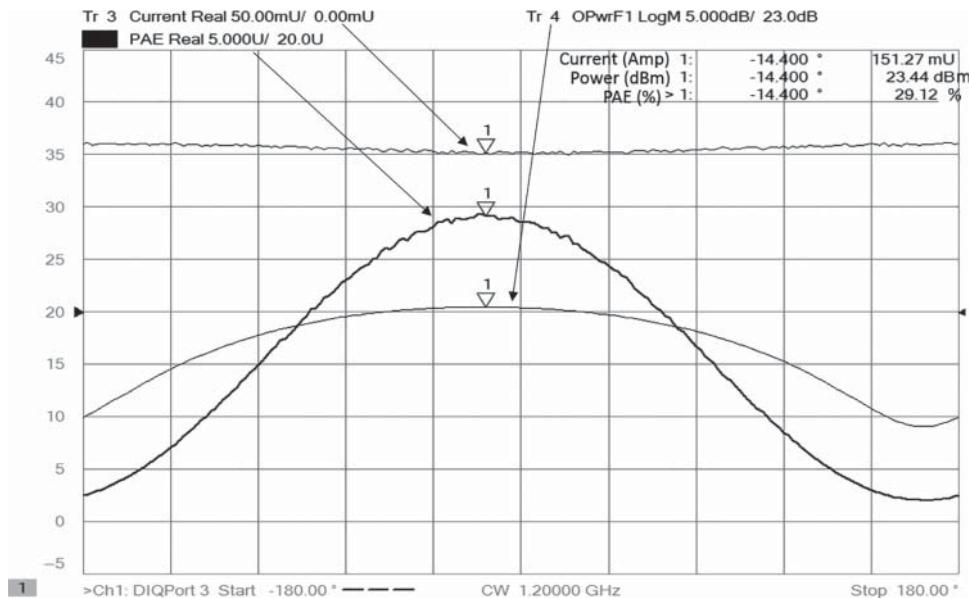


Figure 6.50 Power and PAE versus phase of the input.

From the power sweep, one can pick a point of optimum PAE. We will use the point to investigate the effects of phasing on the input of the amplifier. Figure 6.50 shows a plot of the output power, and PAE of an amplifier with the peaking source swept from -180° to $+180^\circ$, with respect to the carrier amplifier, at the input power found to be optimum for PAE in the previous slide. Here we can see that 0° split does not provide the best PAE, which indicates some unsuspected phase shift in the matching circuits or even in the amplifier modules themselves. By sweeping the phase, we can see what value will provide the optimum PAE. In a similar way, once the optimum phase is found, one may also sweep the input power to either the carrier or peaking amplifier to see if an equal split at the input is the optimum choice. Also shown is the output power and the total current. Interestingly, with phase shown at the marker, the current is slightly reduced, likely because the apparent load change of the peaking amplifier makes the load appear more optimum for the carrier amplifier.

While this discussion focused on a dual-input Doherty amplifier, with a single peaking section, it is possible to repeat the process with multiple input amplifiers as well, by using additional external sources to provide drive signals.

6.8 X-Parameters, Load-Pull Measurements, Active Loads, and Hot S-Parameters

For amplifiers operating in the linear mode, the effect of the load impedance on the output power is completely described by the S-parameter definition from Eq. (1.17), in particular

$$b_2 = S_{21}a_1 + S_{22}a_2 \quad (6.12)$$

In the case where the amplifier is terminated by some load impedance, Γ_L , this can be rewritten as

$$b_2 = \frac{S_{21} \cdot a_1}{1 - S_{22}\Gamma_L} \quad (6.13)$$

The power delivered to the load is

$$\begin{aligned} P_{del} &= |b_2|^2 - |a_1|^2 \\ &= |a_1|^2 \cdot |S_{21}|^2 \frac{(1 - |\Gamma_L|^2)}{|1 - S_{22}\Gamma_L|^2} \end{aligned} \quad (6.14)$$

Thus, the power delivered depends upon both the load impedance and the S_{22} of the amplifier.

However, at high powers, S_{21} and S_{22} can change with the drive power and with the load impedance so that predicting the output power is not possible using normal techniques. In these cases, the most common method of evaluation is to drive an amplifier to the desired level while modifying the load to determine which load provides the maximum output power and often also the PAE of the amplifier at the load. The load impedance is modified by using an impedance tuner at the output, or more recently by providing an active load. This measurement method is usually called *load-pull*.

6.8.1 Non-linear Responses and X-Parameters

One aspect of the amplifiers' non-linear response is that both the gain and the effective output impedance, and sometimes the input impedance, can change as a function the input drive level, and it may appear to change as a function of load impedance, even at a constant drive power. For most devices, the actual operating point is more complex than can be expressed using standard S-parameters but rather requires a new set of parameters, so called X-parametersTM, to fully define the non-linear behavior of a DUT (Verspecht et al. 1997; Root et al. 2005).

Without going in-depth into the mathematical derivation for X-parameter models, a narrow-band (only considering the fundamental input and not harmonics) definition can be derived that is useful in identifying the reason that the operating point, S_{22} and S_{21} , appear to change with load impedance. The most basic version of X-parameters forms a transfer function of input waves to output waves, but each X-parameter element has a dependency on input drive power. The output power of an amplifier is described as

$$b_2(A_1) = \underbrace{X_{21}^F(|A_1|)P}_{\text{Gain}} + \underbrace{X_{22}^S \cdot a_2 + X_{22}^T P^2 \cdot a_2^*}_{\text{Match}} \quad (6.15)$$

where $P = S_{21}A_1/|A_1| = 1 \cdot e^{\varphi_{A1}}$

The definition of X-parameters includes two terms, a gain term and an output match term. The gain term, X_{21}^F , is essentially the output power of the DUT for any input power, P, that contains the phase of the incident signal. It may seem unusual that X_{21}^F has units of power rather than gain, but this is done for computational efficiency in its use, and for practical purposes it can be considered the same as the gain. The superscript F denotes the fact that this is the response to the input at the fundamental frequency, that is, the response as if there are no

reflections at the input or output at the fundamental and any harmonics, for a particular input power. This could as easily be defined as the gain, but convention has settled on the power as the fundamental response term. Here, A_1 is the large signal input, and the use of a capital letter implies the response of the DUT depends upon the magnitude of this stimulus. However, the second portion uses the lowercase a_2 and implies that the response of the amplifier is substantially linear with respect to power of a_2 .

The second set of terms identified as match consist of X_{22}^S and X_{22}^T (but informally called S_{22} and T_{22}) and form the pair of terms that describe the change in b_2 power as a function of the signal reflected from the load, and together they form the effective output match $S_{22}(|A_1|, a_2)$, which can be called the true “effective Hot- S_{22} ” of the amplifier. Now, the effective value of S_{22} appears to change with the magnitude and phase of the load, but this is only because of the way the S and T portions add and subtract as the magnitude and phase of a_2 vary. The value of X_{22}^S and X_{22}^T are *presumed constant* in the X-parameter model for any particular value of $|A_1|$ and any magnitude or phase of a_2 . This is provided a_2 is small enough to not affect the actual DC operating point of the amplifier, an assumption that is true in many devices. The X_{22}^S changes only slowly, if at all, with the $|A_1|$; at low input powers it devolves to S_{22} , and at high powers it is essentially what is measured as part of a traditional, but incorrect, Hot- S_{22} measurement. But such a measurement ignores the T term, which adds in as a conjugate multiple of the a_2 signal. The T term is created through a high-order non-linear mechanism, similar to an intermodulation product, and its level changes as a second-order power of the input signal. Figure 6.51 shows the relative response of the S_{21} term, the S_{22} term, and the T_{22} term of an amplifier as a response to the input drive level. The S_{21} term shows compression at higher power, the S_{22} term shows some small variation as a function of power, but the T_{22} term shows substantial variation following the 2:1 curve expected of a higher-order product. In fact, at some levels it becomes larger than S_{22} and thus has a larger effect on the output power than does the S_{22} term.

To understand how the conjugate operation of a_2 on the X_{22}^T affects the apparent total reflection from the DUT, consider the vector diagram in Figure 6.52, where a_2 is the result of some non-matched load. The figure on the left shows the total signals coming from the DUT, as a result of the X_{21}^F (large light error), the $X_{22}^S \cdot a_2$ (smaller darker arrow), and the $X_{22}^T \cdot a_2^*$ (also small darker arrow); the vector sum of these terms (constructed with the light dashed

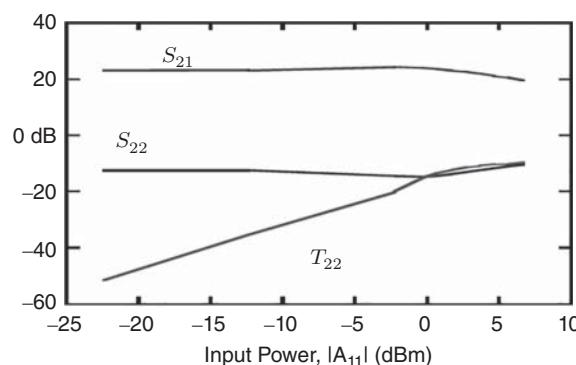


Figure 6.51 S_{21} , S_{22} , and T_{22} terms as a function of input power.

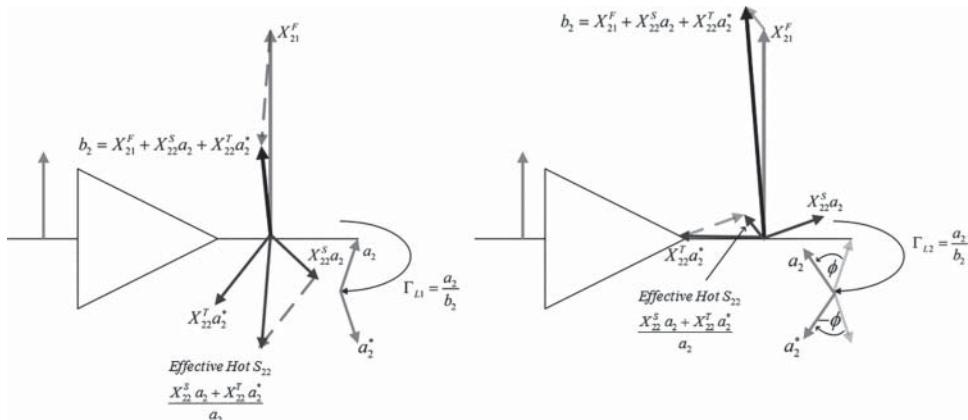


Figure 6.52 The vector effect of a_2 and a_2^* on apparent reflection.

arrow) forms what could be called the true “effective Hot- S_{22} .” The total power, b_2 , is the vector sum of all the X terms, illustrated by the dark black arrow on the figure. The figure on the right has identical X-parameters, but now the phase of the load reflection, a_2 , is changed, as might happen through rotating the phase by adding a length of transmission line.

For this second figure, the relative value of $X_{22}^S \cdot a_2$ versus the fundamental output, X_{21}^F , shifts as a_2 shifts in phase. This is expected and occurs just the same as in a linear device. But the $X_{22}^T \cdot a_2^*$ shifts in opposite phase as a_2 shifts, causing the relative amplitude of the effective DUT reflection to change as the phase of the load changes. In this way, the effective Hot- S_{22} (S_{22} under the condition of forward drive) does appear to change with the load, and one might infer that as such, the operating point of the amplifier changes with the load. But in fact, the S and T terms are constant; only their vector-sum changes due to the conjugate phase behavior of the T-term and S-term interaction, which causes a change in the apparent response of the effective Hot- S_{22} . This also causes the b_2 value to change in a manner that is not consistent with a fixed value of S_{22} .

For amplifiers that are pre-matched with impedances relatively close to 50Ω , used in narrowband applications such as cellular phone handsets, the fundamental response may be sufficient to provide good agreement between modeled behavior and actual response, and predictions made for maximum power and efficiency based on X-parameter simulations may yield matching networks that give similar results between simulation optimization and real application. However, for “bare” transistors, those that require substantial matching to transform them into a 50Ω environment, the fundamental X-parameters measured at 50Ω may not sufficiently characterize the DUT. Instead, load dependent X-parameters may be required, where the X-parameters become a function of both the amplitude of the input signal, $|A_1|$, and the load reflection coefficient, Γ_L . In such a case the X-parameters provide a kind of perturbational or differential characterization around each load, with the S and T terms describing the behavior of the DUT at each load point characterized.

In addition to the load dependency, many amplifiers also have a dependency on the reflection of the harmonics at the input and the output of the DUT and a dependency on the DC operating point. Each harmonic term has its own S and T terms that characterize how any

particular harmonic (say the l^{th}) generates an output at any other harmonic or fundamental (say the k^{th}), and from any port (perhaps the j^{th}) to any other port (as in the i^{th}). Thus, in general, X-parameters have four indexes, two of which describe which port and of which harmonic the input signal is incident to the DUT, as in a_{ik} , and two of which related to which port and which harmonic or fundamental is scattered from the DUT, as in b_{jl} , so that, in general, the X parameters are represented as

$$\begin{aligned} b_{ik} = & X_{ik}^F(DC, |A_{11}|, \Gamma_2)P^k + \sum_{j,l} X_{ik,jl}^S(DC, |A_{11}|, \Gamma_2)P^{k-l} \cdot a_{jl} \\ & + \sum_{j,l} X_{ik,jl}^T(DC, |A_{11}|, \Gamma_2)P^{k+l} \cdot a_{jl}^* \end{aligned} \quad (6.16)$$

This generalized equation describes a multidimensional dependency of frequency and amplitude of the input signal, DC operating point, and output impedance point. The method for determining the X-parameters is beyond the scope of this book, but in general an X-parameter extraction is performed for each of the independent variables to determine the multi-dimensional response. During simulation, the X-parameter result of any arbitrary input encompassed by the data is determined by a multidimensional interpolation of the X-parameter data.

6.8.2 Load-Pull, Source-Pull, and Load Contours

Load-pull measurements create a map of the output power of an amplifier as a function of the output load. If an amplifier were responding in a linear way, this map would form circles centered on the maximum output power, which would occur at a load impedance equal to the conjugate of S_{22} of the amplifier.

In practice, an amplifier operating in a non-linear range does not produce maximum power at the same impedance as predicted by the linear measurement of S_{22} , but rather the effective output impedance moves as the power is increased and, in the past, the impedance for maximum power could only be assessed using a load-pull system. From the previous section, it is now understood that the apparent change in output impedance with increasing drive power is a consequence of the match portion of the X-parameters. As such, the need for direct load-pull measurements may be reduced in the future as the understanding and quality of X-parameter characterizations improve. However, in many cases, particularly for very high power or under the conditions of complex modulation, load-pull characterization remains a good option for determining the effects of output impedance on the performance of an amplifier. Load-pull systems are classified as mechanical, active, or hybrid, depending upon how the load impedance state is created at the output of the DUT.

6.8.2.1 Mechanical Load-Pull Systems

Mechanical systems, or true load-pull, use an impedance tuner at the output of the DUT as an impedance transformer, transforming the VNA port 2 impedance into some other impedance on the Smith chart. Some systems use multiple tuners or multisection tuners to supply a desired match for the fundamental and harmonics. Harmonic load-pull, used to determine

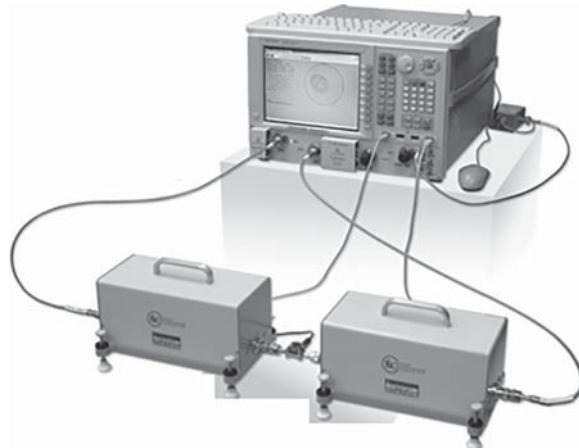


Figure 6.53 Illustration of a source and load-pull system.

the impedance match for harmonics to optimize amplifier performance, is a relatively new and active area of research. Recently some studies have shown that the impedance of the source at harmonic frequencies can also have a significant effect on amplifier performance, so source-pulling has been more common as well. An example of a source-pull and load-pull system is illustrated in Figure 6.53.

Source-pulling is often used for noise parameter characterization. It is also sometimes used to investigate the stability of amplifiers, especially in an out-of-band region. For these investigations, the source tuner is set to a high reflection value, and the output of an amplifier is monitored on a spectrum analyzer to look for oscillations or even a rise in the noise floor, which indicates a tendency to oscillate.

For investigation of maximum gain or maximum power transfer at the fundamental, it is not at all necessary to do actual source-pulling as the effect of a source-pull is simply to change the effective voltage applied to the input of the DUT. This could just as easily be performed by computing the output power from a $50\ \Omega$ source that would result in the same voltage and adjusting that voltage. This is identical to doing an active source-pull on the source and would yield the same results as a tuner for CW measurements. However, the same cannot be said for harmonic source-pull, which must be relative to the second harmonic generated in the DUT, so the VNA must provide an actual effective reflection coefficient at that frequency, although the reflection can be created by another active source.

6.8.2.2 Active Load-Pull

Active load-pull or active tuning is an alternative to using mechanical tuners and relies on utilizing a second source driving the output of the amplifier, the amplitude and phase of which is controlled to create an reflected a_2 trace that provides the desired load reflection coefficient, as illustrated in Figure 6.54.

In active load-pull, all four waves from the DUT must be monitored. The process starts with applying an input signal to the DUT at the desired power level, and monitoring the

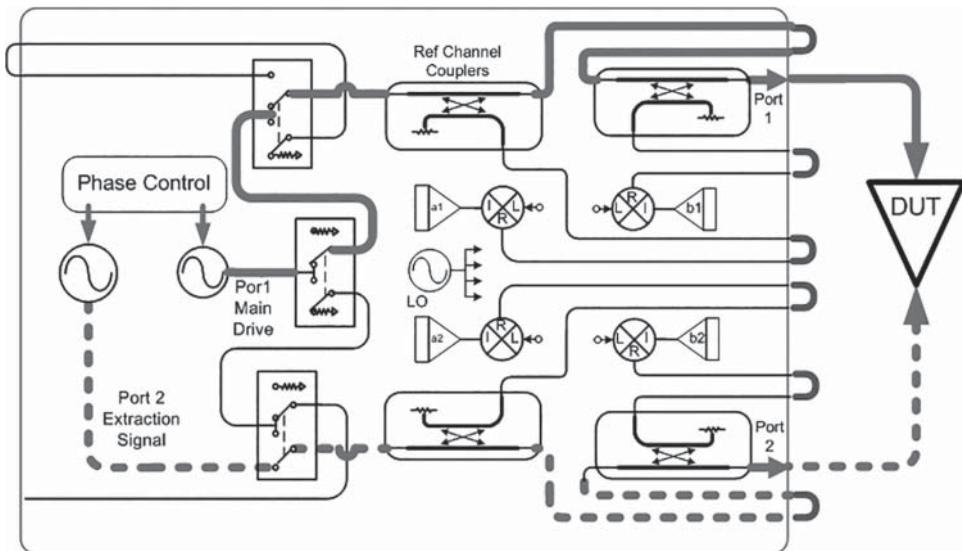


Figure 6.54 Block diagram of a VNA with active load-pulling.

output waves b_2 as well as the raw reflection for the VNA load, a_2 . The effective reflection coefficient is computed from these waves, and a second source at the output is activated to provide an additional signal to add or subtract to the raw reflection to create the desired reflection coefficient. The S-parameters are computed from the two stimulus signals applied using the formulation from Eq. (1.21). From the first measurement, the value of S_{22} is not known, so the effect of applying the second source at port 2 on the b_2 power is not known. To solve for all four S-parameters, a change in the output signal must occur, so it is common to change the phase of a_2 to provide a second set of equations to solve. After solving for the S-parameters, a good estimate of S_{22} is established, and a new estimate can be made for the second source amplitude and phase. This process is repeated until the correct value of effective reflection coefficient is achieved within predefined limits.

Note that the S-parameters computed in this manner are truly “hot S-parameters” provided that the change in a_2 at the output is small enough to not affect the operating point of the amplifier. The S_{22} value gathered is the total effective reflection coefficient from the amplifier’s output and represents the sum of the X_{22}^S and X_{22}^T for that particular value of load, again provided that the change in the a_2 wave is near that load value. In fact, this is a similar process to that used to determine the X-parameters of a DUT, which differ only in that multiple values of the a_2 wave are applied to allow solving the X-parameter defining Eq. (6.15).

One simple but informative example of an active load-pull response is looking at the effective reflection coefficient, or effective Hot- S_{22} , while sweeping the load impedance around the Smith chart. An example is shown in Figure 6.55 where the one trace (marked Γ_{Load}) shows the value of the active load. Also indicated on the chart is the value of a lower power measurement of S_{22} of the amplifier (the trace being nearly constant), and a locus of points illustrating the value of the effective Hot- S_{22} as the phase of the load is rotated around the Smith chart. In this case, the magnitude of the load was set to match the magnitude of the

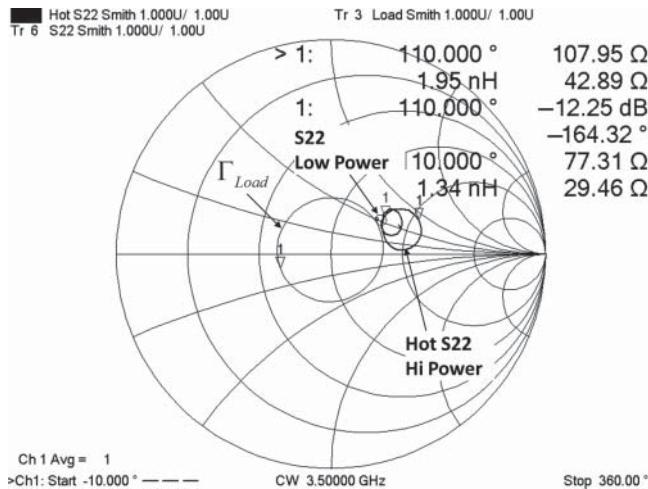


Figure 6.55 Active load-pull showing output power and effective Hot- S_{22} .

linear S_{22} , or about -12 dBc ($\Gamma = 0.25$). It is interesting to see the trajectory of the effective Hot- S_{22} as the load circulates. Because there are two vectors moving for effective Hot- S_{22} (S and T terms or the a_2 and a_2^* signals), there will be two loops in the effective Hot- S_{22} for one loop in the load response. This shows the complex nature of the true effective Hot- S_{22} , but one must realize that the underlying S and T terms in fact do not change with load reflection.

Active load-pull has some advantages and disadvantages relative to the passive or mechanical system. The key disadvantage is that the output power supplied from port 2 must be the same as the power from the DUT if the active load is to reach the edge of the Smith chart, or full reflection. The key advantage is that it is often much faster, perhaps up to 100 times faster, than a mechanically tuned system. An additional advantage is that the active load will not induce any oscillations in the device as oscillations occur when some load state applies an impedance that lies in an unstable region of the amplifier. The oscillations start because thermal noise is amplified by the near infinite gain, producing the oscillation signal. But in active load, there is no load reflection (other than the raw VNA reflection) applied to the amplifier except at the exact frequency of the CW source, and the noise enhancement due to unstable gain is limited by the maximum power of the port 2 source used to create the active load. Once that maximum is reached, the effective active load is limited, and any oscillation is suppressed. That does not mean one cannot detect these regions of instability; in fact, these will be areas on the Smith chart where active load-pull will fail to achieve the desired value of reflection coefficient. This gives the user a strong hint that the DUT will be unstable in that region of the Smith chart.

6.8.2.3 Hybrid Load-Pull Systems

In some systems, a mechanical tuner for the fundamental is combined with active tuning for the harmonics, providing fine/fast tuning of the load. For very high-power amplifiers, it may not be practical or possible to create an active signal that can drive the DUT to the

desired state. In these cases, a mechanical tuner is used as a prematching network to create an approximate value of the desired load impedance. In fact, when X-parameters are measured at impedances other than $50\ \Omega$, this is the method used to create the load impedance states. The second source is used to do a small load-pull around the load impedance, by which the X-parameters are extracted.

In other applications, the mechanical tuner is used to create a load match at the fundamental, and the active load is used to provide the harmonic load impedance states. In such a system, the mechanical tuner can provide a high-power reflection at the output, and the power level for the active load, which is relative to the power at the harmonics, can be much lower, making such a system quite practical. One attribute of many mechanical tuners is that while they have a high reflection at the fundamental frequency, they provide essentially a low loss transmission line path at other frequencies, allowing the active source to be driven through the tuner at a different frequency, such as at the harmonics.

6.8.2.4 Power Contours

Load-pull measurements in the most classic form are typically analyzed to show the impedance point of maximum power along with the locus of points representing lines or contours of constant power on the Smith chart. Power data from dozens or hundreds of impedance points are gathered to generate these contours, “connecting the dots” between impedance points that share a common power. Similar power contours can be created in simulation from an X-parameter model, with remarkably good correlation to actual measured results. As an example, the schematic shown in Figure 6.56 demonstrates the circuit needed to simulate a load-pull experiment utilizing an X-parameter data set of an measured amplifier.

A comparison between simulated load-pull based on an X-parameter model of an amplifier and the actual load-pull results is shown in Figure 6.57. From this we can conclude, that in some way, the X-parameters contain the information of the dependency on load reflection for determining maximum power transfer. Section 6.8.3 will explore this dependency in detail and introduce a new result based on recent work.

6.8.2.5 Efficiency Contours

The load impedance applied to the DUT can also affect the PAE of an amplifier, especially one in a non-linear range of operation. If the power, voltage, and current of an amplifier are measured, and PAE computed as a function of the load impedance, then load contours of PAE can be created as well as max power. While they tend to be similar (maximum efficiency will occur near maximum power), there can be some differences, and observing the power and PAE contours is a quick way of seeing the optimum load for a particular amplifier. These contours can also be created versus second or third harmonic load, or source load at second or third harmonic. An example of PAE contours is shown in Figure 6.58.

6.8.3 Hot S-Parameters and True Hot-S₂₂

In Section 6.8.1, the term “effective Hot-S₂₂” was introduced, with the understanding that under non-linear conditions there is not a single value for the effective reflection of

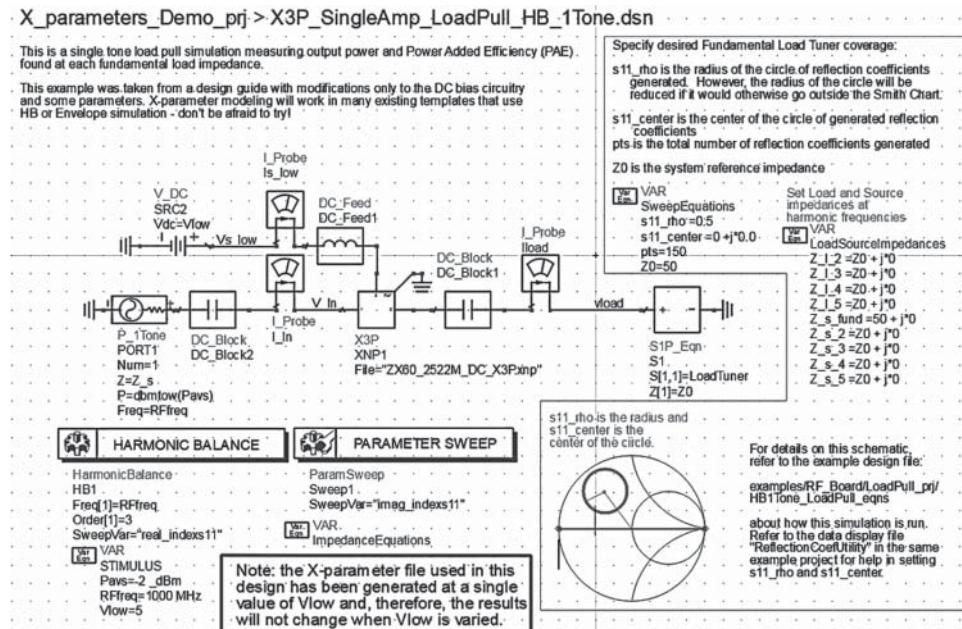


Figure 6.56 Schematic for an X-parameter based Load-pull simulation.

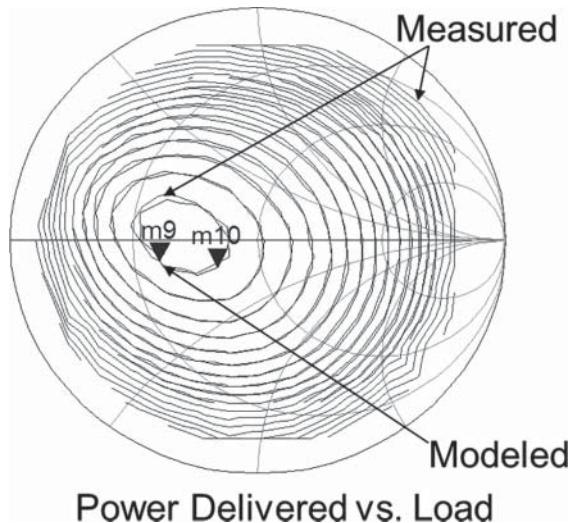


Figure 6.57 Comparing X-parameter simulated load-pull with real load-pull values.

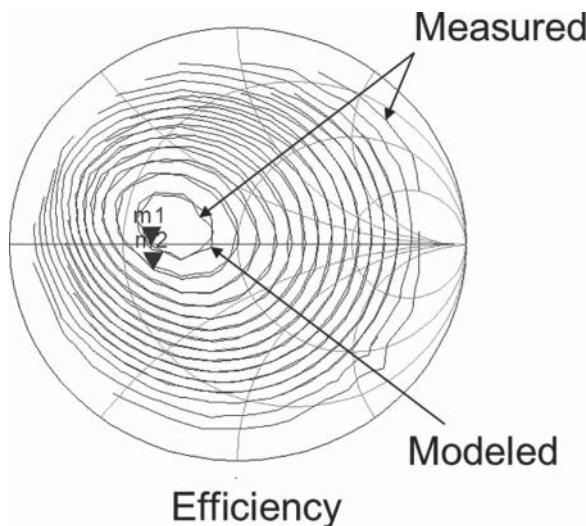


Figure 6.58 PAE contours versus load impedance.

the amplifier under test, but rather the effective reflection depends upon the phase of the load-match facing the amplifier, and a change in the phase of the load reflection will change the effective Hot- S_{22} or effective reflection of the amplifier.

In some papers, a singular value for Hot- S_{22} was attempted to be discerned using a method of driving the amplifier into a non-linear condition utilizing one source at the desired frequency and then injecting another source into the output of the amplifier, at a slightly offset frequency, and measuring the S_{22} of this second signal. I call this the “traditional Hot- S_{22} ,” and while it provides essentially the correct result for a linear amplifier, as soon as the amplifier is driven into a non-linear state, this traditional Hot- S_{22} is almost always completely wrong. To understand the reason for this, we can look at the actual signals coming from the amplifier under these drive conditions. Consider the block diagram in Figure 6.54, but in the case where the second source is slightly offset in frequency, this is the measurement block diagram for a traditional Hot- S_{22} acquisition.

6.8.3.1 Evaluating the Spectrum of Traditional Hot- S_{22}

Some VNAs, such as the PNA-X, can be used as spectrum analyzers, and we utilize the directional couplers at port 2 to separate the reverse injected signal (essentially the active load) from the signal that is reflected from the amplifier, following the active load configuration of Figure 6.54. For a first test, only the source at the input is turned on, and we can see the spectrum of the output signal from the amplifier in Figure 6.59. This shows the output power in the Marker 1 readout, for the case where the amplifier is driven with about -30 dBm input signal, thus showing a gain of about 30.5 dB.

New, let us turn on the port 2 source, set to 0 dBm power, slightly offset in frequency, but also disconnect the amplifier from the output so port 2 is open. Since the test port is open, there is a total reflection and the reflected signal shows a 0 dBm value, as shown in Figure 6.60.

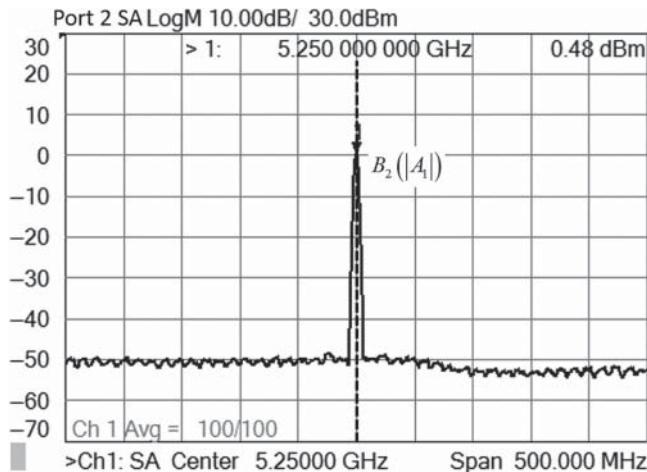


Figure 6.59 Output (b_2) spectrum of an amplifier with a single input signal.

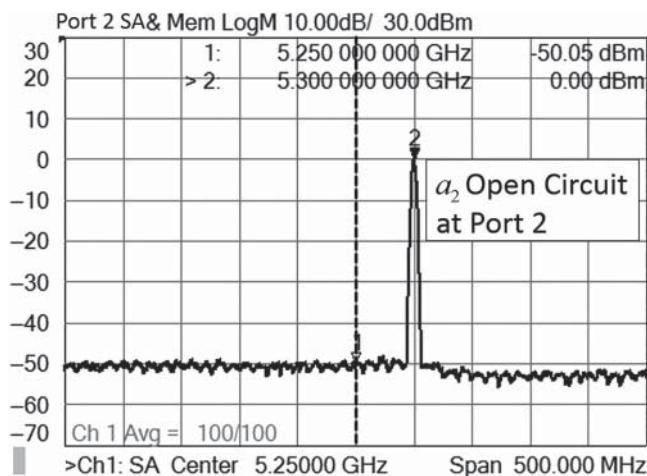


Figure 6.60 Traditional Hot- S_{22} measure of a total reflection from an open port 2, slightly offset in frequency.

If we were to put a load on port 2, this signal would drop below -40 dBm, indicating a good return loss. Here we rely on the raw directivity of the port 2 test port coupler to separate the incident wave from the reflected wave. In the figure, they are the same since the reflection applied is an open (total reflection).

For the next step, we will reconnect port 2 of the VNA to the output port of the amplifier, shown in Figure 6.61. This represents the traditional Hot- S_{22} measurement, and we can see by the change in the b_2 signal (as represented by Marker 2) that we are measuring the S_{22} of the amplifier, under the condition of forward drive, but here the input signal is low enough so that the amplifier is still in a linear condition. If we use the standard VNA application to measure

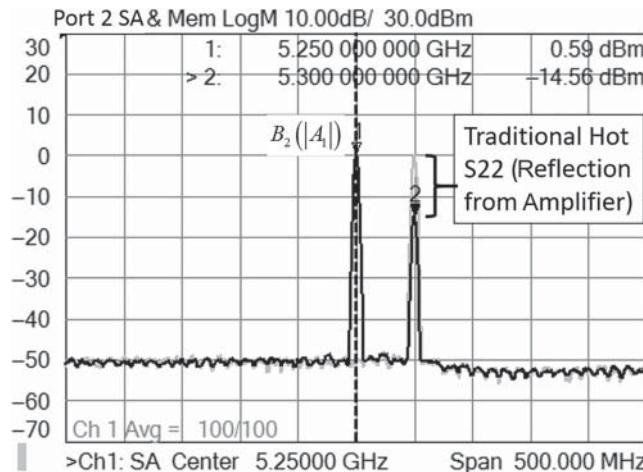


Figure 6.61 Traditional Hot- S_{22} , showing the output reflection of the amplifier in the signal that is offset from the main signal.

the S_{22} , it measures the same value. This is the traditional Hot- S_{22} measurement and works for linear amplifiers. But it provides an incomplete and, in most cases, totally wrong value for the effective output impedance of an amplifier when the amplifier is driven into a non-linear state.

The value shown in the figure is, in an actual sense, the average impedance of the output impedance over the period of the measurement. Since the offset frequency is using one sine-wave signal to measure the behavior of the amplifier, which is driven at a different sine-wave frequency, and sine waves of different frequencies are orthogonal functions, we can say that the measured traditional Hot- S_{22} is the average impedance at the offset frequency. But, and this is a critical point, it does not tell us anything about the behavior of the amplifier at the actual input driving-frequency, or to be more precise, it does not tell us what the effect of load mismatch causing a reflected wave at the driving-frequency will have on the power delivered to the load. Because reflections at the actual driving frequency are not orthogonal functions, the average value will not be valid.

We can label these signals according to Eq. (6.15), the fundamental X-parameters, where the large signal at the fundamental is the $X^f(2,1)$ term, and the reflected signal, offset in frequency, is associated with the $X^s(2,2)$ term. Since the amplifier is operating linearly, there is no $X^r(2,2)$.

Finally, let us raise the input power so that the amplifier is near compression and thus operating in a non-linear way. Figure 6.62 shows the output spectral response when the amplifier input is driven with a sufficiently large signal to cause the amplifier to operate in non-linear way. We now see a kind of companion signal arise out of the noise floor on opposite (left) side of the port 2 drive frequency.

The amplitude of this signal rises in a non-linear, nearly third-order response to changes in the port 1 input drive-signal amplitude but is linear with respect to changes in the port 2 reflected signal power. This signal, which we can think of as *Transposed* from the right-side port 2 drive signal, is associated directly with the X^{r22} term in Eq. (6.15). (A side historical

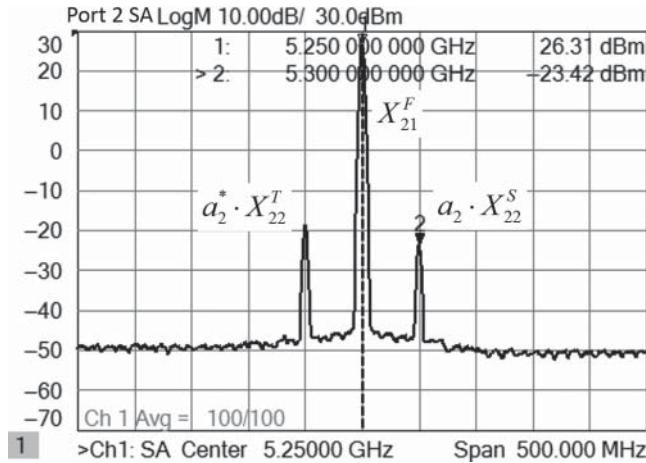


Figure 6.62 Spectrum of traditional Hot- S_{22} with a non-linear amplifier.

note: while the term “Transposed” does describe the signal, the originator of the X-parameters has stated that T is simply the letter to follow S, but still this author likes the connotation of a transposed signal.)

Note in this case, the amplifier is driven with a port 1 input drive of 0 dBm and resulting in an output power of 26.53 dBm; thus, it is nearly 4 dB in compression compared with the previous linear measurements.

The genesis of this $X'(2,2)$ signal is a mixing of the non-linear fundamental-output large-signal with the offset frequency port 2 signal to create a kind-of intermodulation product at exactly the negative of the offset frequency. The mixing of these signals is realized through the time-variant non-linear output impedance of the amplifier operating at the fundamental drive signal and commutating the port 2 injected signal at that offset frequency rate. It is analogous to a mixer where the fundamental frequency is acting like the local oscillator, causing a time varying impedance in the mixer, and operating on the RF input signal.

In the case shown in Figure 6.62, we can see the negative offset frequency of the signal associated with the $X'(2,2)$ term matches the positive offset frequency of the signal associated with the $X^s(2,2)$ term. Clearly, if we decrease the offset frequency of the port 2 signal associated with $X^s(2,2)$, it moves closer to the $X'(2,1)$ signal. The $X'(2,2)$ signal offset will also decrease, and it will become closer to the $X'(2,1)$ signal. In a similar manner, if we change the phase of the port 2 drive signal, the phase of the effective reflection associated with the $X^s(2,2)$ will change in the same direction, but the phase of the $X'(2,2)$ will change in the *opposite* direction. Changing phase of the a_2 reflected signal is really the same as changing its frequency very slowly. This is the genesis of the conjugate term in Eq. (6.15). If the offset frequency is set to zero (which is the case for a true-reflection at the $X'(2,1)$ frequency from a load), there still exists two terms associated with the non-linear reflection, and they can have independent amplitudes and phases, and their relationship depends on the phase of the signal applied to port 2. So, as shown in Figure 6.52, this is why the effective output impedance of an amplifier changes with phase of signal reflected from the load and why the traditional Hot- S_{22} does not represent the effective reflection of a non-linear amplifier.

6.8.3.2 Computing True Hot- S_{22}

Recently (Root et al. 2017) work in the area of non-linear amplifier behavior had a breakthrough in the relationship between load-pull response, X-parameters and the concept of a Hot- S_{22} value. As discussed earlier and in Section 6.8.2.2, the effective reflection changes with the reflected a_2 wave in manner described by Eq. (6.15). One can also directly write the power delivered to a load as

$$P_{del} = |B_2|^2 - |a_2|^2$$

with

$$B_2(A_1, a_2) = \left\{ X_{21}^{(F)}(|A_1|) e^{j(\varphi_{A1})} + \left[X_{22}^{(S)}(|A_1|) + X_{22}^{(T)}(|A_1|) e^{j(\varphi_{A1}-2\varphi_{a2})} \right] a_2 \right\} \Big|_{A_1=\text{constant}} \quad (6.17)$$

where $\varphi_{A1} - 2\varphi_{a2}$ represents the conjugate of the a_2 term

From this, one might take a partial derivative of the power delivered to a load, with respect to the a_2 wave reflected from a load, to find the optimum reflection that maximizes the power delivered to a load as

$$\begin{aligned} \frac{\partial}{\partial a_2} (P_{del}) &= 0, \\ \frac{\partial}{\partial a_2} \left(\left| X_2^{(F)} e^{j\varphi_{A1}} + X_{22}^{(S)} \cdot a_2 + X_{22}^{(T)} \cdot a_2 \cdot e^{j(\varphi_{A1}-2\varphi_{a2})} \right|^2 - |a_2|^2 \right) &= 0 \end{aligned} \quad (6.18)$$

The solution to Eq. (6.18) was found (Root et al. 2017) to be

$$a_2^{Opt} = \frac{X^F X^{S*} (1 - |X^S|^2 + |X^T|^2) + X^{F*} X^T (1 + |X^S|^2 - |X^T|^2)}{(1 - |X^S|^2 - |X^T|^2) - 4|X^S|^2 |X^T|^2} \quad (6.19)$$

where the subscripts for ports are not shown for the X^F , X^S , and X^T terms, for clarity's sake.

From this result, one can find the B_2 wave for the maximum power as

$$B_2^{Opt} = X_2^{(F)} e^{j\varphi_{A1}} + X_{22}^{(S)} \cdot (a_2^{Opt}) + X_{22}^{(T)} \cdot (a_2^{Opt}) \cdot e^{j(\varphi_{A1}-2\varphi_{a2})} \quad (6.20)$$

and find the reflection that produces the maximum power delivered to the load as

$$\Gamma_{Opt} = \frac{a_2^{Opt}}{B_2} = \frac{X^{S*} (1 - |X^S|^2 + |X^T|^2) + e^{-j2\varphi(X^F)} X^T (1 + |X^S|^2 - |X^T|^2)}{(1 - |X^S|^2 - |X^T|^2) + 2e^{-2j\varphi(X^F)} X^S X^T} \quad (6.21)$$

These remarkable results indicate that the load for maximum power transfer depends only on the X^S and X^T parameters of the amplifier, and if one evaluates Eq. (6.21) under the condition of linear operation, where X^T approaches zero as

$$\begin{aligned} \Gamma_{Opt} \Big|_{X^T \rightarrow 0} &= \frac{X^{S*} (1 - |X^S|^2 + |X^T|^2) + e^{-j2\varphi(X^F)} X^T (1 + |X^S|^2 - |X^T|^2)}{(1 - |X^S|^2 - |X^T|^2) + 2e^{-2j\varphi(X^F)} X^S X^T} \\ &= \frac{X^{S*} (1 - |X^S|^2)}{(1 - |X^S|^2)} = X^{S*} \end{aligned} \quad (6.22)$$

we find that the optimum load for maximum power becomes the conjugate of $X^s(2,2)$. And we know in the linear case $X^s(2,2)$ approaches S_{22} as $X^t(2,2)$ approaches zero. This result is, of course, to be expected as it is well established that for a linear device, maximum power occurs when the load has the conjugate of the output impedance of the device.

From the result of Eq. (6.21) we can observe two key points.

- 1) The X-parameters contain the information on the value of load impedance that results in the maximum power transfer. This is already evident in Figure 6.57 where simulated and measured match-dependent power contours match very well, when an X-parameter model is used to simulate the load-dependent power effects. With Eq. (6.21), the relationship between load reflection and maximum power delivered is explicit in terms of X^s and X^t .
- 2) The result of Eq. (6.21) demonstrates a unique value of load for maximum power transfer. We can use the analogy to a linear result to now define a new term: true Hot- S_{22} , or simply Hot- S_{22} , as being that single impedance, the conjugate of which results in maximum power transfer. In this way, the concept of Hot- S_{22} retains the fundamental meaning of S_{22} as it applies to linear devices. We can further expand this concept to all S-parameters to create what is now called Hot S-parameters.

6.8.3.3 Computing Hot S-Parameters

From this understanding of X-parameter measurements and result of Eq. (6.21), we can derive the definition of Hot S-parameters. As described earlier, the definition for Hot- S_{22} is

$$\text{Hot_S22} = \Gamma_{Opt}^* \quad (6.23)$$

In an analogous manner, note that S_{21} in a linear sense is the gain of an amplifier driven from a zero reflection source into a zero reflection load. We can define Hot- S_{21} similarly and recognize that the $X^f(2,1)$ term represents the power delivered to a zero-reflection load (meaning $a_2 = 0$) so

$$\text{Hot_S21} = \frac{X_{21}^F(A_1)}{|A_1|} \quad (6.24)$$

Note that Hot- S_{21} does depend on the magnitude of the input signal, unlike linear S_{21} ; since it is defined as a function of the large-signal A_1 wave, the value of source impedance will not affect the definition.

Similarly, Hot- S_{11} can be defined as

$$\text{Hot_S11} = \frac{X_{11}^F(A_1)}{|A_1|} \quad (6.25)$$

Finally, Hot- S_{12} is defined as

$$\text{Hot_S12} = X_{12}^S(A_1) \quad (6.26)$$

While there does exist the possibility of a $X^t(1,2)$ term, in practice we find that it is negligible. To determine the actual value of $X(1,2)$ one would need to characterize the reverse power at port 1 while varying the phase at port 2.

6.8.3.4 Examples of Hot S-Parameters

It is instructive to investigate Hot S-parameters for some example amplifiers. For a first experiment, an amplifier is measured across a swept frequency range, at both low power and a power sufficient to drive it into non-linear operation. The results are shown in Figure 6.63. The plot shows four measurement results, two at low power, where the amplifier operates in a linear manner, and two at a higher power. The normal S-parameters are characterized under full 2-port error correction. The Hot S-parameters are derived from fundamental X-parameter data, as described earlier.

In the low-power, linear-response case, the gain is the same (within 0.02 dB) between the normal S-parameters and Hot S-parameters, as expected. When the input power is increased to the high-power operating point, near the center band, the normal S-parameters show substantial ripple, but the Hot S-parameter shows a smooth trace that we can associate with an overall non-linear or “Hot- S_{21} .” For this amplifier, the gain increases as drive level is increased until it reaches a saturation point, as illustrated in the power sweep of Figure 6.64. The cause of the ripple in the normal S-parameters can be traced to the error correction algorithm as it corrects for the S_{22} of the amplifier and the load match of the test system. This is one of the chief benefits of utilizing Hot S-parameters: only Hot S-parameters properly correct for mismatch errors associated with a non-ideal load match of a test system. In this case, we can infer that the effective S_{22} of the amplifier is changing with drive level (as is the gain), and the normal, linear, 2-port error correction does not compensate for the changing S_{22} with drive power. Remember, normal S-parameter calibration turns off the input (port 1) signal when the

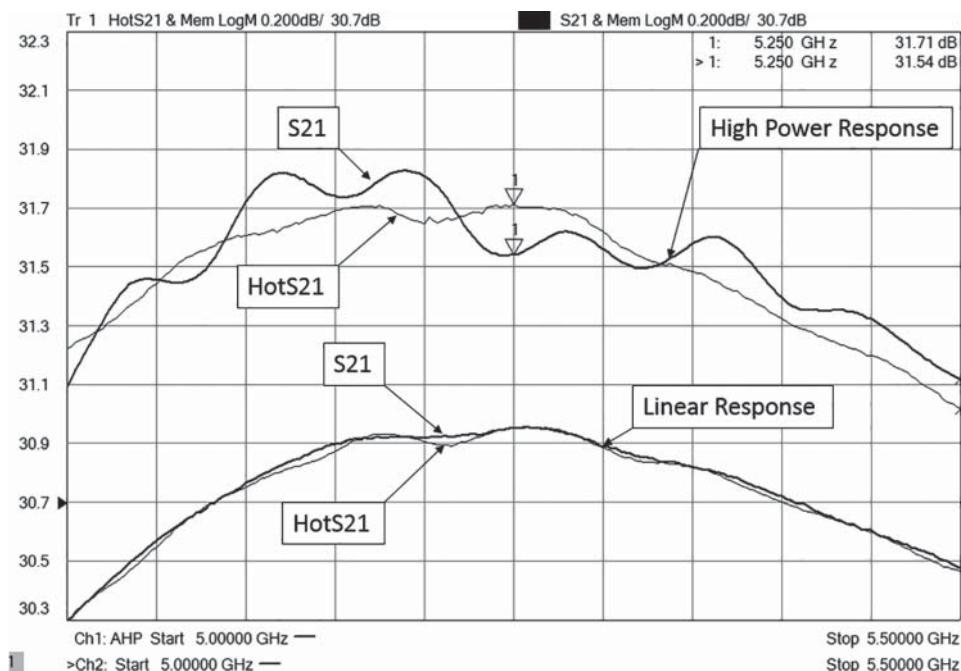


Figure 6.63 Swept frequency response, linear and high power, showing normal and Hot S-parameters.

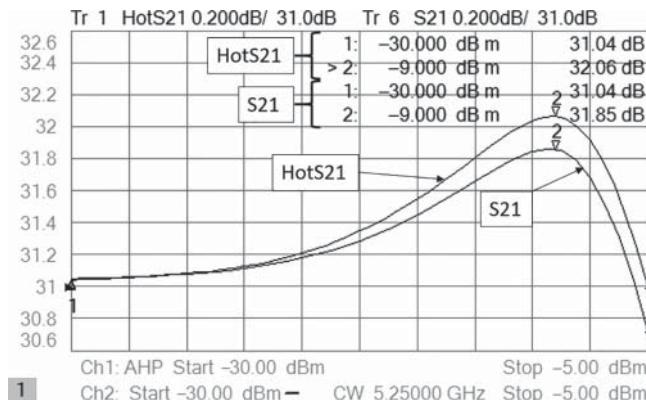


Figure 6.64 Gain versus power for an amplifier, comparing traditional 2-port correction with the Hot- S_{21} result.

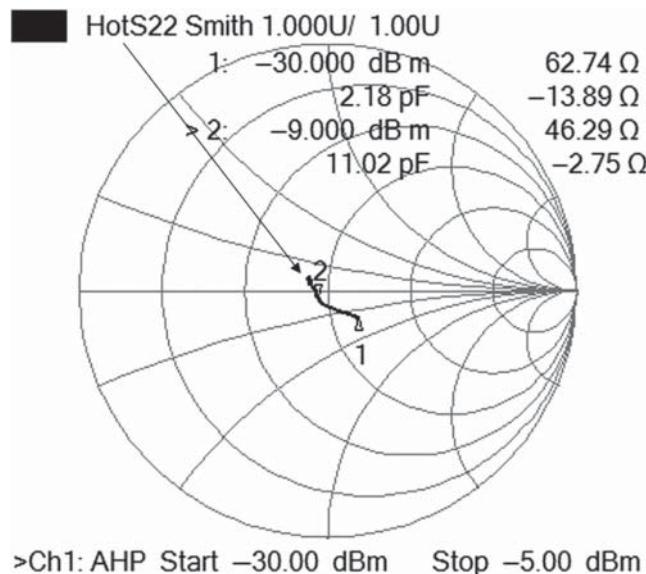
measuring S_{22} , so the amplifier driven with a high power will show a different effective output impedance during the forward (S_{21} and S_{11}) sweep than during the reverse (S_{12} and S_{22}) sweep.

To understand more fully the effects of changes in Hot- S_{22} , consider the results of performing a power sweep on an amplifier, driving it into a non-linear condition while monitoring the gain. Figure 6.64 shows the non-linear response of an amplifier with two different measurement methods. One uses traditional VNA 2-port error correction, where the S_{22} is measured with port 1 turned off. The other shows the Hot- S_{21} and computed in Eq. (6.24). For this particular frequency, the gain shows a greater peaking and a higher 1 dB compression point when measured with Hot S-parameters (utilizing X-parameter representation) compared with the traditional 2-port error correction.

The reason for this improvement becomes clear when one looks at the plot of Hot- S_{22} , as a function of drive power, as shown on the Smith chart in Figure 6.65. Clearly the Hot- S_{22} is becoming closer to 50Ω , and thus the mismatch loss at this frequency is becoming less lossy, and the gain is higher than would be reported by a traditional VNA using 2-port error correction. In this case, the low-power impedance is on the order of 62Ω , but at higher power, the impedance improves to about 46Ω . Another way of saying this is that the error correction algorithm is correcting for the low-power S_{22} , when we see the Hot- S_{22} is completely different.

This is even more clear looking at the magnitude of the Hot- S_{22} , as shown in Figure 6.66. Here note that at the point in Figure 6.64 where the gain difference is large, the Hot- S_{22} differs from S_{22} , measured by normal 2-port methods, by the largest amount.

From Figure 6.67, it is clear that the Hot- S_{22} formulation provides an entirely different result than the “traditional Hot- S_{22} ” configuration (which essentially measured just the $X^s(2,2)$ term), when viewing the X-parameters in the figure. Here we see the $X^s(2,2)$ term is only slowly changing with drive power, and this is what would be measured using the “traditional Hot- S_{22} ” method, which utilizes the offset frequency S_{22} measurement method. In contrast, the $X^t(2,2)$ term shows a dramatic increase in level as the RF drive is increased and, at high power, becomes even larger than the $X^s(2,2)$ term.



>Ch1: AHP Start -30.00 dBm Stop -5.00 dBm

Figure 6.65 S_{22} and Hot- S_{22} for a power sweep.

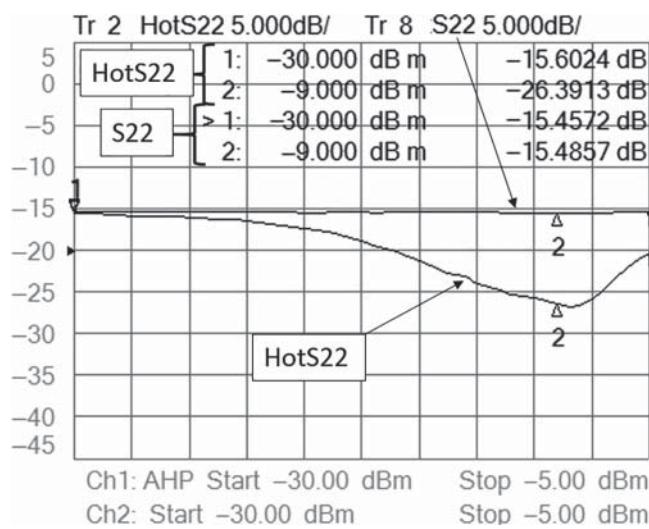


Figure 6.66 LogMag of S_{22} and Hot- S_{22} in a power sweep.

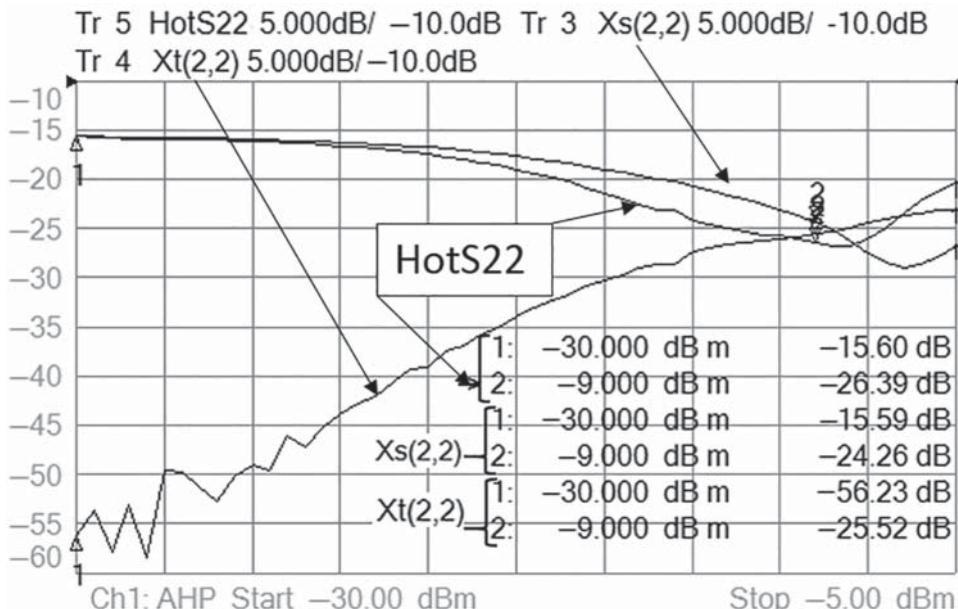


Figure 6.67 $X^s(2,2)$ and $X^t(2,2)$ versus drive power.

It is this large $X^t(2,2)$ term, canceling the effective reflection of the $X^s(2,2)$ term, which makes the Hot- S_{22} dip at high power and gives a different reading for the Hot- S_{21} versus traditional S_{21} .

6.8.3.5 Hot- S_{22} , Hot S-Parameters, and System Performance

Hot S-parameters methods, and particularly Hot- S_{22} measurements, have several applications when considering test methods and system performance.

The principal benefit for making Hot S-parameter measurements is to properly remove the effect that load-match variation in the test system has on the measure gain and measured power of a DUT. In particular, the power under large-drive conditions, sometimes called the *saturation power*, will have a substantial error of up to ± 1 dB or more, if the DUT is measured on a test system with imperfect match. Hot S-parameters properly measure the output power and gain under these large-drive (*hot*) conditions.

As second benefit of Hot S-parameter measurements, particularly Hot- S_{22} , is to determine if the effective output match of a DUT is optimized to drive a Z_0 matched load. The Hot- S_{22} is a measure of the match that produces maximum power and maximum gain, and thus if the Hot- S_{22} is well matched, one can determine that the amplifier is properly designed.

Further, in some cases, the exact match of an amplifier is not as important as the consistency of match of the amplifiers used in the system. If the match of the amplifier is important to system performance at lower power, it is likely more important at higher power. Changes in Hot- S_{22} across a wafer can be indicative of process variation, since changes in

Hot- S_{22} are closely related to limiting behaviors in the amplifier such as current saturation or voltage breakdown. For manufacturers of large-volume products, limiting the variation of the effective output match through production testing using traditional load-pull methods is not practical, and we have seen that “traditional Hot- S_{22} ” does not produce complete picture of the performance of the amplifier. The new Hot- S_{22} methods presented here have the benefit of being extremely fast as well as being mathematically correct and rigorous.

Hot S-parameters have a final benefit of producing a fundamental X-parameter file that can be used in simulation or modeling to understand in a complete way how an amplifier’s non-linear behavior can affect system performance. Unlike previous attempts (e.g. the Power Dependent S-parameter file, P2D, found in ADS), the X-parameter characterization has been shown to properly account for non-linear behavior across a wide range of devices and has become the gold standard for non-linear characterization.

6.8.3.6 Limitations of Hot S-Parameters

There are some limitations when using Hot S-parameters that should be understood. Foremost is that this measurement method is designed for use with amplifiers that are nominally matched, rather than as a method for evaluating an amplifier embodied by a poorly matched, single transistor. Many high-power and high-efficiency amplifiers are designed from transistors with very low impedances (often only an ohm or two) and must be matched to 50Ω through the use of a matching network. Often the initial design phase requires the use of a mechanical impedance tuner to determine the matching required. Additionally, the harmonics of the amplifier must be properly terminated or rather reflected, as their remixing with the main signal can produce significant improvements in the power or efficiency, and for this purpose either multiharmonic tuners or active harmonic tuning is required. All of these methods are supported in advanced software products like Keysight’s Non-Linear-Network-Analyzer (NVNA).

Hot S-parameters are valid under the assumption that the extracted X-parameters are measured at an impedance near 50Ω and are valid within a region around 50Ω that is representative of the match seen by the amplifier under test. Put another way, they are valid for normal amplifiers tested in normal systems.

Once the initial design is finalized, or in the case of an MMIC that might include integrated matching and even multitransistor designs in the implementation, then the amplifiers are essentially pre-matched and have a nominally Z_0 input and output impedance. Hot S-parameters are targeted at measurements of these amplifiers. They are useful in determining if the amplifier as designed is operating properly, if there is any defect in an amplifier, and in removing non-ideal test-system effects associated with mismatch of the test system.

6.9 Conclusions on Amplifier Measurements

An introduction to a wide range of amplifier characterization measurements was presented in this chapter, with comprehensive details focusing on gain and match corrected power. A few key points should be emphasized: the time spent pretesting the amplifier under test and optimizing the VNA setup is very well spent in avoiding mistakes in biasing and setup, even in preventing damage to the DUT and VNA through overdrive conditions. For the most

part, modern VNAs can perform substantially all the basic characterization measurements required on amplifiers including distortion measurements, in the form of harmonics and IMD, noise and noise figure measurement, power and efficiency measurements, and even load-pull measurements. A key advantage of using a VNA is that the high quality of calibration typical of S-parameter measurements can be extended to other measurements, resulting in both high speed and low error in the results. These methods, including some advanced de-embedding techniques, allow high quality measurements even with devices that have non-standard connections such as on-wafer and in-fixture devices.

References

- Keysight Technologies. (n.d.) Stability Circles Equation-Editor function courtesy of Andy Owen, Keysight Technologies
- Root, D.E., Verspecht, J., Sharrit, D. et al. (2005). Broad-band poly-harmonic distortion (PHD) behavioral models from fast automated simulations and large-signal vectorial network measurements. *IEEE Transactions on Microwave Theory and Techniques* **53** (11): 3656–3664.
- Root, D.E., Verspecht, J., and Xu, J. (2017). Closed-form solutions to large-signal PA problems: Wirtinger calculus applied to X-parameter. In: *2017 12th European Microwave Integrated Circuits Conference (EuMIC)*, 212–215. Nuremberg: <https://doi.org/10.23919/EuMIC.2017.8230697>.
- Shoulders, Robert E., and Loren C. Betts. (2008) Pulse Signal Device Characterization Employing Adaptive Nulling and IF Gating. Agilent Technologies, assignee. Patent 7340218. Print.
- Vendelin, G.D. (1982). *Design of Amplifiers and Oscillators by the S-Parameter Method*. New York: Wiley. Print.
- Verspecht, J., Bossche, M.V., and Verbeyst, F. (1997). Characterizing components under large signal excitation: defining sensible ‘large signal S-parameters’. In: *49th IEEE ARFTG Conf. Dig*, 109–117. Denver, CO: IEEE.

7

Mixer and Frequency Converter Measurements

7.1 Mixer Characteristics

Mixers and frequency converters are the key components that make radar, wireless, and satellite communication systems possible. The requirements on the characterizations of mixers and converters match closely to those of amplifiers, including frequency response and phase linearity of these frequency converters, output power and compression levels, noise figure, distortion and harmonics, and, unique to frequency converters, higher-order mixing products, local oscillator (LO), radio frequency (RF), and image rejection.

Just as amplifiers can be categorized by application, frequency conversion circuits can be separated into classes of devices grouping similar attributes and test requirements. At the broadest levels are mixers, which are simple 3-port devices that have an RF, intermediate frequency (IF), and LO port. These are most commonly created as a single-balanced or double-balanced device that uses a large external signal known as the LO to drive a set of diodes into on-off conduction over the period of the LO waveform.

A single balanced mixer is shown in Figure 7.1. This simple mixer is balanced only at the LO port, and the LO signal causes the mixer diodes to conduct during the positive half-cycle of the mixer. The balanced LO signal is created through the use of a transformer (at lower frequencies) or a BALanced–UNbalanced (balun) circuit at high frequency, often consisting of coupled transmission lines. Current from the RF source flows in the IF load during this conduction, with the $\frac{1}{4}$ wave lines keeping the RF drive from shorting the LO signal at the diodes. Essentially the RF signal is commutated, sampled, or chopped at the LO rate. Fourier analysis of the output waveform of the mixer will show frequency content at the sum and difference of the RF and LO signals and their harmonics. Since the LO is balanced, very little LO signal is present at the output. In the illustration, the lower-frequency IF can be seen in the repetitive pattern of the IF output waveform.

A single balanced mixer is quite simple and is often used at very high frequencies, but it has the disadvantage that the conduction is only 50%, so it has lower conversion efficiency than other models. Its main advantage is that since the RF and IF ports are not balanced, no balun is required, which can allow them to be quite broadband with simple construction.

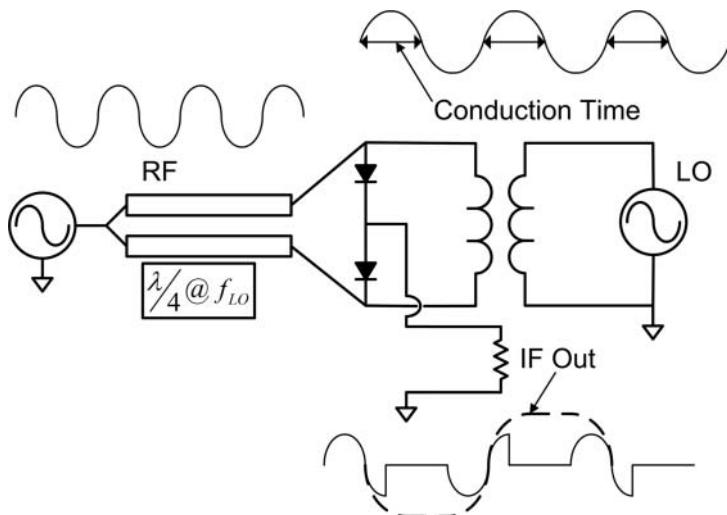


Figure 7.1 Input, LO, and output wave forms from a single-balanced mixer.

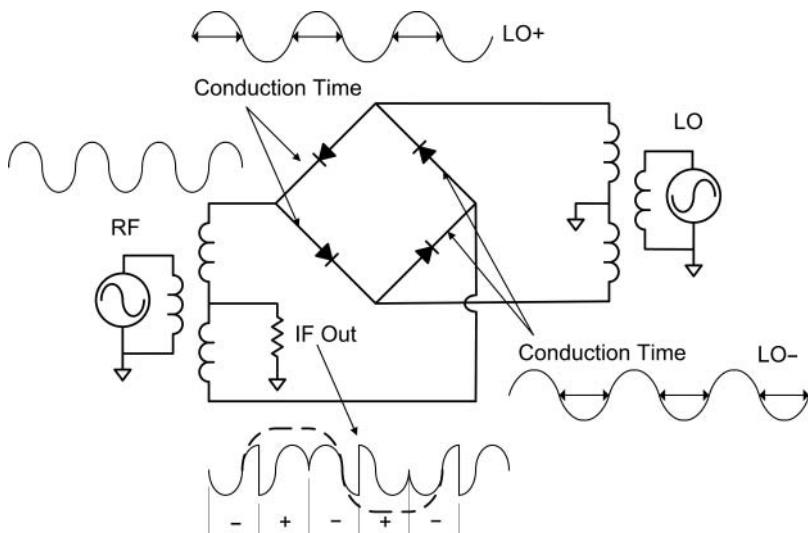


Figure 7.2 Conduction of a double-balanced mixer.

For most work in RF and microwave frequency ranges, conversion efficiency and isolation are more important than a simple design, and double-balanced mixers provide improvements in both areas. A typical double-balanced mixer (Maas 1986) is shown in Figure 7.2, with an illustration of the conducted signals at the output. The LO waveform alternatively turns on sets of diodes, and with each LO half-cycle, the RF signal's sign is changed as it appears at the IF port.

Double-balanced mixers have natural isolation of the transformer on each of the RF and LO ports, isolating these signals to the IF port. One variation on the double-balanced mixer is splitting the RF and LO signals to two sets of diode rings, with inversion on the RF path, and combining the IF signal (which are now out of phase) in a third transformer. This is sometimes called a double doubly balanced mixer, dual-ring mixer, or triple-balanced mixer. This has an advantage of splitting the RF signal between two mixers, thus reducing the RF in each mixer and improving the linearity of the overall mixer by 3 dB. The trade-off is more complexity and 3 dB higher LO power requirement for the same performance.

Other mixer forms include image reject, or single sideband (SSB) mixers, that utilize two mixers, with 90° of phase shift on two of the mixer ports. The classic image-reject mixer uses two mixers with a 90° phase shift on each of the RF and IF path, and an LO with an equal phase splitter to each LO input. These have the disadvantage of requiring 90° hybrids or phase shifters on each of the RF or IF path, as shown in Figure 7.3 (upper), which generally have limited frequency response.

Another configuration for image reject mixers have a 90° phase shift on each of the LO and IF outputs so that the signal at the combined output suppresses the one-sideband while

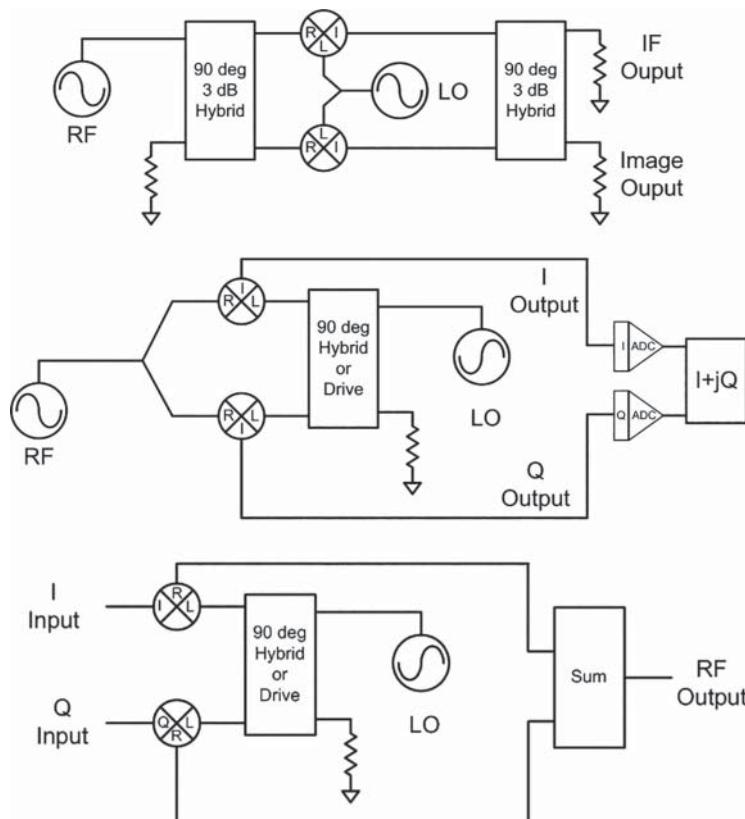


Figure 7.3 Image reject and IQ mixer topologies: standard topology (upper), topology used in digital demodulators (middle), IQ modulator or up-converter (lower).

enhancing the other. This version is often found in monolithic microwave integrated circuit (MMIC) mixers, where the LO phase shift is created by multiplying the LO, or using a higher frequency LO, and then dividing it in such a way to create 90° offsets; alternative forms use adjustable phase shifters to generate the offset on the LO port. The IF hybrid is often replaced by digitizers on each output (called the I and Q output), the 90° summation occurring numerically, shown in middle diagram of Figure 7.3.

The up-converter form of this mixer has two inputs, with assumed 90° phase shift, that are often used for complex modulation schemes where the output signal is centered directly on the LO, shown in the lower diagram of Figure 7.3. These are sometimes referred to as I/Q modulators as the LO signal is not translated in frequency, but rather any waveform on the I/Q input ports modulate the LO signal to generate the same baseband envelope waveform at the modulator output. In these mixers, the LO suppression is critical specification. Details of I/Q mixer measurements are discussed in Section 7.10.

7.1.1 Small Signal Model of Mixers

Even though the very nature of mixers is strongly non-linear, for the most part the transfer of signal from the RF port to the IF port is linear in its behavior and can be modeled much like S-parameters in an amplifier. Though the frequency is translated, it is linear in the sense that doubling the input voltage will double the output voltage, and if one applies a small modulated signal to the RF input, the same modulated signal will appear at the IF output, without distortion. The non-linear nature of the diodes produces the chopping of the RF signal, which translates it to the IF frequencies but does so in a linear way. Considering just the first Fourier component of the LO as a sine-wave input signal, multiplied by an input cosine signal, the mathematical representation is

$$\cos(\omega_{In}t) \cdot \sin(\omega_{LO}t) = \frac{1}{2}(\sin[(\omega_{In} + \omega_{LO})]t - \sin[(\omega_{In} - \omega_{LO})]t) \quad (7.1)$$

Thus, the output will have frequency elements at the sum and difference of the two input signals. Of course, the LO signal has many harmonics (all odd if it is symmetrical signal) and so there will be also be outputs at the sum and difference of each of the harmonics of the signals with the input signal as well, sometimes referred to as intermodulation spurs, or higher-order products. Either the sum or difference signal can be created. The sum signal always represents an up-conversion, and by convention the input signal is referred to as the IF signal (the IF signal always being the lower frequency of either the input or the output). If the difference signal is the desired output, the mixer can be either an up-converter (if the input is lower than the difference) or a down-converter (if the input is higher than the difference or higher than the LO). This is shown graphically in Figure 7.4. The upper figure shows that this condition creates both up- and down-converted signal; filtering on the output will determine if this is an up-converter or down-converter. We call this *normal mode*, because an increasing frequency at the input causes an increasing frequency at the output.

The lower plot shows the case where the mixer is only an up-converter, and the output can be either an image ($\text{LO} - \text{In}$) or normal ($\text{LO} + \text{In}$) mode mixer. An important aspect to note is that as the input signal moves up in frequency, the output signal can move up in frequency (call this the standard or normal mode) or down in frequency (call this the image mode or reverse

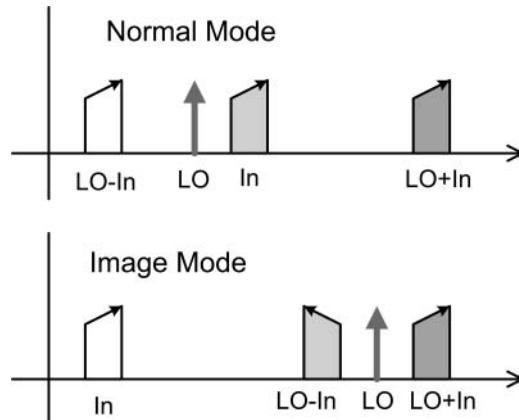


Figure 7.4 Graphical representation of signals at the input and output of a mixer.

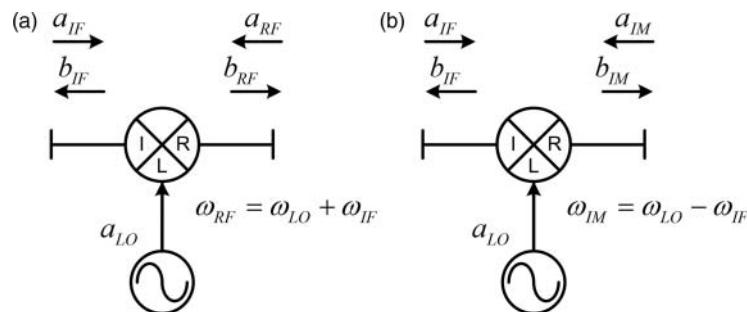


Figure 7.5 Schematic of a normal (a) and image (b) mixers showing incident and scattered waves.

mode). For the image case, the reversal carries over to a change in phase of the input signal: for a standard mixer, a positive change in phase at the input results a positive change of phase at the output; for an image mixer, a positive change of phase at the input results in a negative change of phase of the output. This property is important to consider when determining how to cascade the effects of mixers, filters, and other components, or to remove the effects of cables and connectors during a measurement.

A small signal model based on incident and scattered waves can be developed for a mixer (Dunsmore et al. 2004; Williams et al. 2005) using the definitions shown in Figure 7.5.

For an ideal standard mixer, the input signal is translated to the output with no change in amplitude or phase and no reflection from the port. Signal applied at the output of the mixer is translated similarly to the input frequency with no change in amplitude or phase and no reflection from the output port. Such an ideal mixer can be represented mathematically as

$$\begin{bmatrix} b_{IF} \\ b_{RF} \end{bmatrix} = \begin{bmatrix} 0 & a_{LO}^* \\ a_{LO} & 0 \end{bmatrix} \cdot \begin{bmatrix} a_{IF} \\ a_{RF} \end{bmatrix} \quad (7.2)$$

for a standard mixer where $|a_{LO}| = 1$

Here, the LO is assumed to interact with the mixer in such a way that a change in its power does not change the mixer conversion efficiency. For the normal operating point of mixers, this assumption holds true, so only the frequency and phase of the LO affect the transfer function. This represents an up-converter in the forward direction, and a down-converter in the reverse direction.

The scattered wave (output signal) at the RF port is at a frequency that is the sum of the IF and LO; if either go up in frequency, the RF goes up, similarly for phase. Notice, however, that the scattered wave at the IF port depends upon the incident wave RF port, but if the LO goes up in frequency, the IF goes down; it's similar for the phase, so in the reverse direction the IF response moves as the conjugate of the LO. The ideal response for an image mixer is subtly different and described by

$$\begin{bmatrix} b_{IF} \\ b_{IM}^* \end{bmatrix} = \begin{bmatrix} 0 & a_{LO} \\ a_{LO}^* & 0 \end{bmatrix} \cdot \begin{bmatrix} a_{IF} \\ a_{IM}^* \end{bmatrix} \quad (7.3)$$

for an image mixer where $|a_{LO}| = 1$

If the IF is taken as the lowest frequency, this is also an up-converter, but the image moves down in frequency as the IF moves up. And in the reverse direction the IF moves down in frequency if the image moves up so that conjugation of one of the LO terms and both of the image terms is required for proper treatment of the phase response.

These equations describe the frequency and phase response of an ideal standard mixer and an ideal image mixer, but real mixers will have reflections at the input and output, as well as transmission frequency response, and so must have a more complex description.

The additional response can be characterized in a few ways, as indicated in Figure 7.6. In (a) the non-ideal response is contained completely on the IF side; in (b) the non-ideal response is contained completely on the RF side; and in (c) the non-ideal response is split between the IF side and the RF side, with the forward response assigned to the IF side, and the reverse response assigned to the RF side. All these representations are equally valid for any particular frequency plan (a frequency plan describes the one-to-one correspondence of input frequency

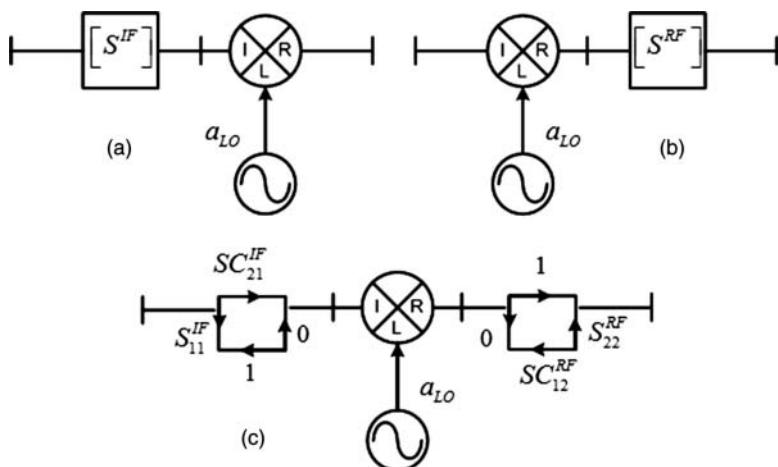


Figure 7.6 Schematic representations of mixers with nonideal responses.

range, LO range, and RF range) and represent an “error box,” which contains the non-ideal elements of the mixer behavior; and one may change from one form to another in a normal mixer. However, for an image mixer, special care must be taken when moving an error box from the input to the output.

From this figure one can define a set of scattering parameters that describe the behavior of these waves for a normal mixer under any load condition. The waves at the IF and RF frequencies for a standard mixer are represented by

$$\begin{bmatrix} b_{IF} \\ b_{RF} \end{bmatrix} = \begin{bmatrix} S_{11}^{IF} & a_{LO}^* \cdot SC_{12}^{IF} \\ a_{LO} \cdot SC_{21}^{IF} & S_{22}^{IF} \end{bmatrix} \cdot \begin{bmatrix} a_{IF} \\ a_{RF} \end{bmatrix} = [S^{IF}] \cdot \begin{bmatrix} a_{IF} \\ a_{RF} \end{bmatrix}$$

or

$$\begin{bmatrix} b_{IF} \\ b_{RF} \end{bmatrix} = \begin{bmatrix} S_{11}^{RF} & a_{LO}^* \cdot SC_{12}^{RF} \\ a_{LO} \cdot SC_{21}^{RF} & S_{22}^{RF} \end{bmatrix} \cdot \begin{bmatrix} a_{IF} \\ a_{RF} \end{bmatrix} = [S^{RF}] \cdot \begin{bmatrix} a_{IF} \\ a_{RF} \end{bmatrix} \quad (7.4)$$

for a standard mixer where $|a_{LO}| = 1$

Even though the S-matrix for $[S^{IF}]$ and $[S^{RF}]$ refer to different frequencies, the elements have identical values, and one can say that $[S^{IF}] = [S^{RF}]$, remembering that they are indexed by different frequencies. Thus, the real response of the mixer can be moved from one side to the other, allowing concatenation of mixer responses with other network elements. Applying this to an image mixer, however, is more complicated and subtle.

A similar scattering matrix can be defined for an image mixer as

$$\begin{bmatrix} b_{IF} \\ b_{IM}^* \end{bmatrix} = \begin{bmatrix} S_{11}^{IF} & a_{LO} \cdot SC_{12}^{IF} \\ a_{LO}^* \cdot SC_{21}^{IF} & S_{22}^{IF} \end{bmatrix} \cdot \begin{bmatrix} a_{IF} \\ a_{IM}^* \end{bmatrix} = [S^{IF}] \cdot \begin{bmatrix} a_{IF} \\ a_{IM}^* \end{bmatrix}$$

or

$$\begin{bmatrix} b_{IF} \\ b_{IM}^* \end{bmatrix} = \begin{bmatrix} S_{11}^{IM*} & a_{LO} \cdot SC_{12}^{IM*} \\ a_{LO}^* \cdot SC_{21}^{IM*} & S_{22}^{IM*} \end{bmatrix} \cdot \begin{bmatrix} a_{IF} \\ a_{IM}^* \end{bmatrix} = [S^{IM}]^* \cdot \begin{bmatrix} a_{IF} \\ a_{IM}^* \end{bmatrix} \quad (7.5)$$

for an image mixer where $|a_{LO}| = 1$

Here again, even though the S-matrixes for $[S^{IF}]$ and $[S^{IM}]$ refer to different frequencies, the elements have similar values, but for an image mixer $[S^{IF}] = [S^{IM}]^*$, remembering that they are also indexed by different frequencies.

For the standard mixer, an overall equivalent circuit can be drawn that moves all of the response from the IF side to the RF side, including the effect of source-match, as shown in Figure 7.7. For this equivalent circuit, the mixer is eliminated and the source changes frequency, but the values for the source-match remain the same.

The same equivalent circuit can be drawn for an image mixer, but with a decidedly different result, as shown in Figure 7.8. Here, moving the IF scattering matrix to the output conjugates all the terms. What is remarkable, and until recently not well understood, is that moving the source to an equivalent version at the image (IM) frequency requires that both the source wave, b_S , and the source-match, Γ_S , must be conjugated. Thus, Eq. (7.5) also tells one how to cascade components in the IF or RF paths before or after the mixer. From this, one can create an overall response including IF effects and mismatch between elements.

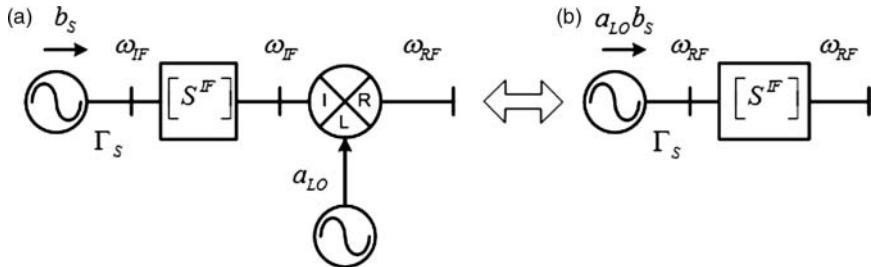


Figure 7.7 Actual circuit (a) and equivalent circuit at the RF (b) for a source and standard mixer.

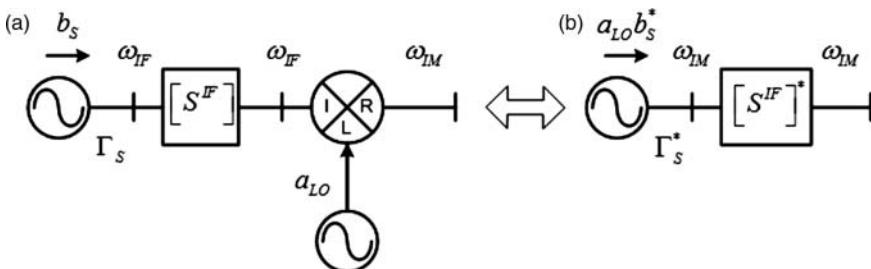


Figure 7.8 Actual circuit (a) and equivalent circuit at the RF (b) for a source and image mixer.

This representation has important use in the computation of responses of reciprocal mixers, used in vector mixer calibration, for image mixer cases, as discussed in Section 7.5.3.1.

Note that if only an input, or only an output, error box is used, then one cannot determine the response except for the case of the exact same LO frequency. For example, if a mixer is characterized across a wide band (say 10 GHz), with a fixed LO (say 15 GHz), and then later evaluated with a narrow band (1 GHz) within the wide band, the error box is only for the single LO frequency. If the error boxes are at the input and output, then one may break apart the individual frequency responses and compute a composite response with different choices of LO frequency provided the range is inside the original characterization range. This presumes the frequency response of the mixer does not change with LO drive frequency, which is roughly correct to a first order.

7.1.2 Reciprocity in Mixers

Reciprocity in mixers has a special meaning when considering the definitions of Eqs. (7.4) and (7.5). For the amplitude response, the mixer is said to be reciprocal if $|SC_{21}| = |SC_{12}|$; this behavior can be verified using standard scalar measurements as described later in this chapter. Since the phase response from the input to the output of a mixer depends upon the phase of the LO, and a shift in the LO phase also shifts the output, it is difficult to explicitly state the phase response of a mixer, as it will change in time, if the input and output are not harmonically related. That is, the phase of the LO, even if known relative to the IF at some particular time, will rotate in time to some other relationship at some later time. However, the LO, RF, and IF have a common factor; they will retain a specific phase relationship. For example, if the

IF is 10 GHz, the LO is 15 GHz, and the RF is 25 GHz, these have a 5 GHz common factor, and their phase relationship will repeat at a 5 GHz rate.

Thus, when referring to phase reciprocity, one generally means that the deviation in the phase response, versus frequency, of SC_{21} should match the deviation in the phase response of SC_{12} , or to be more succinct, the group delay of the SC_{21} should equal the group delay of SC_{12} , for a mixer to be reciprocal. This definition is very useful, as one primary reason for caring about the reciprocal nature of a mixer is in the calibration process for measuring the group delay of a mixer or converter.

7.1.2.1 Notes on LO Phase Response

One interesting aspect of the LO phase response of a mixer can be demonstrated if one splits the source signal to two mixers, fed by the same LO, and compares the IF signals on a coherent receiver, such as the test and reference inputs of a vector network analyzer (VNA), as shown in the upper window of Figure 7.9. The relative IF signals will be stationary in time. If one changes the phase of the LO in the path of one mixer, one will see a resulting phase shift in that paths' IF. Here, the LO is swept through a 1 GHz span, as well as the RF and the IF is fixed. Three cases are presented: the base mixer case, the case where a small delay is added to the RF path, and a case where a small delay is added to the LO path. Since RF and LO both sweep, this represents additional phase change as a function of frequency, and one might expect additional phase slope on the measured phase of the mixer. There are two plots:

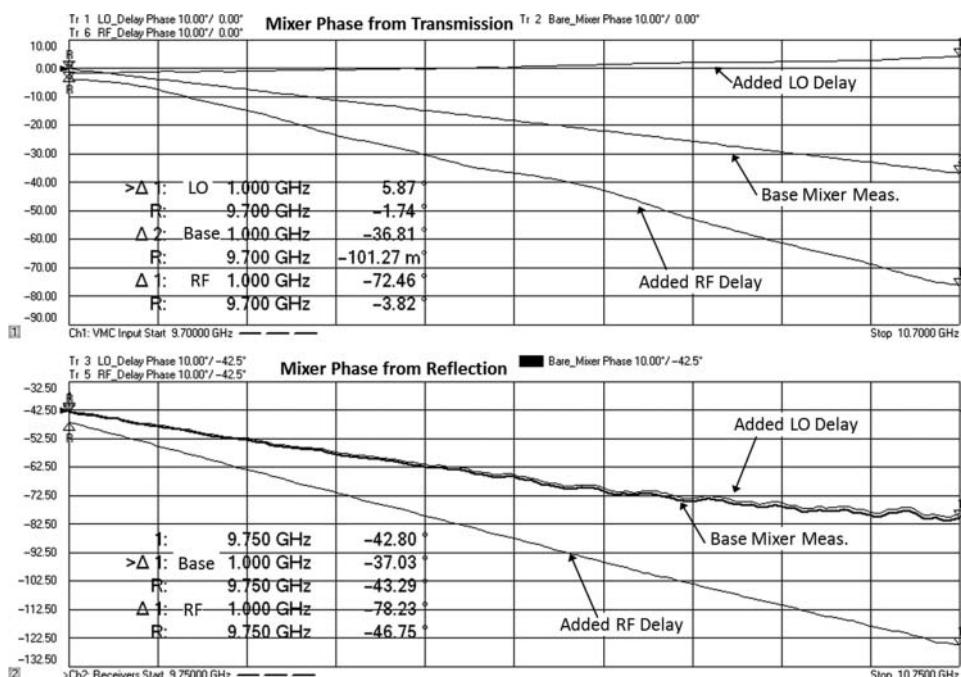


Figure 7.9 Understand the effect of phase shift of the LO: transmission phase (upper), reflection phase, two-way (lower).

the upper plot shows the phase shift for a transmission measurement; the lower plot shows the response where a reflection measurement is made (two-way) through the mixer first as an up-converter, reflecting off a short, and reconverting. For the case of the RF delay, one does see additional slope, of about 36°. The additional delay was about 100 ps: the expected additional phase would therefore be

$$\Delta\phi = \text{delay} \cdot 360 \cdot \Delta f = 100 \text{ ps} \cdot 360 \cdot 1 \text{ GHz} = 36^\circ \quad (7.6)$$

However, when the same delay is added to the LO path, an opposite slope is observed, of about the same value. Thus, the phase shift of the LO is translated directly to the IF signal, but as a conjugate of the phase change. For an image mixer, the opposite result is expected. This is exactly in-line with the behavior described in Eq. (7.4). In this case, the measurement represents the RF to IF conversion, related to SC_{12} term. In the lower window, a characterization of the mixer phase is performed looking at the two-way reflection of a short placed at the end of the mixer, as one looks at the equivalent S_{11} of a mixer at the input, with a short at the output.

Ignoring other match terms, the reflection is essentially the product of SC_{21} and SC_{12} . But the actual response includes the effects of the LO so that the proper description is

$$\begin{aligned} \text{For a standard mixer } S_{11}^{IF} \Big|_{P2_Short} &= a_{LO} SC_{21}^{IF} a_{LO}^* SC_{12}^{IF} \Big|_{S_{11}=S_{22}=0} \\ \text{For an image mixer } S_{11}^{IF} \Big|_{P2_Short} &= a_{LO}^* SC_{21}^{IF} a_{LO} SC_{12}^{IF} \Big|_{S_{11}=S_{22}=0} \end{aligned} \quad (7.7)$$

Thus, if the mixer is reciprocal, the square root of this S_{11} measurement is the conversion loss of the mixer. From this, one can conclude that shifting the phase of the LO does not have any effect on the apparent phase of the S_{11} response of a mixer, as the up and down-conversion create a similar positive and negative shift due to the conjugation function the LO so that any phase shift is canceled. The measurement in the lower window shows exactly this effect. Adding the delay to the RF input shows the same response as the upper window, but adding delay to the LO path shows virtually no change (the slight change is likely due to some small mismatch effect on the LO port). Thus, one cannot determine if the phase of an LO has been shifted by observing only the S_{11} input of the mixer, regardless of the terminating impedance. In one method of measuring mixers (see Section 7.5.3.1), the reflection of a termination on the output is used to determine the two-way conversion loss through the mixer. From Eq. (7.7), it is clear that the phase response of the LO port is not included in such a characterization.

7.1.3 Scalar and Vector Responses

The response of mixers was shown previously to be complex-valued function with amplitude and phase response. For many applications, the only concern is the amplitude response, usually called the *conversion loss*. Other responses include the compression, output power, output harmonics, and spurious products. All of these responses are represented by scalar quantities, and the characterizations are commonly called *scalar mixer measurements*. In the past, most scalar quantities were measured using simple systems of two sources (one for the LO and one for the input) and a spectrum analyzer (SA) (power meters are generally not used in mixer characterization as the other mixing products and LO feed-through cause substantial error).

When mixers are used as part of a communications system, both the magnitude and phase response become important in some formats, so these systems require characterization of the mixer's phase or delay response in addition to amplitude. Since the responses are complex, they are commonly called the *vector response* of the mixer, and measurements are classified as vector-mixer measurements. Until recently, entirely different systems were required for vector versus scalar, with many systems providing only the delay response, often through interpretation of modulated results. Newer techniques, described in the following sections, provide for measuring these complex values with a single system, with high accuracy in both magnitude and phase responses.

7.2 Mixers vs. Frequency Converters

Frequency converter is a term used to describe a system comprised of filters, amplifiers, isolators, and mixers that are combined to create an overall frequency conversion system block. These can have one, two, or even more stages of mixing, with amplification and filtering before and after each stage. These systems are developed to provide rejection of unwanted signal and images, remove or isolate higher-order mixing products, and provide the necessary gain and power required by the overall system design. Due to their specific attributes, the methods of measuring frequency converters are somewhat different from those used on “bare” mixers; the methods described here will be distinguished by the terms *mixers* and *frequency converters*.

Mixers are understood to be lossy or low-gain (if active mixers), without input or output filtering, and have only a single conversion. Passive mixers are often reciprocal, or nearly so, and can convert frequencies with nearly the same efficiency from the RF to the IF or from the IF to the RF. Passive mixers can have substantial higher-order products at the output, with some feed-through elements (such as LO) being as high or higher than the desired output products. An example spectrum from a mixer is shown in Figure 7.10 where two measurements are made, with a shift to the input-frequency to identify spurious products. Shifts in the output products indicate the order of the multiplication that generated the term. Explicitly called out in the picture are some of the harmonics, feed throughs, and spurs. For example, the 2:1 spur, from $Spur_{2:1} = 2 \cdot f_{LO} - f_{RF}$ is shown at Marker 5, and since it is near the IF, it might be in-band. The 3:1 spur is very large, only 13 dB below the main RF out; this indicates that the mixer has substantial conversion efficiency on the third harmonic of the LO. Also seen are many spurs that have second product of the RF, and one that has a third order: Marker 9 highlights the 5:3 spur, which changes power as it shifts down in frequency. Also shown at Marker 4 is the primary sum product of RF and LO. One must realize that all of these spurs at the output can reflect off a non-ideal load and remix inside the converter.

Frequency converters, or simply converters, have filters that protect against unwanted image conversions and typically have active stages that provide substantial positive gain and isolation between stages and can have one or more stages of mixing. The multiple stages of conversion provide a way to create the same conversion as a simple mixer but allow spurious products to be eliminated by filtering between stages. The amplification stages provide substantial isolation in the reverse direction so that converters are essentially unilateral (one-direction) devices.

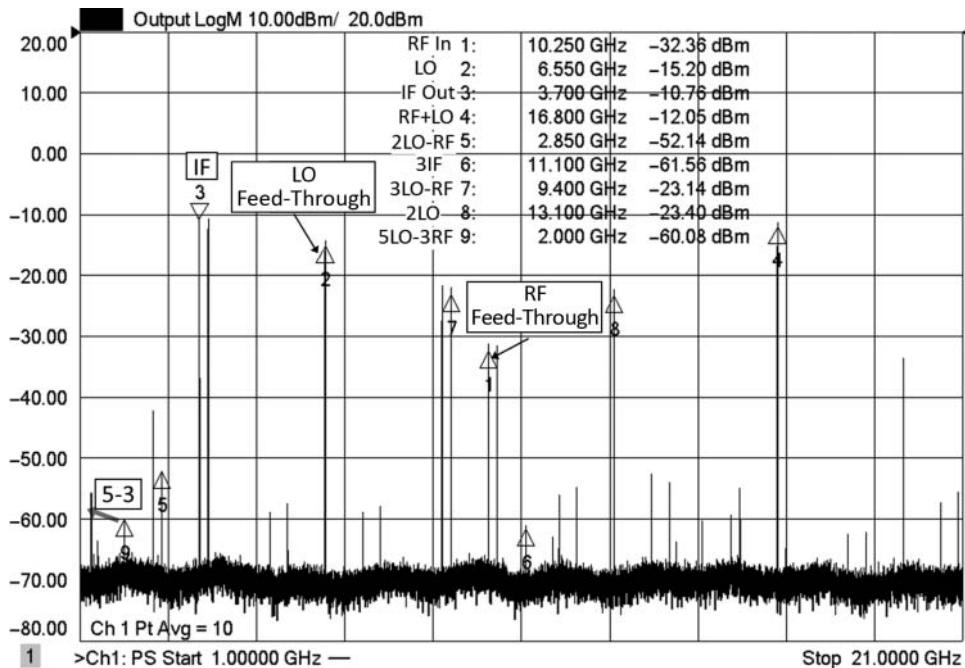


Figure 7.10 Typical output response of a mixer showing harmonics and spurious products.

7.2.1 Frequency Converter Design

While this is not intended as a text on design, it is helpful to understand the principles of converter design to understand how the higher-order products of a mixer are created and how they are eliminated using multiple conversion stages. Consider the diagram in Figure 7.11, which is roughly to scale. The up-conversion of the input (IF) signal is shown above the upper line. On the lower line are shown the harmonics of the IF and LO. Below the upper line are shown the construction of the higher-order products: 2 : 1 and 3 : 2 and 4 : 3 products, and one can see that they must cross the desired output at some the indicated frequencies. In this construction, the crossing frequency occurs when the height of the higher-order product equals the height of the RF output.

This mixer will always have the spurious products, and in measurements of the conversion loss, when the higher-order product crosses the desired product, the conversion gain will show a discrete anomaly, or spur. This is *not* a measurement error; in fact, at that frequency, the mixer conversion loss really does change due to the higher-order product. These products can be eliminated with careful multistage conversions as described next.

7.2.2 Multiple Conversions and Spur Avoidance

If a multistage converter is constructed as shown in Figure 7.12, the major higher-order spurs can be avoided. The first stage consists of an up-converter with a high-side LO so that the

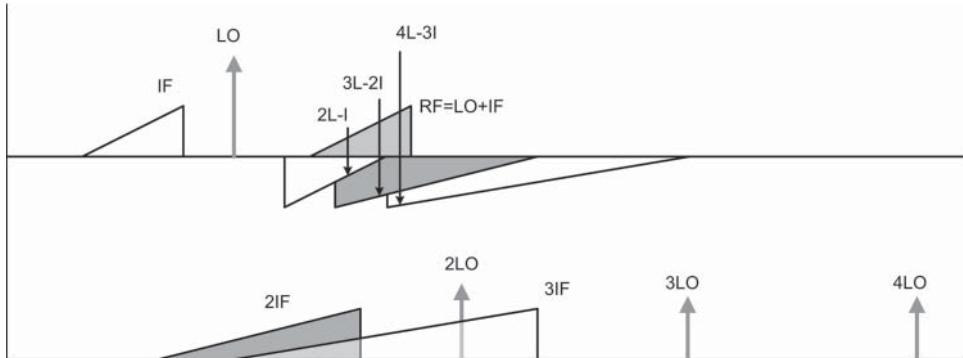


Figure 7.11 Diagram for mixer high-order products.

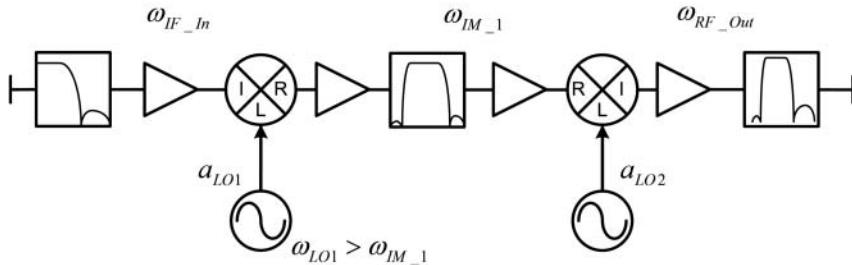


Figure 7.12 Multistage frequency converter.

higher-order products lie outside the desired output. Depending upon frequency, multistage converters can have a wide variety of conversion configurations.

The second stage down-conversion now occurs at higher frequencies so that the higher-order products do not cross the desired output range. The first stage is an image mixer and will cause a phase reversal; if the second stage is also an image mixer (high-side LO), a second phase inversion occurs, and the overall mixer response will be normal. Filtering provided at each stage ensures that the overall response is not sensitive to out-of-band signals; amplification is required to make up for the loss of the multiple stages and filtering. Typically, an input amplifier is used to improve the noise figure of the mixer, but there is often a trade-off between creating a low noise converter and creating distortion by over-driving the input mixer, so the input low-noise amplifier (LNA) interaction with the mixer distortion needs to be well understood. Often there is an input filter to restrict the frequencies at the LNA input. Similarly, between stages, amplifiers are often added to provide the proper gain; the distribution of gain is made such that the noise figure is maintained while preventing higher-order products from causing distortion in the amplifiers.

Adding filters to mixers completely changes the overall response of mixers and greatly affects the measurement techniques. Mixers tend to have poor matches, so the interaction between the filter and the mixer must be carefully controlled and carefully measured. Because

the input match is complicated (due to the combination of input and output filters), there are often tight test specifications to ensure the match is within specified limits. This means that mismatch correction techniques are very important, and the test setup must support high-quality measurements of mismatch.

On the other hand, bare-mixer measurements can have substantial higher-order products exiting all three ports (see Section 8.3.2 for details), and the interaction of these products with the test system can cause substantial errors. In some cases, a down-converter (which has as a desired output the difference term, RF – LO) can have a sum term (RF + LO) that can be large enough to reflect off the VNA load, re-enter the mixer, reconvert to the input, and then remix with the LO, to create a significant amount of IF signal. This will cause an apparent change in the conversion gain, just due to the output match of the test system at the sum frequency. In fact, a class of mixers called *image-enhanced mixers* is particularly designed to improve their conversion efficiency by making use of this image.

For normal mixers, there is no error correction method yet proposed to remove this effect, so the best choice is to minimize the test system mismatch. Thus, with a bare mixer, adding attenuators at the test ports may be required to obtain a sufficiently low mismatch to allow a good measurement. This issue only applies to higher-order products; the mismatch of the desired product with the test system can be characterized and removed, as discussed in Section 7.5.

7.3 Mixers as a 12-Port Device

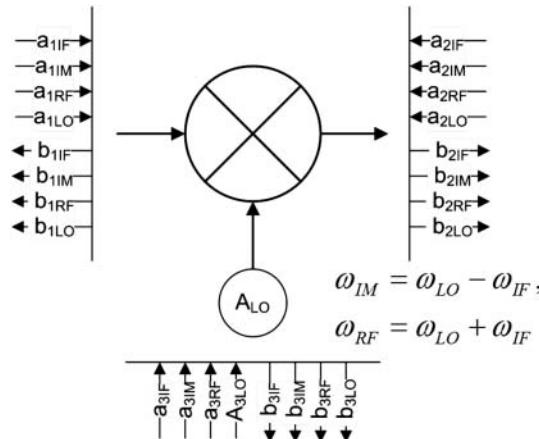
In the most common understanding, mixers are treated as a simple 3-port device, with an RF port having only RF frequencies, an LO port having LO frequencies, and an IF port producing only IF frequencies. This simple understanding is characterized by a conversion matrix that looks like Eqs. (7.4) and (7.5). And indeed, a properly designed converter does appear that way; but a mixer (without filtering and isolation amplifiers) behaves in a much more complex way.

In reality, there will be RF at the LO and IF ports, IF at the RF port and LO ports, LO at the IF and the RF ports, as well undesired products such as IM at the IF and RF ports, and so on, so that even for the first-order products (sum and difference of the LO and input frequency), a mixer should be considered as a 12-port device (Dunsmore 2009). Or rather, it is a 3-port device with four “modes” at each port. Mathematically, each of the frequency elements is treated as a separate input at each of the ports, so the overall response is described by the equation illustrated in Figure 7.13. In fact, no one ever uses this mixer description to any practical effect, but it is an illustrative way to demonstrate all the possible mixing and remixing effects.

7.3.1 Mixer Conversion Terms

7.3.1.1 IF to RF Conversion

The mixer described by the Figure 7.13 is an up-converter; the ports are identified by number and by the frequency present at that port number. If one assigns port 1 as the input, port 2 as the output, and port 3 as the LO, then $SC_{2RF, 1IF}$ is the first-order standard up-conversion gain (sum of LO and IF), and $SC_{2IM, 1IF}$ is the first-order image (difference of LO and IF) up-conversion gain, as illustrated in the figure. Also illustrated are some examples of re-conversion terms.



Re-conversion Due to Input Mismatch

Re-conversion Due to Output Mismatch

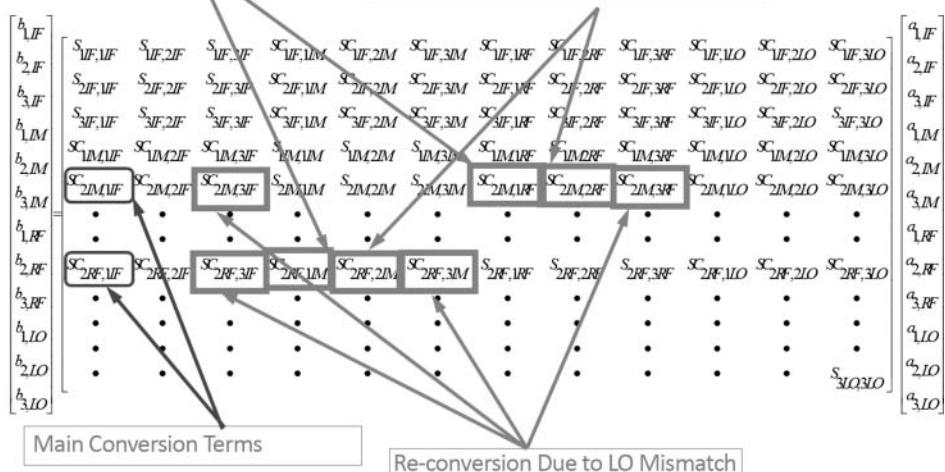


Figure 7.13 A mixer as a 12-port device to describe all first-order products.

7.3.1.2 Reflection and Re-conversion

The re-conversion terms indicate how a reflection or mismatch at one of the ports will generate an RF or IM signal at the output port. These represent only first-order terms; there are sets of these conversion terms for higher-order products as well. These re-conversion terms are an ever-present source of ripple in the conversion gain measurement, and a sample of these terms are illustrated in the following equation.

$$\begin{aligned}
 \text{Errors} = & (\Gamma_{\text{Source}} \cdot (S_{1RF,1IF} \cdot S_{2RF,1RF} + S_{1IM,1IF} \cdot S_{2RF,1IM} + \dots)) \\
 & + (\Gamma_{\text{LO}} \cdot (S_{3IF,1IF} \cdot S_{2RF,3IF} + S_{3IM,1IF} \cdot S_{2RF,3IM} + \dots)) \\
 & + (\Gamma_{\text{Load}} \cdot (S_{2IF,1IF} \cdot S_{2RF,2IF} + S_{2IM,1IF} \cdot S_{2RF,2IM} + \dots) + \dots)
 \end{aligned} \quad (7.8)$$

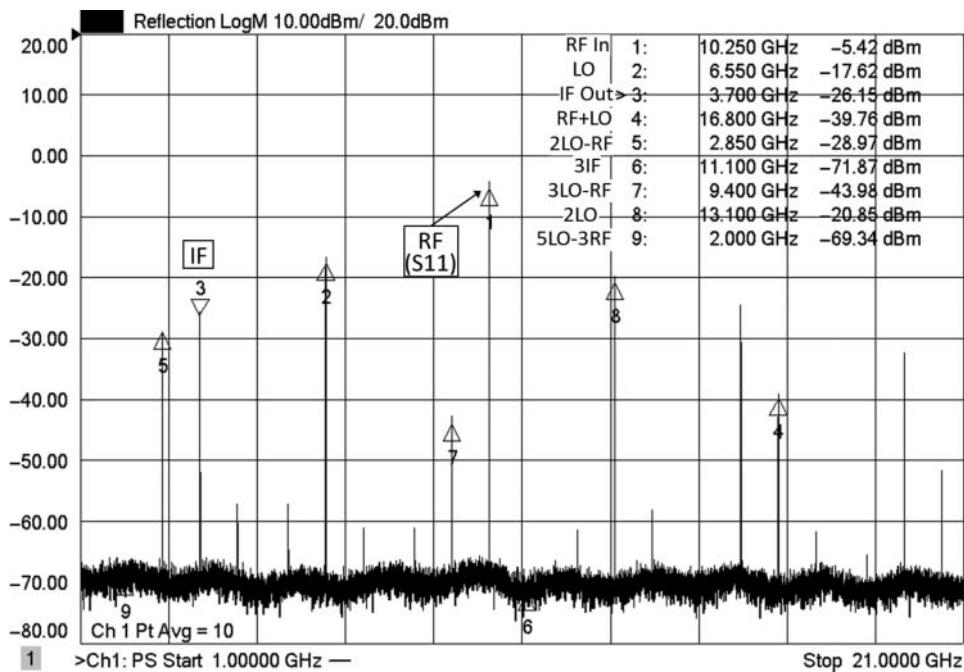


Figure 7.14 Mixer signals emitted or scattered (reflected) back from the input port.

The re-conversion due to the input and output mismatch can be controlled while testing a bare mixer by having very good match at the ports, for example from adding an attenuator at each port. To fully illustrate the nature of mixer products at each port, consider the spectrum shown in Figure 7.14. This shows a spectrum of signals that are emanating or scattered from the *input* port of the mixer; this is not a measurement of the output of a mixer! Of course, the RF input is the highest signal and represents the S_{11} of the mixer (here the input power was 0 dBm, so S_{11} is -5 dB). But notice the rich spectrum of other components; in particular, notice that some higher-order products, such as the 2:1 spur highlighted at Marker 5, emit from the input at an even higher power level than from the output (see Figure 7.10). Reflection of these higher-order products, and re-mixing in the mixer under test (MUT), is a source of errors in mixer measurements.

7.3.1.3 Image Enhancement

Some mixer designs, particularly narrowband applications, take advantage of the reconversion terms to employ a particular impedance matching at the image frequency to reflect it back to the mixer. This reflected signal is reconverted through the $SC_{2RF, 2IM}$ term in Figure 7.13. This term is a measure of conversion gain of the image signal incident on the mixer output (port 2 in this case) to the RF signal exiting the mixer output port. In a normal ideal double-balanced mixer, the input signal is equally converted to RF and image (IM) signals. The conversion loss to the RF can be enhanced by reflecting the IM signal back into the

mixer, where it can reconvert to RF (one may think of it as a double conversion, first to IF and then to RF). Though not widely used in practice, this is an example of reconversion that can have beneficial attributes. However, because phasing becomes very difficult, this benefit of improved conversion loss can turn detrimental as the bandwidth of the mixer is increased, and the phasing causes the output to cancel rather enhance, so that the conversion loss is degraded. As the bandwidth of the mixer is swept, the phasing between the mismatch and the mixer changes, and the enhancement/degradation appears as conversion loss ripple.

7.3.1.4 Conversion on the LO Port

The LO reconversion terms poses a greater problem, because LO drive levels needed for mixer measurements are often high levels, and many systems cannot afford the loss of an attenuator on the LO port. The use of isolators in the LO path does not necessarily improve the situation as the isolators typically are not well-matched at the RF, IM, and IF frequencies, even if they provide a good match at the LO frequency. The input signal can leak to the LO port, and the output and image signals can also appear out the LO port; as they progress down the LO feed and reflect back into the mixer from the LO mismatch, they will be reconverted (in the case of input or IM) or leak (in the case of output) to the output and add to or subtract from the main output signal. As frequency is swept, the phasing of this error signal will cause ripple in the measurement.

In converter designs, a filter is commonly used on the LO port, and this will provide a high reflection at the other frequencies, but if the length between the filter and the mixer is small, the effect on conversion gain will be slowly varying and prevent excessive ripple in the conversion gain. In many cases, an amplifier is placed very near the mixer to boost the LO signal; this has the added benefit of providing isolation to the other frequencies and so prevents these signals from reflecting with large phase shifts (due to longer delays) and thus minimizes ripple.

For a frequency converter, the 12-port mixer model is not needed. Filtering on the RF, IF, and LO ports ensure that only signals at these frequency exit the mixer, so only these signals interact with the test system. That is the primary reason why measurements on frequency converters differ from measurements on bare mixers: the wide range of other products coming from bare mixers interact with the test system match to produce ripple; only by providing good, well-matched ports can high-quality measurements of bare mixers be realized. With frequency converters, which produce only a single frequency at each port, any errors in match from the test system can be characterized and removed from the measurement results. These mismatch correction methods are discussed in the next sections.

7.4 Mixer Measurements: Frequency Response

7.4.1 Introduction

As with any other measurement, the quality of mixer measurements depends upon the quality of the measuring equipment. In the past, it was common to fully depend upon the instrument manufacturer's flatness performance of the source and the receiver (most often a spectrum

analyzer), and it was up to the user to provide post-processing compensation for any cables or connectors used in the measurement.

Legacy VNAs were not used for mixer measurements except in the case of trying to measure relative phase between mixers. In the early 1990s, the first implementation of frequency offset mode (FOM) was developed for VNAs that allowed the source to sweep over one set of frequencies while the receiver swept over a different set of frequencies. The source was phase-locked to the receiver at some offset frequency; for a mixer this amounted to the LO frequency.

This early implementation did allow conversion loss measurements but measured only the output power and relied upon a good source-power calibration for accuracy. Only response-calibrations were available. In the ensuing years, the capability VNAs applied to mixer measurements has dramatically improved, with high-speed measurements and high-quality calibrations, so that now VNAs are the preferred instrument for measuring mixers and frequency converters.

The discussion of frequency response measurement methods is divided into sections, first covering the measurement methods for amplitude response (Section 7.4.2); phase response (Section 7.4.3), which includes some group delay measurements; and group delay response based on modulation methods (Section 7.4.4). In addition, some special consideration is given for Swept LO measurements (Section 7.4.5). This section only discusses the concepts of measurement; calibration for each of these measurements is discussed in Section 7.5. While the calibration of the amplitude measurements is covered in the straightforward way, calibration for phase and delay are more complicated, where most phase-measurements rely on a characterized mixer for calibration, with one exception. All the significant phase calibration methods are discussed in Section 7.5, and each calibration method may be applied to one or more measurement methods.

7.4.2 Amplitude Response

In general, there are two different ways in which mixers are used, and these require two different stimulus setups to provide the characterizations of these mixers. The first is a swept-RF/swept-IF/fixed-LO mode of operation, and this is the manner in which virtually all communication mixers and converters are used. Often, a frequency converter is used a block down-converter to translate many channels or RF signals down to a common IF channel. Sometimes this is referred to as a “fixed-IF” measurement, but that is a misnomer. For each fixed-LO frequency, the RF-to-IF channel response is measured, and the measurement is repeated for many LO frequencies.

In some cases, only the gain at the center of the channel response is measured, at each of the channels defined by the RF and LO steps. In this case, it is still not really a “swept-LO” measurement, but more properly defined as a stepped LO, RF-to-IF measurement. The distinction is important primarily when considering the group delay response of a converter: the definition of delay refers to RF-to-IF transition, and so this implies the both the RF and IF must change (sweep in frequency) in order to have a defined group delay. One *cannot* simply sweep the RF and LO at a fixed IF and measure the phase change of the IF: the mixer definitions given in Section 7.1 show that while the LO phase will cause an phase shift in the IF, it is not related to the channel characteristics and so will distort the measurement of delay.

For delay measurements of a converter, even over a range of LO frequencies, the LO must be stationary during the acquisition of the RF-to-IF transfer function.

In some limited applications, particularly in mixers used in radar systems, the RF and LO are swept together, and in these cases, the relative phase and amplitude changes between the RF and LO are important attributes. In this case, because the LO is not constant, a reference mixer must be used as will be discussed in Section 7.4.3.2. But in this case, only the relative phase can be measured; the delay result is not well defined as the IF frequency does not change at all, and therefore there is no “delta-frequency” at the output from which to form a definition of delay.

7.4.2.1 Fixed LO Measurements

Amplitude response of a mixer is defined simply as the output power divided by the input power, or

$$SC_{21} = \left| \frac{b_{2_OutputFreq}}{a_{1_InputFreq}} \right| \Bigg|_{a_{2_OutputFreq}=0} \quad (7.9)$$

In older legacy analyzers, it was not possible to fully decouple the source from the receiver. Since the source would be phase-locked to the desired input frequency, by inserting the MUT in the path between the source and receiver that was used as the source-phase-lock reference, only the output power could be measured. In this case, it was not possible to measure the power of the input signal and so the value of input power was determined simply from the source power setting. In fact, the earliest implementation would require two different physical connections to first measure the source output power, over the input frequencies, on the reference channel with no mixer connected; this would be stored in memory. Next the MUT was added, and the frequency synthesizer was set to allow offset phase-locking so that the source remained at the input frequency while the receiver was tuned to the output frequency, and the output response was measured. The ratio of the measured response to the memory gives the conversion gain, provided that the source power doesn't drift, and the receiver response is calibrated. This method presumes that the LO is provided separately; it did not require, however, that the LO be locked to the VNA reference. Any offset between the LO reference and the VNA would be absorbed by the receiver phase-lock process.

However, any change in source power would require a reconnection to remeasure the input power. A key limitation here is that the source was not truly independent of the receiver, but could be offset only by putting a frequency converter (the MUT) in the phase-lock path. The HP8753 and HP8720 VNAs were the first commercially available VNAs to provide this capability.

Using modern techniques, the principal method for measuring the amplitude response of a mixer now is to connect the mixer directly from the VNA input to the VNA output. Many VNAs provide additional ports and additional sources; one of these can be used to drive the LO directly. A typical connection diagram is shown in Figure 7.15.

Two sweeps are made, one over the input frequency and one over the output frequency, and the ratio of these responses is computed to give the mixer conversion gain. This requires, of course, that the VNA receiver can be decoupled in frequency from the VNA source, and most modern VNAs provide this capability, either inherently or as an optional upgrade. During the

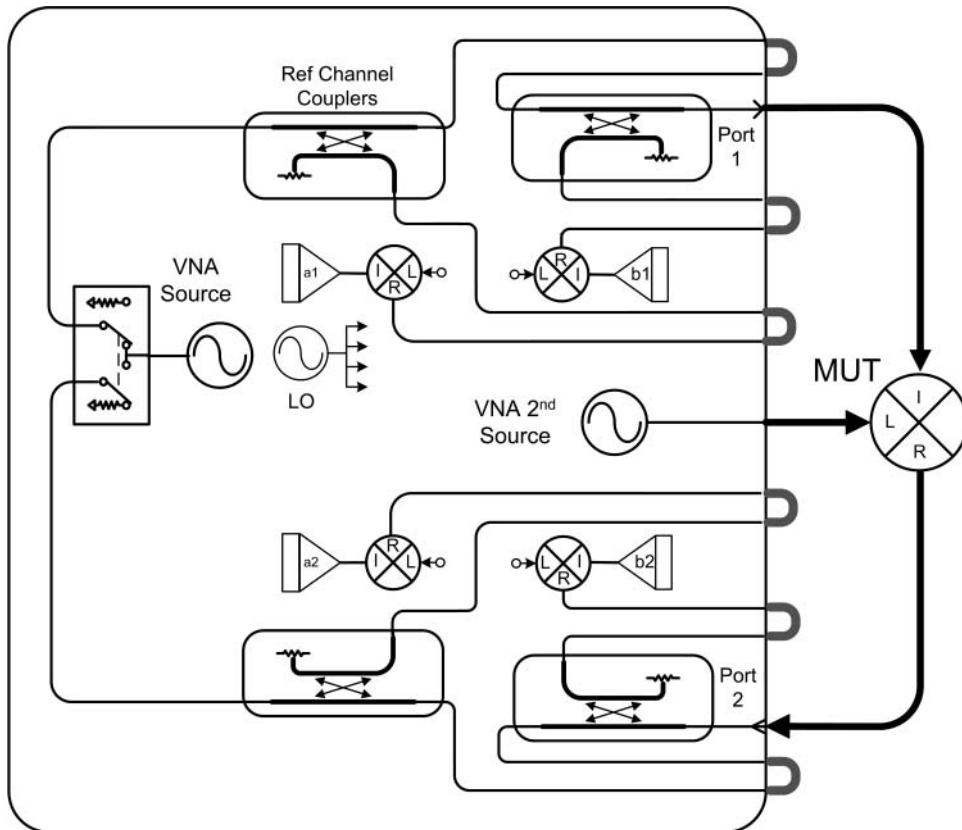


Figure 7.15 Typical connection for mixer measurements.

first sweep, the reference channel measures the input signal, and the input match is measured as well. During the second sweep, the VNA receiver is offset, and the power at the output is measured. If each sweep is performed on a separate channel, the equation editor can be used to display the conversion gain. Alternatively, the receiver can be switched from input to output frequencies on each point, providing a faster response between measurements and reducing the likelihood that the source power has drifted between measurements.

Some VNAs provide application specific software to measure mixers automatically, including input and output sweeps, as well as automatically controlling the LO frequency and power, through a software graphical user interface (GUI). An example of one such mixer measurement GUI is shown in Figure 7.16. In this case, both single and dual conversion measurements are supported.

The response is still dependent upon the input and output match of the VNA, but with additional measurements of the input match at the input frequency and output match at the output frequency, correction for RF and IF mismatch can be accomplished as will be described in Section 7.5. Thus, the complete amplitude response of mixers, as defined by the conversion matrix from Eq. (7.4), may be obtained. An example of a mixer measurement, showing all

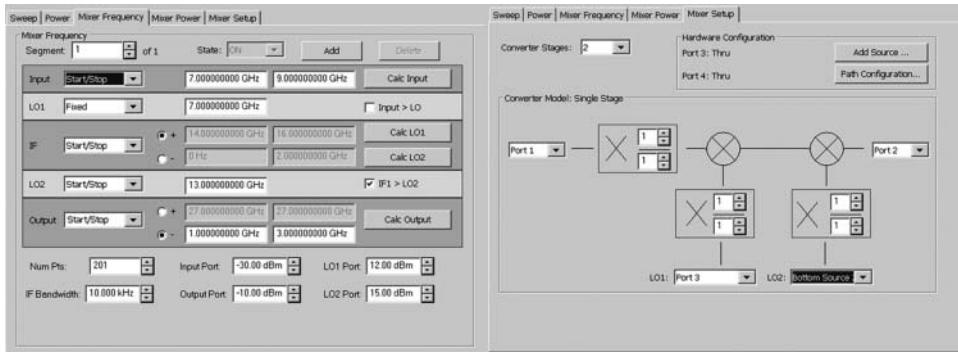


Figure 7.16 Mixer measurement graphical user interface.

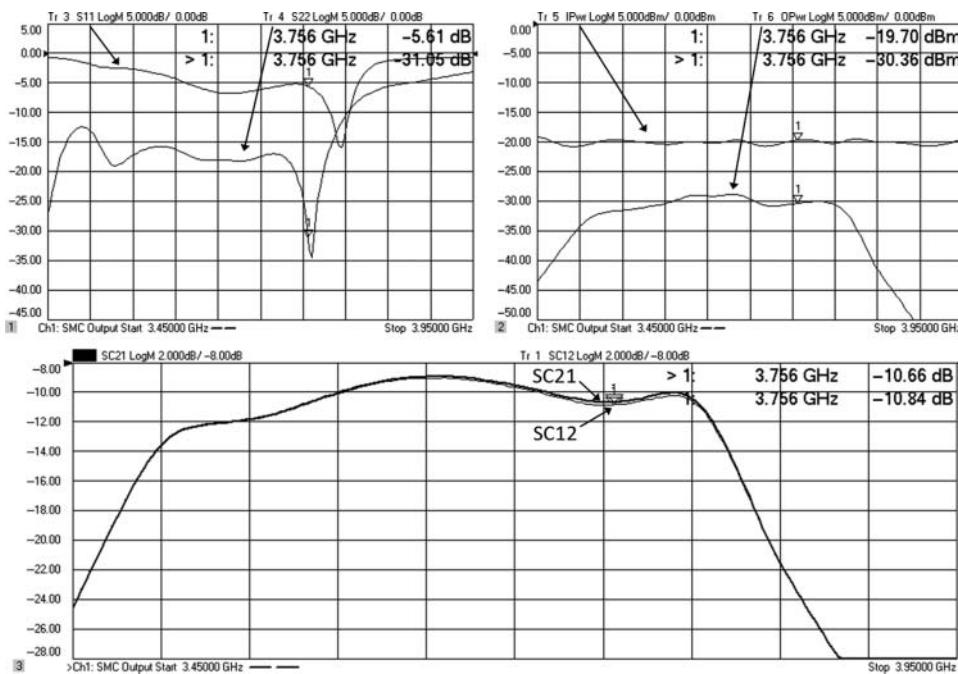


Figure 7.17 A complete mixer “S-parameter” measurements.

the mixer “S-parameters” including forward and reverse conversion, input and output match, as well as input and output power in the forward direction, is shown in Figure 7.17.

This is an example of a single mixer with input and output filters, and the conversion loss of the forward direction and reverse direction are very similar, indicating that this mixer is very nearly reciprocal in amplitude. Measurement of a mixer without any filtering may show it to be less reciprocal, due to re-mixing of higher-order products. Whenever a bare mixer is measured, it is highly recommended to use a 3–6 dB pad on each of the test port cables to improve the port match. In some cases, even if the source and load match are well controlled

for the frequencies used, out-of-band frequencies can see different reflections from the VNA. Adding pads at the test port helps to eliminate ripples caused by the reflection and remixing of these higher-order products.

Some examples of frequency converters will be demonstrated in the following sections on delay measurements (Section 7.4.4) and high-gain mixers (Section 7.9.5). Other measurements besides frequency response, such as gain versus drive, TOI, and noise figure, are covered in Sections 7.6–7.8.

7.4.3 Phase Response

With the increase of the use of complex modulation formats, the phase response of mixers and converters is becoming a key measurement parameter. In the past, measuring the phase (and from that computing delay) of a mixer was difficult. Several methods have been proposed, and until recently, each had substantial drawbacks. In the latest development in mixer phase measurements, a very new method removes many of the previous drawbacks and provides a simple but accurate calibration as well. The following sections describe the various mixer phase and delay measurement methods. All of the measurement methods have one or more associated calibration methods, which are described in Section 7.5.

7.4.3.1 Down/Up-Conversion

The first practical means for measuring mixer delay employed the use of a reconverting mixer so that the input and output frequencies are the same. The overall response of the mixer pair was measured, and the MUT response was inferred by compensating for the effect of the reconverting mixer in some way (see Section 7.5.2). This measurement method is illustrated in Figure 7.18. Here, the MUT contains a built-in band-pass filter, which dominates the frequency response and delay. The key benefit of this method is that it can be used on a standard VNA.

While very straightforward in concept, this method has some substantial difficulties, listed here:

It requires a reconverting mixer that matches the frequency range of the device-under-test (DUT) mixer but operates in the opposite conversion mode; i.e. if one's MUT is an down-converter, the reconverting mixer must be an up-converter.

It requires that the LO be shared between the up- and down-converting mixers; if either mixer has an embedded LO or the MUT is a dual stage mixer, the method is not practical.

For proper results, a band-pass filter must be used between the mixers to remove the image signal, or else the reconverting mixer will reconvert both the RF and the IM signal, resulting in an erroneous response. The effects of the filter response must be compensated for in the overall response, and mismatch between the filter and the mixers, and even between the two mixers in the passband, can lead to substantial errors in the overall result.

Further, higher-order products that fall within the bandwidth of the image filter may reconvert in the second mixer and create an error in the overall response, which cannot be filtered out.

The calibration and accuracy depends upon the characterization of the reconverting mixer; in most cases, this mixer must be reciprocal to allow characterization.

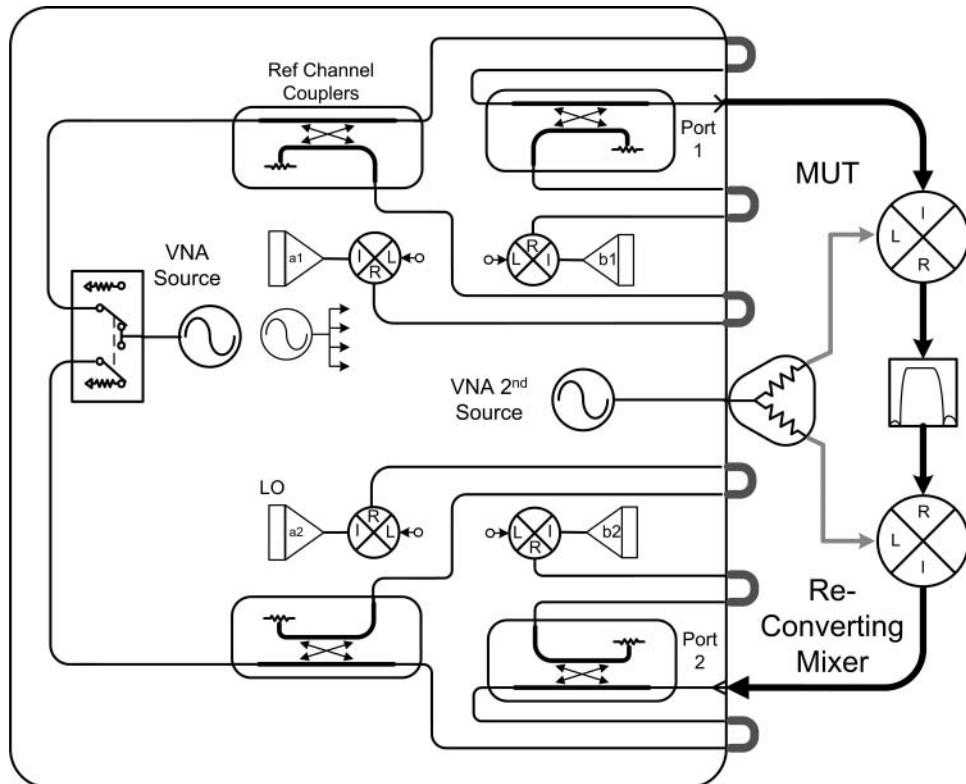


Figure 7.18 Down/up-conversion method for measuring phase.

Because of these difficulties, this method is seldom used except in cases where the VNA does not support mixer measurements, so the input and output frequencies must be the same. It is most commonly used when measurements are made on converters with substantial phase response (large delay); in such a case, the delay response of the reconverting mixer may be ignored. The measurement of phase deviation response of an example mixer, which will be utilized in the demonstration of other methods as well, is shown in Figure 7.19. The calibration method for this measurement uses the reciprocal mixer cal described in Section 7.5.3.1. After a full 2-port calibration, the equivalent S-parameters of the reconverting mixer, referenced to the MUT input frequency, are de-embedded from port 2 to create a calibrated response.

The upper window shows the magnitude response and phase response. In this figure, an equation editor function is used to compute the deviation from linear phase over the bandwidth of the MUT, as illustrated in the upper window. The deviation from linear phase function allows one to look at the phase response just in the passband, without being distracted by the noise or excess response outside of the passband, and will be used here to compare several of the measurement methods.

The lower window shows the group delay response, with some odd-looking delay ripples in the mid-band. The built-in band-pass RF filter in this converter has a smooth delay

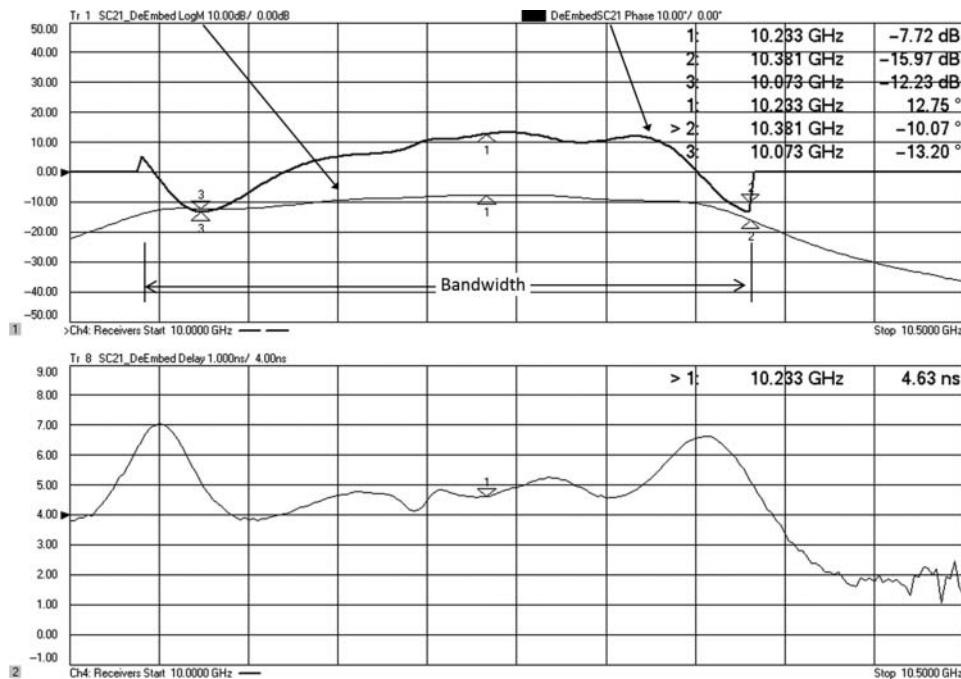


Figure 7.19 Phase response using down/up-conversion.

response here. The extra ripples in this response are from interactions between the MUT mixer and the reconverting mixer.

The details for this measurement are as follows:

1. Pre-characterize the reconverting mixer, as described in Section 7.5.3.1.
2. Configure the down/up-converter path, using the MUT and the reconverting mixer and an image filter. It is usually best to have the VNA tuned to the highest frequency. If the MUT is a down-converter, connect the RF port of the MUT to port 1 and connect the output to the reconverting mixer, so the reconverting mixers goes from the MUT-IF to port 2. If the MUT is an up-converter, connect it to port 2, and connect the reconverting mixer to port 1 with its IF connected to the MUT IF input. It is a best practice when possible to use the lowest frequency to connect the mixer pair, which often eliminates difficulties in removing the image signal.
3. Verify that the VNA can measure the down/up pair; check that the LO is split properly between both mixers. Note, this method will *not* work with a converter that has its own LO unless its reference can be locked the reference of the LO for the reconverting mixer. Check to ensure that neither mixer is compressing.
4. Once this setup is verified, disconnect the mixer pair and perform a full 2-port calibration; this will allow the removal of some mismatch effects at the ports.
5. De-embed the reconverting mixer from its port using standard de-embedding math, referenced to the port frequency. De-embedding is discussed in detail in Chapter 11.
6. Reconnect and measure the mixer pair to obtain the results shown in Figure 7.19.

One note of caution: when choosing an image filter, one may choose a filter labeled with the appropriate bandwidth but one that has a spurious passband at the image frequency (as the author did when first attempting to make the measurement of Figure 7.19). This is especially true for band-pass filters, which are notorious for having spurious passbands around the third harmonic of the filter. If this spurious passband occurs where the image signal (RF + LO) in the mixer occurs, erroneous measurements will certainly result.

With this method, only the frequency response can be reasonably determined for the low-power linear region of the mixer. Measurements of power dependent parameters are difficult because of the series connection between mixers. This limitation can be eliminated using the other methods that follow.

7.4.3.2 Parallel Path Using a Reference Mixer (Vector Mixer Characterization, VMC)

Newer VNAs allow the source and the receiver to be independently tuned (without the external phase lock path required on older analyzers such as the HP 8753), so it is possible to configure a setup with the MUT in the test channel, and a separate, similar mixer in the reference channel, as shown in Figure 7.20. The mixer in the reference channel is used to measure

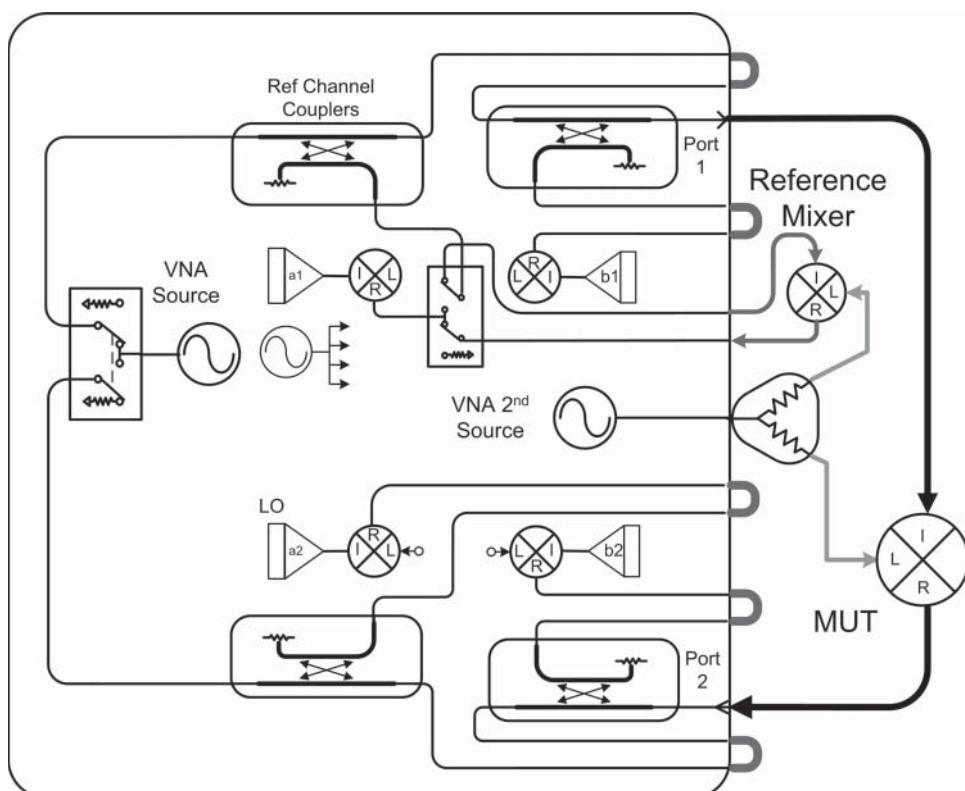


Figure 7.20 Vector mixer measurement system using a parallel path.

the source or input signal but provides the same frequency signal to the reference receiver as the MUT does to the port 2 test-mixer, so the phase relationship of the two represents the phase of the output signal to that of the input signal. The amplitude response can be discerned in a similar way. In the optimum configuration, a switch is provided in the reference path to allow the reference mixer to be bypassed so that normal S-parameter measurements, such as S_{11} , can be performed.

Since the MUT is connected directly between ports, the effects of mismatch between port 1 of the VNA and the input of the MUT at the input frequency, and the mismatch between port 2 of the VNA and output of the MUT at the output frequency can be removed. This is accomplished in a method nearly identical to the full 2-port calibration used in normal S-parameters, with the exception that the S_{12} term is set to zero, as the system cannot measure the reverse conversion loss of the MUT. Calibration of this system follows a normal 2-port calibration method, but requires an extra step of placing a calibration mixer (or cal-mixer) in the thru path to establish the transfer phase of the reference receiver. Details on this are described in Section 7.5.3.1. Once the calibration is completed, the cal-mixer is replaced with the MUT, and the conversion gain is directly measured. The calibration removes the mismatch effects, but only for signals at the input and output frequencies. Other products that exit the mixer can still remix and create ripple on the SC_{21} trace if the VNA port match is not particularly good and if the MUT is not filtered.

The parallel measurement system has several advantages over the down/up method described earlier:

Match correction of the input and output signal can be performed to improve corrected performance.

The reference mixer and cal-mixers have the same conversion direction as the MUT. In many cases, users have several examples of each mixer, so allowing one to be used as the reference when the other is used as test mixer is convenient.

The measurement does not require that the MUT mixer have any filtering; the VNA naturally selects only the desired output frequency.

One need not use the same LO signal to drive the reference and MUT mixer; in fact, they do not even need to be locked to the same frequency reference. In cases where the MUT has its own embedded LO, the VNA receiver can be programmed to track the output signal. Details on this are discussed in Section 7.9.4.

If the signal level to the reference mixer is small enough so that it is linear over all drive powers, then the gain and phase compression versus RF drive can be determined for the MUT. Also, if separate LO drives are used for the reference and the test mixer, then the mixer conversion loss versus LO drive can be determined as well.

An example of a mixer phase deviation measurement is shown in Figure 7.21 (again the phase deviation is only shown over the normal bandwidth of the mixer); the measurement is labeled VC21 for Vector Converter 21, and this method of parallel measurement is sometimes called Vector Mixer/Converter or VMC. Also shown is the response of the same mixer from Figure 7.19, using the down/up method and de-embedding (labeled SC21_DeEmbed). The same second mixer used for the reconversion mixer in Figure 7.18 was used, in its down-conversion mode, as the calibration mixer, and was characterized as described in Section 7.5.3.1. The up/down method with de-embedding shows a response with more

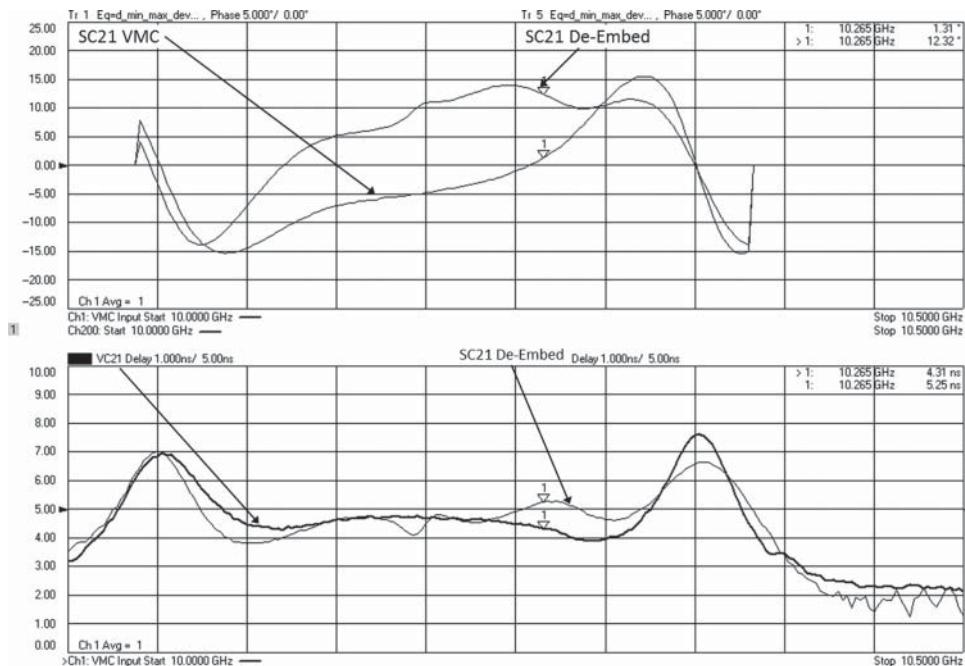


Figure 7.21 Phase deviation for a mixer using the parallel (VMC) method.

ripple, which is very likely an error, compared to the VMC method. The VMC results for group delay and phase deviation match reasonably well with the S_{21} response of the filter that is used in the converter. The fact that the VMC trace is smooth and symmetric is a good indication that it has a better quality of calibration than the up/down method.

The VMC method has been used extensively in high-performance tests of converters and has the advantage that phase noise on the LO is canceled between the reference and test channel; thus, the VMC produces very low noise phase and delay responses. The VMC method provides very good measurements even for a lossy mixer, and one can see from the figure that the group delay measurement is much smoother in the high-frequency, stopband region. The main difficulty is that a calibration is required for every LO frequency used, so that if a converter has many channels, many calibrations must be performed, and for each LO frequency, the cal-mixer must be separately characterized. The difficulties of calibration with multiple LO frequencies are essentially eliminated using the method described in Section 7.5.3.3.

7.4.3.3 Phase-Coherent Receivers

In 2010, new hardware and software innovations provided an advancement in VNA functionality that allowed the sources and receivers of some VNAs (Keysight PNA-X) to change frequency in a phase-coherent manner; this allowed new innovations in mixer measurements that vastly simplified the setup and calibration process.

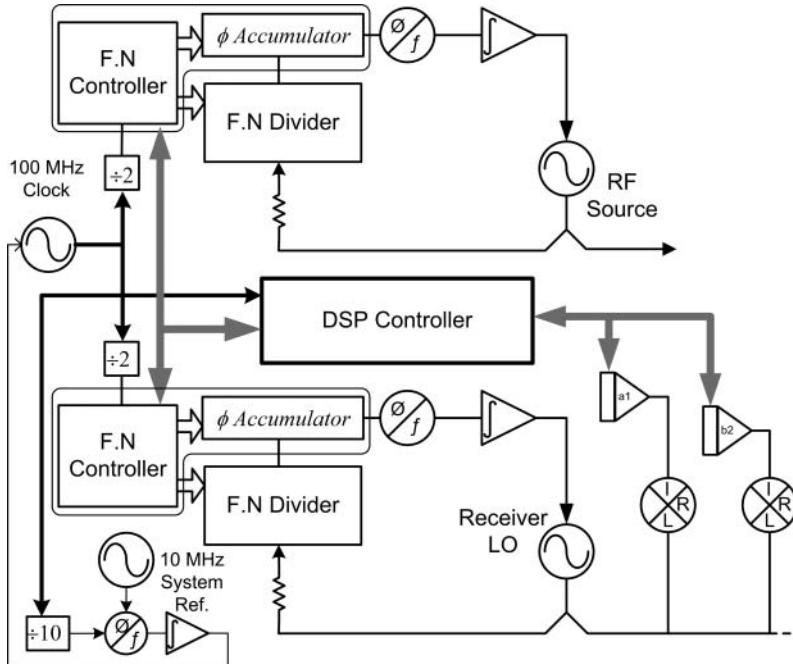


Figure 7.22 The synchronous sweeping is accomplished by a common reference and a digital phase accumulator in the synthesizers of the source and LO.

The measurement system consists of a VNA with integrated sources and receivers, where both the source and receiver frequencies are determined by a pair of frequency synthesizers. One synthesizer provides the source stimulus signal, and the other synthesizer provides the local-oscillator for the receiver. The difference of these two signals represents the IF frequency from the receiver, which is sampled by an integrated digital IF.

The synthesizers in this VNA are unique in that they use a custom fractional-N synthesizer chip-set with an integrated phase accumulator. When programmed to sweep frequency, the phase accumulator is programmed to accumulate a certain amount of additional phase each clock cycle, thus providing a synthesized phase sweep coherent with the system clock. The DSP and digital IF are also locked to a common clock, so each source and LO and the digital IF (not shown) have a deterministic phase relationship over the data sweep acquisition. The high-frequency clock is locked to the system 10 MHz clock. Figure 7.22 shows a simplified block diagram of the fractional-N synthesizer, DSP, and analog-to-digital-converter (ADC) control.

With this system it is possible to measure the absolute phase change across a span of frequency on the reference or test mixer. That is, the magnitude and relative phase of the “a” and “b” waves of a mixer may be directly measured. The mathematics of a frequency converting system was defined in Eq. (7.4). From this we can see that phase of the output (RF) signal depends upon both the phase of the LO signal and the phase of the input signal (IF) as well

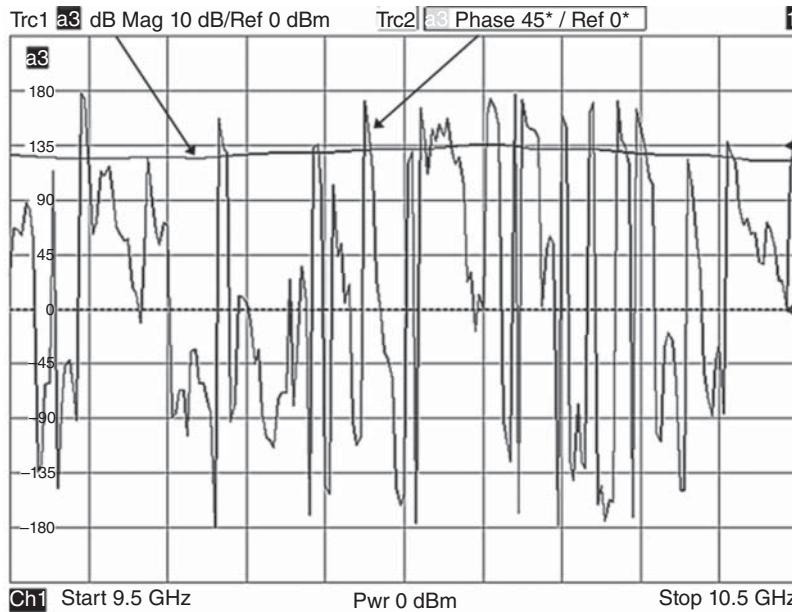


Figure 7.23 Amplitude and phase response of a single receiver in a normal VNA.

as a mismatch term that depends on the reflected RF signal at the output. The output signal from a mixer can now be defined as

$$b_{RF} = a_{LO} S_{21}^{IF} \cdot a_{IF} + S_{22}^{RF} \cdot a_{RF} \quad (7.10)$$

Since it is possible to measure the magnitude and phase of b_{RF} and a_{IF} directly, one can compute SC_{21} of the converter directly; Eq. (7.10) also shows there is a dependency of the phase of a_{LO} . This dependency is not compensated, but for fixed-LO converters the phase of the LO is a constant offset and does not affect the phase or group delay response of the mixer with respect to the information bandwidth. Thus, there exists an arbitrary offset of the phase response, which we take advantage of when applying this method to converters with embedded LOs. Finally, for direct computation one assumes that there is no mismatch at port 2 (the RF port in this case). If $a_{RF} \neq 0$, then an additional step is required to compensate for port 2 mismatch, as described in Section 7.5.3.3.

To understand the differences between this system and a standard VNA, consider the response shown in Figure 7.23 of a single receiver magnitude and phase measurement. The amplitude response is smooth and flat, but the phase response is nearly random across the frequency range.

Since the phase response is not stable on a sweep-by-sweep basis, it is not possible to find a correction method to recover the single-receiver phase response.

Figure 7.24 shows a measurement of reference R1 (a_1) and test B (b_2) phase responses, using a VNA with coherent receivers, in a standard measurement class and applying an

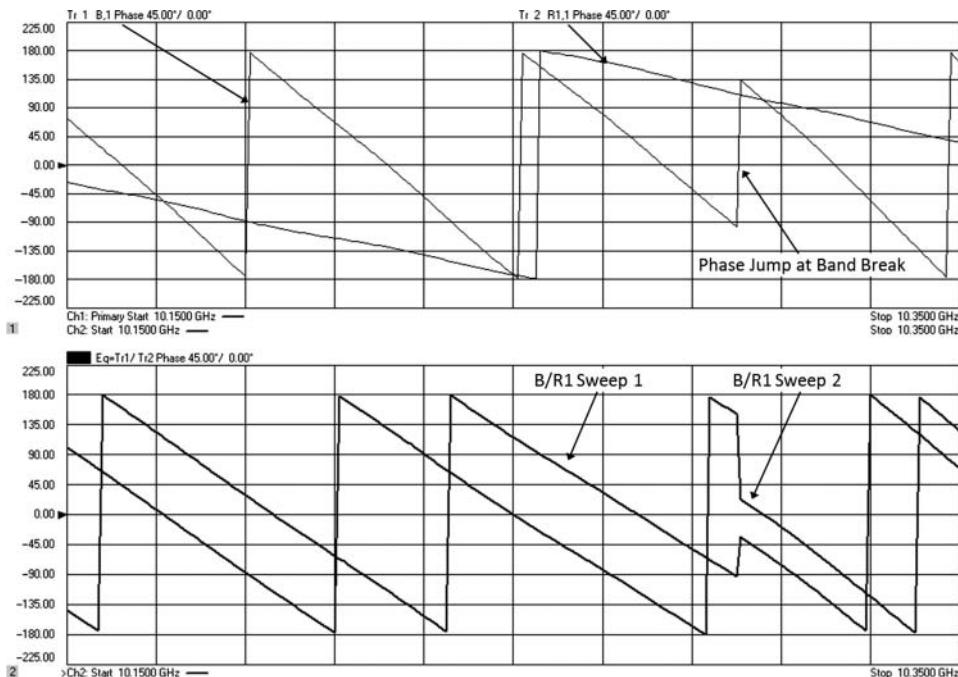


Figure 7.24 Phase response of the B and R1 receivers a standard VNA (upper plot) and on the ratio of B/R on (lower plot) in frequency offset mode. The memory trace illustrates the change that occurs from sweep to sweep.

offset frequency between the source and receiver. The R1 and B receivers are at different frequencies, and the data for each is acquired on different sweeps. The lower plot shows change in phase response B/R1 (b_2/a_1) with data and memory traces representing two different sweeps. The upper window shows the individual phase response of a_1 and b_2 . It is clear that the phase response is not stable, and if one looks at the 10.3 GHz point (for input frequency) in the sweep, there is a sharp discontinuity in the phase response that does not repeat from sweep to sweep; this is due to the 4 GHz band-switch in the output receiver. In a normal VNA mode, this is commonly seen and is a consequence of resetting the synthesizers between bands.

While the phase trace is not stable sweep to sweep, one notices that it does retain a sort of continuous response for each band of the synthesizer. This is in contrast to normal VNAs where the phase response of an individual receiver is completely random on a point-by-point basis (though a ratio of B/R will have a stable phase if at the same frequency). But with a VNA with coherent receivers, the digital IF is coherent with the synthesizers, at least over a band, and the phase relationship of a single receiver is maintained inside a band on a single sweep; this attribute will be exploited for new measurement method.

One recognizes two important features of the traces in Figure 7.24: (i) the phase response at the start of the sweep changes with each sweep, but the point-to-point variation remains constant, except at the band break; and (ii) the phase-slope on each side of the band break

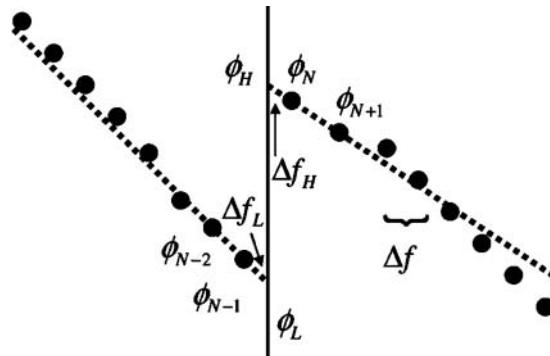


Figure 7.25 Phase stitching at synthesizer band breaks.

remains constant even if there is a phase offset. The new method utilizes these two facts to remove the sweep-to-sweep variation of the phase response of the receivers. First, the phase traces are normalized to the phase of the first point of the sweep. From Eq. (7.10), the phase of the LO adds an arbitrary phase offset, so one can simply assign this phase offset to the first point of the sweep, removing the sweep-to-sweep variation. Second, the new method recognizes that the phase slope at each side of break forms the group delay of the response of the system near the band break, as shown in Figure 7.25. This can be used to extend the phase of the lower band break to ϕ_L , and from the upper side to ϕ_H , thus allowing a method of stitching together the phase offset at each band.

The offset is computed as

$$\begin{aligned} \text{offset} &= \phi_H - \phi_L \quad \text{where} \\ \phi_H &= \frac{\Delta f_H(\phi_N - \phi_{N+1})}{\Delta f} + \phi_N \quad \text{and} \quad \phi_L = \phi_{N-1} - \frac{\Delta f_L(\phi_{N-1} - \phi_{N-2})}{\Delta f} \end{aligned} \quad (7.11)$$

Figure 7.26 shows the result of phase normalization and phase stitching. The input and output signals are labeled as IPwr and OPwr. One can see that the phase response is now continuous, and what is more, the phase response is stable on a sweep-by-sweep basis; the lower window shows the results of two consecutive sweeps, overlapped. This allows one to make un-calibrated phase measurements without a reference mixer, giving a vector version of SC_{21} . But these responses also include the phase response of the source, the reference channel receiver and the test channel receiver and the phase response of the associated cables and connectors. Fortunately, these can be compensated with a few calibration methods, as described in Sections 7.5.3.1 and 7.5.3.3.

Computing from the markers on the input receiver (R1, called out as IPwr) and the output receiver (B, called out as OPwr), one can see that the SC_{21} phase is exactly the difference. Further, on a sweep-to-sweep basis, the phase drifts very little. The lower plot shows the results of two sweep-acquisitions, with only 0.005 degrees of difference. Since the reference and test measurements occur with two different sweeps, the phase noise of the source and LO are not accounted for and will add to trace noise. This can be effectively eliminated by adding more averaging; in this case 10 averages were used.

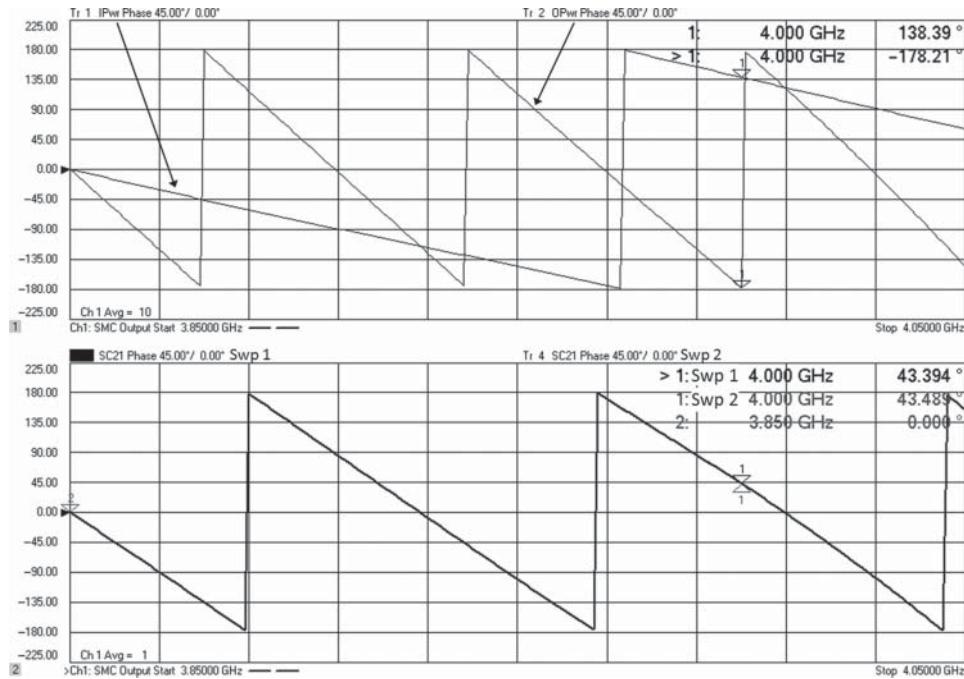


Figure 7.26 The phase response of the IPwr and OPwr, with phase stitching is now stable.

A calibrated measurement of the same mixer used in Figure 7.21 is shown in Figure 7.27 for the coherent receiver test method. Here all three measurements (De-Embedded, VMC, and Coherent Phase) are shown and they compare similarly, with VMC and Coherent Phase giving the very nearly the same answer.

The calibration details for these measurement methods are discussed in Section 7.5.

7.4.4 Group Delay and Modulation Methods

Group delay for mixers are computed in a similar manner as for linear devices, with the exception that the response for image mixers must be computed with the phase conjugated to show the delay as positive number, thus

$$\begin{aligned}\tau_{d_std} &= -\frac{d\phi_{Rad}}{d\omega} = -\frac{1}{360} \frac{d\phi_{deg}}{df} \\ \tau_{d_image} &= \frac{d\phi_{Rad}}{d\omega} = \frac{1}{360} \frac{d\phi_{deg}}{df}\end{aligned}\quad (7.12)$$

Here, the group delay is computed from the phase response of the mixer. There is one additional technique that may be used to compute the group delay of mixers based on modulation techniques.

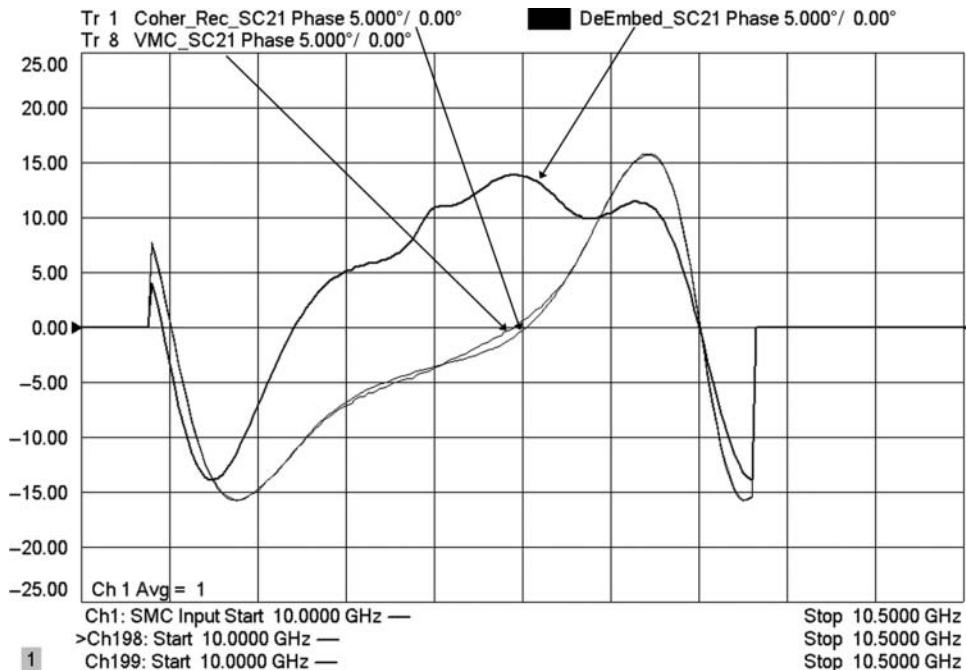


Figure 7.27 Comparison of three methods measurement for phase deviation.

In one form, an AM signal is applied to the input of a mixer, and the AM envelope is detected at the input and the output and compared. Any phase shift of the AM envelope can be associated with the delay of the MUT, and with the delay determined in this way, computing delay is

$$\tau_d = \frac{\Delta\phi_{\text{deg}}}{360 \cdot f_{\text{mod}}} \quad (7.13)$$

In this measurement method, the higher the modulation frequency, the lower the error in measurement provided the phase detection is the same.

An alternative of this method uses a two-tone stimulus to create essentially an AM modulated signal. However, rather than detecting the envelope, a receiver with essentially a dual-channel detection path is utilized to measure the relative phase of two tones at the input and compare them with the relative phase of the two tones at the output. In this way, if there is drift in the frequency of the LO for the MUT, the drift applies to both tones, and the delay measurement is not affected to the first order. Thus, the group delay can be computed by looking for an excess or difference in the change in phase of the measurement as

$$\tau_{d_2\text{-tone}} = -\frac{\Delta\phi_{\text{Input}} - \Delta\phi_{\text{Output}}}{360 \cdot \Delta f_{2\text{-tones}}} \quad (7.14)$$

The calibration of modulation methods rely principally on comparing the MUT response to that of a cal-mixer, much like the parallel method and suffers from similar drawbacks that

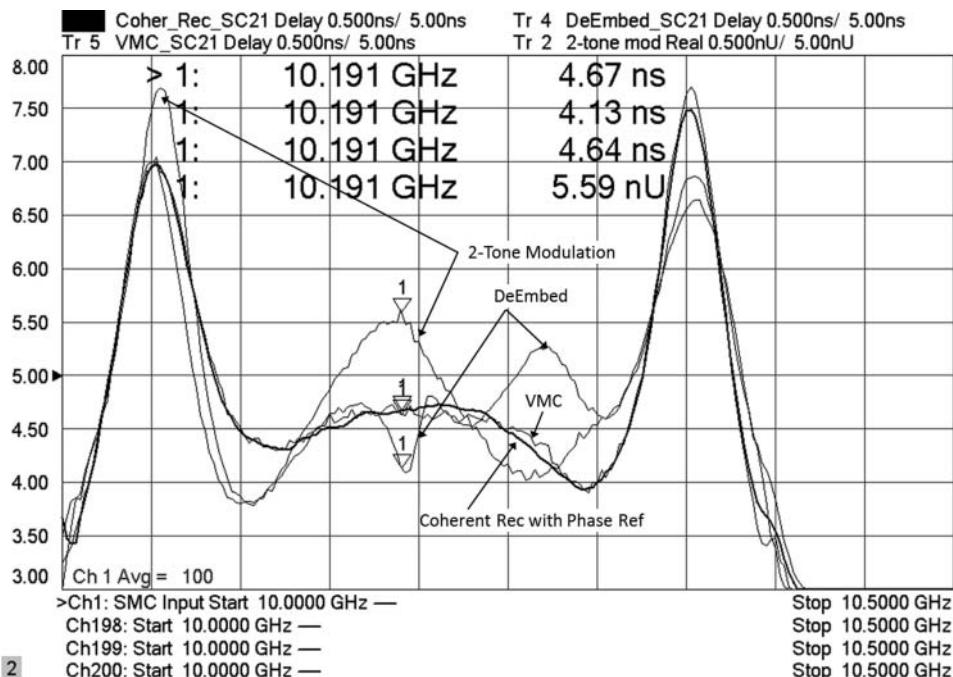


Figure 7.28 Comparison of group delay responses for various methods, each offset by one division.

if the LO frequency changes, a new calibration is required, as well as a new characterization for the cal-mixer.

A comparison of the computed group delay for each of the measurement methods is shown in Figure 7.28. For each method except the modulation method, the delay aperture was two times the point spacing (three-point aperture) as it is the minimum aperture that does not skew the data by one-half a measurement bucket. The modulation method has an inherent delay aperture of the tone spacings or AM modulation frequency. The calibration methods vary for the measurements; the calibration method for de-embedding and VMC utilize the same characterized reciprocal-mixer, de-embedding in the first case and as a cal-mixer in the second. The coherent receiver uses a phase reference calibration (described in Section 7.5.3.3), and the two-tone modulation uses normalization to a cal-mixer similar to the VMC method. All methods except the two-tone method provide for some mismatch correction at the ports, which likely explains the excess ripple shown in the two-tone method. Adding attenuator pads at the port can reduce this effect.

With any of the modulation methods, the phase response can be computed by integrating the group delay response. This gives a reasonable result for phase deviation but cannot be used for any fixed frequency measurements such as phase versus RF drive or phase versus LO drive, as the compression of the phase at each frequency will be similar, and so no change will be apparent in the difference of the phases at the output. Modulation methods also cannot be used in looking at the phase response of a pulse profile applied to a mixer for similar reasons.

7.4.5 Swept LO Measurements

In some cases, the mixer characteristics are defined over a range of RF or LO frequencies, with a common or fixed IF frequency. As described in the introduction of this section, in most cases the fixed IF measurement really refers to the final use case of a converter that has a fixed IF output channel, and the RF-to-IF transfer function is to be determined for each of a set of stepped LO frequencies representing the different channels of operation. In these cases, it is convenient to show the overall effect of RF or LO frequency on the nominal response (usually taken at the center frequency of the IF), so a quasi-swept LO measurement is desired. For most measurement systems, including VNA-based systems, the only change needed is to specify that the IF as a fixed frequency and the RF and LO as swept. One such implementation is shown in Figure 7.29. In older VNAs, it may not be possible to directly control an external LO; in such a case, the LO cannot be stepped synchronously with the RF, and instead a series of CW stepped measurements must be made using some external programming control. This is a common measurement on bare mixers, to allow a simple characterization of the response of each of the LO and RF ports.

A measurement of a wideband triple-balanced mixer with a different combinations of swept LO, swept RF, and swept IF over a wide range of frequencies is shown in Figure 7.30. In the upper windows, the roll-off of the low-frequency response of the mixer conversion gain is due to the LO or RF balun used in the mixer having a limited low-frequency response. By measuring $RF > LO$ (right upper) and $RF < LO$ (left upper), one can infer that the RF corner is lower than the LO corner.

Independently fixing the RF, the LO, and the IF while sweeping the other inputs is one way to determine the bandwidth limitation of the various ports of a mixer. For example, if the LO is held constant while sweeping the RF and IF, one can determine the combination of RF and IF bandwidth characteristics. One can set the RF and IF range at a particular LO frequency and normalize the conversion gain response. After each sweep, the LO can be set higher or lower, keeping the IF sweep constant and thus changing the RF range by the change in the LO. If the entire trace response drops, that indicates a limitation of the LO bandwidth; if the conversion gain changes versus frequency, this indicates a limitation due RF range. A similar method can be used to evaluate the IF range and the LO range. Figure 7.30, lower window,

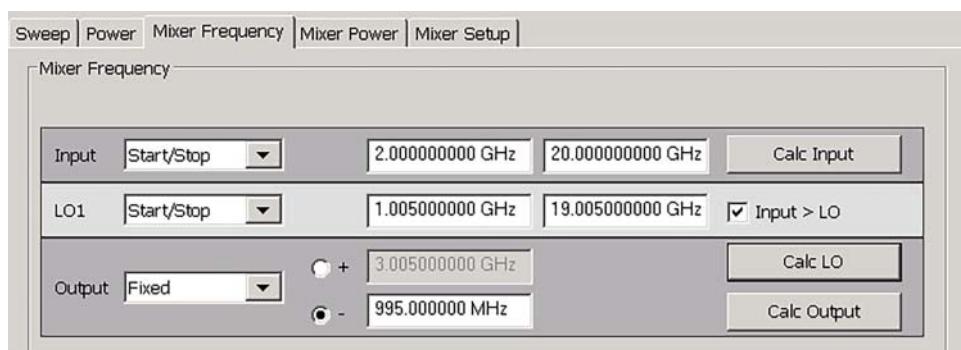


Figure 7.29 Defining a swept LO measurement using a GUI.

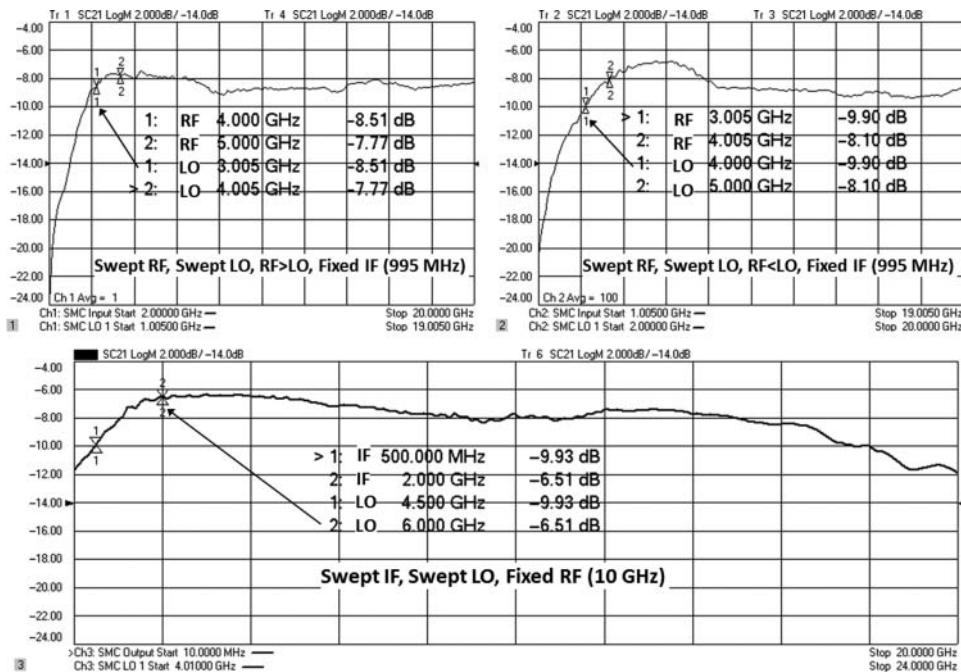


Figure 7.30 Fixed IF, swept RF/LO measurement of a mixer for two different IF frequencies.

shows the effect of changing the IF frequency on the conversion gain. The low-frequency corner moves with a change in LO and IF, with a constant RF range. Since the corner is at 2 GHz IF, but 6 GHz LO, one can presume it is limited by the IF balun.

7.4.5.1 Phase Measurements for Swept LO: Beam Forming and Radar System Matching

In some cases, particularly for mixers used in radar or phased-array applications, it is desired to understand phase response of the IF path as the RF and LO are swept, including deviation from linear phase. That is, in these systems it is desired that the phase response apparent to the IF be identical for each of several converters. In operation, a comparison of the IF magnitude and phase is used for a variety of tasks including detection of objects and direction finding. It is through processing of these responses that tasks such as electronic beam forming are accomplished. The great difficulty in determining this response is that any phase-change to either the RF or LO path will cause an apparent phase shift in the fixed IF. Often, linear phase change (such as adding a line length to one of the paths) is of less concern than deviation of linear phase, but in some systems, the absolute phase-response measurement between several converters is required. New methods make it much easier to perform channel-to-channel phase difference measurements, such as using the “Differential and IQ” application of the Keysight PNA (option 089), which allows direct control and measurement of magnitude and phase of signals at different frequencies.

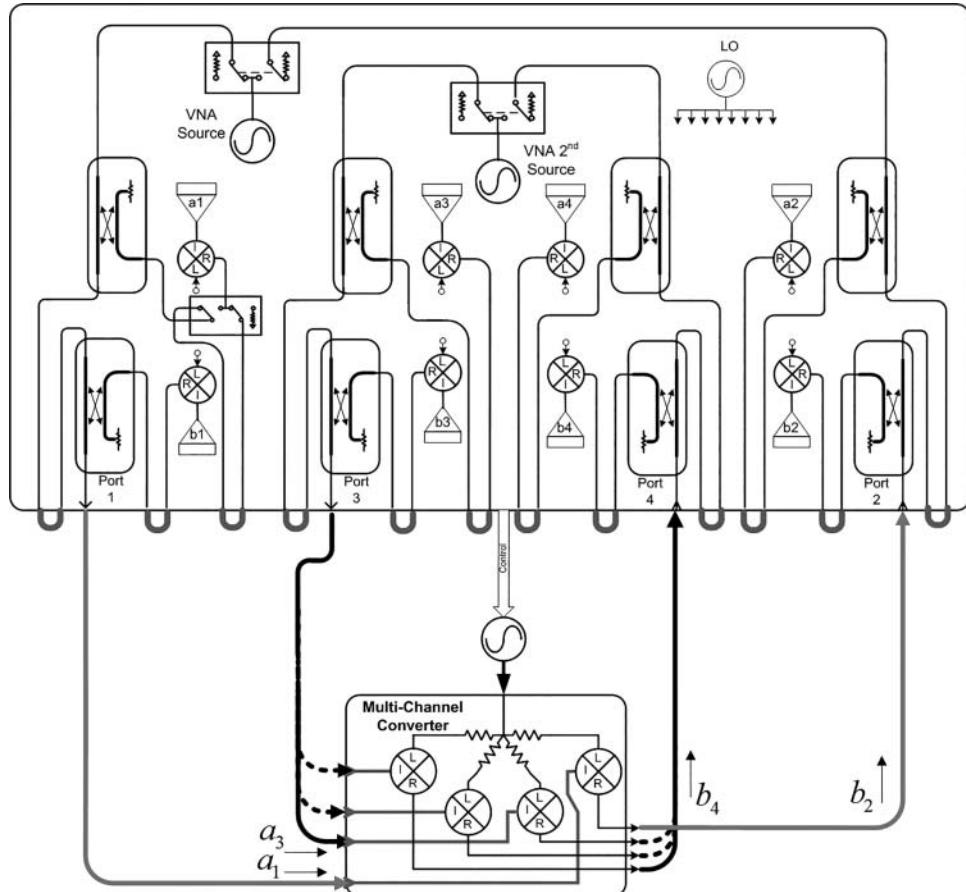


Figure 7.31 Measuring phase response of multichannel fixed IF converters.

Consider the system shown in Figure 7.31. Here we see a multichannel converter, set up to be measured two channels at a time. To accomplish the measurement, the signals at both ports 1 and 3 are turned on, and the ratio of signals a_1/a_3 gives the relative amplitude and phase of the input signal. At each test frequency, the receivers at ports 2 and 4 are tuned to the desired output frequency, and the ratio of b_2/b_4 is measured. Finally, the ratio of the input to the output is computed as

$$\text{Phase_Difference} = \frac{(b_2/b_4)|_{\text{OutputFreq}}}{(a_1/a_3)|_{\text{InputFreq}}} \quad (7.15)$$

The setup configuration for this measurement is shown in Figure 7.32. In this case, the input is a fixed frequency, with a swept LO and swept RF output. We want to measure the phase difference between two distinct paths through the converter; in some cases, this difference is recorded and used to compensate the relative phase of the IF input as to compensate for path differences. In other cases, there is simply a specified maximum path difference that must

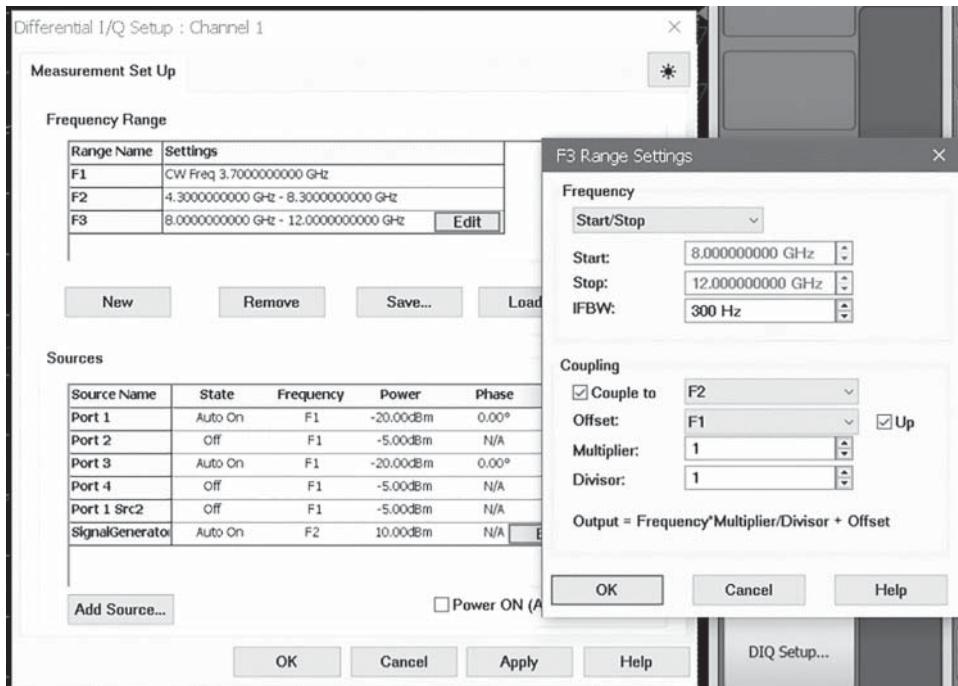


Figure 7.32 Setup for swept LO phase-difference measurement.

be verified. Here the measurement application allows one to set the actual phase difference between the two inputs. In many cases, the actual phase difference does not matter, but for some devices it is desirable to set the phase to a prescribed value. In the case of this example, the two phases are set to be the same, at 0° .

The calibration of this system is relatively simple and relies on a single power-meter-based source calibration, as well as S-parameter calibrations at the IF input frequencies, LO frequencies, and RF output frequencies. Since the measurement determines phase differences at the input, over the input frequencies, and over the output at the output frequencies, there is no need to make cross-frequency phase measurements. Note, the first edition of this book described a method that required multiple VMC calibrations and was dependent on phase characteristics of the calibration mixer; this method requires no calibration mixer.

Figure 7.33 shows a comparison of the phase response of two channels of a multichannel mixer used in a swept-LO, fixed-IF up-converter, as configured in Figure 7.32. The upper plot shows the gain of each path, SC_{21} path and SC_{43} path. The two gains of the matched mixers are practically the same.

The lower plot shows the path-to-path phase difference between the SC_{21} path and SC_{43} path, as a function of the swept RF (and swept LO) frequency. Here the input IF is fixed and both the RF and LO sweep. The relative phase measurement is stable and accurate, because it is measured on each path simultaneously and is not affected by phase linearity of

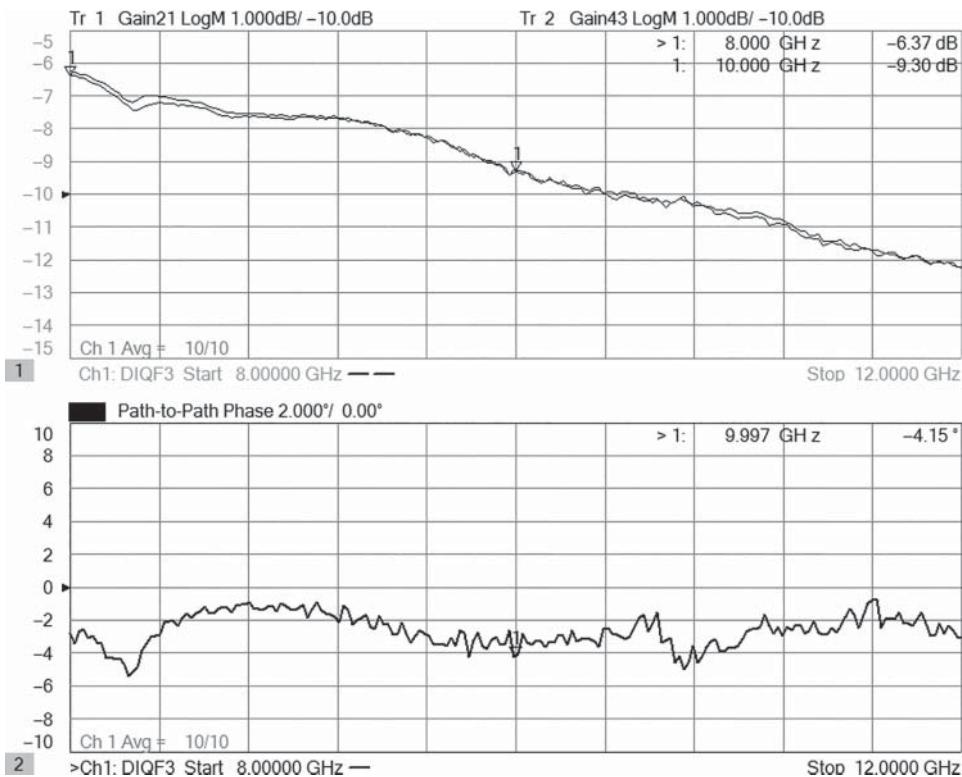


Figure 7.33 Gain comparison and path-to-path phase difference of a multichannel up-converter.

the external LO. But a change in the internal LO path would show a change the path-to-path response. For this multichannel mixer, the path difference has about a 2° fixed offset and about 4° of peak-to-peak variation.

7.4.5.2 Absolute Phase Measurements of Mixers

The ultimate solution would allow a single converter phase response to be acquired so that it can be matched with other converters in a system. But this would require acquiring the phase of the LO in addition to the RF-to-IF phase response; in fact, some of the latest non-linear vector network analyzer (NVNA) systems do acquire RF, LO, and IF amplitude and phase response, so long as the signals remain “on a grid” of a common phase reference base frequency (typically 10 MHz); that is, all of the RF, LO, and IF frequency points must be related by a rational number. But NVNA systems can be a little complicated to set up and calibrate, and the measurement time can be slow. Fortunately, the same application used in the previous section, “Differential and IQ,” can be used in a way to make, for the first time, full absolute phase measurements on a mixer without an NVNA system, by adopting the key component

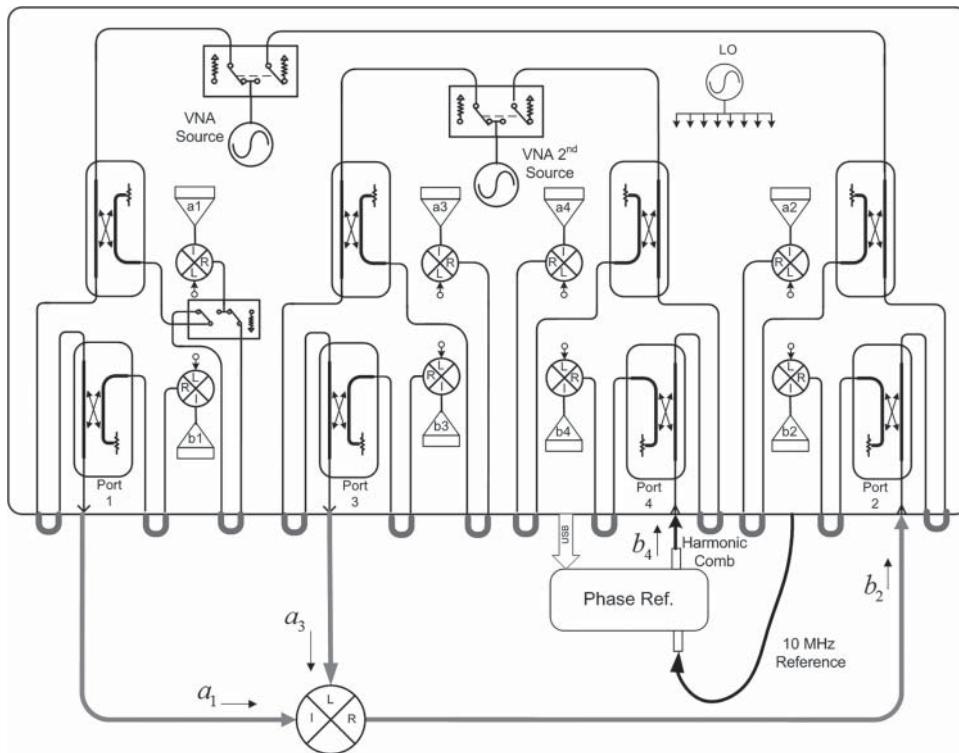


Figure 7.34 Measuring the absolute phase response of a mixer.

of the NVNA: the harmonic comb-generator phase reference. Figure 7.34 shows the setup diagram for making absolute phase measurements on a mixer.

The user setup for such a measurement is similar to that shown in Figure 7.32, but in this case the source from port 3 is set to be the LO frequency, and an external source is not needed. The frequency range and spacing used must have a common factor, and the comb-generator must also be driven at that common factor. The comb-generator in this case is a Keysight U9391C. This produces a comb or set of harmonics that all have the same phase, namely, 180°. In fact, it is really a pulse generator that produces a very narrow, well-defined, negative pulse (in the time domain) that has a comb-like spectral content. The comb-generator can use input frequencies from 10 MHz to 6 GHz and has an internal divider that provides divide ratios of 2, 4, 8, and 16. Thus, if it is driven at 16 MHz with a divide ratio of 16, it will produce a common-factor of 1 MHz. For most operations, the comb-generator is driven by the VNA 10 MHz reference, and it is common to use 10 MHz as the frequency step.

The measurement of the mixer must utilize measurements, magnitude and phase, of the two input signals, IF at port 1 and LO at port 3, as well as the output signal, at port 3, and the phase reference signal at port 4. For each setting of the source and LO frequency, a set of receiver measurements are performed at the IF, RF, and LO frequencies. The key to obtaining the phase response is to utilize the measurement of the phase reference signal to provide a known cross-frequency phase standard at port 4, which is used to normalize the phase of each

of the other ports. The value of SC_{21} is computed, in the case of a normal mixer (where the LO is less than the RF), from these receiver measurements as

$$SC_{21} = \left. \frac{(b_{2_Fout}) \cdot \left(\frac{|b_{4_Fout}|}{b_{4_Fout}} \right)}{(a_{1_Fin}) \cdot \left(\frac{|b_{4_Fin}|}{b_{4_Fin}} \right) \cdot \left(\frac{a_{3_FLO}}{|a_{3_FLO}|} \right) \left(\frac{|b_{4_FLO}|}{b_{4_FLO}} \right)} \right|_{LO < RF} \quad (7.16)$$

From Eq. (7.16) we can see that the ratio of the magnitude of the b4 receiver to the b4 complex value is the inverse of the phase of the comb-generator where

$$\phi_{CombGenerator} = \left(\frac{b_{4_Fxx}}{|b_{4_Fxx}|} \right) \quad (7.17)$$

represents the phase of the phase-reference measured at the b4 receiver. The phase reference has almost perfect phase response, with each frequency having a phase of almost exactly 180° . Normalizing each receiver, at its appropriate frequency, with the phase reference removes the phase ambiguity associated with measuring at different frequencies. For the case of an image mixer, there is small change, where SC_{21} becomes

$$SC_{21_Image} = \left. \frac{(b_{2_Fout}) \cdot \left(\frac{|b_{4_Fout}|}{b_{4_Fout}} \right)}{(a_{1_Fin}) \cdot \left(\frac{|b_{4_Fin}|}{b_{4_Fin}} \right) \cdot \left(\frac{a_{3_FLO}^*}{|a_{3_FLO}|} \right) \left(\frac{|b_{4_FLO}|}{b_{4_FLO}} \right)} \right|_{LO > RF} \quad (7.18)$$

where we must take the conjugate of the LO to properly account for the inverting nature of the response.

The calibration of this setup is straightforward, requiring only a power calibration of 1-port and an S-parameter calibration of the four ports. The power calibration gives an absolute power reference and the S-parameter calibration allows all the other receiver response terms to be computed from the power calibration and the appropriate ratio response term. The phase calibration is inherent in the phase reference (note: the phase reference does contain characterization data of the phase response, but experience shows that the error in the raw performance of the phase reference is on the same order as the uncertainty of its measured value).

Figure 7.35 shows the response of measuring a bare mixer for magnitude and phase (upper left), along with measurements where an approximately 100 ps delay line is added to the LO path (upper right), the RF path (lower left), and the IF path (lower right). Note that since this is a measure of the phase response with a swept LO, taking the derivative will *not* give the delay response, from IF to RF, as the LO contribution will distort the result. However, if the setup was for a fixed LO case, where the IF and RF were both sweeping, then the resulting delay response would be a valid measurement, properly calibrated. Here we can see that the true mixer behavior is measured, and the new mixer, comprised of the mixer plus the short delay line, shows the proper and expected phase response, relative to original bare mixer. For the IF delay (lower right), the phase shift is constant across the sweep (since the IF frequency is fixed) and is only about 120° . For the case where the delay is added as part of the mixer LO, the delay shifts more as frequency increases, because the phase-shift from delay in the

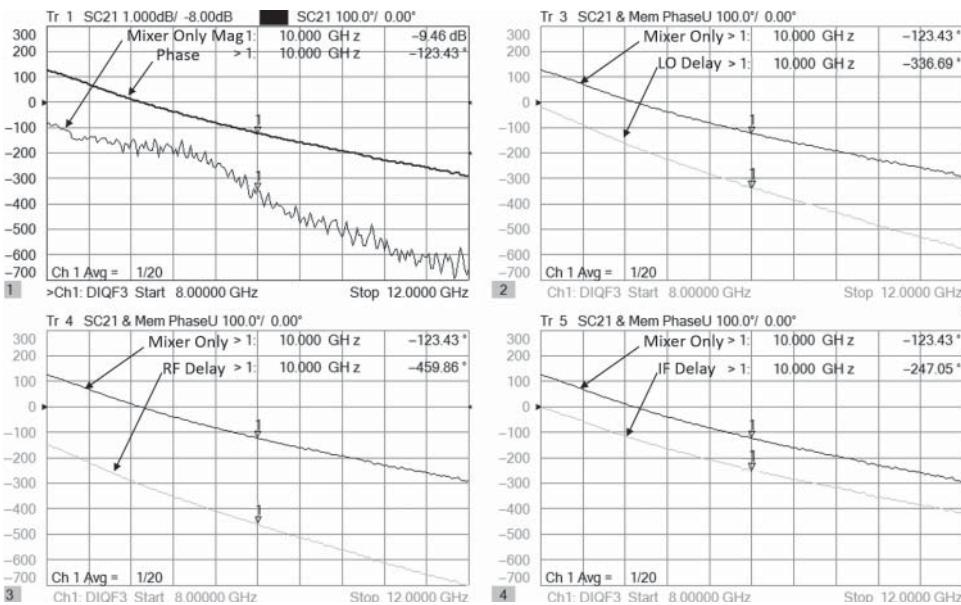


Figure 7.35 Absolute phase response of a mixer under swept LO conditions.

LO port increases the mixers phase response as the LO frequency is increased. Finally, the delay added to the RF (output) port shows even greater phase shift, because the delay effect at the output of the mixer is at an even higher frequency. With this method, individual mixers may be independently measured and phase matched.

7.5 Calibration for Mixer Measurements

Each of the measurement methods for mixers can have one or more calibration methods applied. For amplitude response, the most direct and accurate measurement is using a power calibration on scalar-mixer-calibration (SMC) measurements, as described next. Accuracy of the down/up method and the VMC method, for both magnitude and phase, depend upon the accurate characterization of a calibration or reconversion mixer.

For phase calibration, one must depend upon reference to calibrated mixer for most measurement methods; the lone exception is the coherent-receiver measurement that can be calibrated independently for power and phase, relative to a known phase reference (discussed in detail in Section 7.5.3.3). All other methods extract a phase-response from a presumably reciprocal mixer, through various methods, each of which is described in detail.

7.5.1 Calibrating for Power

The calibration for power measurements in a mixer utilizing the SMC methods are performed independently from the calibrations for phase and utilize methods identical to those described

in Section 3.7.2. Of course, the power calibration is needed for the reference receiver over the input frequencies and the test receiver over the output frequencies. The typical method for creating these power calibrations is to do a segmented enhanced power calibration, over both the input range and output range. This generates a complete calibration map for both the forward direction and the reverse direction, and some implementations of the SMC calibration method support measuring the full four S-parameters of a MUT including the forward and reverse conversion, as was demonstrated in Figure 7.17.

The calibration proceeds in two steps:

1. A power meter is connector to port 1 of the VNA to measure the incident power from the source at both the RF and IF frequencies. At the same time, the reference receiver is calibrated to the same level as the power displayed on the power meter.
2. A full 2-port S-parameter calibration is performed from port 1 to port 2 over the RF and IF frequencies. This acquires the input and output match correction terms, as well as the transmission loss between the two ports.

From these two measurements, a complete calibration of the VNA for measuring power at all ports is computed, as described next. Any connectors or adapters used between the power meter and VNA ports can be removed by an additional 1-port calibration at the power meter reference. This allows one to use a coaxial power meter for the power calibration step and then use any other calibration method, including on-wafer calibration, for the S-parameter step. The mathematical details are as follows: the mathematical notation changes slightly for the mixer calibration as the responses of the inputs and outputs are at different frequencies, so the tracking terms must be associated with the particular ports' frequency of interest. The input frequency power correction is performed as

$$a_{1A_MatchCor} = \frac{a_{1MUT_In}}{RRF_{In} \cdot (1 - ESF_{In} \cdot \Gamma_{1MUT})} \quad (7.19)$$

where $RRF_{In} = \frac{a_{1M_Cal_In}}{P_{Meas}(1 - ESF_{In} \cdot \Gamma_{PwrMeter})}$

Here, the ESF term is determined during a 1-port or 2-port calibration performed at the input frequencies.

The output power correction is similarly determined over the output port frequencies as

$$b_{2A_MatchCor_Out} = \frac{b_{2MUT_Out}}{BTF_{Out}} \cdot (1 - ELF_{Out} \cdot \Gamma_{2MUT_Out}) \quad (7.20)$$

where $BTF_{Out} = ETF_{Out} \cdot RRF_{Out}$

Here the BTF term is computed based on the transmission tracking from input to output, over the output frequency range, and the reference receiver tracking, also over the output frequency range, computed as in (7.19) but for output frequencies.

The correction for conversion gain is computed as

$$SC_{21_Cor} = \frac{b_{2MUT_Out}}{a_{1MUT_In}} \cdot \frac{RRF_{In}}{BTF_{Out}} (1 - ESF_{In} \cdot \Gamma_{1MUT})(1 - ELF_{Out} \cdot \Gamma_{2MUT_Out}) \quad (7.21)$$

This applies to standard mixers, but a small change is necessary when dealing with image mixers, to account for the effect of frequency and phase reversal. For image mixers, the correction is

$$SC_{21_Cor} = \frac{b_{2MUT_Out}}{a_{1MUT_In}} \cdot \frac{RRF_{In}}{BTF_{Out}} (1 - ESF_{In} \cdot \Gamma_{1MUT}) (1 - ELF_{Out}^* \cdot \Gamma_{2MUT_Out}^*) \quad (7.22)$$

Thus, the gain computation for a mixer and an amplifier are very similar, with the difference being only in a small portion of the mismatch correction term associated with the loop-term $SC_{21}SC_{12}ESF_{In}ELF_{Out}$. Unlike low-loss lines, mixers always have either an associated loss or an amplifier with isolation, which makes the loop term a very small effect.

7.5.1.1 Split-Port Cal for Amplitude Response

In some cases, it is not possible to perform the S-parameter thru step between the input and the output of a SMC calibration, usually because one of the ports is a connector (such as waveguide) that does not pass both the input and output frequencies. If the input connector is waveguide, then the BTF at the output cannot be computed. (On the other hand, if the output is a waveguide connector and the input can pass the waveguide frequencies, then it is possible to add an adapter to make an unknown through calibration.) One way to deal with incompatible connectors is to remove the waveguide adapter from the port requiring a waveguide connection to the mixer, and then perform the calibration entirely in coax, before replacing the adapter and de-embedding the adapter. In the past, this was the only method available for dealing with mixers that have a waveguide port. However, in the case of testing at mm frequencies, the waveguide test port is often a mm-wave extension head (sometimes called *mm-wave multipliers* or simply *mm-wave heads*). In this case, the mm frequency is generated in the head, and it is not possible to remove it to get to a coax port or to calibrate it for a mixer measurement with normal methods.

A new method, sometimes called *split-cal*, allows a mixer calibration by splitting the power and receiver calibration into two separate power and reflection measurements. The basic concept is to perform a power meter calibration and a 1-port calibration on each port. The port 1 power meter and reflection calibration allows the computation of the input power correction exactly as indicated by Eq. (7.19). The new technique then applies the same steps to the output port, at the output frequencies, as well to find the Reference Response Reverse tracking term, RRR_{Out} (the tracking response of the a_2 receiver), over the output frequency range as

$$RRR_{Out} = \frac{a_{2M_Cal_Out}}{P_{2_Meas}(1 - ESR_{Out} \cdot \Gamma_{PwrMeter_Port2})} \quad (7.23)$$

From these measurements, the output $b2$ response tracking term is computed as

$$BTF_{Out_SplitCal} = RTR_{Out} \cdot RRR_{Out} \quad (7.24)$$

The error correction math becomes

$$b_{2A_MatchCor_Out} = \frac{b_{2MUT_Out}}{BTF_{Out_SplitCal}} \cdot (1 - ESR_{Out} \cdot \Gamma_{2MUT_Out}) \quad (7.25)$$

The overall correction for SC_{21} gain becomes

$$SC_{21_SplitCal} = \frac{b_{2MUT_Out}}{a_{1MUT_In}} \cdot \frac{RRF_{In}}{BTF_{Out_SplitCal}} (1 - ESF_{In} \cdot \Gamma_{1MUT}) (1 - ESR_{Out} \cdot \Gamma_{2MUT_Out}) \quad (7.26)$$

Here the source-match replaces the load match, as it is not possible to measure the load match during the calibration process, because there is no way to supply a signal from another port. An alternative method could use the MUT measurement of a_{2_Out}/b_{2_Out} to determine the load match during the active measurement and correct for it in that way. However, in many cases, the difference between the source-match and load match are very small, especially if there is any loss between port 2 DUT connection and the reference coupler.

7.5.2 Calibrating for Phase

In the late 1990s, a few new techniques for mixer phase measurements were developed. In one example, three mixers were used in three different measurements to extract the behavior of each mixer. In another, the response to reflections at a mixer output formed the basis of calibration. Recently, new methods based on non-linear VNA techniques have been developed that promise to widely replace other methods due to the simplicity and flexibility of the calibrations. The details of each method and best practices of using the new methods are described next.

7.5.2.1 Three-Mixer Method

One of the first methods proposed to characterize the phase response of a mixer using VNA measurements made use of three measurements on pairs of mixers. This method is used to find the phase of a calibration mixer. Because it is a multistep approach, it is not very suitable for characterizations of mixers in a normal test sense. This method is illustrated in Figure 7.36. In this method, the MUT is Mixer C. Mixer B is another mixer that can up-convert the signal from Mixer C to the same input frequency as Mixer C. Mixer A is a *reciprocal-mixer* that can be used both as a replacement for the MUT mixer (in the first line of the illustration) or as a replacement for Mixer B (in the second line of the illustration).

Since the input and output are at the same frequency, full 2-port calibrated S-parameter measurements can be used to determine the gains in the three cases, G_1 , G_2 , and G_3 respectively. The overall response for each gain (excluding mismatch effects) is

$$\begin{aligned} G_1 &= SC_{21_A} \cdot S_{21_IF} \cdot SC_{21_B} \\ G_2 &= SC_{21_C} \cdot S_{21_IF} \cdot SC_{12_A} \\ G_3 &= SC_{21_C} \cdot S_{21_IF} \cdot SC_{21_B} \end{aligned} \quad (7.27)$$

where S_{21_IF} is the loss of the IF filter. If the image filter is not used and the mixer creates nearly equal image signals, the overall gain either can be doubled (if the image adds in-phase) or can go to zero (if the image is out of phase and exactly equal to the standard conversion signal). This is an unacceptable error and must be reduced by using the image filter.

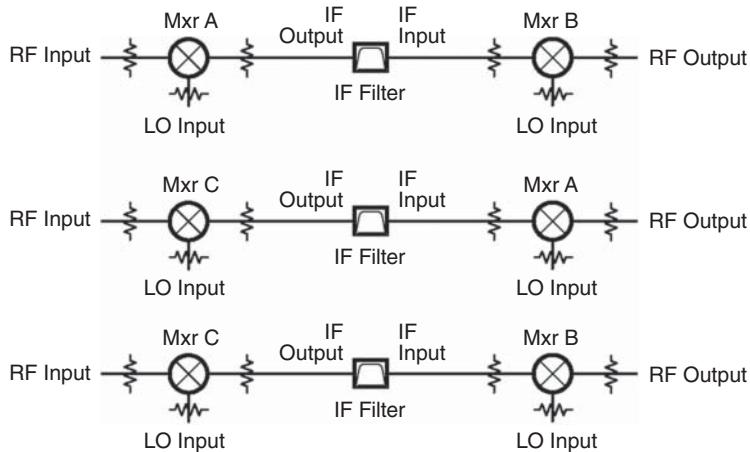


Figure 7.36 Three-mixer method of for measuring mixer phase on a VNA.

From Eq. (7.27), the conversion gains of any of the mixers are computed as

$$\begin{aligned} SC_{21_A} &= \sqrt{\frac{G_1 G_2}{G_3 S_{21_IF}}} \\ SC_{21_B} &= \sqrt{\frac{G_1 G_3}{G_2 S_{21_IF}}} \\ SC_{21_C} &= \sqrt{\frac{G_2 G_3}{G_1 S_{21_IF}}} \end{aligned} \quad (7.28)$$

Since all the gains are complex numbers, the phase response of the mixer can be computed as

$$\phi_{21_C} = \frac{\phi_2 + \phi_3 - \phi_1 - \phi_{21_IF}}{2} \quad (7.29)$$

In the computation of the square root function for a phase response, care must be taken to choose the appropriate root (for details, see Section 11.3.1.1).

While this is a straightforward method of determining the phase of a mixer, it does require several steps and requires at least one mixer to be reciprocal. Further, in the method described, the gain excludes the effects of mismatch between the mixers and the IF filter. If the match of each were measured independently, it would be possible to include some of the mismatch effects, but typically a set of attenuator pads is used to lower the mismatch between the mixers. In such a case, the attenuator pad loss is lumped into the filter loss. And of course, one of the major difficulties is that three mixers must be used, and it might be non-economical to create three mixers, especially at high frequencies.

An example of applying the three-mixer technique is shown in Figure 7.37. Mixer A in this example is reciprocal and will be used as the cal-mixer in examples of the next section as well, and was extracted as the de-embedding mixer in Figure 7.19, as described in Section 7.5.3.1.

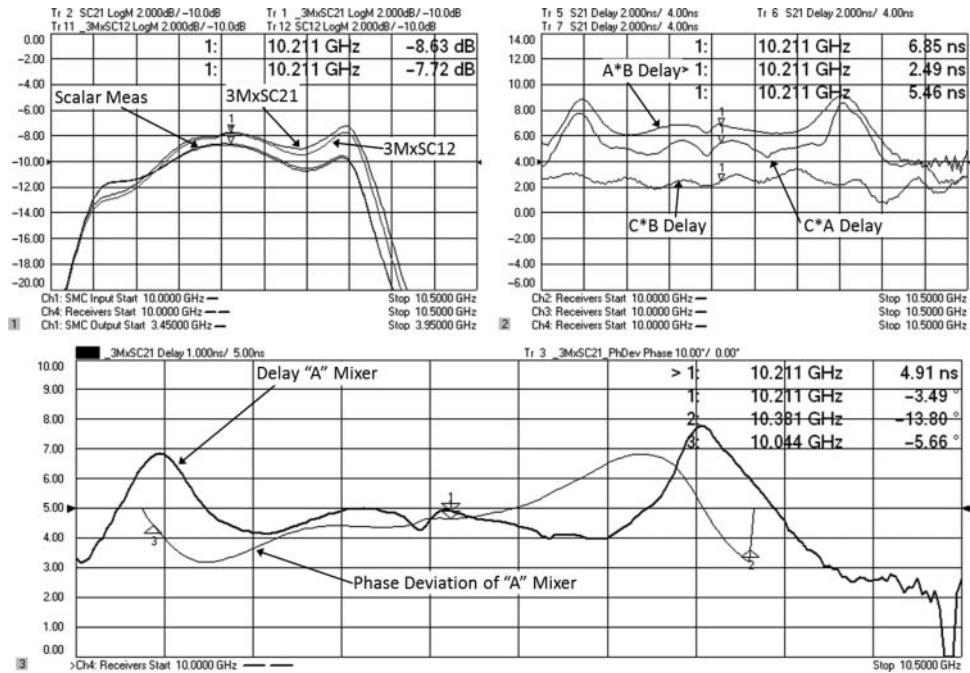


Figure 7.37 Amplitude response of Mixer A measured using SMC and three-mixer method (upper left), delay of three mixer pairs (upper right), phase and delay of just Mixer A (lower).

It includes the IF filter so that it is not necessary to add an additional IF filter or compensate for its S_{21} . Rather than showing phase for the mixer pairs, the delay of each mixer pair is shown instead, with the understanding that it was derived from the phase response for each. The delay provides a nicely intuitive feeling for the relative response of each mixer.

In the figure, the upper-left widow displays the magnitude result of a scalar measurement of the SC_{21} and SC_{12} of the “A” mixer. Also shown is the extracted up and down-conversions, using the three-mixer approach. It is clear that there is some substantial error (on the order of 2 dB or more) in this approach. The upper-right plot shows the group delay for each of the mixer-pair measurements. Mixer “A” is strongly filtered, and the “B” and “C” mixers are not so strongly filtered, and they show a flat response when measured as a pair. The lower plot shows the group delay of SC_{21} extracted for mixer “A” with the three-mixer method, as well as the deviation from linear phase (restricted to the passband region of the mixer). The maximum deviation here measures about $\pm 18^\circ$. It is interesting to note that a mid-band ripple in the delay trace is apparent here, just as in the up/down method of Section 7.4.3.1. The three-mixer approach is just as susceptible to high-order products affecting the result as the down/up approach.

From these results, any of the mixers responses can be computed, and then any of the mixers, with their associated data, could be used as calibration mixer. However, the ripple in the delay measurement suggests that there are some errors in the characterization. In the next two sections, improved methods for mixer characterizations are described, which don’t show these error effects.

7.5.3 Determining the Phase and Delay of a Reciprocal Calibration Mixer

Several of the mixer measurement methods depend upon using a known mixer as a calibrated thru standard. The difficulty is in determining the qualities of the known mixer. One method utilizes reflections from the output of the mixer to infer the two-way response of the mixer and then presumes the mixer is reciprocal to compute the one-way response. A similar approach is used in a second method that is related to the unknown thru calibration.

7.5.3.1 Reciprocal Cal Mixer Reflection Method

In the reflection method for characterizing a mixer, the presumption is that the conversion loss in the forward and reverse directions are the same, and not too great. In this method, a filter must be added to the output of the mixer that reflects or absorbs the undesired image and passes the desired product with low loss. A succession of reflection standards is placed on the output of a mixer/filter pair, and the input match is measured for each reflection standard. Essentially, a 1-port calibration is performed at the output of the mixer/filter combination, as illustrated in Figure 7.38.

In this case, the IF⁻ signal is the conversion that is desired; the IF⁺ signal (from the sum of RF and LO) will reflect off the IF filter rather than passing through and so will not appear at the mixer output. For up-converting mixers, the filter should pass the IF⁺ signal and reject the IF⁻ signal.

From this one can see that for each standard, there are up to four major contributors to the overall reflection response: Reflection from the mixer S_{11} of the RF signal (marked as RF), reconverted reflection from the IF⁺ signal (which reflects off the stop band of the filter, marked IF⁺), reconverted reflection off the cal standard (marked $IF^- \cdot \Gamma_{Short}$, for example), and reconverted re-reflected signal off the cal standard and S_{22} of the cal

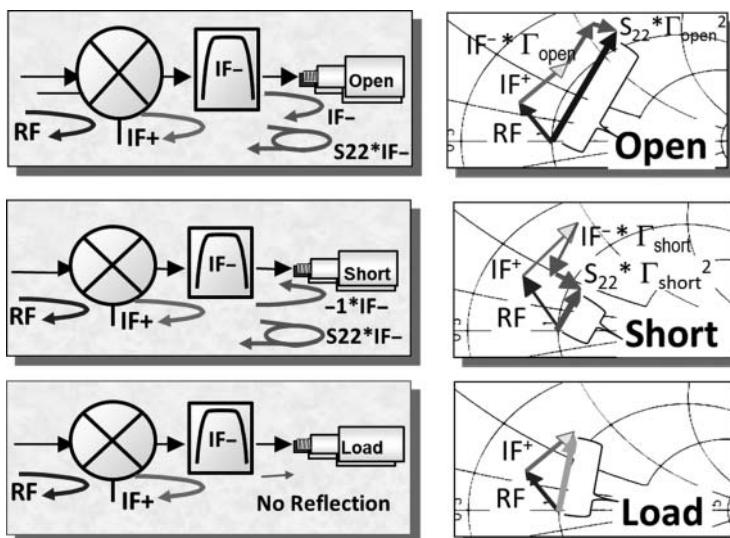


Figure 7.38 The reflection method of mixer characterization is essentially a one port cal at the output.

mixer/filter combination (marked $S22 * \Gamma_{Short}^2$, for example). These are illustrated as vectors in the right half of the figure. If a 1-port calibration performed using each of the standards, the resulting error terms can be mapped to the mixer response. For each standard, the S_{11} and IF+ reflections are unchanged, and in the construction of a 1-port cal, they represent the directivity error term; this is S_{11} of the mixer. The two-way transmission through the mixer is essentially the average of the open and short response and so represents the reflection tracking term, or $SC_{21} * SC_{12}$. And the difference between the open and short is the source-match term of the 1-port cal, or S_{22} of the mixer.

In practice, this is measured with a non-ideal network analyzer, so if a 1-port calibration is performed before the mixer and a second 1-port calibration is performed after the mixer, adapter characterization math of the equation (see Chapter 11 for details) can be directly applied to yield

$$\begin{aligned} S_{11_MUT} &= \frac{(EDF_{MUT} - EDF)}{[ERF + ESF \cdot (EDF_{MUT} - EDF)]} \\ S_{21_MUT} = S_{12_MUT} &= \frac{\sqrt{ERF \cdot ERF_{MUT}}}{[ERF + ESF \cdot (EDF_{MUT} - EDF)]} \\ S_{22_MUT} &= ESF_{MUT} + \frac{ESF \cdot ERF_{MUT}}{[ERF + ESF(EDF_{MUT} - EDF)]} \end{aligned} \quad (7.30)$$

Here again one notes the presence of a complex square root term, and the same cautions apply as in the example of Eq. (7.28). In the case of the second 1-port calibration, after the cal-mixer, the calibration kit terms must be used for the output frequencies and not the normal values of the input frequencies. If an Ecal is used, then the values of the Ecal must be taken from the output frequency range; in general, this capability is not readily available to the user, so one must depend upon the manufacturers' embedded functions to perform a mixer characterization using an Ecal.

Some notes on this method: the mixer needed is essentially the same as the Mixer A in the previous section, that is, a mixer that is reciprocal. In this case, the IF filter is embedded in the mixer response, so the mixer/filter combination is always considered as a whole. Any reconversion from the undesired image is captured in the overall response; thus, the response may have substantial ripple, but as long as the true value of the ripple is captured in this characterization, the actual value of the ripple is of no importance. An example of a mixer characterized in this way is shown in Figure 7.39 and compared with the previous three-mixer method.

In the figure, the upper-left plot shows the SC_{21} amplitude comparison; the SC_{21} extracted using the reciprocal mixer is almost identical to the SC_{21} response using SMC calibration, with the three-mixer method having the largest difference. In the upper-right plot, the deviation from linear phase is shown; the reciprocal method shows a nearly ideal response to the original filter shape. Similarly, in the lower plot, the delay of the reciprocal extraction method shows a smooth and symmetric delay response as would be expected from the filter used, compared with the three-mixer method.

Following the illustration of Figure 3.46 showing phase error as a function of amplitude error, one might postulate that if the amplitude response of a mixer is reciprocal within some limit, the phase response is similarly reciprocal within the limits proposed by Eq. 3.103. Thus, the SMC measurement of amplitude of SC_{21} and SC_{12} gives an idea of the quality of

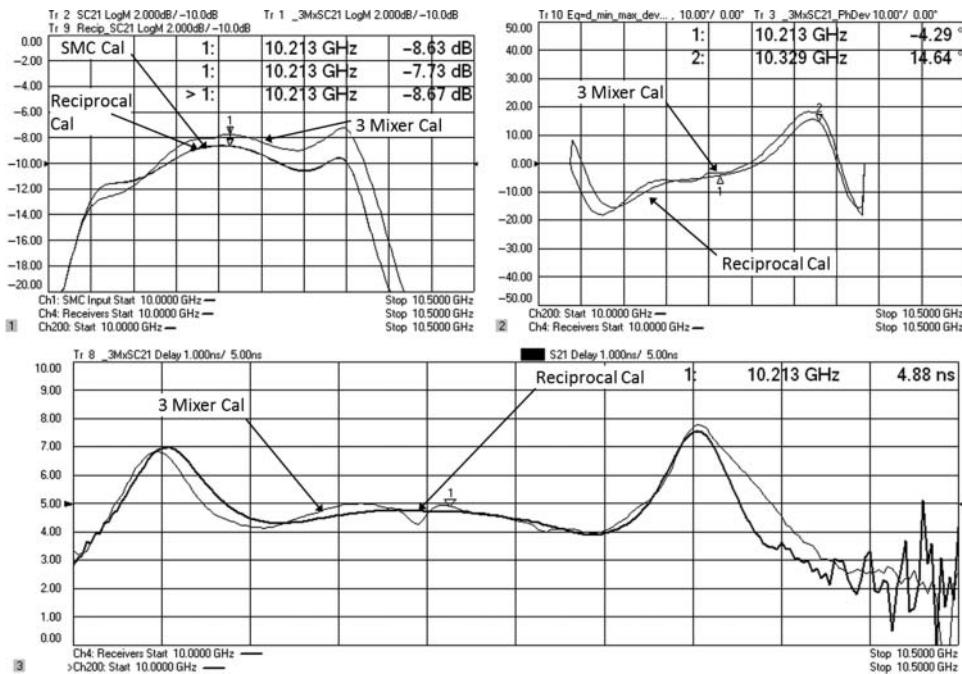


Figure 7.39 Mixer characterized using the reflection method.

the reciprocal calibration. In this case, the amplitude reciprocity error is less than 0.09 dB, implying the phase reciprocity is less than 0.6°.

Some notes on this characterization method: since the reflection off the mixer input is often the largest component of the signals measured and the method requires this value to be consistent between standards, it is critical that there is no drift or instability in the measurements of S_{11} of the mixer. This characterization method works best when the cal-mixer is connected directly to the VNA test port, with *no* intervening test cable. Also, loss in the cal-mixer will increase this effect; for one-way conversion loss less than 10 dB, there is usually no problem. If the loss is between 10 and 15 dB, extra care is needed, with lower IF bandwidths, and more averaging is necessary to obtain a good result. This method generally does not yield good results if the one-way loss is greater than 15 dB. This is evident in the noisy response of the delay in Figure 7.39 in the out-of-band region. In such a case, the method of the next section may be used. Since stability is important, using an Ecal to generate the 1-port calibration at both the input and the output removes the error associated with connector repeatability.

7.5.3.2 Reciprocal Cal Mixer Unknown Thru Method

An alternative method for characterizing a reciprocal calibration mixer takes as its basis the unknown thru calibration described in Chapter 3. For this method, a configuration is created that allows the signals from the MUT to be converted in such a way that the measurements of an unknown thru calibration can be performed. In one implementation, switched mixers in the forward and reverse reference paths provide a reconversion of the incident signal so a phase

measurement can be performed in each direction. In another implementation, a pair of mixers is added behind one of the test port couplers to reconvert the source and receiver signal at that port. In each of these instances, the LO oscillator must be split three ways to provide the same signal to each of the three mixers. The chief advantage of this method is that it performs better if the cal-mixer has greater than 15 dB insertion loss. However, because of the complexity of the setup, neither implementation of this method has seen wide acceptance. Interested readers are referred to reference material at the end of this chapter for further information.

All three of the previous methods required reciprocal mixers, and the characterization must be performed all over again for each new LO frequency. However, recently a new method of calibration for vector mixer measurements has been introduced, based on the coherent receiver test approach, that greatly simplifies the calibration process and does not require any mixers at all, other than the MUT that is eventually tested; this method is introduced in the next section.

7.5.3.3 Phase Reference Method

With the coherent-receiver measurement system described in Section 7.4.3.3, it is possible to measure the absolute phase change across a span of frequency on an input or output signal of a mixer. Thus, the magnitude and relative phase of the “a” and “b” waves of a mixer may be directly measured, and the mixer conversion response can be computed directly from Eq. (7.10), and from this we can see that phase of the output (RF) signal depends upon both the phase of the LO signal and the phase of the input signal (IF) as well as a mismatch term that depends on the reflected RF signal at the output. With the coherent-receiver system, it is possible to measure the b_{RF} and a_{IF} directly, as well as recognize that a_{RF} is defined by any reflection off the test port 2 loading the mixer. Thus, one can compute SC_{21} of the converter directly. Equation (7.10) also shows a dependency upon the phase of the LO. This dependency is not compensated for, but for fixed-LO converters, the phase of the LO is a constant offset and does not affect the phase or group delay response of the mixer with respect to the information bandwidth. From this one can compute the SC_{21} (in this case, of an up-converter) directly as

$$SC_{21} = \frac{(b_{RF}/BTR^{RF})}{a_{LO}(a_{IF}/RRF^{IF})} \cdot \left(\frac{1}{1 - S_{22}^{RF} ELF^{RF}} \right) \cdot \left(\frac{1}{1 - S_{11}^{IF} ESF^{IF}} \right) \Big|_{|a_{LO}|=1} \quad (7.31)$$

Because the phase of the LO is not discernible, a selected frequency point (typically the center frequency) is selected to be a reference phase point and is set to a constant value, usually 0° .

From Eq. (7.31) one notes that the reference-receiver response-tracking on the input and the test-receiver response-tracking on the output must be individually determined. From an understanding of standard S-parameter correction, one can recognize that

$$ETF^{RF} = \frac{BTF^{RF}}{RRF^{RF}} \text{ and } ETF^{IF} = \frac{BTF^{IF}}{RRF^{IF}} \quad (7.32)$$

Thus, if one can determine the phase response of either of the reference or the test channel over the RF and IF frequencies, then one can find the corresponding channel’s phase through the S-parameter transmission tracking term.

In previous correction methods, the calibration separated the magnitude response of the a and b receivers by measuring their responses independently using a power meter as a reference. Then a calibration mixer was added and the overall response of the system was measured. Independently, the source and load match of the system were measured, and finally a value for the phase of (BTF^{RF}/RRF^{IF}) was computed by solving (7.32) for a measurement of a known mixer above for that ratio. The amplitude was computed independently using the power meter calibrations. This method works well but requires knowledge of the calibration mixer used, and determining this is a substantial problem that the phase reference method eliminates. With previous methods, the calibration was made with a particular calibration mixer tuned to a particular LO frequency. Any other choice of LO frequency required a new calibration and new determination of the ratio of (BTF^{RF}/RRF^{IF}) for the calibration mixer for each particular LO frequency. The phase-reference method finds the individual phase response of each receiver without resorting to a known mixer.

This method is developed from an entirely different approach to calibration that is borrowed from the area of NVNA. NVNA systems allow measuring the magnitude and phase of signals and their harmonics by using a harmonic comb generator as a phase reference into an additional channel and comparing the phase of the harmonics to the phase of this known comb. The waveform of the signal can be accurately reconstructed from the fundamental and harmonics of the signal. Key to this reconstruction is knowing the relative phases of the harmonics of the comb generator, used as a phase reference. The phase of these harmonics can be accurately measured and traced to national standards with very small errors. Figure 7.40 shows the phase and group delay response from one such phase reference, a Keysight U9391

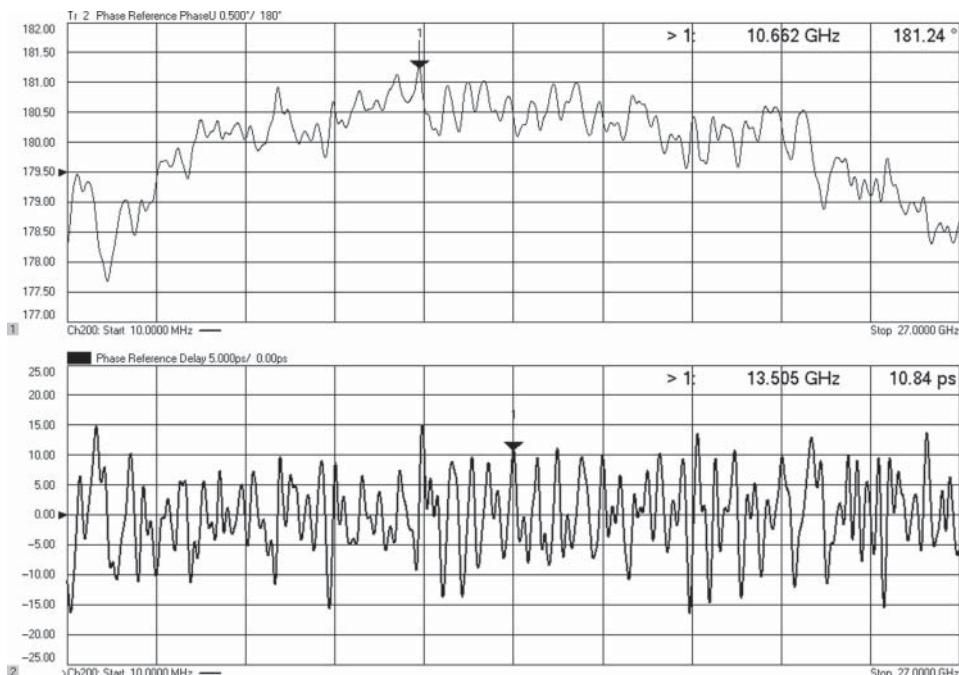


Figure 7.40 The phase and delay response of a phase reference.

model, with the calibrated phase data. This provides the incident signal, $a_{Phase_reference}$, to the receiver input on the VNA.

Since the measurement process has some noise, some smoothing may give a better result when using this phase reference in practice.

This phase reference provides the source of a signal, essentially a pulse, which delivers a known phase of a fundamental and all the harmonics. Since the coherent receiver system can measure the phase of a single channel response, these two elements are combined to create a single-channel measurement receiver with calibrated phase. In the calibration process, the source for the phase reference is the 10 MHz reference oscillator in the VNA. The output of the phase reference is connected to the b_2 receiver of the VNA during calibration, as illustrated in Figure 7.41. Here, one sees that the VNA is configured with a reversed port 2 coupler to reduce the loss to the receiver. This is necessary for best performance as the individual comb-teeth signals are very small, on the order of -60 dBm. Thus, the coupler loss is removed from the measurement, which increases the signal to the receiver, but also adds an offset to

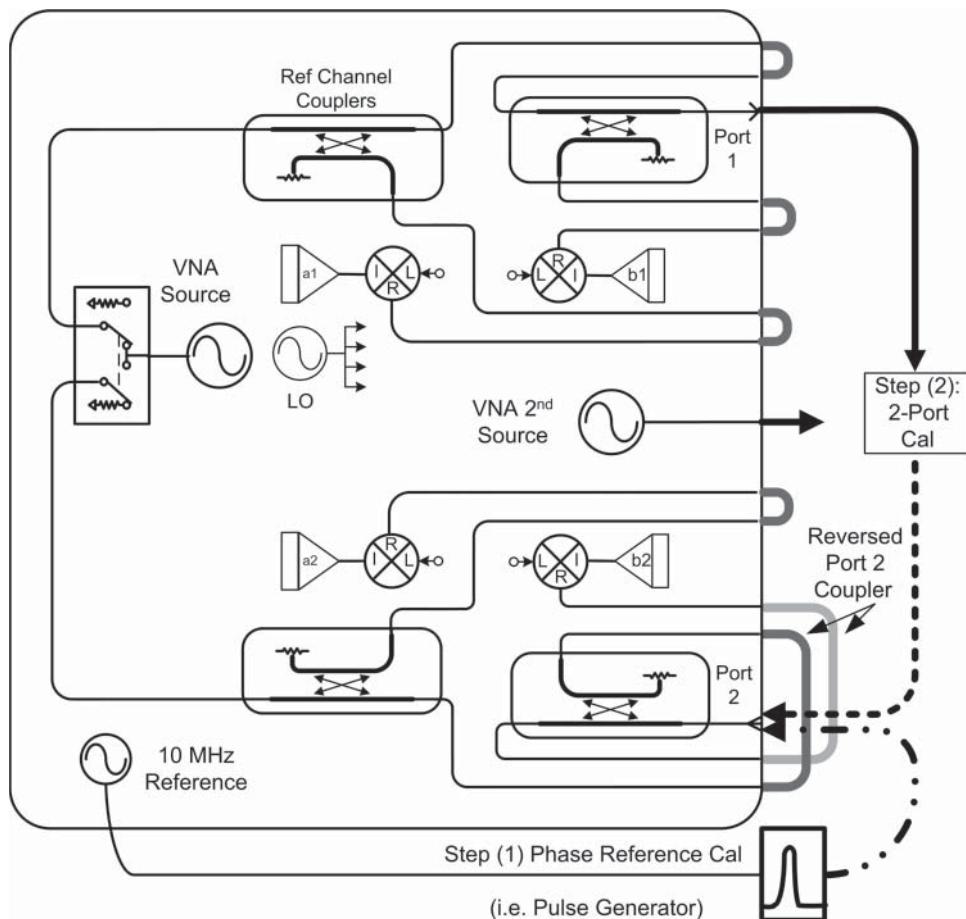


Figure 7.41 Calibration of the VNA using a phase reference.

the measurement. A full 2-port calibration is performed as a second step, which effectively measures the offset loss of the coupler through-path. This offset will be compensated in an additional calibration step.

The frequency stimulus is set to measure exactly every 10 MHz, which is the input reference signal into the phase reference. The b_2 receiver measures the phase reference (unlike the VNA, no additional phase reference is needed). The result of this measurement is shown in Figure 7.42. In this example, the phase reference measurement is performed from 100 MHz to 26.5 GHz. If the MUT is low-gain, the coupler may remain reversed. If the mixer is high-gain, the coupler should be put into the normal configuration and a second 2-port calibration performed. In the figure, the response is about 13 dB higher than the actual power (much more at low frequencies) as the VNA factory calibration is compensating for the expected coupler roll-off, which is not present when the coupler is reversed. The phase reference calibration of the receiver is then compensated by the difference the S_{21} tracking terms, which captures the change in loss of the test port coupler.

These results show approximately -50 to -60 dBm power response and phase response going from 0° to more than $250\,000^\circ$ over 26.5 GHz. It is more convenient to show the phase response in terms of either group delay or deviation from linear phase. Remember that this is the raw amplitude and phase response and represents the combination of the phase response of the phase reference as well as the phase response of the b_2 receiver including the directional-coupler response in front of the b_2 receiver, and the phase response of the VNA mixer's LO (which of course also affects the measured phase in the VNA final IF).

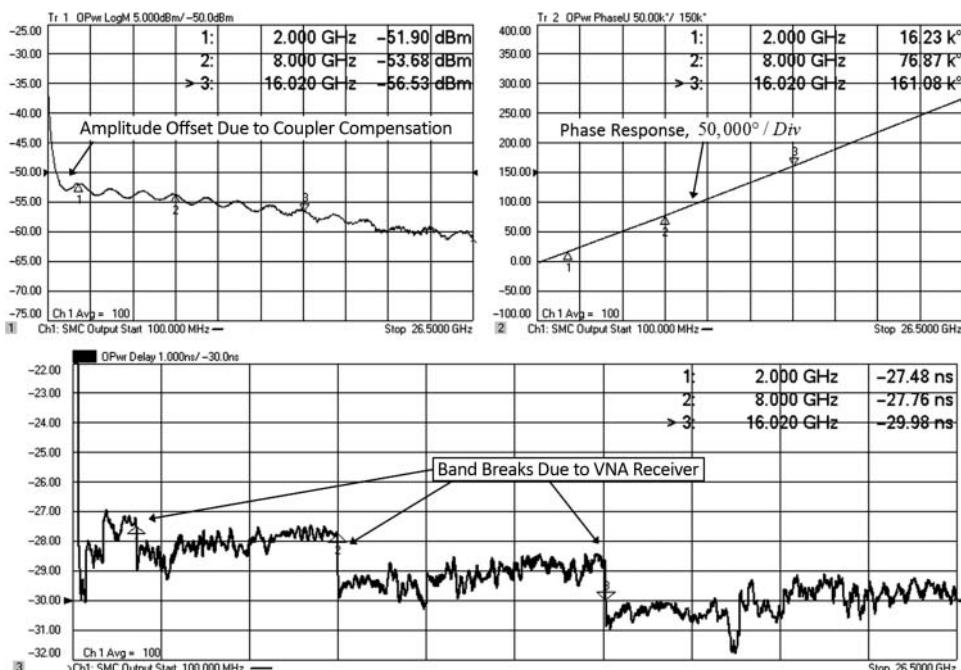


Figure 7.42 Measured amplitude and phase response of the phase reference at the b_2 receiver, including reversed coupler effect.

The delay of the phase reference itself is very small and very flat across frequency, with a delay deviation less than 10 ps. Thus, the phase response in Figure 7.42 is almost entirely due to the receiver response, plus some nominal delay from due to cable lengths. Further, the fine-grain response of the delay is commensurate with the amplitude response variation measured on the receiver. The phase response of the phase reference, as well as its mismatch, can be removed, solving for the VNA receiver tracking response as

$$BTF = \frac{b_{2_PhRefMeas} \cdot (1 - ELF \cdot S_{22_PhRef})}{a_{Phase_reference}} \quad (7.33)$$

where $a_{Phase_reference}$ is the power and phase of the phase reference signal. The breaks in the measured delay Figure 7.42 are really from the VNA receiver, and not from the phase reference.

This receiver response is computed from the previous measurement and is shown in Figure 7.43. The amplitude error is quite small as there is already a factory-based amplitude correction on many VNAs.

In practice, the raw S_{22} of the phase reference is measured during its connection step. The next step is to perform an S-parameter calibration and extract the port error terms, which allow the S_{22} of the phase reference to be corrected, as well as providing the load match error term needed in Eq. (7.33). Also, the transmission tracking term, ETF, is determined during the S-parameter calibration. Since this is done at every 10 MHz, over the entire VNA frequency range, every RF and IF frequency span is included. Finally, from BTF and ETF, the RRF is computed. While the phase reference amplitude information is available, the power calibration described in Section 7.5.1 is usually more accurate and traceable than the phase reference amplitude and should typically be used.

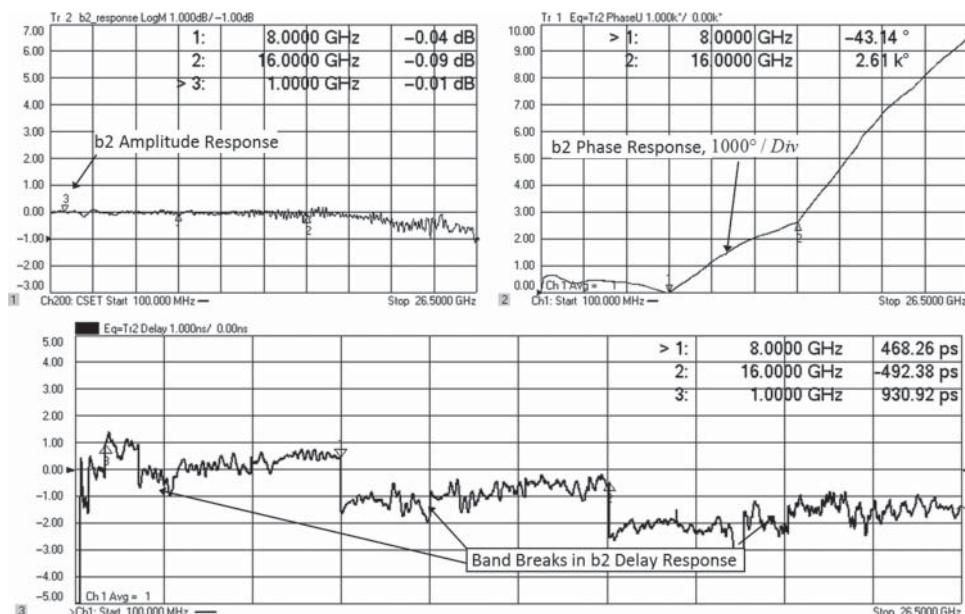


Figure 7.43 Amplitude, phase deviation, and delay of the b_2 receiver in normal mode.

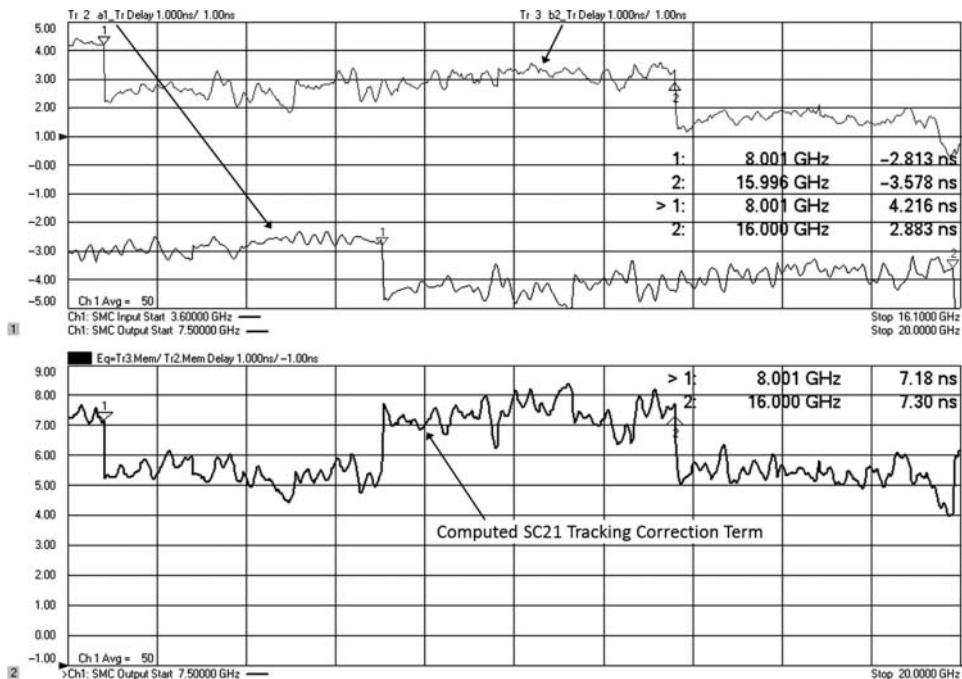


Figure 7.44 Upper: individual input and output response of the VNA receiver for mixer frequencies; lower: combined delay correction term for this setup.

Thus, the response of each receiver, both magnitude and phase, is individually determined; the response of the overall system is easily computed for any mixer condition. Figure 7.44 shows three traces; the top two are the b_2^{RF} tracking and a_1^{IF} tracking computed for a mixer with a 6.5–26.5 GHz input and a 1.5–21.5 GHz output. Also shown is the computed ratio, which is the SC_{21} tracking correction term (BTF^{RF}/RRF^{IF}).

These are formatted in delay for clarity and plotted versus mixer output frequency. Note the discrete jumps in the correction array. Investigating this further, one can plot the reference and test receiver tracking as a function of the frequency seen by the each receiver. Figure 7.45 shows the tracking responses in this way, and we see the discrete jumps are found at the same frequency for each receiver.

This implies a common cause to the jumps in delay between the reference and test receiver of the VNA and is a particularly interesting result. The common factor is the LO used to drive both receivers. The VNA mixers must also act according to the response of Eq. (7.4), and thus their phase response must also contain any response from the LO. Inspection of the details of the VNA used shows that the LO is created by multiplying and dividing a base 2–4 GHz oscillator. After each multiplication path there is a filter, so it is entirely reasonable to expect discrete changes at these multiplier bands, and that is exactly what is shown in the figure. Of course, each receiver also has a fine-grain response associated with its particular conversion loss and any ripple or response in the measurement path before each particular receiver.

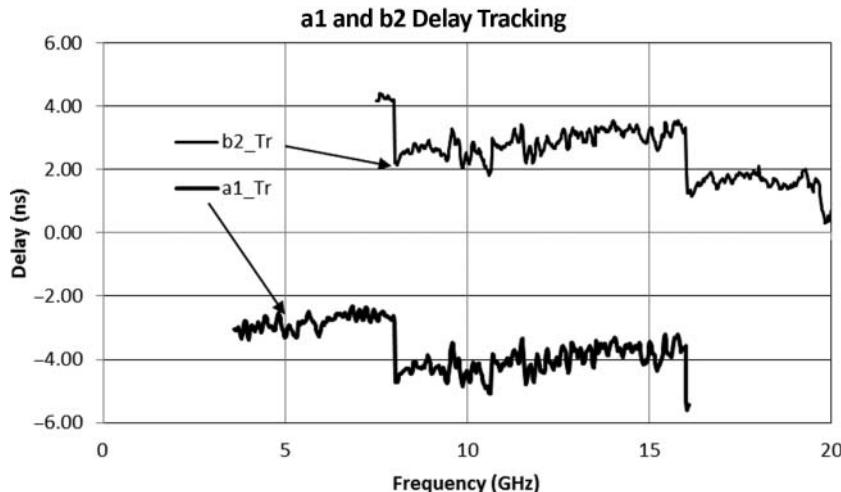


Figure 7.45 Receiver tracking for mixer test plotted as a function of VNA receiver frequency.

Finally, the overall mixer result is shown in Figure 7.46, before and after correction. For this example, a wideband mixer with small delay is used; this frequency range cuts across some of the VNA receiver band-crossings. This shows the remarkable result that the corrected mixer response is flat and on the order of 400 ps delay. This is very near the expected value based on the reciprocal mixer approach used in Section 7.5.3.1 after accounting for the filter used. And this corrected result did not use any other mixer for calibration or measurement.

As another comparison, measurements of the mixer/filter combination used in the previous two calibration methods, along with the phase-reference calibration method, are shown in Figure 7.47. This is the only method whose calibration is independent of any other mixer.

Note that the “reciprocal cal” mixer approach is used in both the VMC correction method and for de-embedding the reconversion mixer in the up/down measurement method. The best response is certainly from the phase-reference calibration, in that it is the smoothest response and has the lowest out-of-band noise at high frequencies. In the passband, it compares almost identically to the reciprocal cal method, with a variation less than 85 ps. In contrast, the three-mixer method varies by about 500 ps.

And in a final comparison, the LO frequency of this mixer combination is changed, and the resulting delay and gain measurements are shown in Figure 7.48. This is the only method that allows changing the LO frequency of test without re-calibrating the system or re-characterizing the calibration mixer.

In this comparison, the input frequencies are kept constant, and different LO frequencies are used that result in different output frequencies. In one case, the LO is shifted 100 MHz from 6.55 GHz to 6.45 GHz. In another case, the LO is moved from low-side to high-side, resulting in a reversal of the original output frequencies. With the small shift in LO frequency, the delay is nearly the same (the delay of the RF filter is slowly varying across this range, with a slight slope, explaining why the double peak in delay is not exactly equal). When the LO is moved to the high-side, at 13.45 GHz, the group delay response of the input filter is identical, but

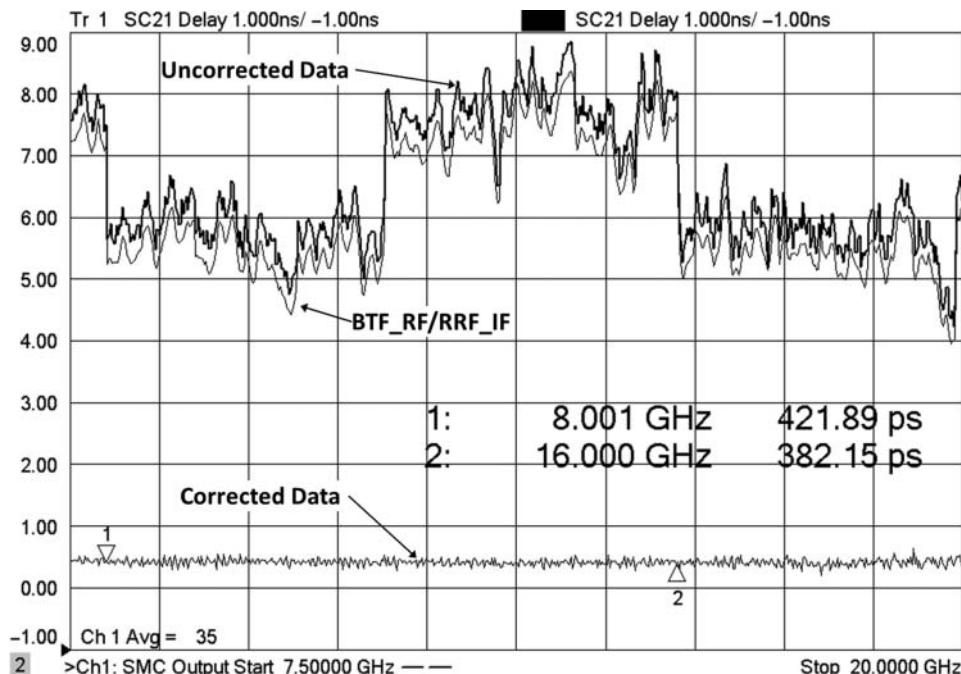


Figure 7.46 Raw and correct delay response of a mixer.

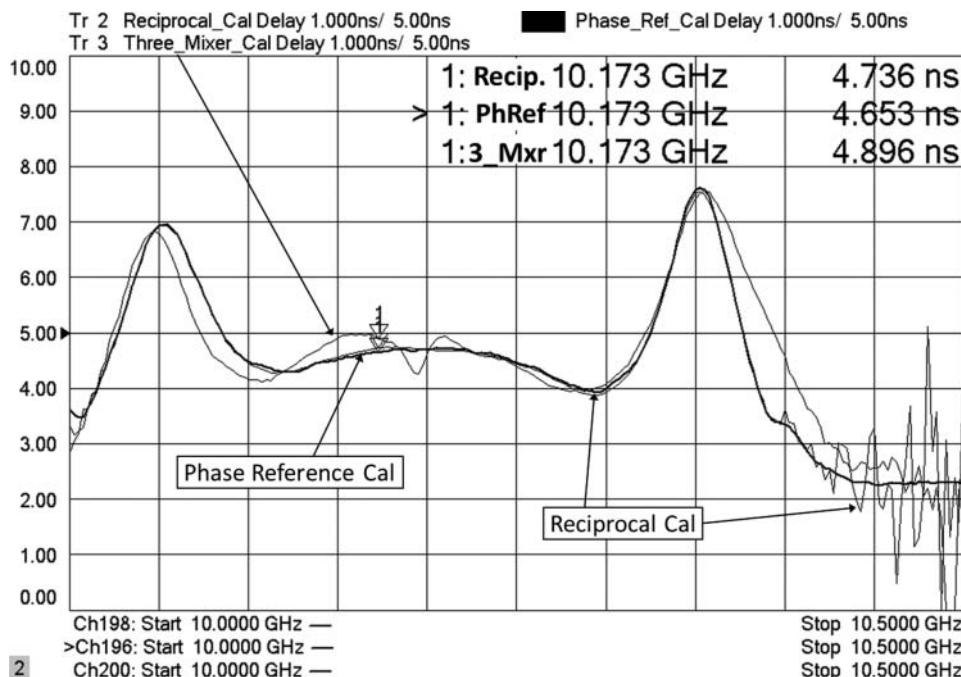


Figure 7.47 Comparing the three mixer-phase calibration methods.

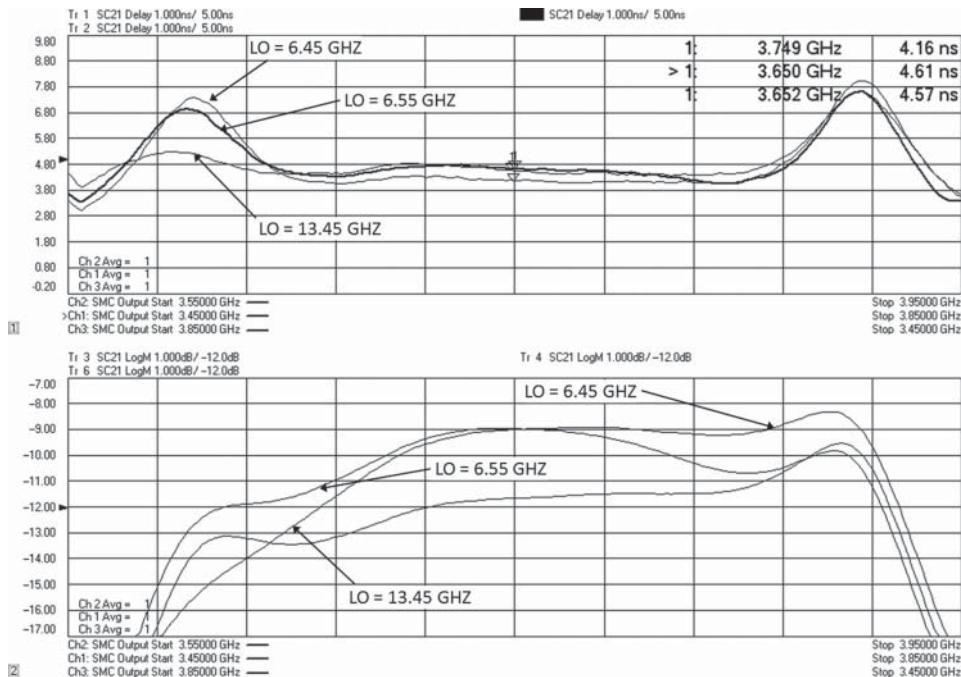


Figure 7.48 Delay and gain measurements with different LO frequencies.

the response of the output filter is flipped, resulting in a big change in delay response. This is mirrored by an equally large change in the slope of the amplitude response (lower window). As one might expect, the amplitude response shows a slowly varying change on the left side of the response, which is typically associated with lower peak group delay.

7.6 Mixers Measurements vs. Drive Power

Just as with other active devices, the behavior and characteristics of mixers depend upon the drive level applied to their ports. Mixers, however, have both a high-level drive (the LO) and a low-level drive (the signal input), and the mixer response is affected by both. In trying to optimize the operation of a circuit or system, determining the best operating point of a mixer requires characterizations versus both the LO drive and the RF drive of the mixer, as discussed in the following sections.

7.6.1 Mixer Measurements vs. LO Drive

7.6.1.1 Fixed Frequency Response to LO Drive

Normally, a mixer is designated for operation at some particular LO level; mixer manufacturers often give some range of LO drive level over which the operation of the mixer, particularly conversion loss and distortion, is presumed to be relatively constant. Often the

presumption is that higher LO drives will always give lower distortion or better conversion loss, but depending upon the details of the mixer design, this may not be the case. As such, it is often useful to perform tests of mixer characteristics over the LO power range.

With a traditional approach, these measurements are implemented using RF sources and spectrum analyzers and are tedious to perform as either the settings of frequency and power are made manually or a user must write their own program to control the individual instruments, gather the data, correct the data, and present the results. In fact, many RF test engineers started their careers writing programs for mixer test, including the author.

Today, many modern VNAs have FOMs or frequency-converter applications (FCAs) that essentially automate all the required controls, provide calibration, and directly present the significant results. This makes it easy and practical for a user to evaluate and understand the behavior of their devices, using essentially the same setup that was created for amplitude characterization of the MUT, by simply changing the sweep modes and the stimulus settings.

Figure 7.49 displays a set of mixer measurements at a fixed frequency as a function of the LO drive level, for a simple mixer. The SC_{21} gain measurement is performed using the built-in FCA function. The S_{33} measurement shows the LO match of the mixer, also as a function of LO drive. In this case, the LO is obtained from the internal second source of the VNA. Since the LO response does not depend upon the RF signal applied, one could also measure LO match on a simpler, 2-port VNA by applying a load to the input and output ports of the MUT, connecting the LO port to port 1 of a VNA, and sweeping the drive power. If the VNA does not have sufficient drive power to test the MUT over the entire LO drive range, the high-power setup described in Figure 6.29 can be used to increase the drive for this test.

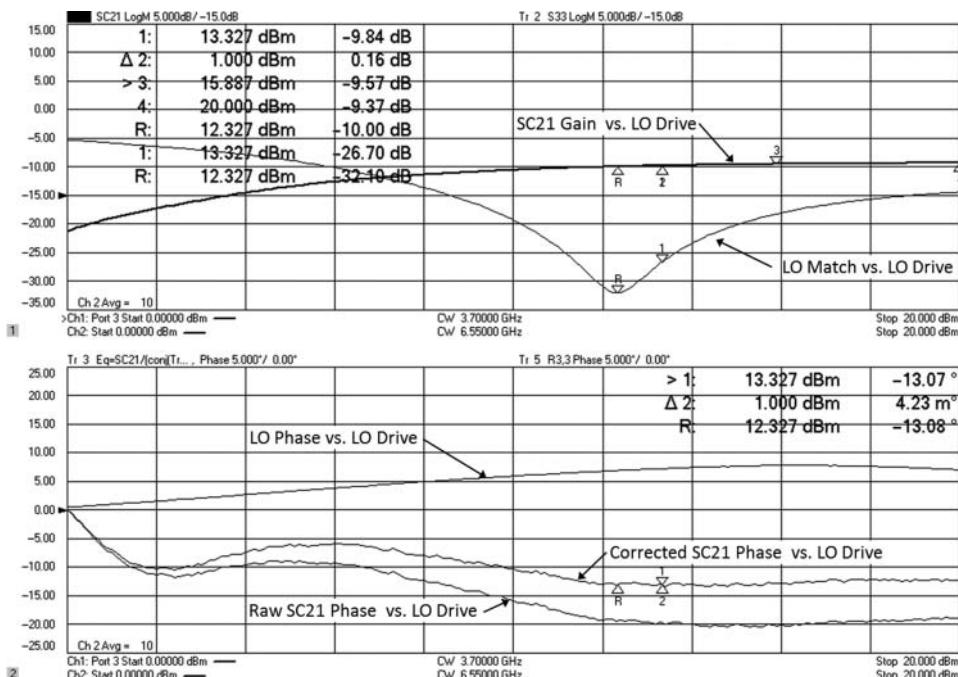


Figure 7.49 Mixer parameters SC_{21} and S_{11} , and S_{33} (LO match) while making an LO power sweep.

What is interesting in this response is that from the SC_{21} response, one can see that there is a definite curvature of the LO-drive for this particular mixer. Other mixer types may show such an optimum LO drive power for highest conversion gain, depending upon the internal limiting effects of the diode or switching structures in the mixer. The flatness and breadth of the SC_{21} response to LO drive gives one a good sense of this mixer's sensitivity to LO drive level, and with a broad and flat response, one can see that the assumption that LO drive does not, to the first order, affect the SC_{21} conversion gain is valid.

Also interesting to note is that the LO input match of the mixer depends strongly on LO drive, and the point at which the match is best is used as a reference point to determine the change in SC_{21} versus change in LO drive power, as indicated with Marker R in the upper window. Marker 2 shows that a 1 dB change in LO drive from the point where the LO match is minimum results in a 0.16 dB change in SC_{21} .

One can sometimes correspond the optimum drive level for minimum conversion loss with the LO match versus LO drive. The S_{33} plot clearly shows when the LO drive is sufficient to "turn-on" the mixer diodes; at lower LO drive levels the diodes represent a high impedance, and thus much of the LO signal is reflected. In this same region, the conversion loss is very poor. As the LO power is increased, the mixer diodes turn on and at some point provide a matched impedance. At even higher LO drive levels, the impedance of the mixer diodes becomes even smaller and so the impedance may no longer be matched.

The lower window shows the SC_{21} conversion phase versus LO drive level. Unfortunately, this "raw" measurement cannot be completely trusted as the LO phase can also change with LO drive level. If an external source is used, the LO can experience step changes in phase if the external source steps through source attenuator settings. Here, the internal source is used, and the reference receiver on the LO port can provide a measure of the change in LO phase versus LO drive. This trace is used, with the equation editor, to compensate the raw SC_{21} phase trace and produce a "corrected" SC_{21} phase response with respect to LO drive. The formula used must also take into account that the LO phase is conjugated because the RF is greater than the LO, so changing the LO phase causes an opposite change in the output phase. The formula used for this is

$$SC_{21_Corr} = SC_{21_Raw} \cdot \frac{|a_{LO}|}{a_{LO}^*} \quad (7.34)$$

This formulation allows correcting for the phase of the LO without its magnitude affecting the result. It is interesting to note that after correction the phase response is nearly constant above the reference marker, where the LO match is optimized.

7.6.1.2 Swept Frequency Response to LO Drive

The previous section demonstrated that understanding the effect of LO drive level on mixer performance is an important consideration when optimizing system performance but was restricted to measurements of a single LO and RF frequency. When a mixer is used over a broad set of frequencies, it is important to understand how to optimize the LO drive for an overall response. This is simply done with one of two measurements. In the first case, the LO is fixed, and the frequency response across an RF and IF band is measured and normalized. Next the LO drive power is varied, and the measurement results are compared. In this way,

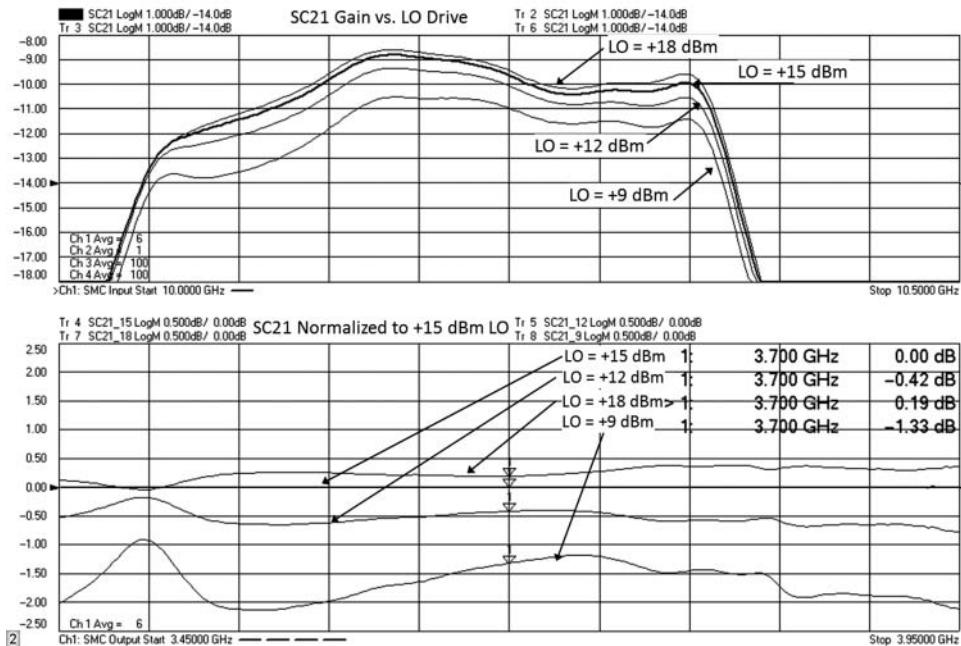


Figure 7.50 LO power effects on SC_{21} over a range of RF/IF frequencies.

the LO power's effect on the broadband conversion loss can be determined, as illustrated in Figure 7.50. The upper window shows the actual SC_{21} over frequency with different LO drive levels. The lower window shows the same data, all normalized to the response of SC_{21} for a +15 dBm LO drive level. Clearly, this mixer has different responses to LO drives at different RF frequencies. One cause of this might be the change in match that the LO sees as a result of the input and output filters that are attached in this simple converter case.

Alternatively, a different manner of characterization is needed for a block converter, where the LO frequency changes as the channel being converted changes. In a similar manner to the lower window in Figure 7.50, the LO power is changed during an SC_{21} measurement, but in this case the LO and IF are swept, and the RF remains fixed, as shown in Figure 7.51. This measurement shows the channel conversion gain as a function of the LO frequency, for different LO powers, all normalized to the conversion gain for an LO power of +15 dBm.

Here again one can see that there is some trade-off in the LO drive for maximum conversion as the IF frequency changes. In this case, the LO drive of +9 could be used at some frequencies but has substantial degradation at certain frequencies across the band. For example, at the center frequency (LO = 8 GHz), there is less than 0.5 dB variation in SC_{21} gain for LO power changes from +9 to +18 dBm; however, at 9.085 GHz, there is almost 4 dB of variation over the same LO power change. Characterizations of this kind allow designers to fully understand the effects of optimization on overall system performance. It is clear that for this mixer, the single frequency characterization of Figure 7.49 is not sufficient to understand this mixer behavior over the wide range of LO frequencies used.

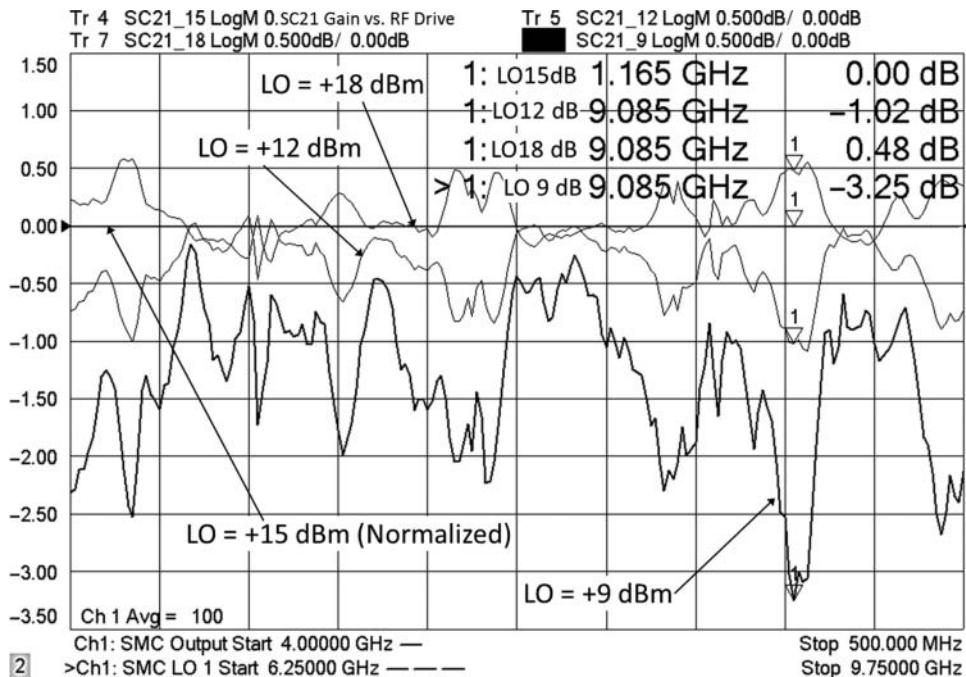


Figure 7.51 RF conversion gain versus frequency for various LO drive levels for swept LO measurements of a block converter.

In this figure, one of the traces uses the output frequencies as its x-axis, and the others use the LO frequency. It is convenient to have the option to select the x-axis and to have both scales available on the same plot. Also, note that there is a reversal in the frequency axis for the RF output; this is caused by the fact that the LO is less than the RF so that when the LO increases, the IF decreases in frequency.

7.6.2 Mixer Measurements vs. RF Drive Level

The LO drive level used for a mixer application essentially sets the operating point for that mixer, just as setting the bias level of an amplifier sets the operating point of the amplifier. One aspect of mixer characterization is looking at the conversion gain of the mixer as the RF drive level is changed. Just as with LO drive characterizations, this can be done in fixed frequency or swept RF frequency.

7.6.2.1 Swept Frequency Measurements vs. RF Drive

While the CW compression measurements give an intuitive feel for the mixer compression curve, one may find that the compression of a mixer changes with RF frequency. As such, it is a good idea to start the investigation of mixer compression by measuring the change in conversion-gain versus change in RF power, over a swept RF frequency, in a similar manner

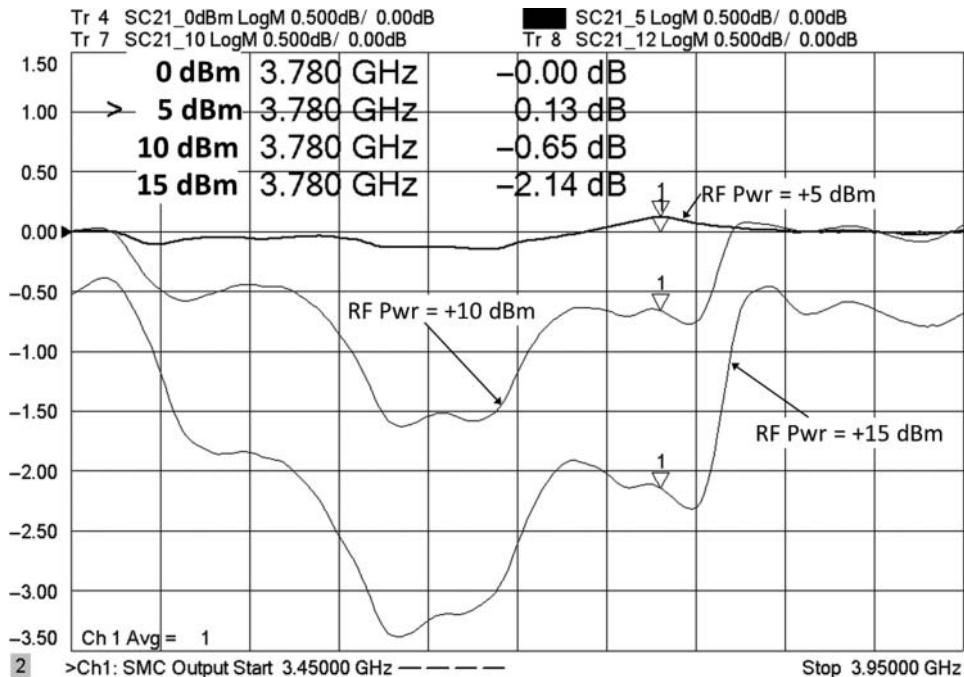


Figure 7.52 Normalized conversion gain versus frequency for various RF drive levels.

in which the amplifier swept frequency compression measurements were performed. For the same mixer used previously, at a fixed LO drive level of +12 dBm several sweeps of conversion gain were performed for RF powers of 0 , +5 , +10 , and +12 dBm, as shown in Figure 7.52. Here the conversion gain is shown normalized to the 0 dBm RF drive level.

From this swept frequency response, one finds an interesting point at 3.78 GHz, where the higher RF power (+5 dBm) causes an improvement in the conversion gain. This frequency point can be further examined by performing a fixed-frequency, swept RF power measurement, as shown in the next section.

7.6.2.2 Fixed-Frequency Measurements vs. RF Drive

The compression characteristics of mixers, that is, how a mixer conversion gain changes with signal (RF) drive level, is characterized in a manner quite similar to amplifiers, but with slightly different attributes.

Just as the LO drive can be varied at a fixed frequency, while monitoring the SC_{21} parameter, so too can the RF drive be varied to test the linearity of the mixer or converter. In Figure 7.53, the results are shown for testing conversion gain versus RF drive power, for several different LO powers. The figure shows both gain compression (in the upper window) and phase expansion (in the lower window) as a function of RF input drive level. As expected, the 1 dB compression point for a mixer, as well as the linear conversion loss, does depend upon the LO drive used.

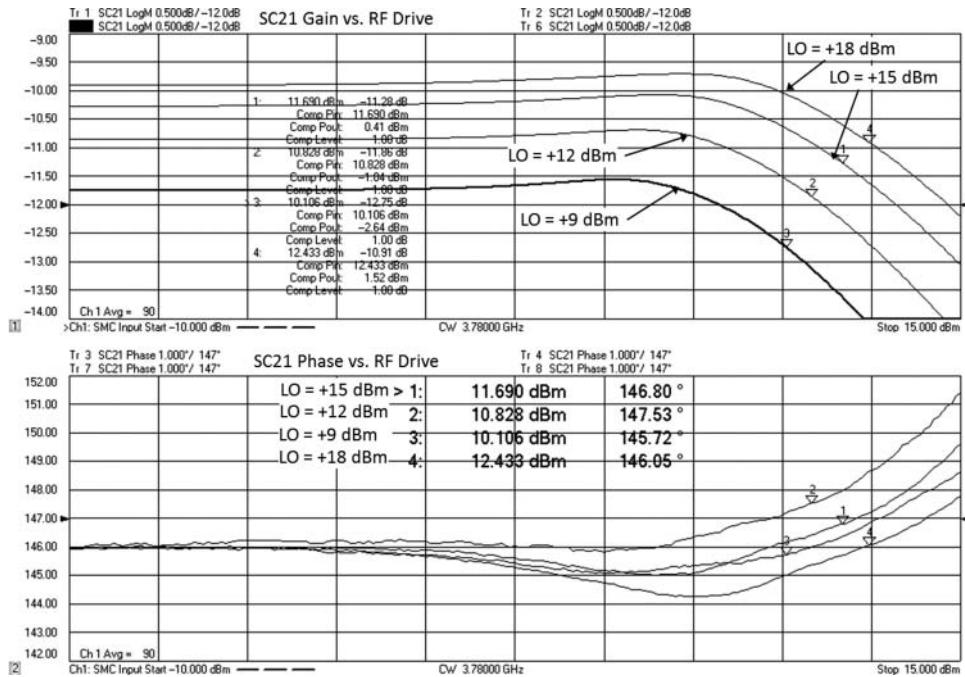


Figure 7.53 Conversion gain vs. RF drive for several different fixed LO powers.

As a very rough rule of thumb, the 1 dB output power compression is found at approximately 10–15 dB below the LO drive level. From the compression curves shown, the linear region of a mixer operates around 10 dB below RF compression level, or about 20–25 dB below the LO drive level. Note that because of conversion loss, the input power for compression is much higher. This is an important point to note when performing tests that depend upon the device behaving with linear gain, such as noise figure.

Another interesting detail on mixer compression is evident here: for the LO drive levels, the conversion gain may first increase as RF drive is increased before it starts to go into compression, as demonstrated in this figure. This effect is more pronounced at lower LO powers. This effect is hinted at in Figure 7.53; at other RF frequencies, a normal compression curve is observed, but at 3.78 GHz, some expansion was noted in that figure. While some linearized amplifiers have a hint of this behavior, it is quite common to find this behavior in mixers. One likely reason is that for low LO drive levels, the mixer diodes are barely turned on and have high loss. Increasing the RF drive causes a change in the bias point of the diodes and effectively adds to the LO drive level to turn on the diodes harder, thus reducing their resistive loss. This change in diode operating point causes the conversion gain to increase at moderate RF drive levels, especially when the LO is under-driven or “starved.” Starving the LO is often a trade-off that is accepted in a system design in order to improve the overall DC efficiency and reduce the complexity of a mixer, by eliminating LO driver amplifier stages.

Interestingly, starving the LO on a high drive mixer may give either better or worse performance than using a lower LO drive mixer in the same applications. Higher-level LO mixers have higher barrier diodes (or more diodes in series) and thus tend to be more linear when

operated with the same RF input level as lower barrier diodes, provided that sufficient LO drive is delivered to allow the mixer to operate in its normal mode.

7.6.2.3 Automated Gain Compression Measurements on Mixers (GCX)

For many mixers, a key specification is the input power that produces 1 dB of gain compression, just like in amplifier testing. While the value can be obtained by repeating the previous fixed-frequency or swept frequency measurements, some manufacturers have created purpose-built applications that directly provide the swept frequency 1 dB compression (or any other chosen level of compression). These setups are almost identical to a combination of the gain compression application (GCA) for amplifiers and the frequency converter application (FCA) for mixers, which form the gain compression application for mixers (GCX; X in a circle is a common symbol for mixers). For these applications, the calibration is identical to the calibration performed for an FCA measurements (in fact, the same calibration can be used) but with the added stimulus parameter of start, stop, and linear power. The measurement starts by sweeping the frequency response of a mixer at the linear power, to provide a reference gain. Next, the power is iterated from the start power (which may be higher than the linear power to shorten the measurement time) to the stop power. Just as in GCA (see Section 6.2.4), the sweep can be iterated in a smart way to find exactly the 1 dB compression point quickly, or a two-dimensional sweep of frequency and power can be performed for every point, generating a 2-D data set that is evaluated for the 1 dB compression point.

Figure 7.54 shows the results of several automated gain compression tests; each one with the mixer driven at a different fixed LO drive level. Evaluations of this sort can be useful

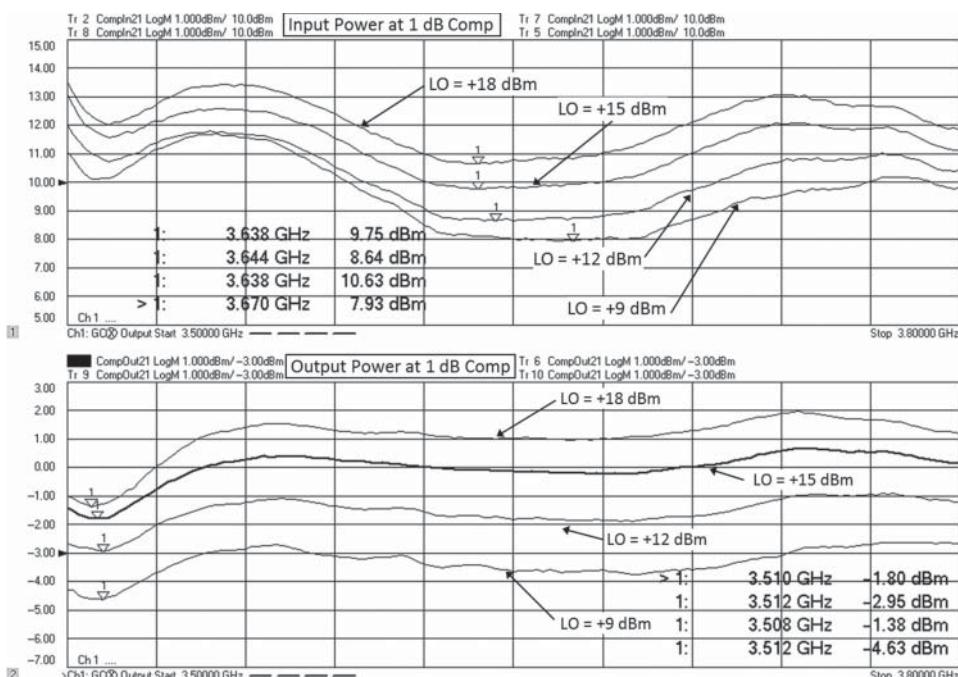


Figure 7.54 Automated GCA measurements for the same mixer with various LO drive levels.

in comparing mixers from various manufacturers or in optimizing system performance of a particular mixer. The upper window shows the input power for 1 dB compression for a variety of LO powers; the lower window shows the output power for 1 dB compression for the same set of LO powers. For this mixer, changing the LO power from +9 to +18 dBm improves the worst-case 1 dB compression point by about 3 dB.

As with amplifiers, gain compression is one measure of the linearity of a mixer, and non-linear behavior will lead to distortion in the output signal. A key measure of distortion is the two-tone intermodulation distortion (IMD) measurement, presented in the next section.

7.7 TOI and Mixers

Distortion measurements in mixers and frequency converters are similar to those of other active devices; however, the terms used are sometimes confused with other mixer measurements. One principal distortion measurement is the two-tone third-order IMD measurement, sometimes called TOI. As with amplifiers, TOI sometimes refers to the third-order intercept point and so IMD is used instead. Unfortunately, the terms *intermodulation* or *intermods* or *mixer intermods* are sometimes used to refer to higher-order products of the RF and LO. Usually the context makes the usage clear, but this book will refer to intermodulation as the mixing of two signals at the RF (or IF) input of a mixer, and the mixing products from RF and LO will be referred to as higher-order products. Both the RF and LO signal levels will affect the measured IMD, as discussed in the next sections. Figure 7.55 shows a typical setup for measuring mixer IMD using a VNA. It is also possible to do the same measurement using three independent sources and a spectrum analyzer, but many modern VNAs provided purpose-built applications, such as Swept IMD (IMDX), that make measuring IMD measurements while sweeping frequency, or power, very easy.

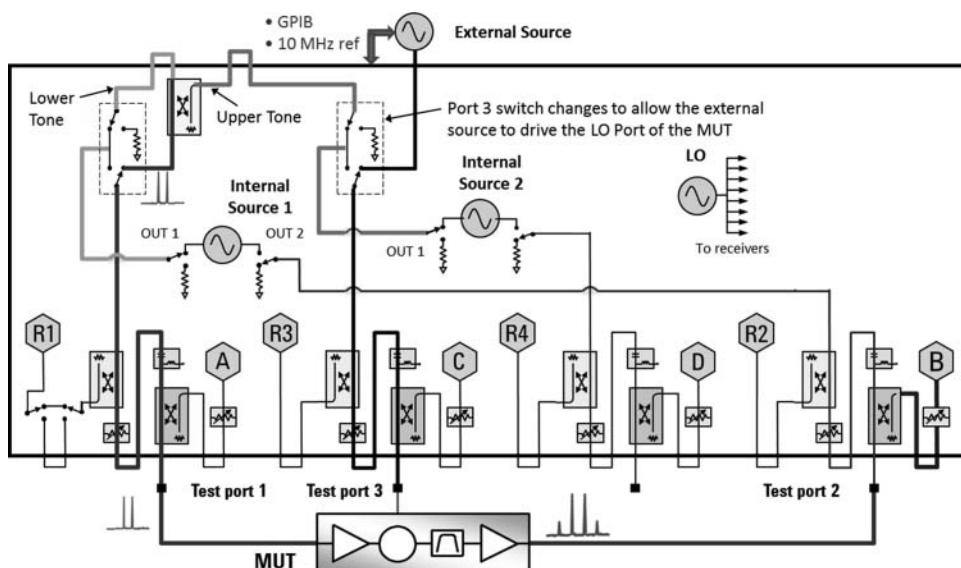


Figure 7.55 Setup for measuring mixer IMD.

For all IMD measurements, the same cautions apply as described in Section 6.6.5 on optimizing measurements. Principally this means ensuring that the sources don't create their own intermods into the mixer, and the power of the two tones hitting the measuring receiver are sufficiently attenuated so that the receiver's own intermods don't affect the result. In the previous block diagram, there are attenuators on the sources and the receivers to optimize each for a particular measurement. Using a coupler instead of a power splitter lowers the available two-tone power, but greatly increases the isolation between sources. A switch on the source going to port 3 allows flexibility in which source is used as an IMD source. With this configuration the VNA source, with its high sweep speed, can be used as a second tone for IMD measurements. By changing the switch position, the same source can drive the LO of the MUT for swept LO measurements. Finally, by routing the external source through the port 3 test port, the power and phase shift of the LO signal can be monitored on the R3 receiver.

Often, IMD is a critical measurement for receiving mixers used in low-power and low-noise applications. As such, a common figure of merit is the input referred third-order intercept point, or IIP3. Since this is so common (unlike amplifiers that are typically specified with output IP3 or OIP3), the following examples all refer to the input intercept point. Output intercept point is easily computed by adding subtracting the conversion loss from the IIP3 value.

7.7.1 *IMD vs. LO Drive Power*

Because the LO power applied to a mixer sets, in some sense, the operating point of the mixer, one finds that the IMD response of the mixer depends strongly on the LO drive level. Figure 7.56 shows the third-order and fifth-order IM product, in dBc, along with the measured output power and conversion loss, and the input referred third-order intercept point (IIP3). This point changes as both the IM product power and the gain of the mixer changes with respect the LO drive power.

What is interesting in this figure is that the LO power for optimum third-order performance, in terms of IIP3, has two values, nor is the optimum power for the third order products the same as for higher-order products; in fact, the IM5 is actually higher than the IM3 for the LO drive of +9 dBm. A spectrum plot of the mixer at these stimulus conditions (RF = -5 dBm, LO = +9 dBm) illustrates this point clearly, as in Figure 7.57.

For an LO drive level of 9 dBm, the IIP3 level peaks and does not return to the same value until the LO drive is up to about +16 dBm. The IIP3 in the region of +12 to +15 dBm (the normal operating range of this mixer) is not as good as at a lower LO drive, for this particular RF drive level. However, one cannot say that a low LO drive of +9 dBm is good for all RF power levels; this must be investigated independently, as discussed in next section.

Finally, the same LO power is not necessarily best across the frequency span of the mixer, as shown later in Section 7.7.3.

7.7.2 *IMD vs. RF Power*

The level of IMD created from an RF input follows the same rules in mixers as it does in amplifiers, provided that the input is sufficiently low. Normally changing the input power of the two tones each by 1 dB will increase the power in the third-order product by 3 dB,

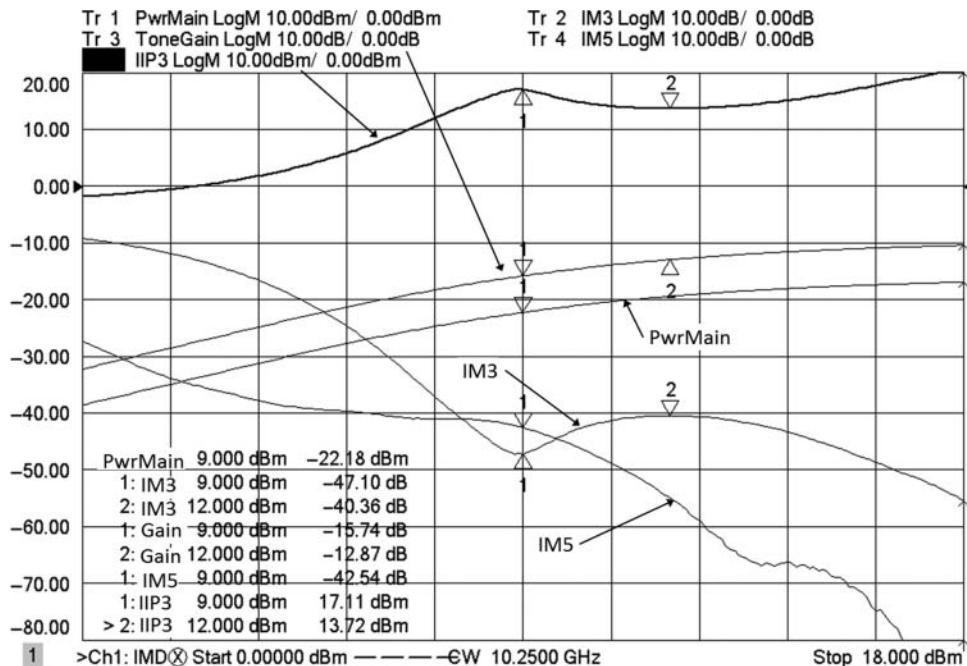


Figure 7.56 Third- and fifth-order IM product versus LO power, as well as output power, IIP3, and gain.

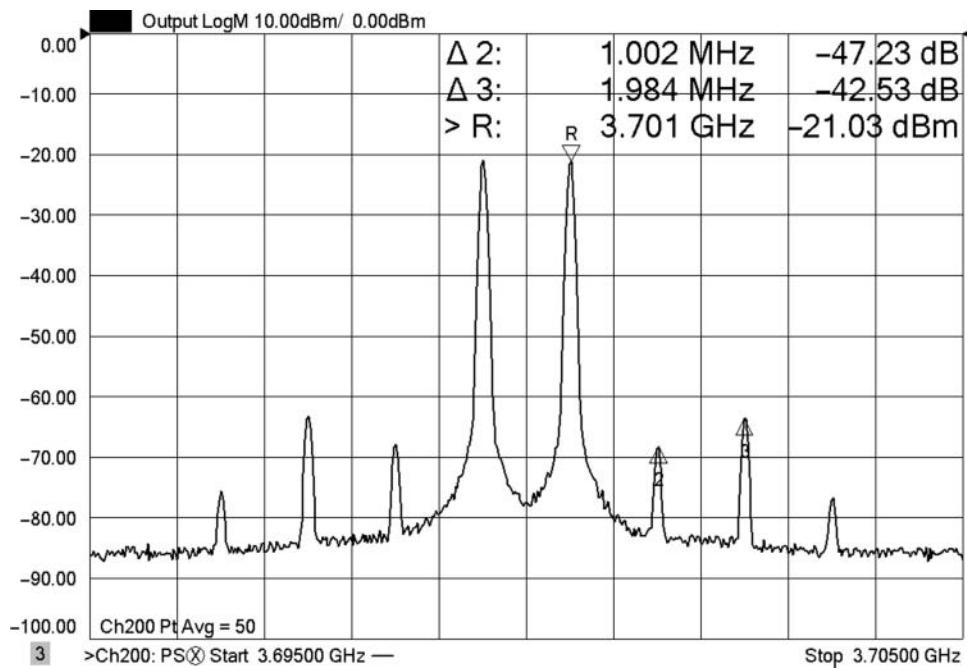


Figure 7.57 Spectrum plot of IM products at RF = -5 dBm, LO = -9 dBm.

fifth-order product by 5 dB, and so on. But this rule will break down in many mixers as the RF level gets high, especially if the mixer has a lower LO signal (starved LO) as the RF signal helps to turn on the mixer diodes. For this reason, it is often necessary to make a swept RF-power IMD-measurements on a mixer to understand the actual IM level at the input power of interest.

Figure 7.58 shows the result of making a swept RF-power measurement on a mixer while monitoring the third- and fifth-order products. Also shown is the IIP3 and the gain for this mixer, as well as the output power. Here again one notes that the IM powers follow the normal rules until the RF drive becomes significant relative to the LO power. Highlighting in the plot is the IM3 value for a gain compression of 0.11 dB: this follows a rule of thumb that the IM3 for 0.1 dB compression is approximately -40 dBc. Marker 1 is used on the tone-gain trace, reference to the R marker, to find the -0.1 dB compression point. Since the markers are coupled, the IM3 marker shows the relative IM3 level at that point.

Another remarkable attribute of Figure 7.58 is the behavior of the IIP3 as the RF power is increased. As described, this mixer follows the behavior that the mixer distortion does not always increase with increasing RF power. In fact, the intercept point (one measure of relative distortion) is highest at an RF input level of -0.6 dBm. The value of intercept point is also strongly affected by the LO power.

Figure 7.59 shows just the main power, the third-order power and IIP3 for the same RF power sweep range, but this time for two different LO drive-powers. It is common in mixer measurements to display many of the mixer characteristics such as gain, compression, and

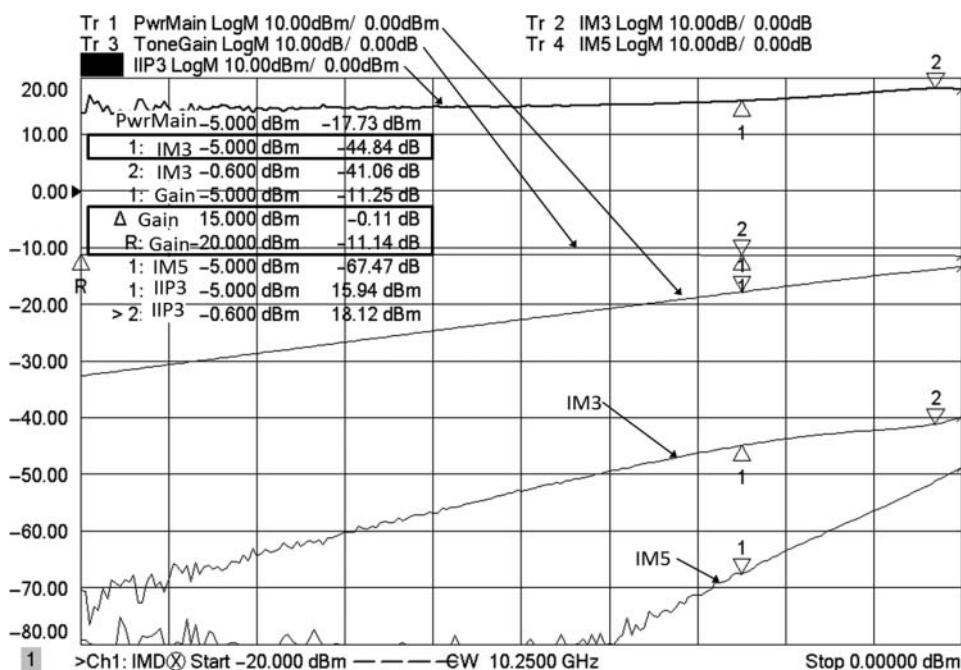


Figure 7.58 Third- and fifth-order IMD power versus RF power, as well as IIP3, output power, and gain.

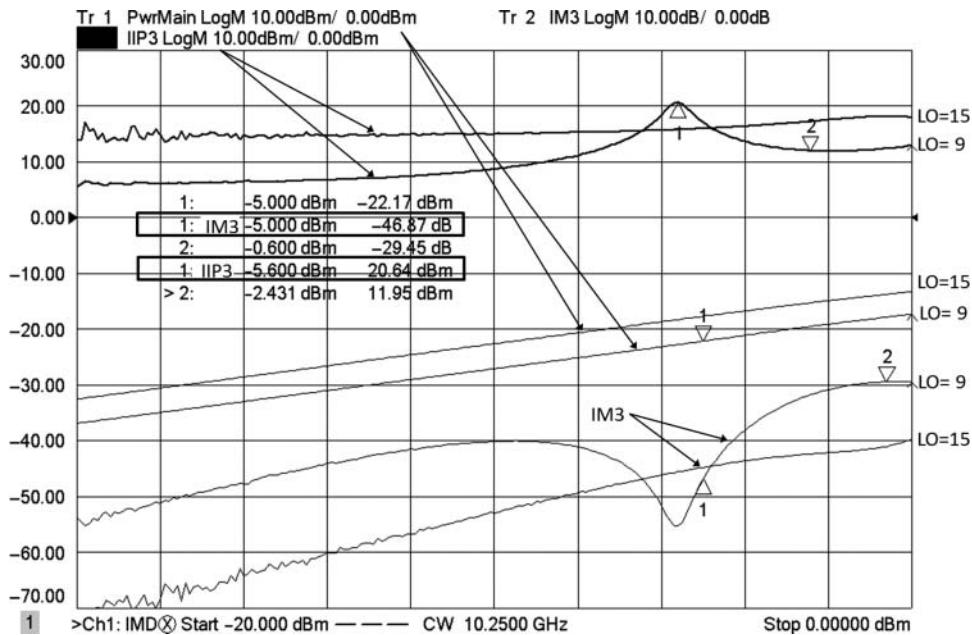


Figure 7.59 Swept RF power IM3 and IIP3 with different LO drives.

distortion, as a function of several LO drive powers so that designers can make trade-offs in the system performance between LO drive levels and other attributes. This is especially true in low-power receiver circuits where LO drive power may be one of the largest contributors to battery drain.

In this figure one notes that the maximum input-intercept-point is not for the higher LO drive case, but for the case where the LO is almost starved. At +9 dBm LO, this mixer gives a region of RF powers where the IIP3 is high, but just outside this region the IIP3 is much worse than the higher LO drive condition. It is for this reason that system designers must have distortion characteristics as a function of several variables, to show the true behavior of a part. This particular mixer could be advertised as having greater than +20 dBm IIP3 at +9 dBm, which is true, as long as the RF drive power is very carefully controlled. Such behavior is not normally expected and would come as a surprise to system designers.

7.7.3 IMD vs. Frequency Response

Since many mixers are inherently band-limited due to the nature of the baluns used in the mixers, or filters used in the construction of frequency converters, it is often necessary to characterize the performance of a mixer over its band of operation. While there may be little change at the center of the band of operation, there can be substantial performance difference at the band edges. Also, it is sometimes desired to use a mixer beyond its specified or nominal frequency range, and in such a case it is important to understand how the mixer's performance is degraded.

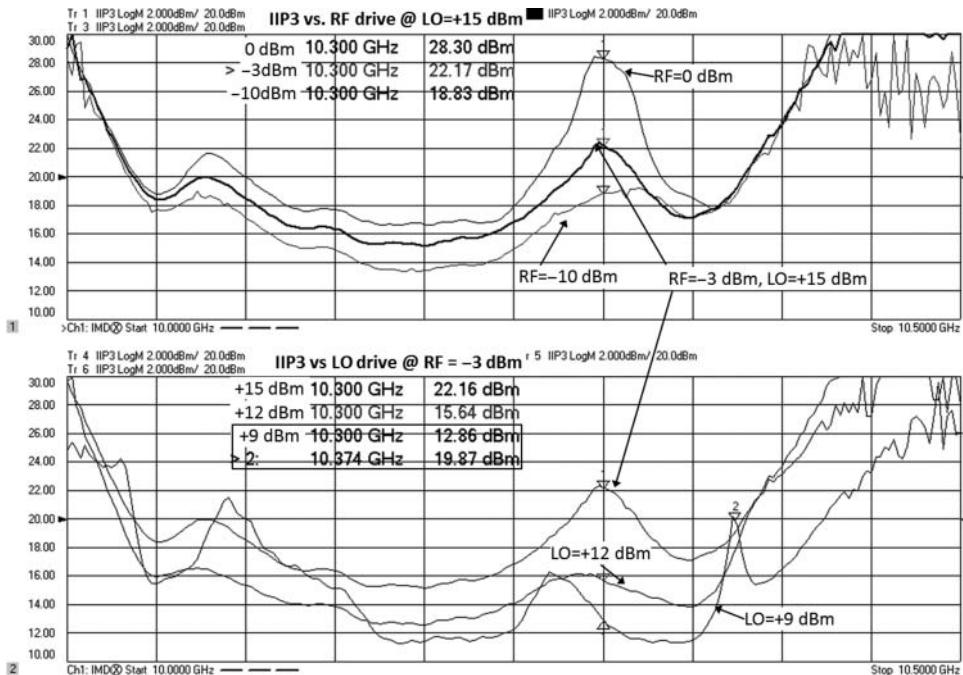


Figure 7.60 Upper: mixer IIP3 and gain versus frequency for three different RF power level; lower same mixer for three different LO power levels.

Figure 7.60 shows two examples of swept frequency IMD. In the upper window, the LO level is held constant, and three different traces of mixer IIP3 are shown versus RF drive level. The lower window shows a similar measurement, but this time the RF power is held constant, and the LO drive level is changed while the IIP3 is monitored for the three different LO powers as the RF and IF are swept in frequency.

In this plot, it is noteworthy that for this mixer, in the upper window, driving a higher RF drive level produces a higher IIP3, much higher at some frequencies such as the point highlighted by the marker. This implies the mixer is becoming more linear at higher RF drive levels; such a behavior is consistent with the results of Figure 7.58. Similarly, the change in IIP3 versus frequency for different LO drive levels shows some unexpected results in the lower window. The behavior at Marker 2, on the traces for LO = +9 dBm, shows dramatically improved IIP3 or a narrow region even for the low-level LO drive. Again, this behavior is anticipated in the results from Figure 7.59 and even from the compression measurements from Figure 7.52.

In practice, any of the RF, LO, or IF may be fixed or swept. When making IMD measurements on mixers, it is a good practice to set up and calibrate over a broad band so that the distortion performance of the mixer can be characterized over any of the frequency range inputs or outputs of any of the ports, as well as a wide range of RF and LO powers. From the measurement results shown in this and the previous sections, it is clear that the distortion

characteristics of mixers do not always follow simple rules, especially for mixers with starved LOs. Thus, the designer is required to make a wide range of mixer characterizations to fully understand the mixer behavior in a system.

7.8 Noise Figure in Mixers and Converters

Mixers and frequency converters are often the first component after the antenna in a receiver system, and as such the noise figure of the first converter predominantly sets the noise figure of the system. Thus, noise figure measurements on frequency converters are an essential aspect of their full characterization.

Most frequency converters have filters that keep unwanted images out of the response, and have amplifiers either before or after the embedded mixer as well, and so have overall gain and excess noise at the output above the kT_0B noise floor of a passive device. Thus, characterizing the noise figure a frequency converter is quite similar to that of an amplifier, where one measures the excess noise, then determines the gain of the MUT, and essentially computes the noise figure from these two values. Y-factor or cold-source techniques both work reasonably well on frequency converters, with image-protected input; however, the Y-factor often produces erroneous results if a mixer or converter does not have an input filter that protects against image conversion. The details of noise figure measurements for mixers are described in Chapter 9.

7.9 Special Cases

There are several special cases in characterizing mixers and frequency converters that fall outside the simple measurements described so far. These include mixers with multipliers in the RF or LO chain, measuring higher-order products, measuring group delay on mixers with a swept LO, measuring frequency converters with an embedded (built-in) LO, and converters with very high gain. Details for improving measurements of these special cases are discussed next.

7.9.1 Mixers with RF or LO Multipliers

Some frequency converters, particularly high microwave or mm-wave converters, are driven with lower-frequency LOs and have an explicit or implicit multiplier in the mixer structure. In many cases, the multiplier is doubler or tripler circuit to provide a higher frequency LO, but in some cases, the mixer is operated in second or third, or higher harmonic mode of the LO frequency (odd orders being more common in balanced mixers as even order products are substantially suppressed). In an ideally switched mixer, the third order conversion gain is down about 9 dB from the fundamental, and the fifth order is down about 14 dB, if no explicit multiplier is used.

Measurements of mixers with built-in multipliers in the LO are very straightforward and only require taking into account the multiplying effects of the LO port when setting the RF,

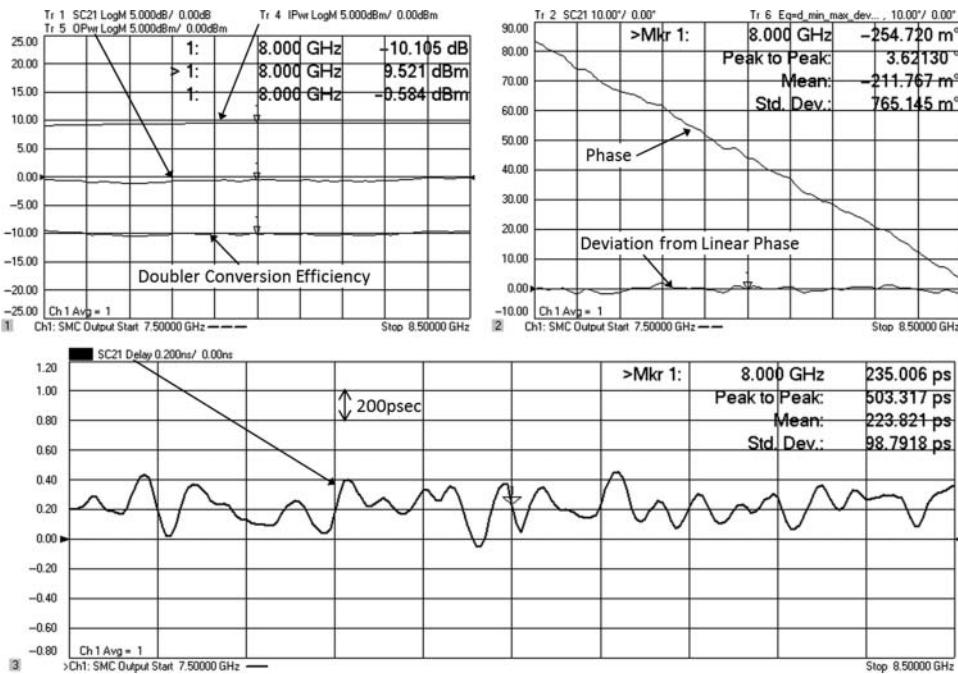


Figure 7.61 Magnitude, phase, and delay response of a frequency doubler.

IF, and LO frequencies. In many VNA mixer applications, an entry is available for providing the LO multiplier value, and the programming of the LO frequency automatically takes into account the effect of the multiplier.

In some cases, particularly in mixers used in radar applications, there can be multipliers in the RF path as well. Again, many VNA applications provide for this case. However, this does make the computation of phase response and group delay more difficult.

A degenerate case is that of a straight frequency-doubler, which is like a mixer that has the input and LO driven with the same frequency. The output of frequency in such a case moves twice as fast as the input frequency and makes computation of group delay a little more complicated. In the case where a multiplier is used in the RF path, the phase of the output signal must be divided by the multiplying effect in order for the group delay to be computed properly, relative to the sweeping input frequency. Calibration for phase or delay measurements is particularly difficult as there are no reciprocal devices that can be used as frequency multipliers in one direction and dividers in the other. One cannot use a mixer driven with the same frequency in the RF and LO, as sweeping the LO frequency produces an additional phase shift in the mixer (see Eq. (7.4)) that is not part of the RF to IF path.

Currently, one of the few ways to measure the delay of a frequency multiplier is to use the coherent-receiver method of Section 7.4.3.3 calibrated with a phase-reference method (Section 7.5.3.3). An example magnitude, phase, and group delay measurement of a small, solid-state frequency-doubler is shown in Figure 7.61. This system was calibrated as per the phase-reference method and treated as a mixer with a multiplier of 2 in the RF and an LO

frequency of 0 Hz. The response shows a small delay that is nearly equivalent to the physical size of the converter.

One can verify that the delay measurement is valid by adding a known delay at either the input or output port. If the phase multiplication is properly accounted for, then the change in delay for the doubler will be correct whether the delay element is added at the input or the output. If one used a normal VNA measurement in FOM, one may see that twice the effective delay in the output as in the input, thus demonstrating an error in the measurement.

7.9.2 Segmented Sweeps

One feature of a VNA that is particularly convenient in making mixer measurements is the ability to define arbitrary segments of frequency sweeps, with the sources and receivers being able to be set independently for each segment. This allows a complex characterization of a mixer to be completed with a single-sweep data acquisition and a single calibration. (The alternative, of course, is setting up many channels, each one representing a single segment condition, but this also requires a separate calibration for each channel.) Specialized display modes in segmented sweeps provide for easy comparison between segment conditions. Some particular examples using segment sweep are discussed in the next sections.

7.9.3 Measuring Higher-Order Products

For mixers and frequency converters used as receivers, it is normal to see some higher-order products cross into the band of interest as the receiver is swept across frequency. Many manufacturers specify the level of higher-order products for particular frequencies, and these are often displayed in an $N \times M$ array when the frequency associated with the array position is $M \cdot LO - N \cdot RF$; typically the array elements are in units of dBc below the $M = N = 1$ element, which is the power of the primary IF. Products where $M = N$ are the harmonics of the IF. Using traditional sources and spectrum analyzer techniques, one can directly measure these elements, but the measurements tend to be slow as the SA must be programmed individually for each frequency.

With a VNA, the segmented sweep function can be programmed so that the RF and LO frequencies are fixed at a single value or swept over a single range of frequencies, but each segment of measurement has the receiver swept over the mixer product. Thus, the higher-order products can be directly measured. If one of the segments represents the IF, the other segments can be compared using a second channel to measure the output power and equation editor to compute dBc values. The segment table for a mixer measurement with a 10–10.5 GHz RF, 8 GHz LO, and 2–2.5 GHz IF is shown in Figure 7.62.

An example measurement of this type is shown in Figure 7.63, using a segment sweep with the x-axis set to point spacing. Thus, each segment will appear as a direct point order, regardless of measurement frequency. In this display, the RF products are measured at 0 dBm input and –10 dBm input; it is interesting to note that the lower RF power affects the product power only for products that are a multiple of the RF frequency and not when they are a multiple of the LO frequency. Trace 2 is acquired on a separate sweep, where the output power is measured for each segment (same value as the first segment of trace 1).

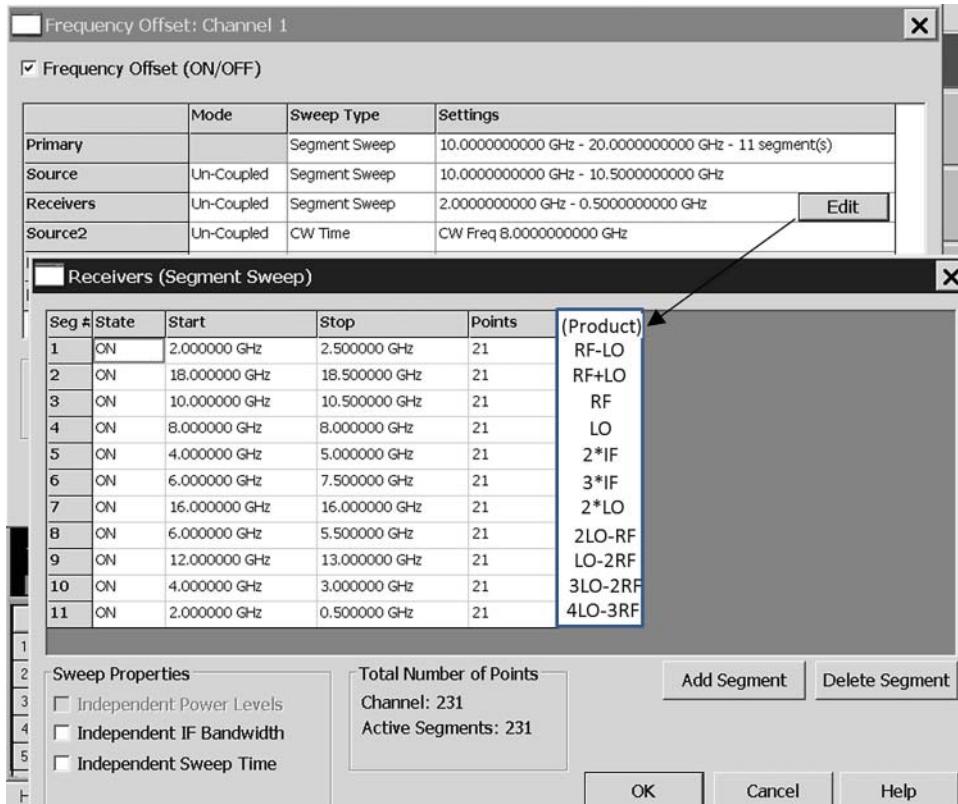


Figure 7.62 Segment table for higher order products measurement.

One caution when setting up measurements to measure higher-order products is that these products can sometimes go out of range of the measuring instruments (either too high or too low) and so the valid data is not available for all product frequencies. In some cases, it will be up to the user to program segments so that the higher-order products remain only in-band. Also, some products may land exactly on other products especially at band edges if round numbers are used. For example, the start of band for the 4:3 spur (last segment) exactly aligns with the main IF tone, and so one sees a large spike in the last segment measurement. A similar effect occurs on the 3rd harmonic of the IF, which exactly lands on the 2:1 spur. Here, the 2:1 spur is higher than the third harmonic. Of course, the LO feed-through products increase with lower RF drive (in dBc terms), as the IF output power is lower but the LO feed-through power remains the same.

From this measurement setup, it is easy to do other evaluations. For example, it is common wisdom (though often not correct) that increasing the LO drive should make the mixer more linear, and therefore a larger LO drive may lower the spurious products. In Figure 7.64, the LO power is changed while keeping the RF power at a constant level of 0 dBm. It is remarkable that many of the higher-order products actually go up with higher drive. In fact, this makes sense for any products that are a multiple of the LO drive; the harmonics of the LO increase with increasing LO drive, as is clearly demonstrated by the LO feed-through elements.

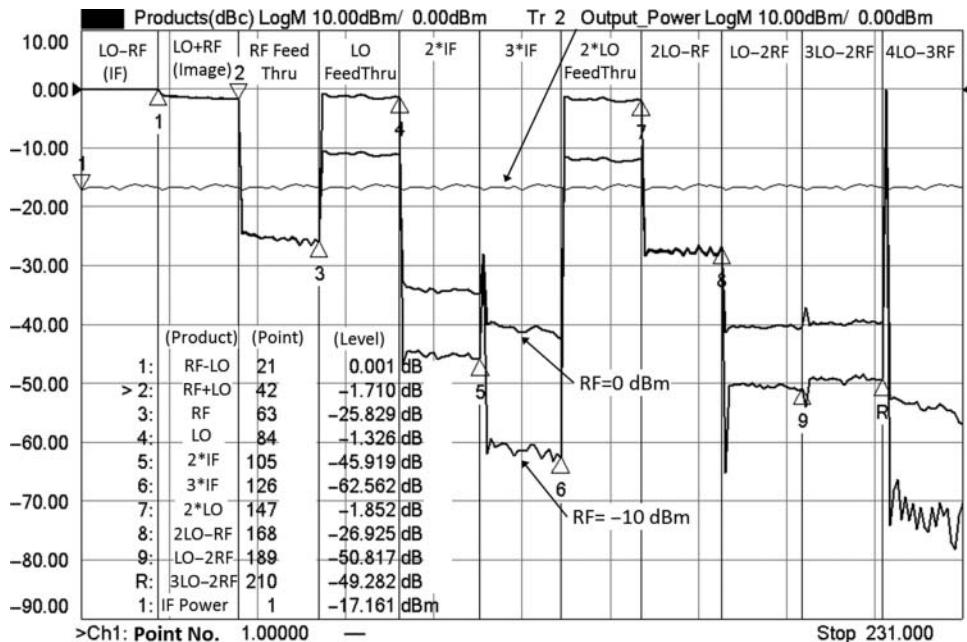


Figure 7.63 Higher-order products measured using overlapped segments.

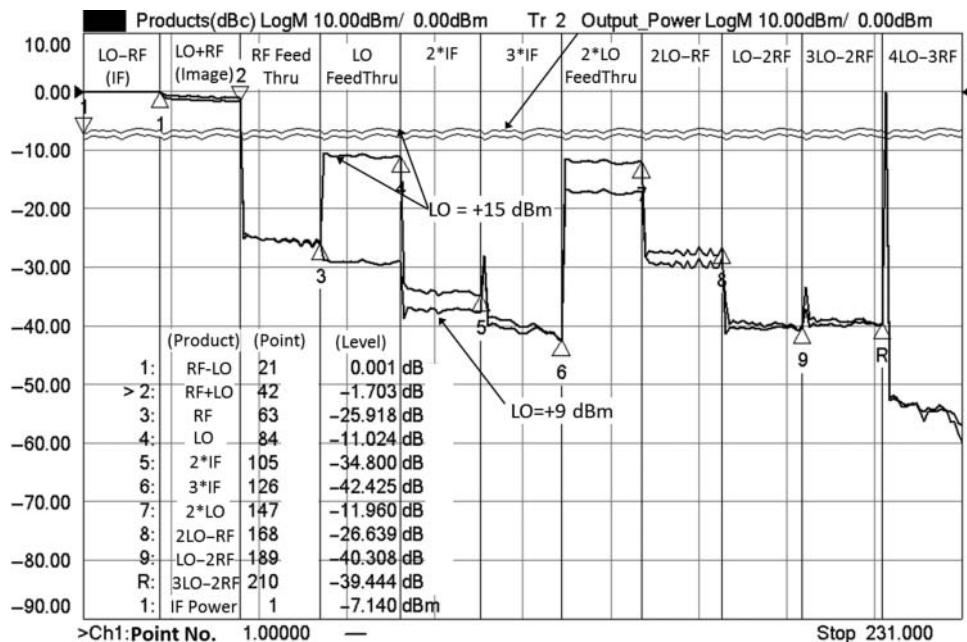


Figure 7.64 Higher-order products as a function of LO drive power.

In fact, the only products that are reduced with higher LO drive power are higher-order products of the RF signal, with the IF output third-harmonic (3^{rd} IF) showing the most (though meager) improvement. The second harmonic of the IF, in fact, gets worse with higher LO drive.

Using similar techniques, a wide range of mixer product measurements can be performed simply and quickly. The entire measurement cycle time for this measurement is less than 250 ms. Because each segment can have an independent IF BW, one can decrease it for higher-order products down to as low as 1 Hz and obtain more than 100 dB of dynamic range (approximately -120 dBm noise floor) for these measurements, without slowing down the measurements of the other products.

Newer methods allow one to more directly specify the measurements of spurious by specifying a Differential and IQ Mixer Application (DIQ) (discussed in detail in Section 7.10.1.2) whereby a single sweep of a source frequency can be defined with several different sets of receiver frequencies. In this way, a complete trace of each high-order product can be displayed as the RF frequency is swept. Figure 7.65 shows the result of sweeping the RF input, while holding the LO constant, and measuring at each point the power at the output associated with the frequency of the higher-order product (or mixer spur). In the figure the swept frequency response is shown for several orders of spurious signals.

In a similar way, the spurious signals may all be measured relative to other parameters. For example, in Figure 7.66, the spurious signals are measured as a function of the LO drive

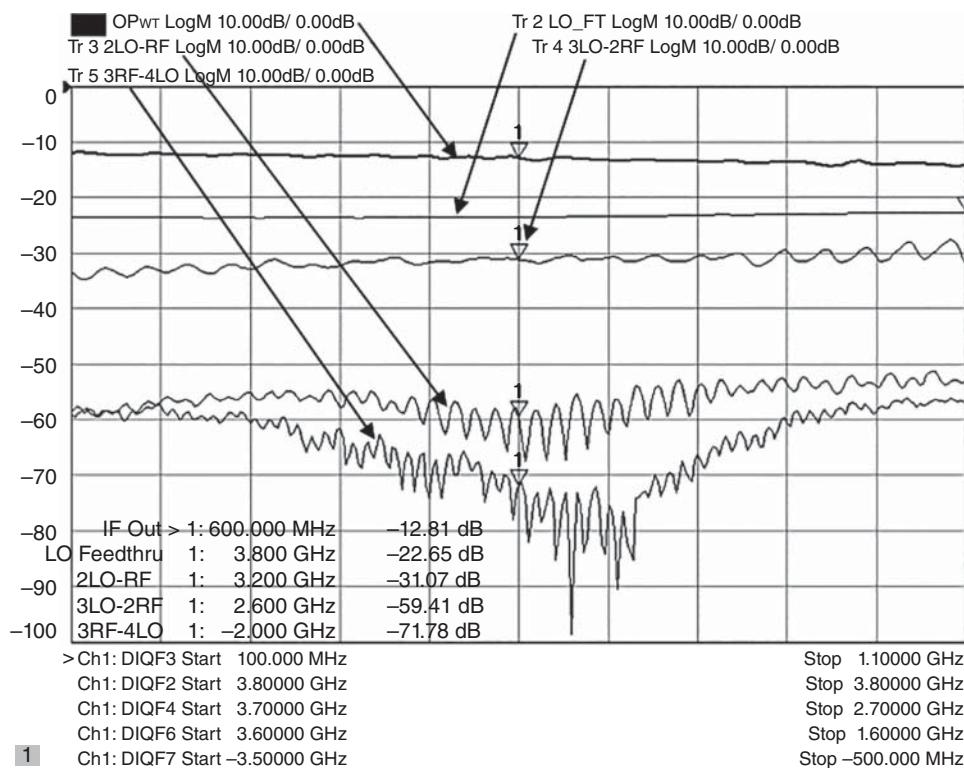


Figure 7.65 Using DIQ to measure higher-order products.

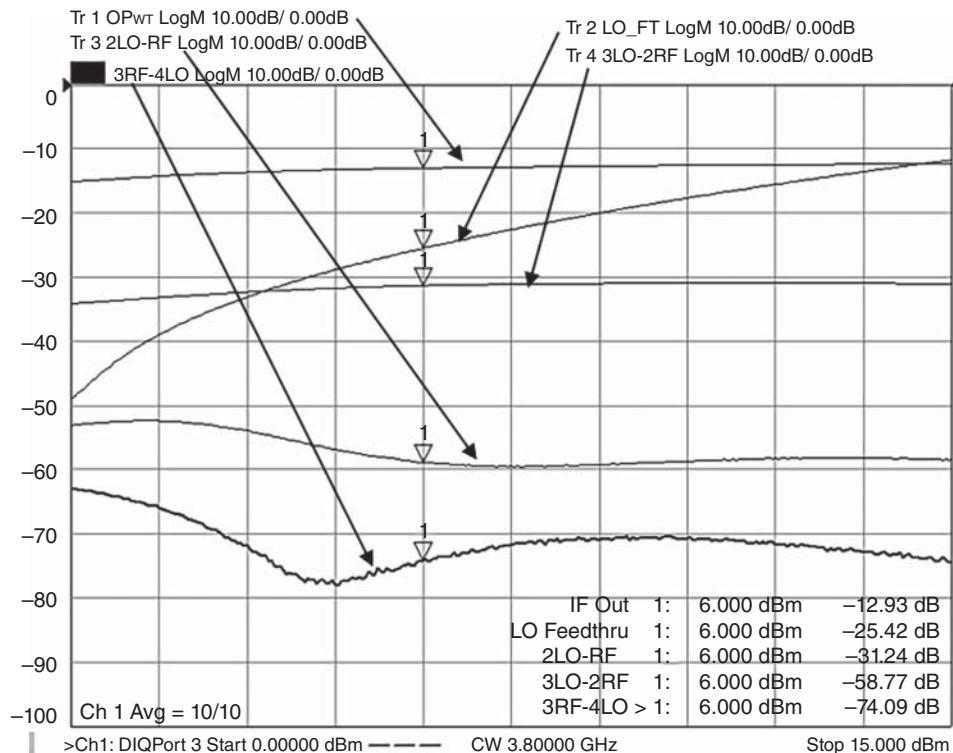


Figure 7.66 Measure of spurs versus LO drive power.

power. The 0:1 spur (zero times RF minus LO, or LO feed-through) shows a strong change in output versus its own drive level. The higher-order spurs actually have an optimum LO drive for lowest spurious level.

7.9.3.1 Swept LO Group Delay

Broadband mixers are often tested with sweeping LO and RF frequencies at a fixed IF frequency so that a single measurement can show the behavior of a mixer as a block down-converter. This works fine for amplitude response, but sweeping the LO frequency does not allow valid phase or group delay responses to be measured. Using segment sweeps can solve the problem of swept LO measurements for group delay by creating segments with a fixed LO, swept-RF, and swept-IF measurement, over all of the LO frequencies. While tedious to do using the GUI, it is a straightforward setup using remote-programming commands and will provide the proper measurement of the delay at each LO frequency. An example of such a measurement is shown in Figure 7.67, where a portion of the segmented frequency table is shown along with the group delay result. In this case a 201 point LO sweep was desired, so 201 segments are required.

In this measurement example, the VNA used is a Keysight PNA, and it treats the delay format somewhat differently in a segment-sweep mode that is particularly useful in this application. As the frequency spacing, and even LO frequency, can be different in each

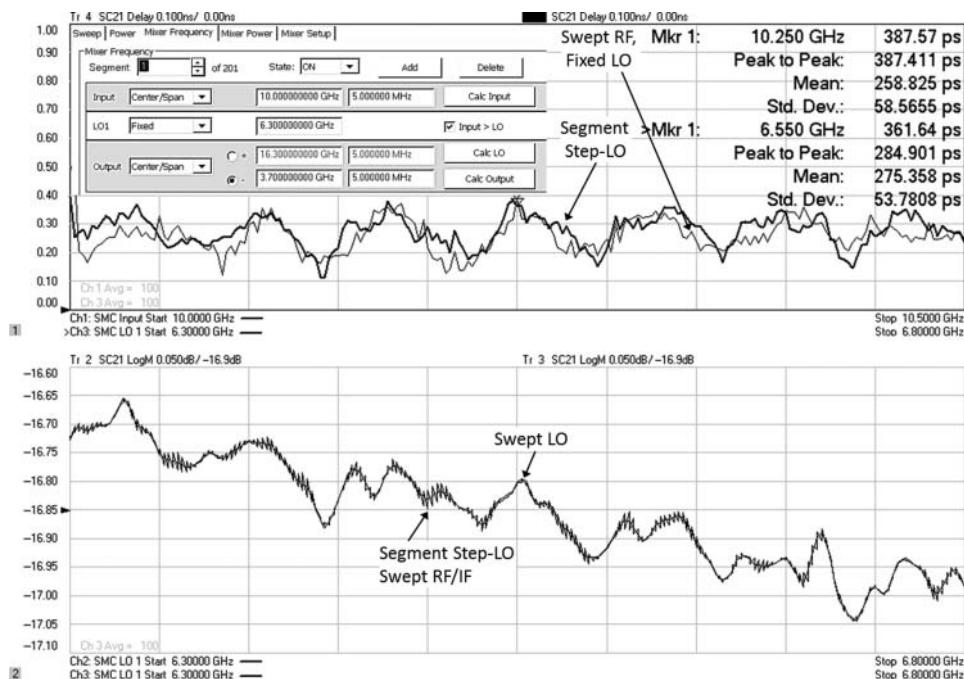


Figure 7.67 Segment sweep provides a way to properly create a swept LO delay measurement (upper plot). The lower plot shows the magnitude response for normal sweep and segmented sweep.

segment, the group delay computation is performed on a segment-by-segment basis, and so the delta-phase/delta-frequency computation occurs only in a segment and doesn't cross segment boundaries. In this example, exactly two points are used in each segment, with the frequency spacing of the two points at in each segment providing the delay aperture. Each segment computes the delay independently, and as in normal delay formatting, either the first or last point is repeated so that the delay format has the same number of points as the underlying measurement. Thus, each segment computes a different delay for the different LO, and for a two-point segment, both points are assigned the same value. Displaying the segmented sweep with the LO frequency as the x-axis provides for a direct measurement of group delay versus swept LO frequency, for a single fixed-IF measurement. The calibration of this measurement is best performed using the phase-reference method, as it does not require any wideband reciprocal mixer, but any of the phase measurement methods can be used as well. Also shown in the upper window is a measurement of the converter delay with a fixed LO, sweeping the RF and IF. The fact that both traces are very similar (within 100 ps) indicates that the RF frequency range contributes most to the ripple and response of the group delay.

The lower windows shows an interesting cross-check of the segmented sweep. One trace is for a swept-RF, swept-LO, and fixed IF measurement, and it shows the amplitude response. The other trace is for the segmented sweep. The trace is exactly as expected, as the segment sweep measures the RF to IF response, at each LO frequency, but sweeping the RF from just below to just above the normal RF frequency from the standard measurement. This allows computation of the delay response at each stepped LO frequency, and the fact that the

sawtooth wave exactly crosses the normal swept-LO response trace indicates a valid setup, measurement, and calibration for this test case.

7.9.4 Mixers with an Embedded LO

All of the mixer measurements presented so far rely on absolute knowledge of the RF, IF, and LO frequencies. Some frequency converters (especially ones used in satellite applications) have their own internally supplied LO, and further, this LO is not accessible and may not have a common 10 MHz reference for the use with the test system. This measurement scenario is commonly referred to as an “Embedded-LO” use case and further implies that the LO frequency may not be well known. Even in the highest quality instruments, without locking the common references together, their internal RF frequencies may differ by many kHz of frequency. Consider that a source with a frequency accuracy of 1 ppm will have an error of 30 kHz at 30 GHz.

When measuring a converter with an embedded-LO, the IF frequency can be offset by this difference in frequencies, when the embedded-LO is not derived from the same reference as the test system source. For amplitude measurements, this usually doesn’t provide much of a problem because the IF BW or resolution BW of the measurement receiver can be widened to always capture the IF signal (although it will lead to higher trace noise). In phase measurements, however, even 1 Hz of offset will give 360° of phase change per second of measurement time, and thus virtually no offset can be tolerated between the defined frequencies in the test equipment and the frequency of the embedded-LO.

One technique that can overcome the problem of LO offsets in some modern VNAs is the use of a software signal tracking and phase locking mechanism. In this technique one point of the sweep, typically the center point, is designated as the software-locking point. Before each measurement sweep for mixer response occurs, some background sweeps are performed where the LO frequency is determined with great accuracy, and in a very fast manner. This is accomplished by first taking a coarse sweep of the receiver while holding the input frequency fixed. This gives an approximate IF frequency (to the resolution of the receiver’s IF BW filter) from which a first estimate of the LO frequency is derived, as illustrated in Figure 7.68. Next, the receiver is fixed-tuned to the estimated IF frequency, and a phase-versus-time sweep is performed, in a so-called precise mode. The slope of this response is the frequency error between the estimated LO frequency and the actual frequency. The resolution of this sweep is the trace noise of the phase sweep divided by the sweep time. The sweep time is set by the inverse of the receiver bandwidth and the number of measurement points. If the phase change is large, the offset frequency is re-computed and another phase-versus-time sweep is acquired; this continues until the desired offset error is achieved. In practice, an offset frequency of less than 0.3 Hz is required to have less than 100 ps of error in absolute delay for a typical measurement. In this implementation, the IF bandwidth is chosen so that 0.3 Hz offset produces less than a 1° change in phase. The phase-versus-time sweeps are also shown in Figure 7.68. For this implementation, the acquisition of the LO frequency can happen very quickly, perhaps less than 50 ms, so that even a drifting LO can be tracked on a sweep-by-sweep basis.

The performance of a software-locking loop is quite good, and at these rates it emulates quite well the hardware locking mechanism of tying the 10 MHz reference of the test equipment to that of the converter’s LO, as demonstrated in Figure 7.69. Here a mixer

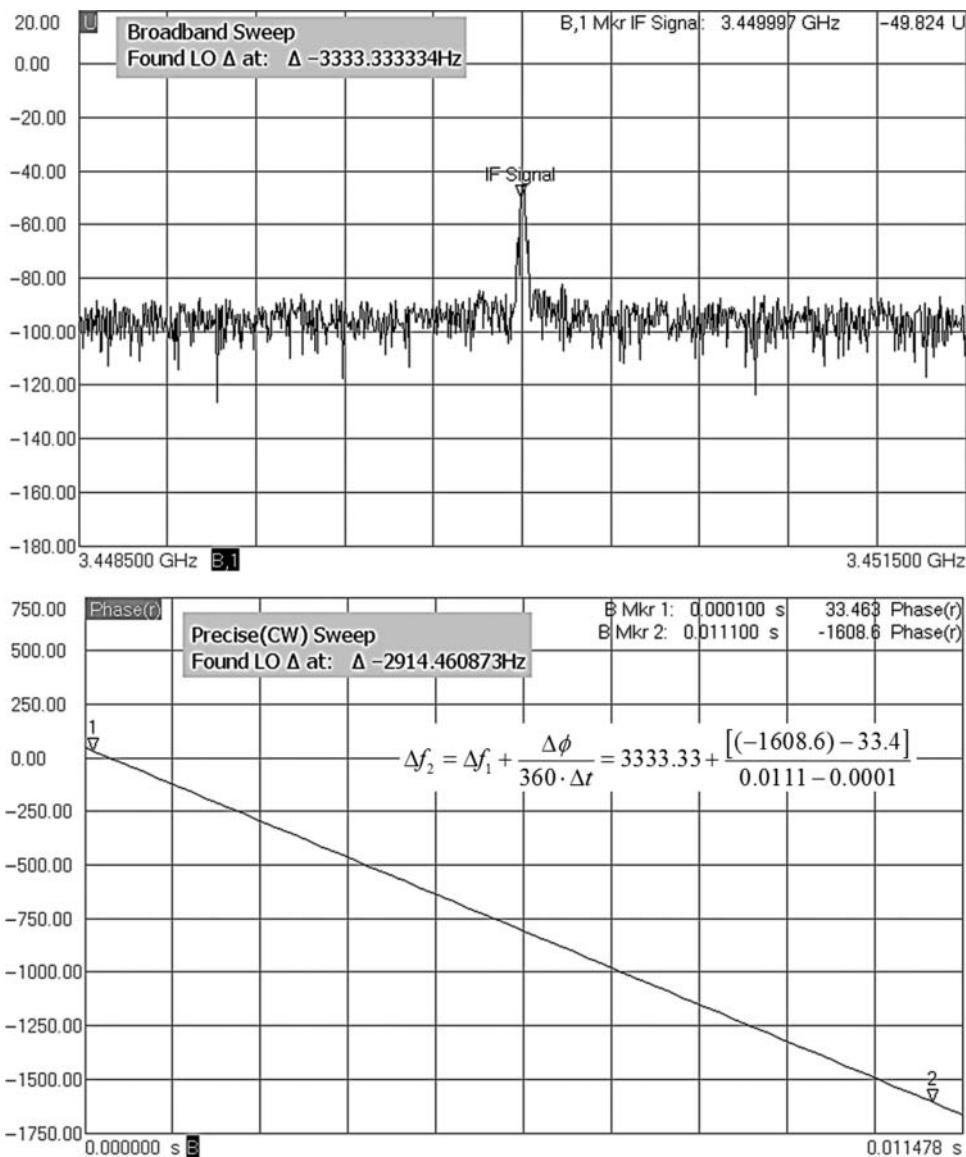


Figure 7.68 Background acquisitions for software locking of an embedded LO.

that has an embedded LO that does have a 10 MHz reference input is measured in the two conditions of software locking and locking the 10 MHz references together. The measured group delay values are identical, and the trace noise of the group delay traces, which is a measurement of the stability of the LO locking, is nearly identical, as shown by the trace statistics computed for the center portion of the measurement. Marker 2 is measuring the maximum deviation between two sweeps, both with a 10 MHz lock. Marker 3 is measuring

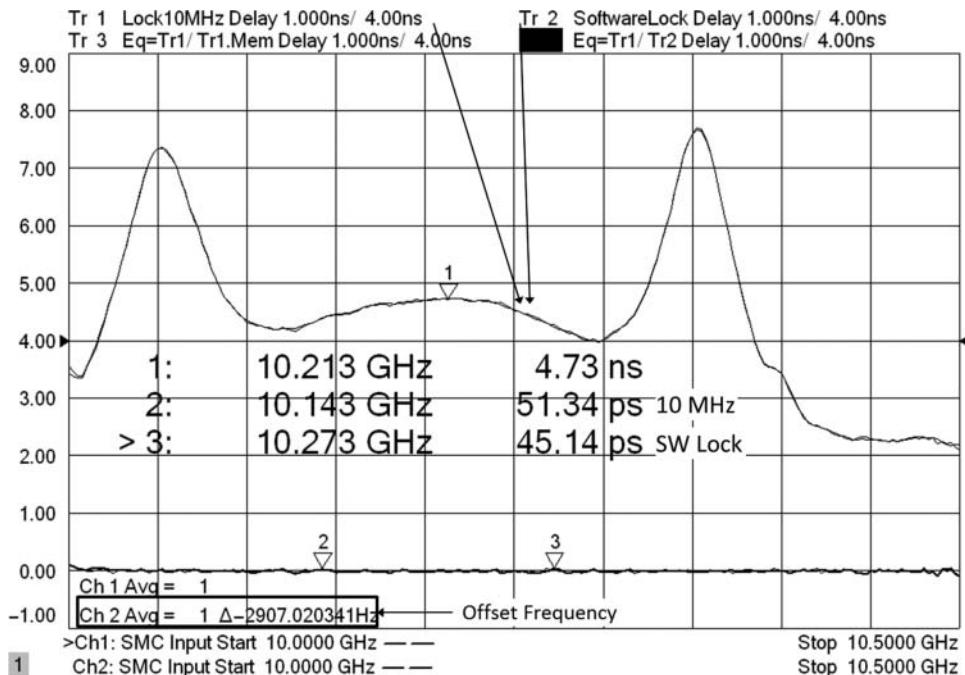


Figure 7.69 Comparison of software versus hardware locking for embedded LO measurements.

the maximum deviation between a software lock sweep and a hardware lock sweep; in fact, the maximum deviation is less than that between two 10 MHz locked sweeps.

Also shown in the Channel 2 display is the offset frequency when the 10 MHz signals are not locked together. This drifts 2–5 Hz over a few seconds, and 5–10 Hz over several minutes. This software locking technique provides an alternative to the modulation measurement technique described in Section 7.4.4 and can be performed with a standard VNA, without requiring a modulated or two-tone source.

7.9.5 High-Gain and High-Power Converters

Another use case that requires special consideration is the measurement of converters with either high gain or high power or measurements at the end of lossy or long test cables. In these cases, special care must be taken to avoid overdriving the converter, and as such, some of the match correction techniques described in Section 7.5.1 can cause more harm than good. One implication of a high-gain converter is that the input power is typically quite low; some satellite converters have input powers below -100 dBm . In this case, the S_{11} or input match measurement can be very noisy indeed. For high-power converters, the output signal may need to be attenuated so that the output does not overdrive the measurement receivers. This attenuation can cause the measurement of S_{22} to become noisy and thus to embed that noise on the SC_{21} trace through the output match correction. In these cases, turning off the input and

output match corrections can improve the measurement results, and some VNA measurement applications provide this as an optional choice in correction method.

Generating an amplitude calibration with a high-gain converter can be problematic, as the typical method for calibrating the input power is to use a power meter to measure the source power at the RF frequency. But there doesn't exist any commercial power meters that can measure a signal at -100 dBm , so a modification to the normal techniques needs to be used.

First, for very low source power levels, it is best to reverse the port 1 test-port coupler, to lower the drive signal while improving the noise performance of the S_{11} measurement. Next, the VNA source attenuator needs to be set to a level such that the source automatic-level-control (ALC) will provide the desired output power at its minimum level, then the power meter acquisition can be performed at its maximum level. For most commercial power sensors, even with high-sensitivity heads, the minimum power is -70 dBm , and one should stay 10–20 dB above the minimum power to avoid noise in the power reading from degrading the power calibration. This means the minimum power for calibration is on the order of -60 to -50 dBm . Most modern VNAs provide about 40 dB of ALC range, so the minimum power for measurements can achieve the -100 dBm level. The details of these setups for low-power source drives are discussed in Section 6.3.

An alternative method for calibrating the input power at low powers when high-sensitivity power sensor is not available, is to calibrate at a higher power using a higher attenuator setting and then change the attenuator setting after calibration. The error due to the change in attenuator setting can be de-embedded from the final measurement by characterizing the difference in attenuator settings using a standard VNA channel, over the input frequency range. This is accomplished by performing a 2-port calibration in the low attenuator setting (not 0 dB) and then, without changing any other setting, change just the source attenuator and save the S2P file. This now contains the error in the attenuator setting (there is slight error in the source-match that can usually be ignored if the low attenuator setting is not 0 dB). The mixer calibration is performed at the low attenuator setting, and then the attenuator can be changed and the S2P file de-embedded from the mixer calibration to create a high-quality calibrated measurement at a very low source power setting. This is similar to the method described in Section 3.14.3.

When measuring high power mixers, it may be necessary to add an attenuator to the output of the mixer to reduce the power applied to the receiver. Calibration in this case can take one of two forms. First, the mixer calibration can be performed with the attenuator in place. The difficulty with this method is that if the mixer is also a high-gain mixer, the source drive during calibration will be quite low and the added loss of the attenuator can cause substantial trace noise in the calibration response. In such a case, the only option is to increase averaging or reduce IF BW during calibration.

An alternative method is to first perform a 2-port calibration at high power as described earlier and then calibrate with the attenuator in place, changing the source attenuator after calibration.

7.10 I/Q Converters and Modulators

I/Q converters describe a class of mixers commonly used for creating or receiving complex modulated signals. Under this class of converters, the signaling information is contained in

both the magnitude and phase, or more precisely, the real and imaginary parts of the signal. The real part is called in *I* or in-phase portion, and the imaginary part is called the *Q* or quadrature-phase. *I* and *Q* signals are rotated from each other by 90°.

While I/Q modulators and down-converters have been around for a long time, their importance has grown as the new 5G standards have allowed very wide, multicarrier signaling. The key advantage to an I/Q converter is that the bandwidth of the *I* and *Q* channels can be one-half that of the signaling bandwidth. At higher signaling rates, the cost and complexity of the digital-to-analog converter (DAC) in the transmitter and the ADC in the receiver can become a limiting factor, so reducing the speed-performance requirement by half is of great value and makes the additional complexity of an IQ converter well worth the trouble.

7.10.1.1 I/Q Figures of Merit

There are some key figures of merit that are distinctive to I/Q mixers, and these have traditionally been somewhat difficult to measure using normal methods (sources and spectrum analyzers). For down-converters and I/Q modulators, these are as follows:

I and Q response: This is the frequency response of each of the *I* and *Q* channels. Each can be measured rather simply in using an SMC+phase measurement, as described in Section 7.4.3.3. But each must be measured in a separate channel. In this measurement, the phase response is often typically normalized at one point (usually the center frequency) and so that the relative phase between *I* and *Q* channel cannot be determined in the normal measurement.

I/Q amplitude and phase match: The matching of the *I* and *Q* amplitude and phase response is important in re-constructing the modulated signal, either in generating the up-converted signal or after individual detection of the DAC in each of the *I* and *Q* channels in a down-converter. The ideal case is that the *I* and *Q* channels are equal amplitude response and equal phase response with exactly 90° offset between the *I* and *Q* channels.

Image rejection: When the *I* and *Q* signals are combined in a complex way (as the real and imaginary parts of a complex number), any signals present at the image frequency should be canceled. This is the image-rejection of the I/Q down-converter. In general, for a down-converter, one must measure the *I* and *Q* signal at the same time to ascertain their proper phase relationship. In an up-converter, the image rejection is a measure of a signal present at the image output-frequency relative to the normal output-frequency, when the *I* and *Q* inputs are exactly 90° out of phase. In the past, this was measured using a physical hybrid to generate the 90° quadrature shift, but hybrids are often limited in frequency response. New methods, described next, greatly simplify this measurement.

LO and RF feedthrough: For a down-converter, the LO and RF feed-through might be specified, but they are not usually a problem in the system as they can be easily filtered, typically occurring at a much higher frequency than the IF output. For up-converters, LO feed-through can be very important, as it will appear in the center of a complex modulated signal. In some cases, the modulation format excludes the center-frequency from the modulation scheme, but in many cases this LO feed-through can limit the capability of generating the desired signal. IF feed-through is usually not an issue with an I/Q up-converter.

7.10.1.2 Measurements on I/Q Up-Converters (I/Q Modulators)

The measurement of the key parameters of I/Q up-converters has been greatly simplified by new measurement techniques such as the “Differential and I/Q” measurement class of the Keysight PNA. This application allows, for the first time, to sweep two sources, the I and Q inputs to the modulator, and maintain exactly 90° offset between the two signals, while measuring key outputs of the converter. In fact, one can set up the application to measure, at the same time, the I/Q gain, the I/Q image rejection, and the LO leakage. One can separately measure the I gain individually or the Q gain individually. And, presuming the I and Q channels are not exactly equal, or not offset by exactly 90°, one can find the amplitude and phase offset at any point a particular frequency.

The measurement setup for the I/Q up-converter is shown in Figure 7.70. Port 1 and port 3 are used to drive the I and Q ports, respectively, and a separate LO, in this case a Keysight

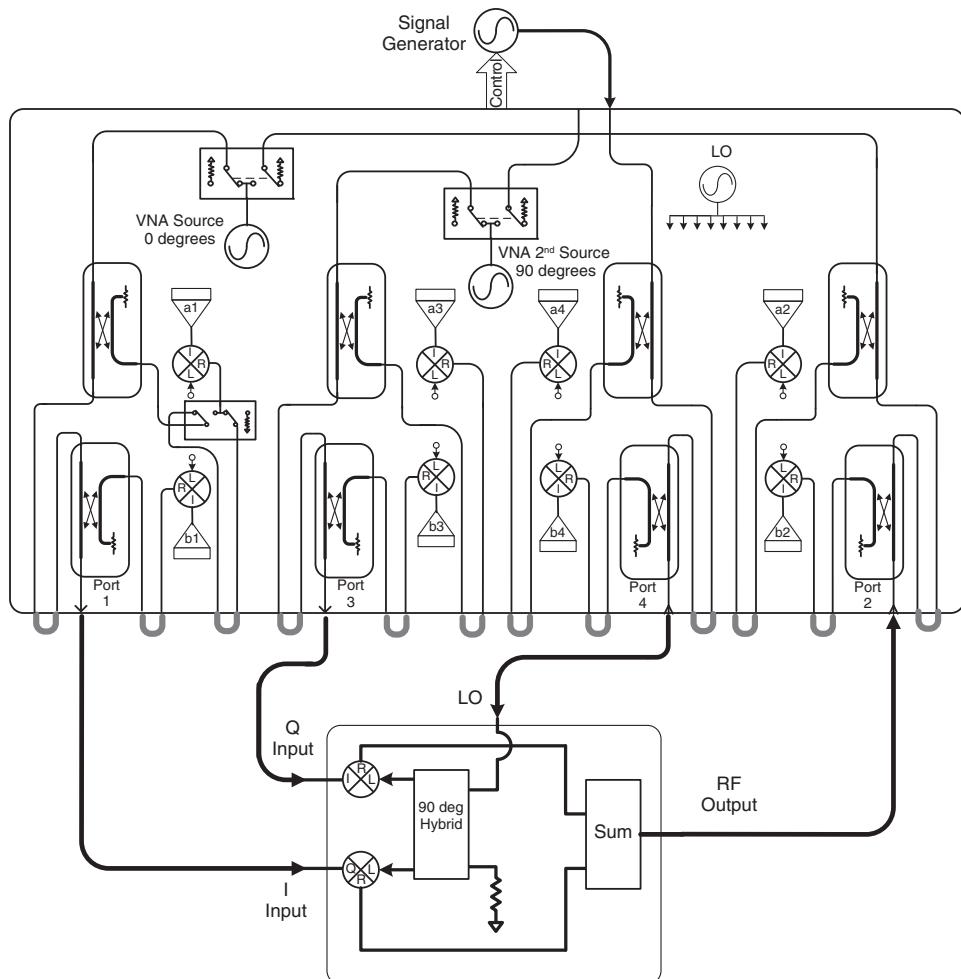


Figure 7.70 Measurement setup for IQ up-converter testing.

PSG, is used to drive the LO port of the modulator. If it is desired to measure the match of the LO input on the I/Q converter, one can route the external LO through port 4 of the VNA (external inputs are provided on some VNAs), and finally the output is measured on port 2 of the VNA.

The sweep setup is similar to that used to measure multichannel converters, but in this case the phase relationship between the I port and Q port must be maintained at 90° as the input frequency is swept. An example of setup user interface (UI) is shown in Figure 7.71. Here four frequency ranges are set. F1 is the input frequency range, and port 1 (I source) and port 3 (Q source) are both assigned to frequency range F1. The external source is controlled by the application and assigned to frequency F2. During the measurement, the I and Q sources are set to the first point of the frequency range of F1 and measured on the a_1 and a_3 receivers. Their relative phase is measured, and any error in the phase setting is determined (including effects of cables and fixturing). Then the sources are reset to adjust their phase, in an iterative way, until they are exactly 90° offset (within the tolerance of the iteration loop, typically 0.5°). During the same iteration the frequency of the external source is set to the first point of F2. If receiver amplitude leveling is turned on, the amplitude of the sources is also adjusted on each iteration. Each iteration measures all the VNA receivers source frequency. Iteration stops when all the criteria, amplitude and phase targets, are met. The receiver values are saved for all the receivers, and then the receivers are adjusted to each of the other frequency ranges, while the sources are held constant, to measure each port at each frequency, and the values are saved.

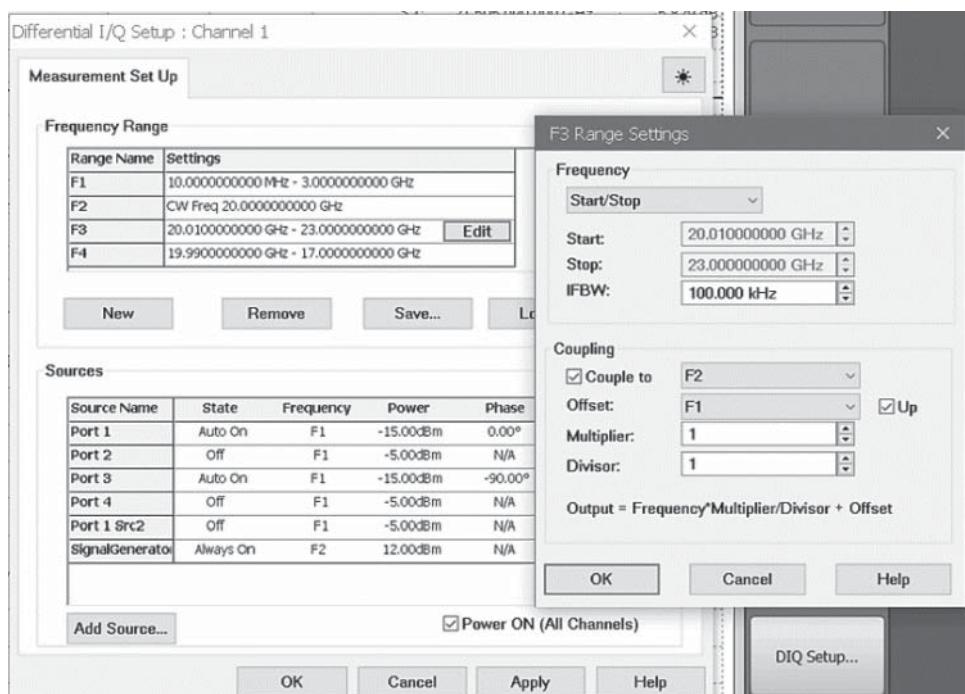


Figure 7.71 IQ up-converter setup UI.

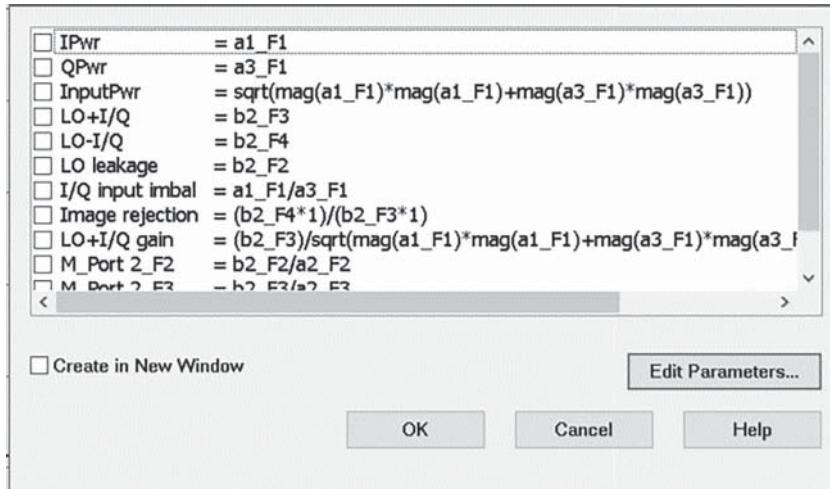


Figure 7.72 Example trace definitions for IQ up-converters.

This occurs for each point across the frequency sweep. At the end of the sweep, any mathematical arrangement of the receiver measurements may be created and assigned a trace name.

Some example trace definitions are shown in Figure 7.72. With this application, the normal configuration for creating traces allows any combination of four receivers, at any of the frequency ranges, in the form of

$$Trace = \frac{(\{ab\}i_Fx) \quad \{+-\times/\} \quad (\{ab\}j_Fy)}{(\{ab\}k_Fw) \quad \{+-\times/\} \quad (\{ab\}l_Fz)} \quad (7.35)$$

Where a and b are the reference or test receivers, i, j, k , and l can be any receiver number, F_x, F_y, F_w, F_z is any frequency range, or the receiver may be replaced by the value 1. So, as an example, the relative balance of the input signals can be defined as

$$\text{I/Q Input Imbal} = \frac{(a_1-F_1) \cdot (1)}{(a_3-F_1) \cdot (1)} = \frac{(a_1-F_1)}{(a_3-F_1)} \quad (7.36)$$

But for I/Q mixers, when describing the total input power, the limitations of the simple entry of (7.35) doesn't provide the necessary mathematical functions. In such a case, one may utilize the equation editor to define the proper equation. In the case of an I/Q input, the total input power is not the direct sum of the I and Q waves, but rather it is the sum of the power applied at the I port and the power applied at the Q port, such that the total input power is defined as

$$\begin{aligned} InputPwr &= \sqrt{|a_1-F_1|^2 + |a_3-F_1|^2} \\ &= |cpx(|a_1-F_1|, |a_3-F_1|)| \end{aligned} \quad (7.37)$$

where $cpx()$ is a complex number formed by the magnitude of the I and Q waves in the real and imaginary positions. Similarly, the power gain for an I/Q mixer, in the up-converted frequency range (that is LO + I/Q frequency), is defined as

$$\text{LO + I/Q gain} = \frac{(b_2 - F_3)}{\sqrt{|a_1 - F_1|^2 + |a_3 - F_1|^2}} \quad (7.38)$$

where F_3 is the sum of the LO plus I/Q frequency, and F_1 is the input frequency. Note that we must form the gain from the output power and input power, from measures of both a_1 and a_3 at the input frequency, but combined as powers.

Some measurement results for an IQ up-converter can be seen in Figure 7.73. In the upper window is an SMC+phase measurement of the magnitude of the I channel and the Q channel gain, measured independently. That is, the Q power is turned off when the I gain is measured (SC21 in the plot), and the I power is turned off when the Q gain is measured (SC23 in the plot) is measured. This is performed on two channels of the VNA. The second plot shows the independent phase response (with 300 ps electrical delay added to flatten the phase). Note that

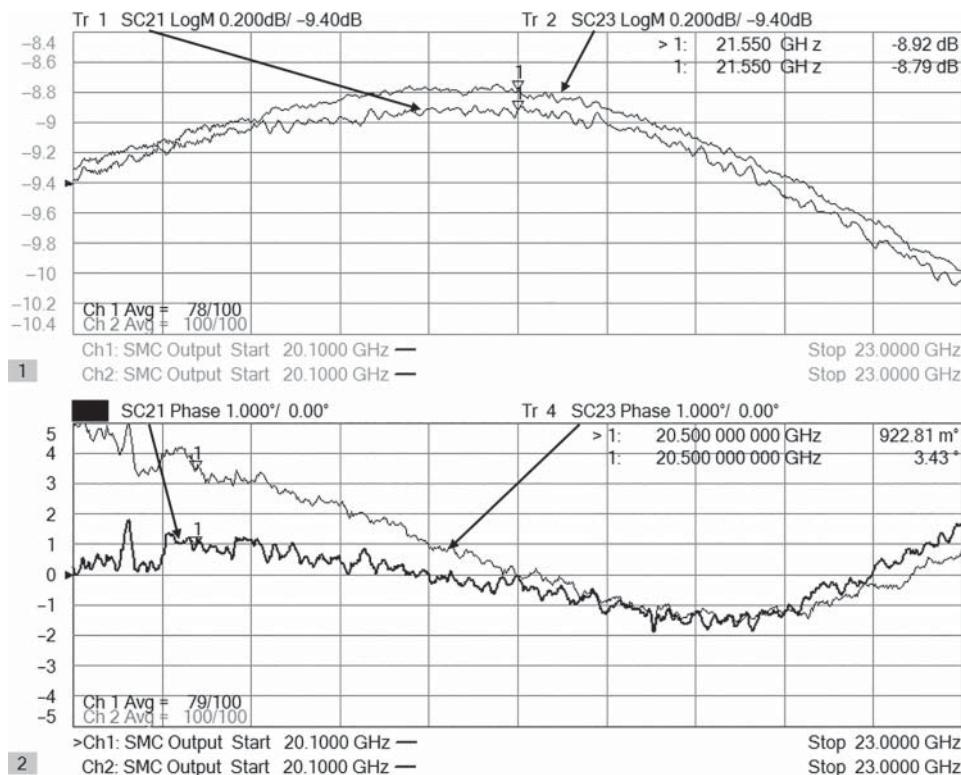


Figure 7.73 Measurement of IQ up-converter.

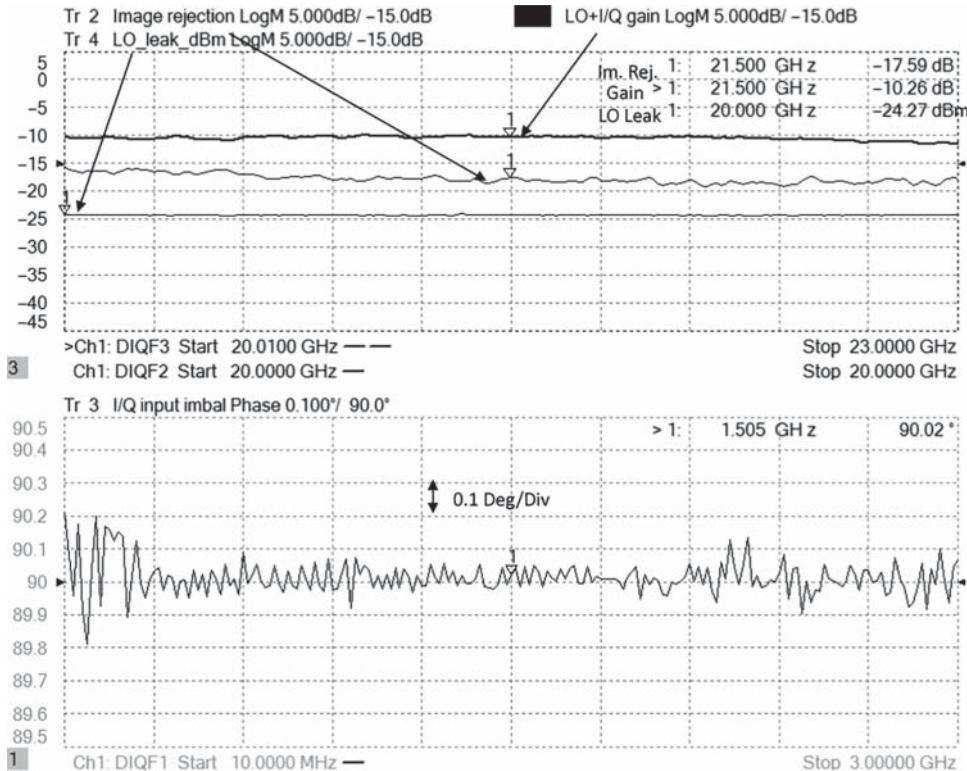


Figure 7.74 LO feed-through, image rejection and gain (upper); I/Q input imbalance (lower).

in this SMC measurement class, for each measurement the phase is normalized to the center frequency point, so while the I and Q phase-deviation response shows the proper deviation across frequency, the absolute offset between I and Q is not available. The amplitude of I and Q agree within about 0.13 dB; the phase offset between the two paths represents about a 4.5 ps delay difference between the paths. There may be an additional phase offset, but since both I and Q phase are normalized at the center point, we cannot use this measurement to determine the overall phase offset between I and Q (where the expected value is 90°). That overall offset can be determined using a different measurement class, as described next.

The DIQ measurement class may be used to directly measure the gain, LO leakage, and image rejection of the converter. As set up in Figure 7.71, the I and Q signals are maintained at a 90° quadrature while the input frequency is swept across the I/Q bandwidth. Shown in Figure 7.74 (upper window) is the I/Q gain (for up-conversion this is called the LO + I/Q gain), the LO feed-through power (in dBm), and most importantly the image rejection (in dBc).

The IQ gain shown, at -5.85 dB, is almost exactly 3 dB higher than average of the I only and Q only gain shown in Figure 7.78. One advantage of I/Q mixers is that they are SSB mixers so the voltage cancels at the image frequency and doubles at the desired frequency.

Thus, the output power is 6 dB higher; but of course the gain is only 3 dB higher as we must consider both the I power and the Q power in the computation of gain as in Eq. (7.38).

Image rejection is a key measure of the quality of the I/Q converter, and it will directly affect the quality of the modulated signal. It is measured as the relative power of the desired signal to the relative power of the undesired signal, and if this mixer is used as a SSB converter, it will be the level of the unwanted sideband. Here we see the image rejection changes as a function of frequency from just less than 15 dB at the low frequency to just about 18 dB at higher frequencies, but overall it is relatively flat. Further, it is not clear what limits the image rejection, which is typically caused by either amplitude or phase imbalance in the I and Q paths.

Shown in the lower window is the I/Q input imbalance, or I/Q phase relationship of the input signal. The source from port 1 and port 3 are phase controlled (within some tolerance) to maintain a 90° quadrature offset, and the plot shows the error in this setting is less than 0.25° across the band. This level of imbalance would lead to better than 40 dB of image rejection, so we must conclude there is some phase offset that would generate this image rejection.

To further understand the limitations of the image rejection, the DIQ application allows one to hold the frequency constant and sweep the phase of the Q relative to the I. Alternatively,

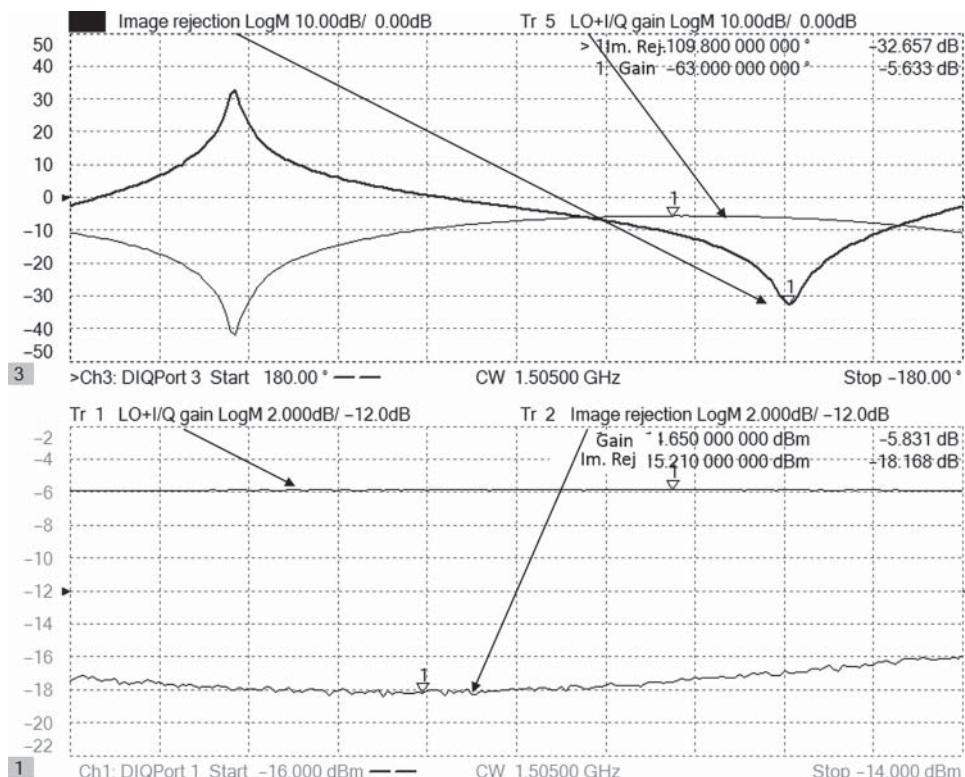


Figure 7.75 Image rejection as a function of amplitude and phase imbalance.

one may also sweep the power of the Q path relative to the I path and thus create a measure of the amplitude imbalance in the mixer. Figure 7.75 shows the result of sweeping the phase of the Q relative to the I, from 180° to -180° , while monitoring the image rejection, in the upper plot. The minimum of the image rejection trace shows the phase that gives an indication of the I/Q offset, and in this case it is offset from the nominal 90° by about 20° at -109.8° . This may be caused by nonideal phase shift in the internal hybrid of the mixer. Note that the point of maximum image signal (poorest rejection) is at a similar offset on the opposite side (approx. $+110^\circ$). Also, errors in the amplitude balance can cause small shifts in the apparent phase offset.

Similarly in the lower plot, the I/Q amplitude imbalance is found by looking for the point of the maximum image-rejection versus I/Q power offset. Here the I channel is swept from -1 to $+1$ dB of nominal, and the Q channel is swept from $+1$ to -1 dB of the offset. Note that the amplitude offset is about 0.2 dB, indicating very symmetrical mixer conversion-loss in the two mixers that comprise this I/Q modulator. This matches very closely with the measured amplitude offset in Figure 7.73 above.

This characterization can be repeated for as many frequency points as desired to determine if the amplitude and phase offset changes with frequency of the I/Q path. In the case of this mixer, it was repeated at the low and high frequencies, and the value of phase shift is almost identical across the I/Q span, indicating that the offset is likely due to some small offset in the LO hybrid path. The nominal offset is about 19° , with about 4° change over the input band of frequencies, matching the phase linearity results from Figure 7.73.

This offset might be caused by some small pathlength difference. The path-length difference, at 20 GHz, can be computed as

$$\text{Delta_path} = \frac{\lambda}{360} \Delta\phi = \frac{c}{f} \cdot \frac{\Delta\phi}{360} = \left(\frac{3 \times 10^8}{20 \times 10^9} \right) \cdot \left(\frac{18}{360} \right) = 750 \times 10^{-6} = .75 \text{ mm} \quad (7.39)$$

Thus, a fraction of a mm path difference in the LO path or the RF combining path would account for this measured phase offset.

Finally, it is possible to correct, in I/Q drive, any linear error in phase, and offset in amplitude, in the I/Q input by sweeping the phase, and amplitude, of the source signal along with the frequency sweep. Sweeping phase as a function of frequency is essentially the same as adding some electrical delay to the path, by noting that the delay from shifting phase linearly with frequency can be represented as

$$\begin{aligned} \tau_{\text{delay}} &= \frac{\delta\theta_{\text{rad}}}{\delta\omega} = \frac{\pi}{180} \frac{\delta\theta^\circ}{\delta\omega} = \frac{\pi}{180} \cdot \frac{1}{2\pi} \cdot \frac{\delta\theta^\circ}{\delta f} = \frac{\delta\theta^\circ}{360 \cdot \delta f} \\ \tau_{\text{delay_discrete}} &= \frac{\Delta\theta^\circ}{360 \cdot \Delta f} \end{aligned} \quad (7.40)$$

So adding a linearly changing phase offset over the frequency range can be used to emulate the effect of adding a delay line. In the case of this converter, the phase offset and amplitude offset were nearly uniform across the band of frequencies. The expected image rejection as a function of the offset in the I/Q, at the 90° quadrature point, is

$$\text{Image_Reject}_{dBc} = 20 \cdot \log 10 (\tan(\Delta\phi/2)) = 20 \cdot \log 10 (\tan(18/2)) = -16 \text{ dB} \quad (7.41)$$

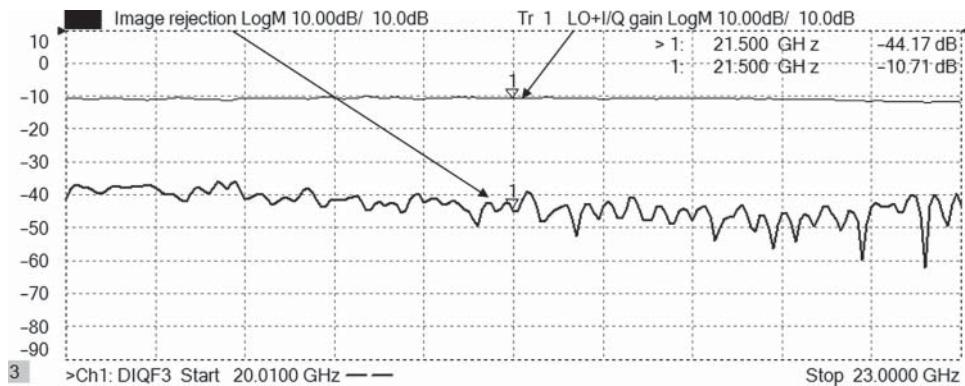


Figure 7.76 Image rejection after a fixed phase and amplitude offset in the I/Q inputs.

which is close to the value observed for image rejection in Figure 7.74. If the input I and Q sources are adjusted with just a fixed offset in phase, and the slight offset in delay, the resulting image rejection is dramatically improved, as shown in Figure 7.76.

Here we see that the rejection has improved to be 20 dB lower over the entire frequency range.

7.10.1.3 Measuring I/Q Down-Converters

I/Q down-converters are the complement of the up-converters, and from a complex-modulated RF input signal they produce I and Q quadrature outputs. In such a configuration, the DAC used to digitize the signal can have a bandwidth of one-half the RF signal bandwidth. This can be a significant savings in cost and power as the cost of a DAC often increases more than linearly with the sampling frequency. Cutting the DAC sampling rate in half by using a direct I/Q demodulator may be the most cost-effective way to obtain a wide bandwidth system.

The measurement block diagram for I/Q down-converter is shown in Figure 7.77. The setup for measurement is very similar to that for an up-converter, but in this case only one source is needed to drive the RF, so the second internal source of the VNA may be used as the LO.

The measurement parameters for an I/Q down-converter are somewhat similar to the up-converter with I and Q flatness being equally important, as well as I/Q balance (amplitude and phase) and image rejection, which has a small complication to allow measurement.

The overall output power and gain of an I/Q down-converter is

$$\text{IQ Pwr} = b_2 F_3 + \text{cp}(0, 1) \cdot b_4 F_3, \text{ where } \text{cp}(0, 1) = e^{j90} \quad (7.42)$$

$$\text{I/Q gain} = \frac{b_2 F_3 + \text{cp}(0, 1) \cdot b_4 F_3}{(a_1 F_1)}, \text{ where } \text{cp}(0, 1) = e^{j90} \quad (7.43)$$

Here, the complex quadrature addition of the I and Q output must be employed. To measure image rejection one must set up two separate measurements. The first one measures power

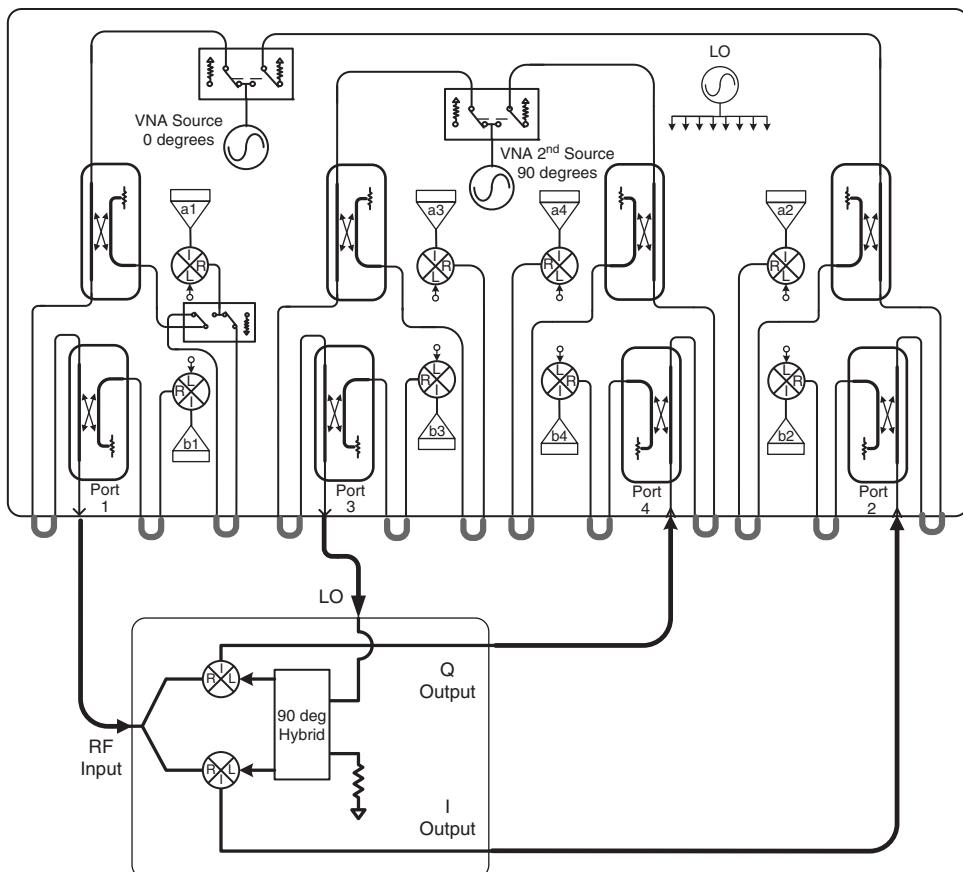


Figure 7.77 Measurement block diagram for an I/Q down-converter.

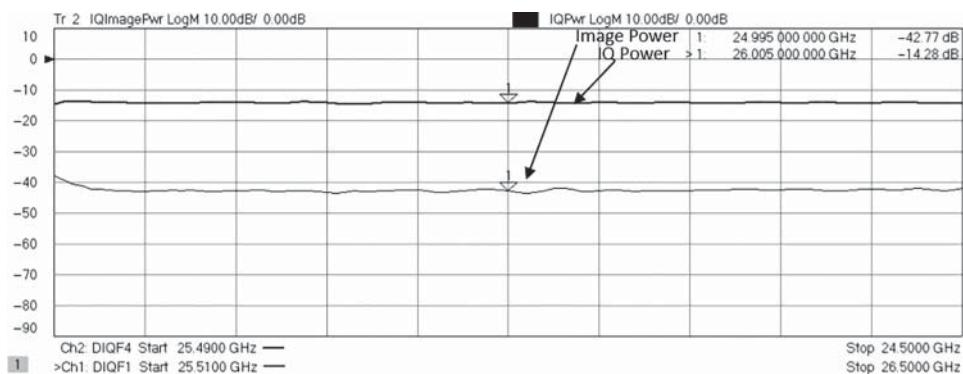


Figure 7.78 IQ power and image power on an I/Q down-converter.

as denoted in Eq. (7.42), where F_3 is the IQ output frequency. For a normal down-converter, the RF frequency is greater than the LO frequency, by the IF frequency. Two channels may be employed to allow a single display of the image rejection. For the second channel, the F_1 frequency is changed to be the image input, where for a normal converter the image frequency will be lower than the LO frequency by the IF. The image power is computed in the same way as (7.42) because there is an inherent 180° shift in the conversion due to the F_1 frequency being below the LO for the image case, and the ratio of image power to normal power is the image rejection. Figure 7.78 shows the normal power and image power for the IQ converter. Here the image rejection is about -28 dBc (the difference between IQ power and image power in the plot).

In addition to overall gain, the individual response of the I and Q channels can be measured, along with their relative response. The relative response of the I and Q channels can be measured directly by taking the ratio of the I and Q outputs as

$$IQ_Bal = \frac{(b_2 - F_3)}{(b_4 - F_3)} \quad (7.44)$$

The measurement in Figure 7.79 shows the absolute response of the I and Q channels (in the upper plot), as well as their relative response. The lower left shows the amplitude imbalance in the I and Q paths, and the lower right shows the phase offset between the I and Q channels.

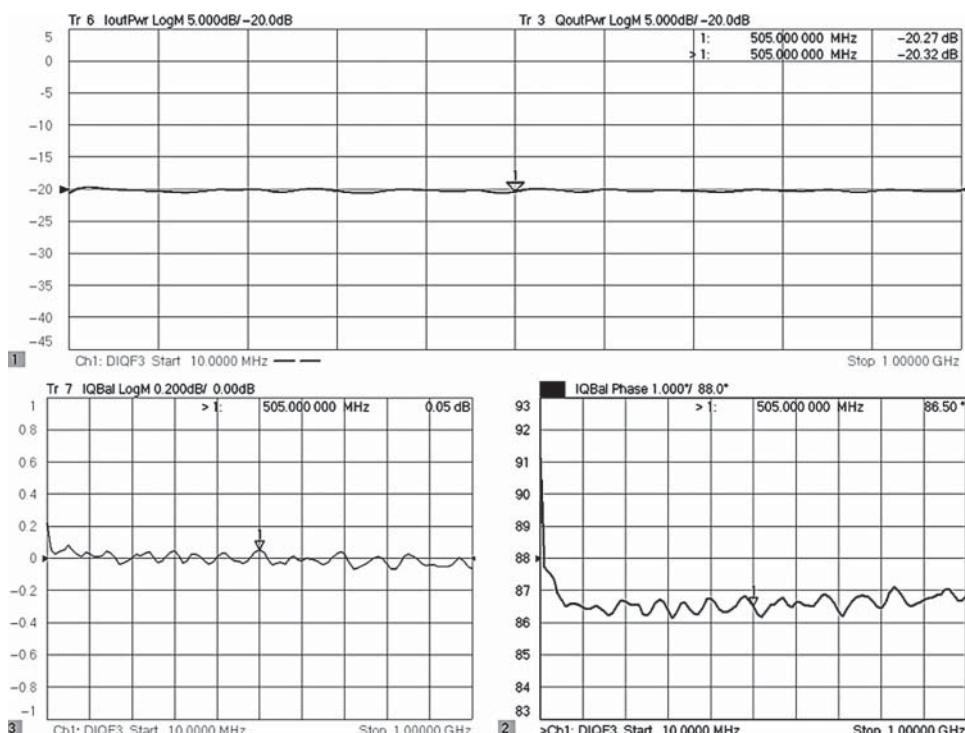


Figure 7.79 I and Q individual and relative amplitude (upper), and phase (lower) response.

Here the amplitude response is nearly ideal with less than 0.1 dB difference in the I and Q output response. The phase response is very flat, with about 2° p-p across the band, but with a nominal offset of about 3.3° . This agrees very well with the measured image rejection. If the amplitudes are the same, the image rejection due to the phase imbalance can be computed as

$$\text{Image_Rej} = 20 \cdot \log(\tan(\text{Phase_Bal}/2)) = 20 \log(\tan(3.5/2)) = -30.3 \text{ dB} \quad (7.45)$$

This value is not the true image rejection, as it must be computed from the ratio of the signal at the input frequency as measured at the output to the signal at the image frequency as measured at the output and so requires two stimulus signals. The slight discrepancy between Eq. (7.45) and the image rejection measured in Figure 7.78 is likely due to the differing response of the mixer at the true image frequency.

As for other parameters, all the normal attributes of a mixer such as gain compression, or variation of other parameters with respect to LO drive or RF drive can be performed on an IQ mixer as well.

7.11 Conclusions on Mixer Measurements

Mixer and frequency converters can have very complex responses and are particularly demanding to characterize. Bare mixers have a wide variety of conversion products that emanate from all the ports, and remixing of these products can cause ripples and other errors in mixer characterizations. The performance of mixers depends in complicated ways on the RF and LO drives, and one must often characterize each desired condition as the behavior is not necessarily well controlled, especially across a wide frequency band.

Besides gain, phase and delay responses of mixers require great care in calibration and measurement. Several different test and calibration methods have been discussed, and the interested reader is recommended to pursue further study through the references provided. The test methods for these parameters are being constantly improved, and great progress has been made in understanding and simplifying converter delay test.

Mixer distortion has likewise seen major improvements in the measurement methods, many of which have been described here. In particular, the ability to quantify distortion as a function of RF or LO drive has been greatly simplified, while the accuracy of the measurements has improved.

Mixer noise figure remains a difficult measurement, and great care is still required in setting up and making these measurements, particularly on low-gain converters. The specific construction of the mixer, image-rejection, and LO noise considerations must all be considered to understand the quality of the NF measurement (see Chapter 9).

Finally, many special cases of frequency-converter testing including high-gain, high-power, and embedded LO have been reviewed. Many of the specialized test methods for these conditions are derived from similar requirements on amplifier test, so one may want to review the material from Chapter 6 to gain further insight into these situations.

While demanding, the theoretical basis for mixer behavior is now relatively well understood, and the test engineer facing some new challenge is advised to return to these first principles. The basics of test remain the same: understand the device through pretest, optimize the measurement system, calibrate with appropriate standards, and analyze the measurements for consistent, reasonable results.

References

- Joel Dunsmore, 2009 “Understanding Uncertainties in Mixer Measurements”, WHWE11 (EuMC 2009 Workshop) Practical Approaches to Achieving Confidence in Microwave Measurements; workshop notes.
- Dunsmore, J., Hubert, S., and Williams, D. (2004). Vector mixer characterization for image mixers. In: *Microwave Symposium Digest, 2004 IEEE MTT-S International*, vol. 3, 1743–1746, 6-11 June. IEEE.
- Maas, S.A. (1986). *Microwave Mixers*. Dedham, MA: Artech House. Print.
- Williams, D.F., Ndagijimana, F., Remley, K.A. et al. (2005). Scattering-parameter models and representations for microwave mixers. *IEEE Transactions on Microwave Theory and Techniques* **53** (1): 314–321.

8

Spectrum Analysis: Distortion and Modulation Measurements

Most modern communication systems can be roughly divided into two general classifications: constant envelope or quadrature amplitude modulated. Constant envelope systems typically employ FM modulation or phase modulation (sometimes called *phase-shift keying* or PSK). These systems do not rely on the relative amplitude of signal to carry information; only the frequency or phase of the signal is used. As such, these systems are quite immune to noise and distortion in the form of gain compression. In fact, the final amplifiers in many of these systems may be fully saturated amplifiers with substantial amplitude clipping. These amplifiers can achieve very high-power efficiencies since they operate in a state of either fully on or fully off. However, because they do not use the amplitude of the signal, they tend to be much less bandwidth efficient, and their use is diminishing as spectrum becomes more expensive and more data throughput is needed in the currently fixed bandwidth systems.

Thus, many modern standards employ some version of quadrature amplitude modulation (QAM) or orthogonal frequency domain multiplexing (OFDM) where both the amplitude and the phase of the signal contain data. In such a case, the variation in relative amplitude of the signal from point to point in time is significant, which also implies the amplitude of a signal will vary considerably over the modulation period of the signal. This is sometimes called the *peak-to-average ratio*. The higher the data content of the signal, that is, the more levels of amplitude that are used, the greater the variation of the peak-to-average ratio, and the greater the negative effect of compression on the modulated signal quality. These distortion effects are found principally in the transmitter power-amplifiers but tend to not affect receivers as the received energy is usually too low to generate much distortion in the receiver amplifier, and many receivers have automatic gain control (AGC) of the receiver amplifiers to ensure they are not overdriven; however, when mobile receivers are closer to the transmitter, they may also be driven near compression especially by energy in adjacent channels. For transmitter amplifiers, the common figures of merit associated with distortion are adjacent channel power ratio (ACPR), noise power ratio (NPR) and error-vector magnitude (EVM); these will be discussed later in this chapter.

That is not to say that receivers are immune to distortion; rather, the distortion that affects receivers is often caused by other undesired signals that self-mix in the amplifier to create

an intermodulation distortion (IMD) product that lands in the desired band of the receiver. A common measurement performed on a receiver to determine how susceptible it is to this effect is the two-tone third-order IMD test. The two-tone IMD is also often used as a figure of merit for general-purpose amplifiers as a measure of the linearity or power handling of the amplifier. The IMD level of an amplifier is closely related to other linearity measures such as gain compression, ACPR, NPR, and EVM.

8.1 Spectrum Analysis in Vector Network Analyzers

Modern vector network analyzers (VNAs) consist multiple, synchronized down-converters with fully digital IF processing and as such can be reconfigured to be used as different instrument types including tuned receivers, noise figure receivers, and spectrum analyzers. Integrated spectrum analysis capability was first widely applied to VNAs around 2010, but the common implementation was constrained by limitations in the basic VNA capabilities, such as maximum number of points and detection methodologies. As the use for spectrum analysis in component test increased, the methodologies for implementation improved as well, resulting in the modern VNAs, operating in a scalar spectrum analysis mode, being in many ways superior to their stand-alone spectrum analyzer cousins. Much of that improvement came in the spectrum analysis data acquisition and processing.

8.1.1 Spectrum Analysis Fundamentals

The early implementations of spectrum analysis consisted of sweeping the VNA receiver in small steps, narrower than the IF bandwidth, to ensure overlapped measured points. Unlike an S-parameter, where the number of points is determined by the user and where it is typical to have an IF bandwidth much narrower than the radio frequency (RF) step size (thus jumping over regions of measurement between the widely spaced points), in a spectrum analyzer, the actual frequency of the signal is presumed unknown and essentially every frequency must be measured in the measurement span. This is referred to as *swept-tuned* or *sweep mode* in spectrum analysis. For example, if you have a 1 GHz measurement span for spectrum analysis and you use a 1 MHz IF bandwidth, you must take approximately 3000 points across the span to capture an unknown signal with an error of less than 0.3 dB. If exactly 1000 points were measured every 1 MHz, then a signal falling between two measurement frequencies would have about a 3 dB amplitude error. In a traditional spectrum analyzer, with a continuous sweep function, the swept time must be slow enough to let the resolution bandwidth filter respond fully before the frequency changes too far. This limits the maximum sweep rate even in an analog spectrum analyzer.

8.1.1.1 Resolution Bandwidth Filters

The receiver detector bandwidth (BW), which when used in a spectrum analyzer mode is commonly called the *resolution bandwidth* (RBW), refers to the ability to discern between two closely spaced signals, where the definition of resolution means to resolve between two signals where the dip just occurs, as illustrated in Figure 8.1. The light traces show just the

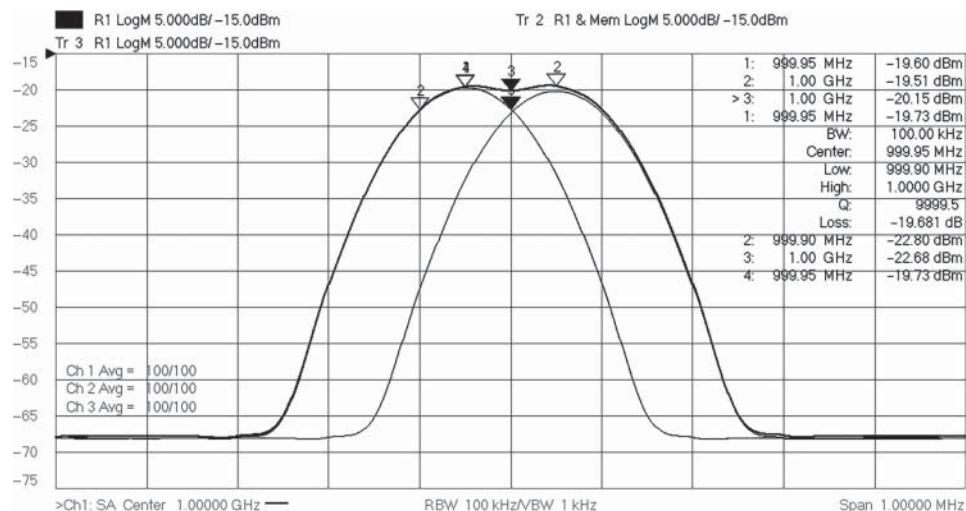


Figure 8.1 Illustrating the resolution of the RBW filter.

lower tone (Tr2) and just the upper tone (Tr3), and they are 3 dB below their respective peaks at the center frequency. But because the tone powers add, the two-tone trace (darker trace, Tr1) shows only the slightest dip. If the two tones were closer than 1 RBW, there would be no dip at all, and one could not resolve that there are two tones.

In fact, this picture occurs only as the average of many sweeps or with a narrow video bandwidth. If one takes a single sweep without averaging, the instantaneous power of the two tones can either add or subtract at the center frequency, causing the trace to jump up and down; but with averaging turned on, the mean value of the signal is obtained. The “bandwidth” markers are turned on for Tr2, which shows the measured BW of the filter is exactly 100 kHz, same as the RBW.

Since all modern spectrum analyzers, including VNA-based ones, use digital filtering, the time needed to acquire the data for the filter is inversely proportional to the bandwidth and depends on the bandwidth shape or equivalently the impulse response shape. A typical VNA filter using a uniform or rectangular response has a response time close to the inverse of the bandwidth but is generally unsuitable for spectrum analysis as the frequency response has sidelobes (spurious pass bands) as high as -13 dBc. Spectrum analyzers typically use either Gaussian filters or so-called Flat-top filters, which have much a flatter response on the top of the filter but have slower responses, between 1.8 and 2.6 times slower than a rectangular filter. These three filters are compared in Figure 8.2; the filter shape refers to the windowing factor in the time domain, where a Rectangular filter has no windowing factor, the Gaussian has an approximately Gaussian shape (approximately because a true Gaussian filter requires an infinite response time), and the Flat-top has a maximally flat in-band response. As a consequence of these filters being digital and implemented with finite filter taps, the filter stop bands are not monotonically going to zero but do have spurious passbands (side lobes) as well as those found in the rectangular window. In fact, these can be considered as convolution of the idealized analytic (e.g. Gaussian) response with a rectangular response of the same length as the total number of taps. Thus, taking the Gaussian filter as an example, while the passband

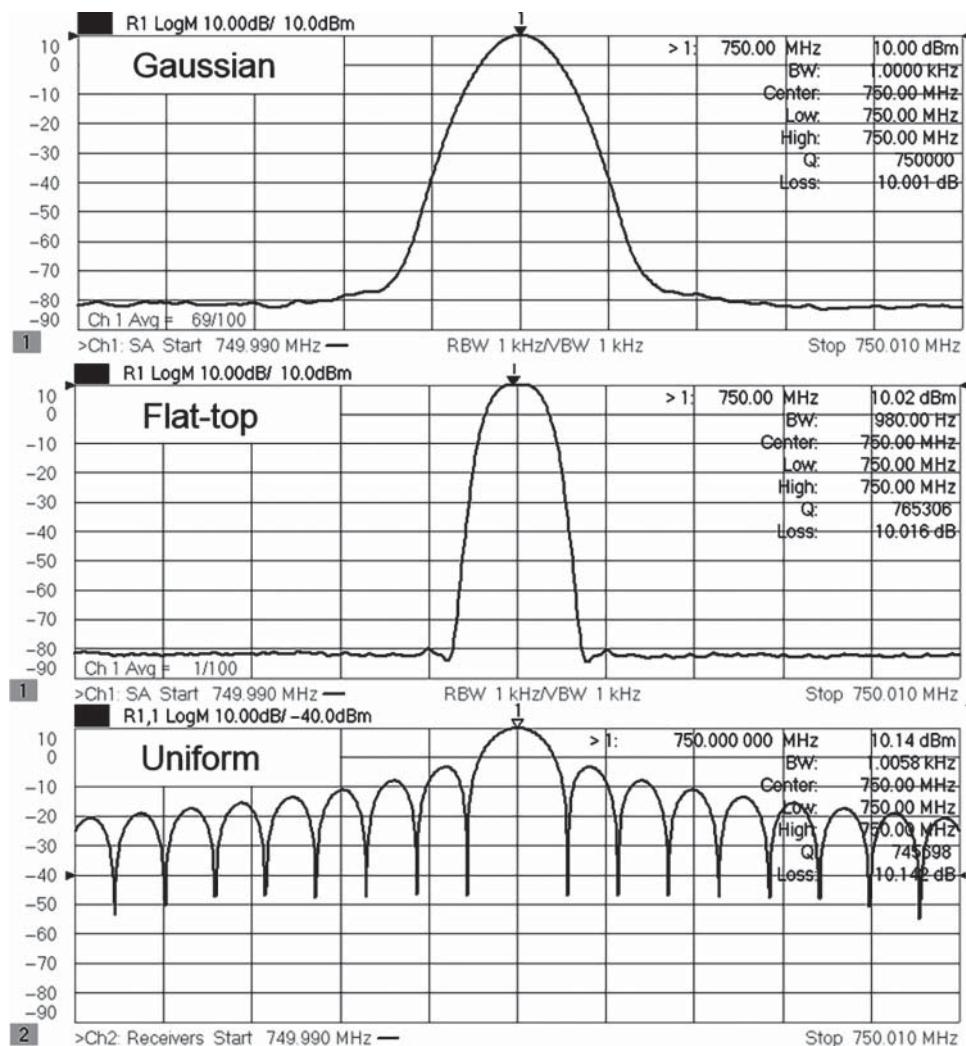


Figure 8.2 Comparison of filter shapes: Gaussian, flat-top, and uniform.

shape is nearly perfectly Gaussian, the stop band shape is determined almost entirely by the effective number of taps chosen. In a typical SA filter, the Gaussian filter length is set to achieve greater than 90 dB side-lobe level in the first side lobe. There is a similar trade-off for the Flat-top filter, where this filter's flatter passband response results in even a larger number of filter taps and thus a slower filter. Because of noise inherent in the measurement system, it is difficult to distinguish the side-lobe level from the noise floor for Gaussian and Flat-top filters, but one can just distinguish side lobes of the Flattop filter. The rectangular window has very large side-lobe levels, approximately 13 dB below the top of the filter.

If two signals are more closely spaced than the resolution bandwidth, then they will appear as a just one signal on the SA display, but because their relative phase changes over time, the

signal will appear to vary in amplitude on a sweep-by-sweep basis; each sweep is acquiring the signals at a different – somewhat arbitrary – time and the relative phase between the signals will be likewise arbitrary.

Two closely spaced signals will have a time varying envelope that is proportional to their spacing. One consequence of having an RBW filter wider than the tone spacing is that it implies the data acquisition time is shorter than the true period of the signal pair. It is as if the SA has sampled only a small time portion of the signal. Since the envelope of the signal varies in time, the samples of the magnitude will also appear to vary sweep to sweep as the SA acquires different portions of the signal on different sweep acquisitions. This sampling effect, commonly seen but not well understood, has significant impact on the measurement of complex modulated signal and will be discussed in detail in Section 8.1.6.

Another aspect of the RBW filter is its ability to lower the apparent noise of the signal going into the SA detector. The broadband noise from the first converter will be filtered out by the RBW filter, and a narrower filter will provide a lower noise power to the detector, which will allow smaller signals to be detected. Of course, a consequence of a narrower RBW filter is that the time to process the signal increases as an inverse proportion of the RBW.

8.1.1.2 Sweeping Speed vs. RBW

One consequence of measuring in a swept-tuned mode is that the measurement time is set by the number of data acquisitions needed multiplied by the time per acquisition. As discussed previously, the number of acquisition points must be on the order of

$$N_{Acq} = \frac{3 \cdot Span}{RBW} \quad (8.1)$$

and the measurement time is on the order of

$$T_{Acq} = \frac{1.8}{RBW} \quad (8.2)$$

(depending upon the filter shape, in this case Gaussian) so the minimum time to measure a particular span is on the order of

$$T_{SweptMeas} = N_{Acq} T_{Acq} = \frac{5.4 \cdot Span}{RBW^2} \quad (8.3)$$

This formulation for swept-mode spectrum analysis means that as the RBW goes lower, more measurement points are needed to cover the span. Early implementations of spectrum analysis in network analyzers had a limited number of display points (e.g. 100 001), which becomes the same as acquisition points, and this maximum number of points in the VNA constrained the ratio of span to RBW. Thus, wide spans with very narrow RBWs were not allowed. The time in Eq. (8.3) represents the theoretical limit. In practice, other aspects such as frequency tuning speed can make the actual sweep time much longer. This applies to traditional spectrum analyzers as well as VNA-based SA modes. In fact, in a traditional spectrum analyzer, the sweep time might be substantially longer due to image rejection circuitry, as discussed in Section 8.1.2. As the RBW becomes small and the per-point acquisition time becomes large, these other attributes become less significant, and the actual sweep time approaches that given in Eq. (8.3).

Further, while clearly expressed in the equation, the reality of the reduction in speed with reduced RBW is sometimes overlooked. An example shows the magnitude of the issue: consider a measurement scenario with a span of 10 GHz, swept with a 3 MHz RBW filter. The total number of acquisitions is about 10 000, and the acquisition time per point is about 0.6 μ s. The total sweep time would be 6 ms, which is very fast. Now narrow the RBW to 3 kHz. The number of measurement points goes 10 million (1000 times more), and the measurement time per point becomes 0.6 ms (1000 times greater), so the overall sweep time becomes 6000 s, a million times slower! This explains why no one used narrow RBW filters on wide sweeps, at least with a swept-mode spectrum analyzer. This is the limitation for all swept-tuned spectrum analysis, whether VNA based or traditional SA based. The Keysight PNA with IMD measurements (option 087) was one of the first VNA systems to offer true spectrum analysis, and it was limited by these attributes of swept-tuned spectrum analysis.

Some more modern SAs have specially calibrated receivers that allow them to compensate somewhat for the amplitude offset associated with sweeping faster than the data acquisition allows, whereby they can reduce the effect of the slowdown at some wider bandwidths, but all slow down dramatically with RBW reduction. This is one of the key issues that brought on the introduction of fast Fourier transform (FFT) spectrum analysis.

8.1.1.3 FFT Mode in Spectrum Analysis

Once digital filters replaced analog filters/detectors in traditional spectrum analyzers, the designers quickly realized that one need not sweep the frequency to capture the spectrum. Rather, with a digital receiver, one can capture the entire spectral content of the signal present in the IF, by capturing the digitized data and applying an FFT. Thus, the span of spectrum that could be captured in one acquisition was limited only by the effective bandwidth of the digitizer used in the spectrum analyzer. However, due to other issues (such as accuracy and image rejection) for normal operations, the spans of FFT acquisitions are in the 10–40 MHz range in current instruments.

In about 2014, VNAs with FFT-based spectrum analysis came onto the market, such as the Keysight PNA option 090. With this introduction, the limitations on data acquisitions were largely removed. No longer was the maximum number of points in the VNA mode a limitation on the FFT SA mode. This also had a dramatic impact on the measurement speed. For an FFT acquisition, the time per acquisition is the same as the time per point for swept-tuned, but instead of getting a single point, the entire BW of the digitizer is obtained. For example, given a digitizer bandwidth of 30 MHz, with an RBW of 3 MHz, 10 FFT “buckets” are acquired representing a frequency every 3 MHz (in fact, due to the filter shape and length, more like 20 FFT buckets are acquired), but the acquisition time is still 0.6 μ s. In the previous example, the time to take the data on a 10 GHz span becomes the spectrum span divided by the digitizer bandwidth times the acquisition time, or

$$T_{MeasDataTaking} = \frac{\text{Span}}{BW_{Digitizer}} T_{Acq} \quad (8.4)$$

In this case, the sweep time becomes about 200 μ s. However, the actual sweep time must include the time it takes to move the local oscillator used in the SA down-converter between FFT acquisitions. In swept-tuned measurements, the local oscillator (LO) is moved continuously. In FFT mode it must be stepped and settled, which might typically take 50–100 μ s.

So the overall time is the FFT acquisition plus the step time, multiplied by the number of steps.

$$T_{FFT_Meas} = \frac{Span}{BW_{Digitizer}}(T_{Acq} + T_{FFT_step}) \quad (8.5)$$

Thus, for this 10 GHz sweep, the total measurement time in FFT mode would be about 17 ms. In traditional spectrum analyzers, there is often a YIG tuned filter in front of the first converter that protects it from unwanted images, and this filter can take milliseconds to settle. Thus, in cases where the RBW is large, the FFT mode is actually slower than the swept-tuned mode.

But to see the real advantage, consider what happens when the RBW is reduced to 3 kHz BW. In this case, there are 333 steps, each acquisition takes 0.6 ms, but each acquisition collects 10 000 FFT buckets. The sweep time includes 50 μ s of stepping time per FFT, so the overall measurement time is about 0.216 s. This is thousands of times faster than the swept tuned in the same RBW condition, which was 6000 s.

One limitation of the FFT-based spectrum analyzer is that the data processing time for an FFT goes up as $N \cdot \log(N)$, where N is the number of FFT buckets. For very large FFT sizes (very small RBW), this time can exceed the data acquisition time, thus further slowing the FFT measurement time.

8.1.2 SA Block Diagrams: Image Rejection: Hardware vs. Software

8.1.2.1 Hardware Image Rejection

Images, that is, spurious artifacts on the display at some frequencies, which look like signals but are really caused by signals at another frequency, are a consequence of the mixing process used to bring the high-frequency signals down to a range where the digital IF can process it. Most spectrum analyzers have circuitry that generally operates in two hardware configurations, depending upon frequency, to remove images from the down-conversion process. Traditional spectrum analyzers used multiple conversion stages, but some modern ones use a single conversion to an IF frequency where the signal is digitized. The image is created because the first converter of the SA has an LO set to some frequency, which can convert signals either an IF above or an IF below the LO frequency. As the LO sweeps through the signal, the signal will appear in the IF twice – once when the LO is below it and once again when it is above it. Only one of these appearances is the desired representation of the signal. Most spectrum analyzers use filtering or multiple conversions plus filtering to isolate the mixer from the unwanted sideband. For example, at low frequencies the input is low-pass filtered and then upconverted by a higher frequency LO to some IF where it is filtered to remove images and then down-converted to an IF frequency that can be detected, as shown in Figure 8.3. This diagram has the advantage that the filters are only fixed tuned, and so the system is quite stable, but it has the disadvantage that the input is “wide open,” and large signals at the input can cause harmonic response in the first converter stage. For example, if there is an input signal at 1 GHz and the LO is tuned to receive the second harmonic at 2 GHz, the input at 1 GHz may cause harmonic distortion in the first converter, and one will see a response at 2 GHz that is not from the signal being measured. The IF filter shown represents the RBW filter in an analog sense and was followed by a logarithmic amplifier that produces a signal proportional

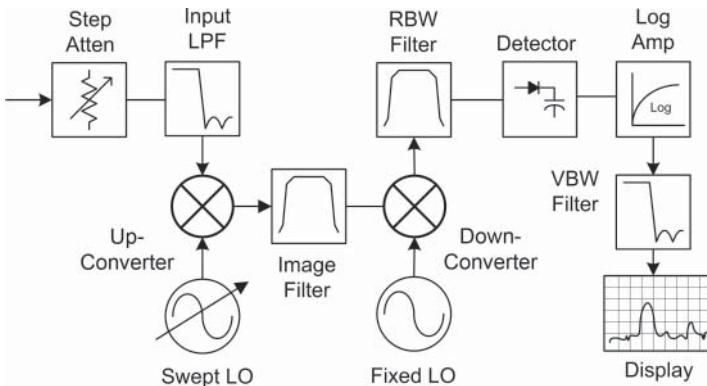


Figure 8.3 Spectrum analyzer block diagram up/down-conversion.

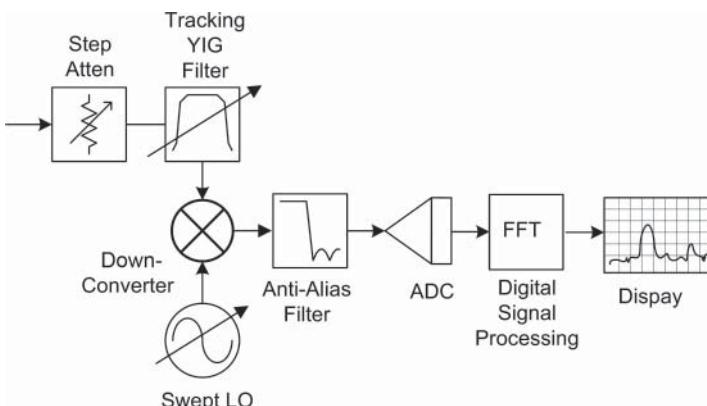


Figure 8.4 Diagram of SA tracking YIG filter, with FFT detection.

to the dB value of the input signal, which was finally detected by an analog detector. This signal was fed to y-axis of a CRT to represent the dB magnitude of the detected signal. There was often an adjustable filter, called the *video bandwidth* (VBW) filter, that could be used to reduce broadband noise from the detector. In modern instruments, everything past the final converter has been entirely replaced by a digital filtering and detection.

But this up/down method is not practical for microwave frequencies and suffers from harmonic generation of the first up-converter when the signal of interest is in the input band of the input filter.

Image rejection at higher frequencies is typically accomplished with a magnetically tuned yttrium-iron-garnet (YIG) filter (but other exotic magnetic materials are used at even higher frequencies), which can be electronically tuned to follow above the LO, as shown in Figure 8.4.

This is called a *pre-selecting filter*, as only signals inside the tuned bandwidth of the filter are present at the input of the first converter (in this case a down-converter), and only the high-side signal is seen by the mixer. As the LO sweeps above the signal, the filter tracks

with it and so the low-side conversion of the mixer, where the LO is above the RF signal, is reduced by the stopband of the filter. This is hardware image rejection. Hardware image rejection is the traditional way in which spectrum analyzers were designed. They have many advantages, chief among them is that the data from the digitized IF represents only the signal of interest. There are some drawbacks to hardware image rejection, however. In FFT mode, the swept-tuned YIG filter is a slow-moving element and often limits the sweep rate of the spectrum analyzer. Also, the filters are limited in their stopband performance, so there are some residual image signals still present in the IF. And if the signal is not time stationary, that is, if it varies over time, the signal must be present when the YIG filter sweeps by it. Depending upon many factors, the sweep rate is typically reduced by 3–10 times due to the YIG filter.

In the block diagram of Figure 8.4, the peak detector, as represented by a detector diode in the previous block diagram, has been replaced by an analog-to-digital converter (ADC), where the entire IF bandwidth can be processed via an FFT to provide spectral content of the IF. In FFT mode, the YIG filter limits the bandwidth that the FFT can achieve, even if the digitizer is a much wider bandwidth. The limitation occurs because the amplitude accuracy is degraded as wider FFTs are performed near the edges of the YIG filter passband.

8.1.2.2 Software Image Rejection

In the VNA-based spectrum analysis, as well as some direct conversion spectrum analyzers such as the Keysight M9393A Modular Spectrum Analyzer, there is no pre-filtering of the signal, and thus the first converter is not image protected in any way. That is, the image signal will have nearly identical level in the IF as the desired signal, as shown in Figure 8.5. This shows the response of the receiver to a CW signal at 1 GHz. This receiver is set to hi-side LO,

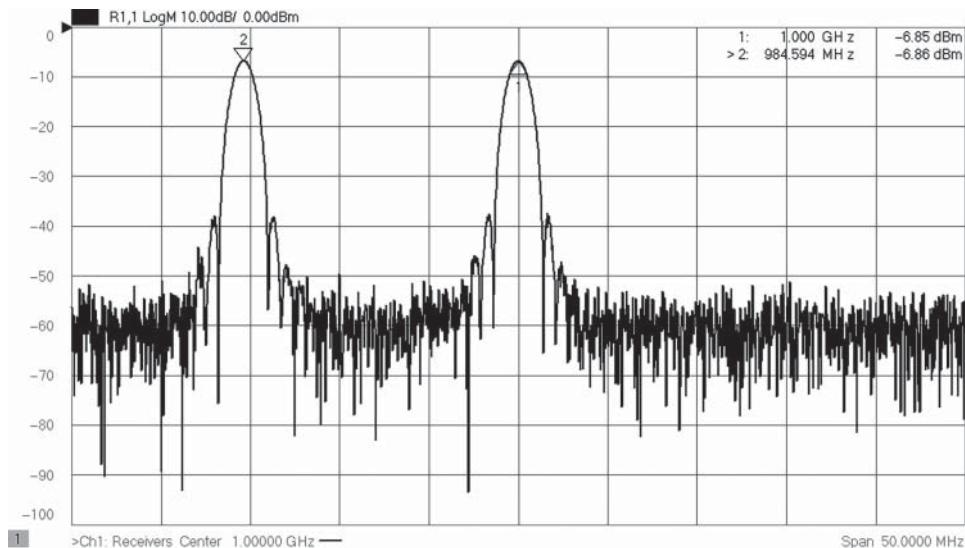


Figure 8.5 Images of a single CW signal, as measured on a non-image-protected VNA receiver.

and in this case the IF frequency is about 7.7 MHz, so to receive a signal at 1 GHz, the LO is tuned to 1.0007 GHz and shows the proper response signal occurs when the LO is above the CW tone. Note the signal at 984.594 MHz. It is an image caused when the sweeping LO is 7.7 MHz below the 1 GHz signal. This is the frequency of the LO for a 984 MHz signal to be measured, but the image of the 1 GHz signal is converted to IF causing the image.

Also notice the side-lobes around each image of the signal. This is caused by attributes of the VNA filter, called the *IF filter*, which is equivalent to the RBW filter and which is designed to measure most quickly and accurately the signals magnitude and phase in the center of the filter but is not appropriate for spectrum analysis. This is similar to the rectangular filter shown in Figure 8.2. To give the same display as a hardware image-protected spectrum analyzer, some means of image rejection is required to remove unwanted images, and a method called *digital image rejection* is utilized.

In the simplest form, the image signal is identified by taking two IF readings for each RF frequency band. The first utilizes the LO below the RF signal (low-side LO), and the second utilizes the LO above the RF signal (hi-side LO). If a signal is an image, it will not appear in the IF in same effective RF bucket for the two cases of LO. If it is the desired signal at the RF frequency, it will appear in both IF measurements. Very early spectrum analyzers utilized a similar method that would allow one to switch from hi side to low side between sweeps; it was called *signal ID*. The image signal would flip from one side to the other of the true signal and could be identified in that way. Modern spectrum analysis, including some PXI spectrum analyzers, as well as traditional spectrum analyzers, used with mm-wave down-converters, and VNA-based spectrum analysis utilize a similar technique by digitally determining if a signal is true or not. This is typically performed by taking the minimum-magnitude of the two IF acquisitions. A set of signals from a complete acquisition (with 2 LO settings) is shown in Figure 8.6. The true signal is the only one that remains in the same relative place in the IF response.

In practice, using just two LO acquisitions can lead to some false images especially when measuring a multitone signal, where the tone spacing might be nearly the same as the IF offset of the two LO acquisitions. In such a case, the image rejection will fail as one or more of the multitone signals may land in the same spot of the IF and be mistaken for a true signal.

However, with proper setting of multiple LOs, the false image issue can be substantially controlled or eliminated. At least one manufacturer, Keysight Technologies, provides several orders of user-selectable image rejection, from none to very high order. The higher order the image rejection, the more overlapped LOs are used to determine the proper signal.

Finally, in the case where the tone spacing of a multitone signal is known, that information can be used to choose best-case LO acquisitions where it is possible to prevent any incorrect image rejection; more about this in Section 8.1.4.

Since multiple acquisitions are required for determining if a signal is an image or not, the digital image rejection can be incorrect in cases of a time varying signal. Time varying signals also present a problem for a traditional swept-mode spectrum analyzer, in that the analyzer may sweep by a region where a time varying (for example pulsed) signal is not on at the instant the SA sweeps past it. FFT-based spectrum analyzers can capture an entire wideband IF acquisition at the same time and so often do a better job of measuring time varying signals if the spectrum analysis span is less than the FFT width. When software image rejection is used, there are some techniques to improve the response to time-varying signals as described in Section 8.1.3.5.

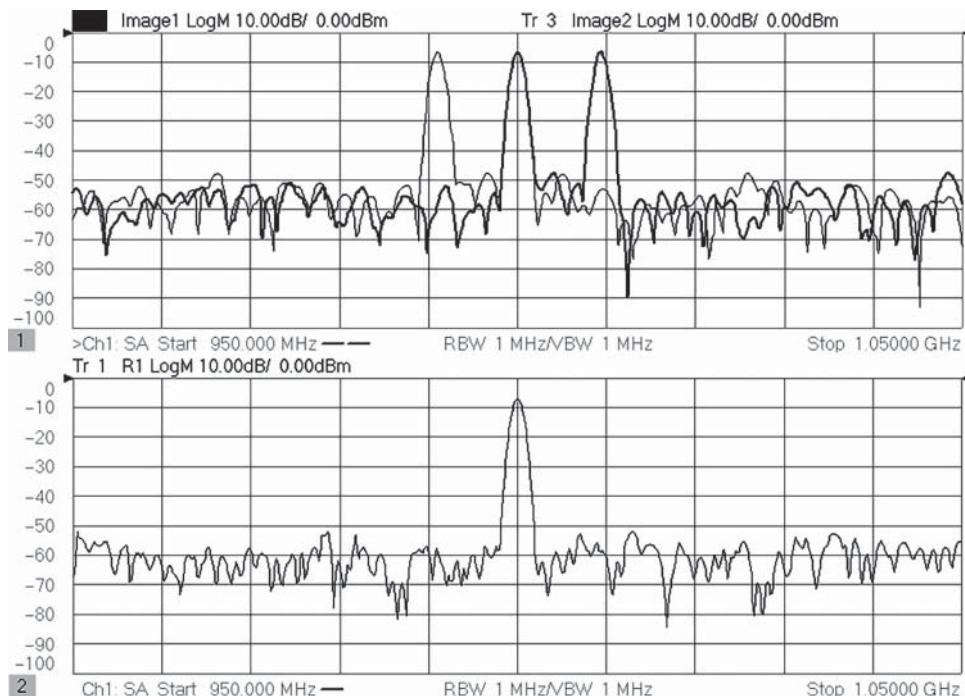


Figure 8.6 Images of two LO settings. Only one signal remains in the same place in the IF.

8.1.2.3 IF Response

One aspect of the FFT-based spectrum analysis that is critical to the performance is the quality of the IF response channel, as well as the correction of the IF response. Unlike swept-mode spectrum analysis, where the signal is intended to always be in the center of the IF frequency, the FFT-based method captures a wide bandwidth of signals all at once. Without IF correction, the raw response will limit the quality and accuracy of the FFT-based spectrum analyzer. IF correction is sometimes performed using nearly ideal signals applied to a wideband SA channel (for example, the Signal Optimizer product from Keysight, for use on the wideband UXA spectrum analyzer). Figure 8.7 shows the uncorrected and corrected response measuring a signal sweeping across the FFT span. The corrected response, shown in Figure 8.8, provides less than 0.1 dB of amplitude error and sets the overall specification for the SA mode accuracy.

While the YIG-tuned filter provides for hardware image protection, the RF response of the filter combines with the IF response of the spectrum channel to set the overall amplitude accuracy. In Figure 8.9, the amplitude accuracy of a traditional YIG-tuned SA response is compared for different FFT IF bandwidths, as well as in a swept-tuned case. For narrow FFT width (400 kHz), the accuracy is quite good. As the FFT bandwidth is increased (40 MHz), it exceeds the YIG filter BW, and amplitude errors become apparent.

Figure 8.10 shows the overall amplitude flatness for a wide and narrow FFT width with the IF calibration applied. As with the traditional spectrum analyzer, the wider FFT has somewhat higher error than a narrower FFT. The error is associated with the quality of the IF

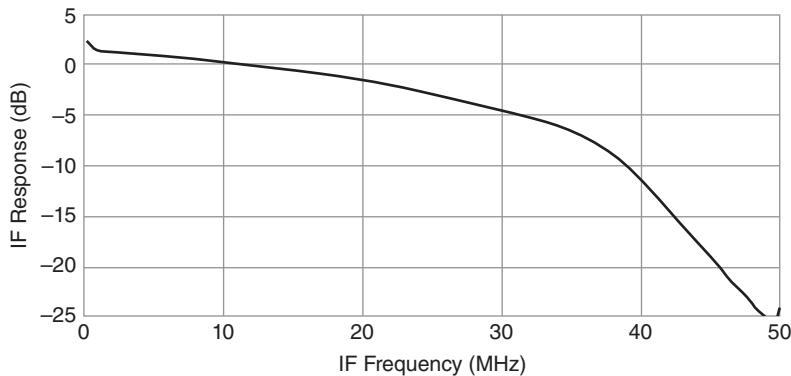


Figure 8.7 Raw IF response of a VNA-based spectrum analysis mode.

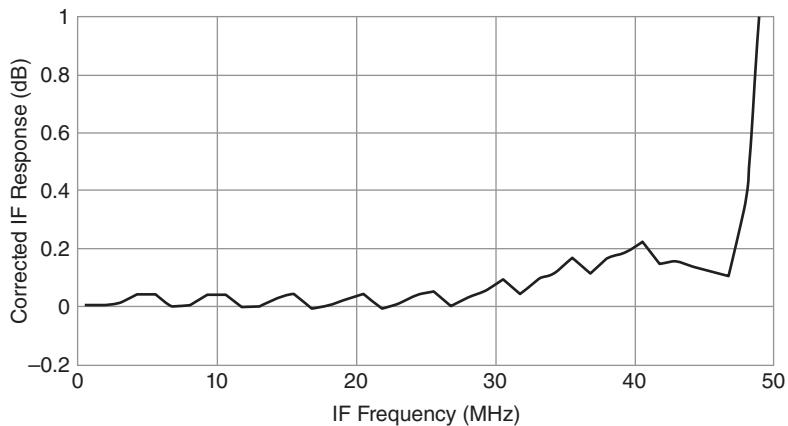


Figure 8.8 Corrected IF response.

calibration, and a wider IF width will accumulate more error, so for best accuracy, choose a narrow FFT width.

The absolute accuracy of a spectrum analyzer (as measured at the test port) includes RF effects before the first converter and IF response effects after the first converter. The RF effects can be removed using a RF response calibration (as discussed in Section 8.1.5). In general, traditional spectrum analyzer block diagrams provide better accuracy when used in the up/down-converter mode, as the signal path is stable and can be accurately calibrated. Where a YIG based preselector is used, the accuracy is degraded both because of the limitations of characterizing the YIG tuned filter (which can have different bandwidth characteristics at different frequencies) and the inability to accurately tune the YIG filter to the same point in the band on repeated sweeps. Finally, when used in a larger test system including cables, connectors, and fixtures, the mismatch that occurs between the external elements and the internal SA circuits is greater when using a YIG-based image rejection block diagram.

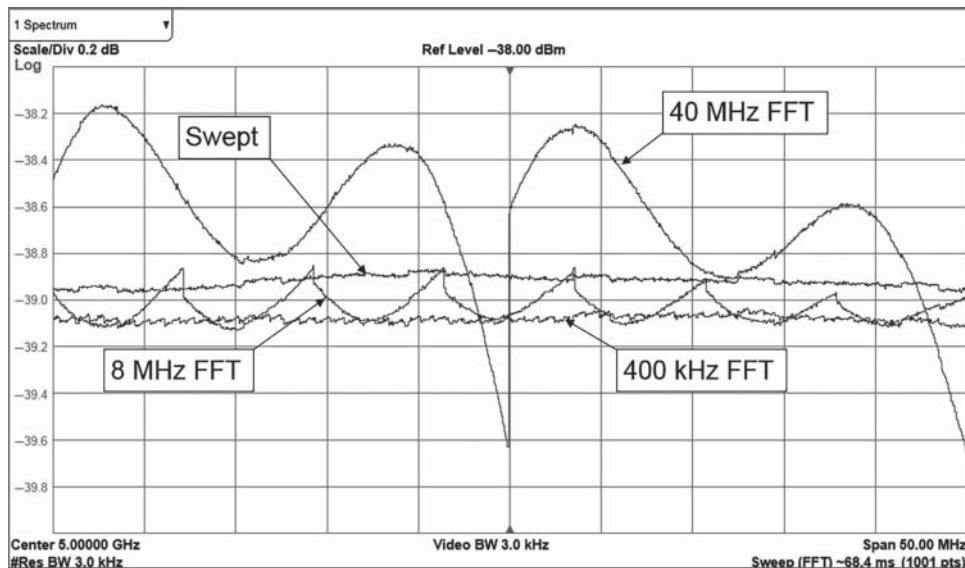


Figure 8.9 Comparing amplitude accuracy for different SA channel FFT widths.

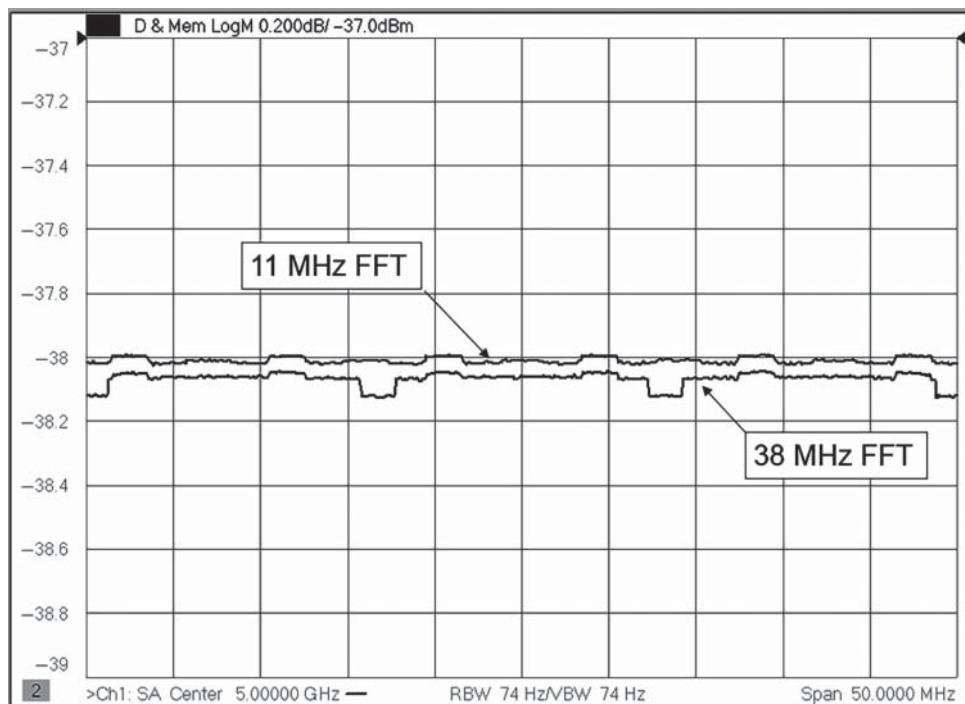


Figure 8.10 Response of VNA-based FFT SA with different FFT widths.

8.1.3 *Attributes of Repetitive Signals and Spectrum Measurements*

For most component test needs, a VNA provides the best method for characterizing the component's attributes. Gain, match, and output power are always most accurate with match-corrected VNA techniques. Even some distortion measurements, for example two-tone IMD, are best performed with VNA capability (see Section 8.1.6) However, with the importance of characterizing a component's effect on wideband modulated signals used in modern communication systems, the use of spectrum analysis to verify performance criteria such as ACPR, EVM, and NPR has been required. For the most part, the analysis is performed using only the received signal from the test component, and the signal analyzer is set to completely determine the signal quality without any foreknowledge of the input signal. While this is necessary when looking at unknown signals, it is an unnecessary restriction when evaluating an individual component's qualities in a controlled test environment. In particular, when an arbitrary waveform generator (AWG) is used to play back a predetermined waveform, utilizing some a-priori knowledge of the waveform can provide a means to greatly improve the quality of the component measurement.

8.1.3.1 **Understanding Modulated Waveforms**

A great majority of the modulated test signals, in fact nearly all used in R&D and manufacturing test, are created from playing a waveform in a signal generator with an internal AWG. Despite the name, the signals created are not in fact arbitrary but are specifically designed to represent some required test condition. For example, the signal may be a frame of long-term evolution (LTE, a current form of cell phone protocol) or a representation of additive white Gaussian noise (AWGN). In these cases, a waveform is created that represents a portion or time segment of the signal. The wider the bandwidth of the signal, typically the shorter the duration (as wider bandwidth signals required data files that grow directly with the bandwidths). One exception is the use of a pseudorandom bit generator, often used for bit error rate testing (BERT), where the repetition of data pattern can be very long indeed. With some exceptions, the focus of the rest of this chapter will be on modulated signals used to test components where the signal is repetitive, that is, created from a waveform played back through an AWG.

For most cases, the waveforms have periods on the order of $1\ \mu\text{s}$ to $10\ \text{ms}$. During the test process, often only a small portion of the test signal is captured at a time, and thus response of the device-under-test (DUT) to the signal is obscured by this under-sampling of the test signal.

Consider the signal shown in Figure 8.11. This is a swept-mode measurement of a band-limited AWGN signal, with $100\ \text{MHz}$ BW, captured on a traditional spectrum analyzer (in this case a Keysight UXA). On repeated sweeps, the amplitude at any particular frequency point varies by $10\text{--}20\ \text{dB}$. This signal looks truly noise-like, but looks can be deceiving, and the observer here is clearly deceived.

The term AWGN refers to the pattern of the noise as measured in the time-domain. This means the time-domain value of the amplitude of the signal is noise-like and can take on any value in a random way. But this is a band-limited AWGN signal, so the maximum deviation

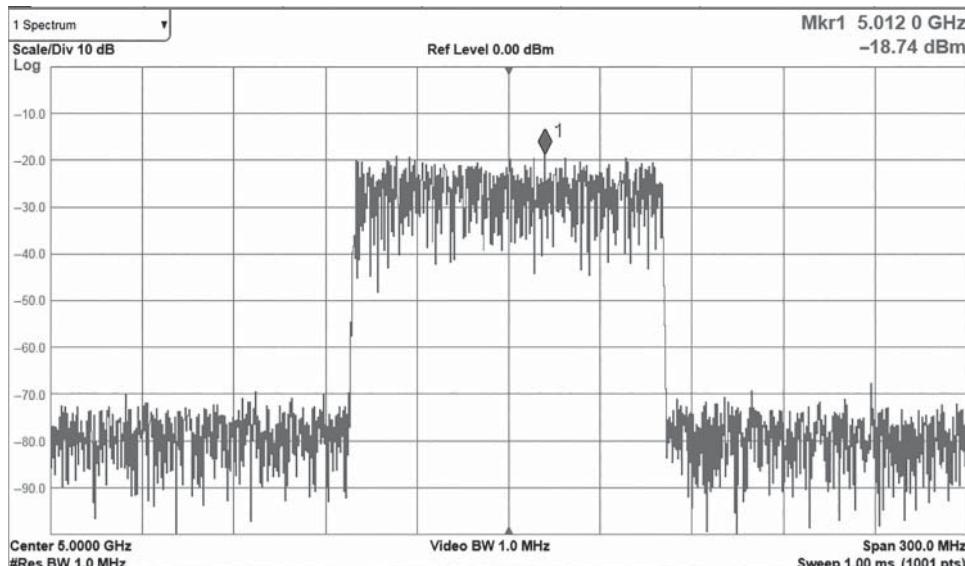


Figure 8.11 Swept-mode SA measurement of an AWGN signal.

(that is, the derivative of the signal) is limited by the bandwidth. One can think of this as adding a filter to an infinitely wide AWGN signal, where the filter has a charge/discharge limitation and even if the input signal can have an infinite rise time, the output of the filter is limited by the delay and impulse response of the filter.

From direct observation we can say that Figure 8.11 is absolutely *not* the correct spectral response of a AWGN signal. The spectral content of an AWGN signal is known to be flat with frequency, with a constant noise power density equal to the power of the signal divided by the bandwidth. The fact that the spectrum appears to vary from sweep to sweep (and time to time) at the same frequency point is an observational error that is seldom understood. Why? Likely because this is called a *noise* signal, and an uninformed observer will see a noisy (time varying) spectrum that meets their (incorrect) expectation of the spectrum of a noise-like signal. To repeat, AWGN has a flat (constant) and unchanging power spectrum even while it has a changing noisy time-domain waveform.

Band-limited AWGN signals are typically created in an AWG (it can be confusing discussing an AWGN created by AWG, so it is common to refer to an AWG as simply an “Arb”), by creating a multitone spectrum with constant amplitude and random phase. This produces a time-domain signal that closely mimics the time power density of a Gaussian noise signal. In fact, the *w* (white) in AWGN means the spectrum is uniform (flat) in the frequency domain. So, by definition, we must agree that a band-limited signal cannot be an AWGN signal because the band limitation means the spectrum is not flat from zero frequency (DC) to infinity. But in common usage, a band-limited noise-like signal with a flat spectrum is referred to as an AWGN signal.

8.1.3.2 Complementary Composite Distribution Function (CCDF)

The complementary composite distribution function (CCDF) is a means for gauging the time-variant nature of a signal. It is often used to determine if two signals have significantly the same or different statistical signal properties.

The cumulative distribution function (CDF) shows the probability that a sample of a signal is less than some value. The x-axis is the threshold value, and the y-axis is the percentage of time the signal is smaller than the threshold. Figure 8.12 shows the CDF for an AWGN signal and a quadrature phase-shift keying (QPSK) signal.

We define the CCDF as 1-CDF and typically display the results on a log-log axis. This gives a more easily understood view of the signal statistics and allows one to understand more intuitively the signal characteristics such as peak-to-average ratio. The reference point is the left of the curve and represents the average value of the signal. The farther to the right the CCDF curve extends, the greater the peak-to-average ratio, which also means the greater the overhead needed in a power amplifier above the average power of the signal, to avoid distorting the signal. The CCDF of a AWGN signal and a QPSK signal, with equivalent occupied bandwidths, are shown in Figure 8.13, along with the theoretical curve of an ideal AWGN signal. Of course, the AWGN signal from the Arb is not ideal, because it is not infinite in length, but rather it is a finite length waveform. But one can see it follows the ideal Gaussian curve with less than 1% deviation. The peak-to-average signal is limited by the waveform length and in this case is about 9.4 dB. At that place on the curve, there is only one point at that peak value, and then no other points, so the percent drops to zero. This represents one out of 4001 points in the time record, about 0.024%. Note that the AWGN signal is applied to the carrier, and the phase of the carrier can change randomly compared to the phase of the AWGN modulating signal, and this can cause slight differences in the peak value, so turning on and off the modulation will cause the single peak point of the time-domain trace to change,

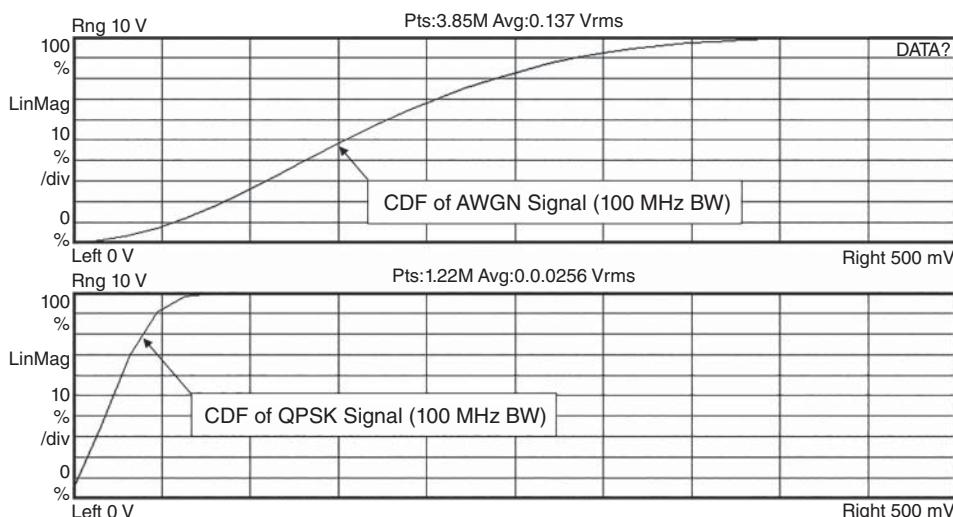


Figure 8.12 CDF curve for an AWGN signal and a QPSK signal.

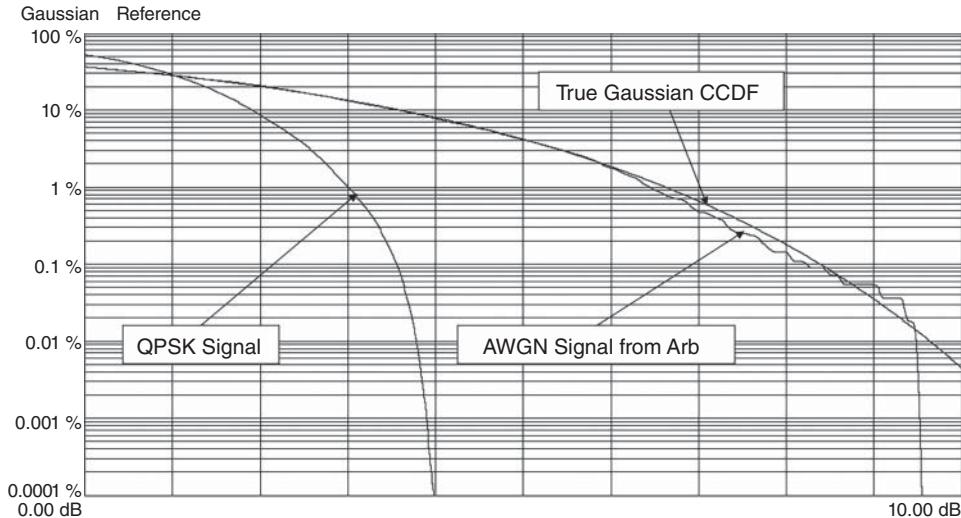


Figure 8.13 CCDF cures of AWGN and QPSK signals.

typically 1 dB or so. This will cause the last point of the CCDF curve to jump back and forth 1 dB on the x-axis each time the modulation is turned on and off relative to the carrier.

Also shown is a QPSK signal CCDF curve. QPSK is a constant amplitude modulation, but filtering can generate some AM, so we see a small peak-to-average ratio still. In this case the peak-to-average ratio is only 4 dB. Generally, the more information that is applied to the signal, that is the higher the data rate, the larger the peak-to-average ratio.

8.1.3.3 Generating Noise-Like Signals

In the case of the signal shown in Figure 8.11, the signal was created from 4001 equally spaced tones with random phase. The time-domain response of this signal does indeed look noise-like, and can be seen in Figure 8.14. The upper portion is a time-expanded view of the full time waveform. From the time-domain measurement, we can see that it has a peak power of 4.61 dBm and an average power of -4.72 dBm for a peak-to-average ratio of 9.34 dB. The peak-to-average noise value is a commonly used characteristic of a modulated signal. For a true-noise like signal, the value should be infinite (noise can take on any value), and so it lacks much meaning in practice. A better choice would be describing a noise-like signal by its statistics such as standard deviation. An AWGN signal should have a Gaussian distribution of power (as the name implies). When created by an Arb, the peak value is limited by the length of the waveform, which also sets the maximum possible number of tones used, and mean power. The peak is obviously the largest value in the time domain waveform, and the average is set by mean of the waveform. In the limit a sine wave has a 3 dB peak-to-average ratio (where we are referring to the peak-envelope power), and an impulse signal (a single non-zero value of the time waveform at maximum output and all others at zero) will have a peak to average of

$$PAR_{Max_dB} = 10 \log(N) \quad (8.6)$$

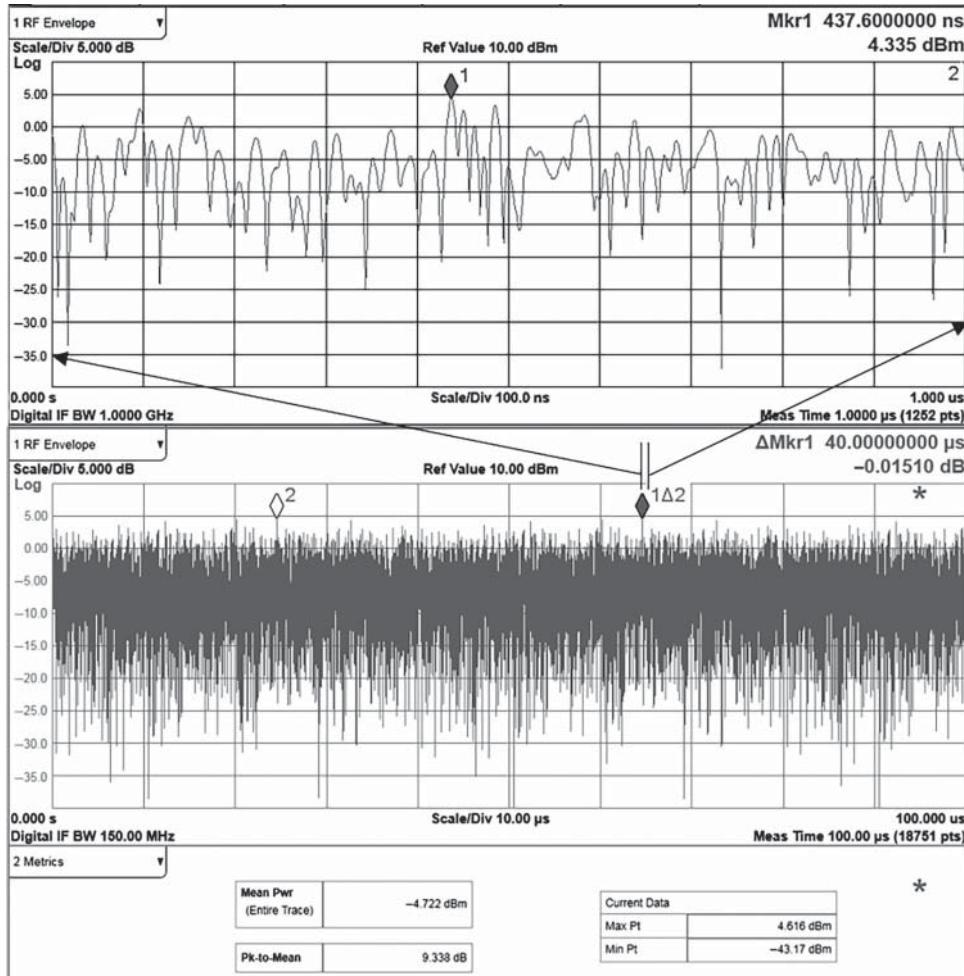


Figure 8.14 The time-domain amplitude (in dBm) of an AWGN signal.

where N is the number of samples in the waveform. If the impulse is band-limited, then the time waveform will have side-lobes associated with the band-limiting. N will also be the number of tones created in the multitone waveform, since the frequency spectrum is the FFT of the time-domain spectrum.

An example of a multitone signal with constant (zero) phase is shown in Figure 8.15 and is comprised of a 101-tone signal with zero phase. The signal is band limited, so the impulse has the expected side lobes one associates with a rectangular window in the frequency domain. While this signal has the same spectral magnitude as the AWGN signal, it has much higher peak-to-average values, in fact, 20 dB peak to average, exactly as predicted in Eq. (8.6).

In the measurement of the AWGN signal in Figures 8.13 and 8.14, the signal is captured using an acquisition time of 100 μ s, equivalent to a 10 kHz detection bandwidth. The AWGN

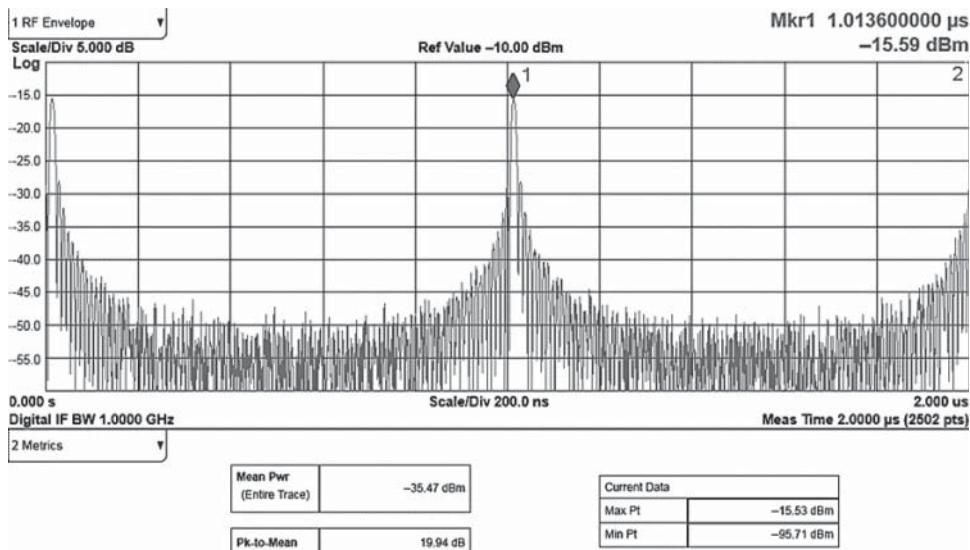


Figure 8.15 Arb signal with 101 tones, zero phase, in the time domain.

signal has a duration of $40 \mu\text{s}$. In the swept-mode spectrum analysis in Figure 8.11, each frequency display point represents exactly one acquisition of a spectral point, with an acquisition time of $1 \mu\text{s}$, and this represents only $1/40$ of the signal. And as the SA sweeps along the frequency axis, each new frequency point displayed represents a new capture of this modulated signal. On the next subsequent sweep, the signal is again acquired in $1 \mu\text{s}$ samples, with each frequency point representing a different time sample, and this is asynchronously sampled with the first sweep. It is no wonder the spectrum seems to vary and be noise-like: the continuous sweep response represents thousands and thousands of different samples of the modulated signal each at a different time. The repetitive nature of the signal is apparent when looking at subsequent peaks represented by Markers 1 and 2 in Figure 8.14. The delta time is exactly the waveform period. The upper plot shows a zoomed-in display in time, and the waveform shows a smooth response we can associate with the bandwidth limitation of the band-limited AWGN signal. It is similar to what one would expect taking a wideband noise signal and passing it through a physical filter.

Contrast Figure 8.11 (flip back a few pages), measured in swept mode, with Figure 8.16 where the FFT mode is employed on the same spectrum analyzer. In this case the FFT acquisition bandwidth is set to about 8 MHz, and the “chunks” of FFT points representing an individual acquisition over one 8 MHz FFT span are highlighted where their edges are discernable. This spectrum is as entirely valid as the spectrum in Figure 8.11, or more properly stated, this spectrum is as equally invalid as the other spectrum in that neither one of them is a true representation of the modulated signal. Both have been distorted (in different ways) by the acquisition method used. Here the “chunks” of the FFT can be observed by looking for discrete transitions in the frequency domain, such as illustrated by Markers 1 and 2 in Figure 8.16.

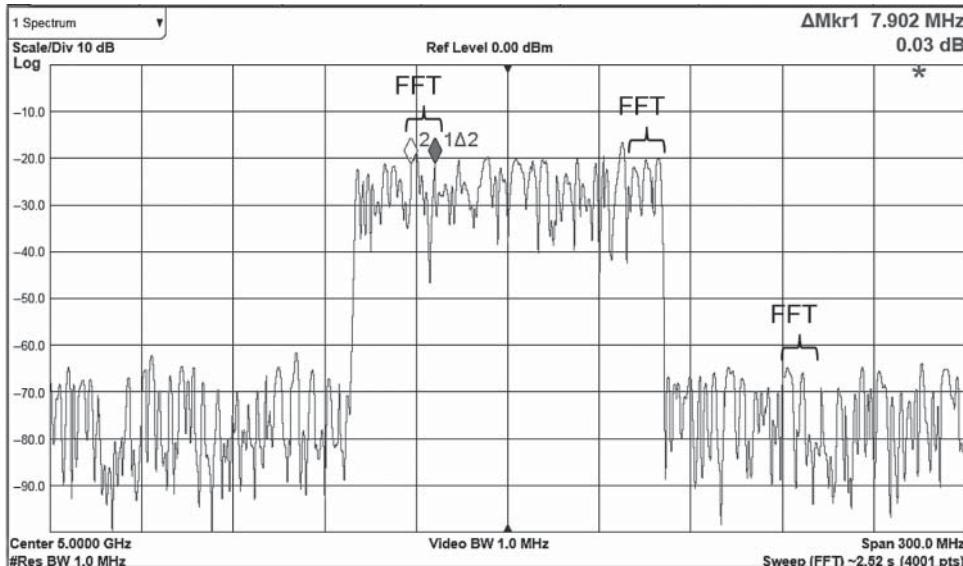


Figure 8.16 FFT mode SA measurement of an AWGN signal.

The trouble lies in the challenge that the swept mode cannot acquire the entire spectrum at one time and so represents different times of the signal at different frequency points in the spectrum. In a similar way, the FFT bandwidth is not sufficient to capture the entire signal at once and so must be stepped through the frequency span of the signal. Thus, it also measures different signals at different times, where each “chunk” of the FFT represents a single acquisition of a portion of the signal, each at a different time as the FFT moves across the frequency domain of the signal. From this context, the swept mode could be considered as the limit case of a single-point FFT where the acquisition BW is the same as the RBW. To repeat the point, Figures 8.11 and 8.16 have identical SA settings, measuring identical signals, but present a different spectral picture. So, where does this leave the observer? Neither method properly captures the spectrum of the signal, and both spectrum appear different. Which is right? The answer is neither.

In fact, for any waveform played back by an AWG, the spectrum of the wave is entirely static and never varies from time to time. This is because a signal played back from an Arb can be represented as a periodic signal where the period is one over the time length of the waveform. The spectrum is truly time stationary, as is true of all periodic signals and can be computed exactly from the FFT of the time waveform. While this is clearly true, many engineers – even, or especially, experienced ones – will insist that a AWGN signal has a noisy spectrum, because that is what their experience has observed on a traditional swept-tuned spectrum analyzer. So if the spectrum appears to vary from sweep to sweep, how are any measurements made to be stable and give consistent results? The answer in the past was to either slow the sweep time down, resulting in what appears to be a time-stable response, or average a large number of sweeps. Figure 8.17 shows the swept-mode response when the sweep time is slowed from 1 ms (equivalent to 0.2 μ s per frequency-point acquisition) to about 80 ms (representing 80 μ s per frequency-point acquisition). This spectrum is entirely

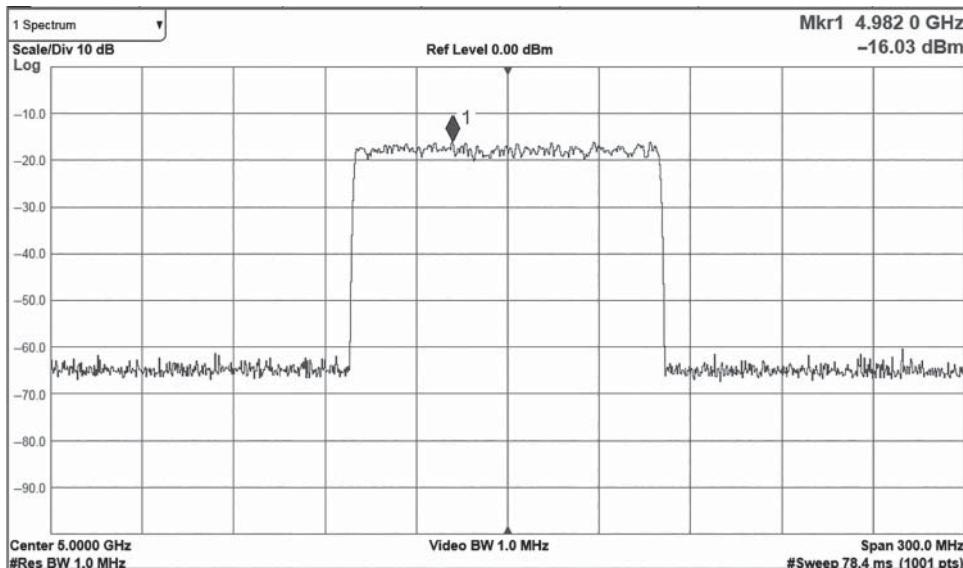


Figure 8.17 Swept-mode SA measurement with slow sweep time.

stable, with variation from sweep to sweep of less than 0.05 dB and, unfortunately, is also an entirely invalid representation of the AWGN spectrum.

8.1.3.4 Effects of Sweep Time

In the case of Figure 8.17, the spectrum becomes stable because the SA is set to “peak-detect” mode, and data acquisition time is long enough ($80 \mu\text{s}$) to encompass the entire duration of the AWGN signal ($40 \mu\text{s}$). Thus, for each frequency display point, the entire AWGN signal is continually measured over about two cycles of the waveform, and the true peak value is recorded. On each new sweep, the same peak value is found. But this is not the true spectrum of the AWGN as the peak-detection is a non-linear function and thus distorts the true spectrum of the signal. In fact, if the band-power is measured of this spectrum, one will see it is much higher than the total power of the AWGN signal, in the cases where the band power measurement is performed in peak-power mode.

However, at least here the signal appears time-stationary and presents the same response sweep to sweep on the spectrum display. This gives us a clue to how to properly capture the spectrum of such a modulated waveform: the secret is time.

8.1.3.5 VBW Effects

In all the previous discussion of modulated spectrums, the SA acquisitions were all made using the same RBW and VBW while comparing the effects of sweep mode (swept-tuned versus stepped FFT) and sweep speed. But the effect of RBW and VBW can be even more substantial, and proper settings of these values will reveal the true spectrum, whereas

non-optimum settings can mask the spectrum as its response may be distorted by the convolution of the spectrum to be measured and the response of the SA. The RBW determines the acquisition time for each frequency point in the swept mode, or each FFT chunk in the FFT mode. Even when slowing down the sweep, the acquisition time of the RBW remains the same. In the case of Figure 8.17, 40 RBW acquisitions occurred for each frequency display point (due to very slowly sweeping across the frequency span), with the acquisition having the largest peak value retained by the peak detector. The effect of changes in VBW are a bit more subtle and require some historical reference to understand the implementation. In this section, the effects of changes in VBW and RBW on the observation of the spectrum are considered.

In the simplified block diagram of Figure 8.3, the detector is represented by an RBW filter, a diode detector, and a VBW filter. But modern instruments, for at least 40 years, have used fully digital detectors, as shown in Figure 8.4. Historically, the purpose of the VBW filter was to remove broadband noise detected by the diode from the video signal sent to the CRT. The video filter in effect produced a CRT signal that was essentially a moving average of the input signal. In swept-tuned spectrum analysis, this analog filter was replaced with a digital filter that had a similar effect. If there was a lot of broadband noise at the input of the detector, the CRT would show a wide, noisy trace. It was difficult to discern a small signal embedded in this noise. Narrowing the VBW filter would reduce this noise, and an indication of the small signal would be clearly present. But, if the spectrum analyzer swept too fast, the peak of the signal would not be properly detected, so the sweep speed of the SA is reduced to accommodate this effect. A digital VBW filter is represented as an infinite-impulse-response (IIR) filter with many taps. The sweep speed must be low enough to process a signal through all the taps of the filter, which is in fact identical to slowing down for an analog filter's impulse response.

But things become a little unexpected when using a spectrum analyzer in FFT mode. In FFT mode the signal to be processed cannot be represented as a time-continuous signal passing through a filter to a CRT. In fact, an entire band of frequencies is collected at one time, as processed by the FFT. A realization of a VBW filter was not possible because there was no video signal to process. However, the noise in the detection would produce a noisy trace still, so some other method was needed to detect a small signal, maybe an impulsive signal, in the presence of noise. Thus, the VBW filter in FFT mode is realized by averaging multiple FFT acquisitions before the peak detection algorithm is applied. This noise reduction technique provides a similar response to the traditional swept-tuned VBW noise reduction. Even though the FFT-based spectrum analyzer presents a VBW feature with a VBW value, this bandwidth value is converted to an integer number of FFT acquisitions. This conversion depends on the ratio of RBW to VBW, and in some Keysight analyzers is

$$N_{FFT_Aquisitions} = \text{Round} \left(\frac{RBW}{VBW} \right) \quad (8.7)$$

That is why, in some cases, small changes of the VBW bandwidth value do not produce any discernable effect: the computed number off FFT acquisitions, being integer valued, don't change with small changes in the VBW bandwidth value entered. The same AWGN signal measured in Figure 8.11 (with a 1 MHz VBW) is measured using a VBW filter value of 20 kHz in Figure 8.18. Here it is easily seen that the spectral content appears to change from that of Figure 8.11. And the peak values are much lower. This is because the VBW filter is processed before the peak detector, so we are seeing the peak of moving time-averaged signal (the averaging occurring in the VBW filtering process).

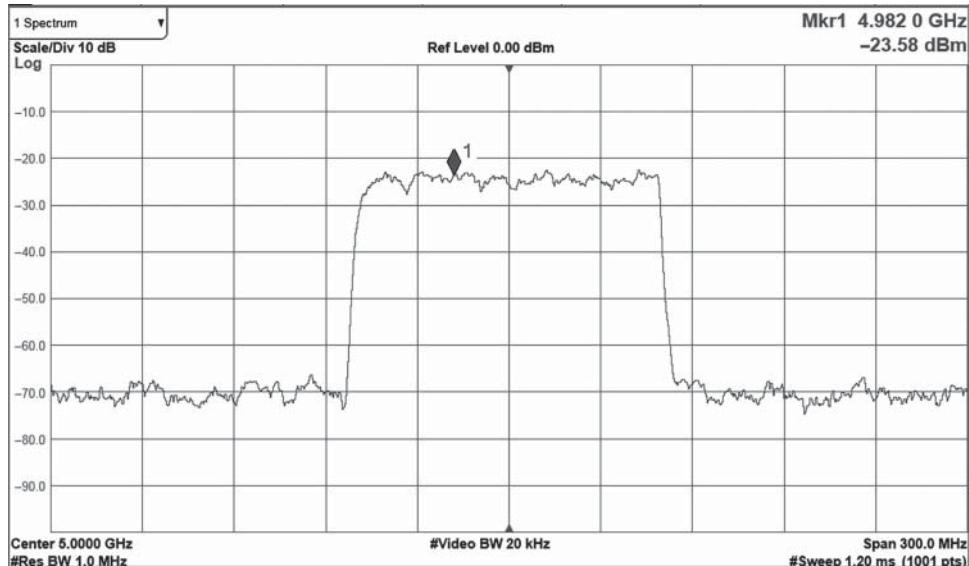


Figure 8.18 Swept-mode measurement of AWGN signal with 30 kHz VBW.

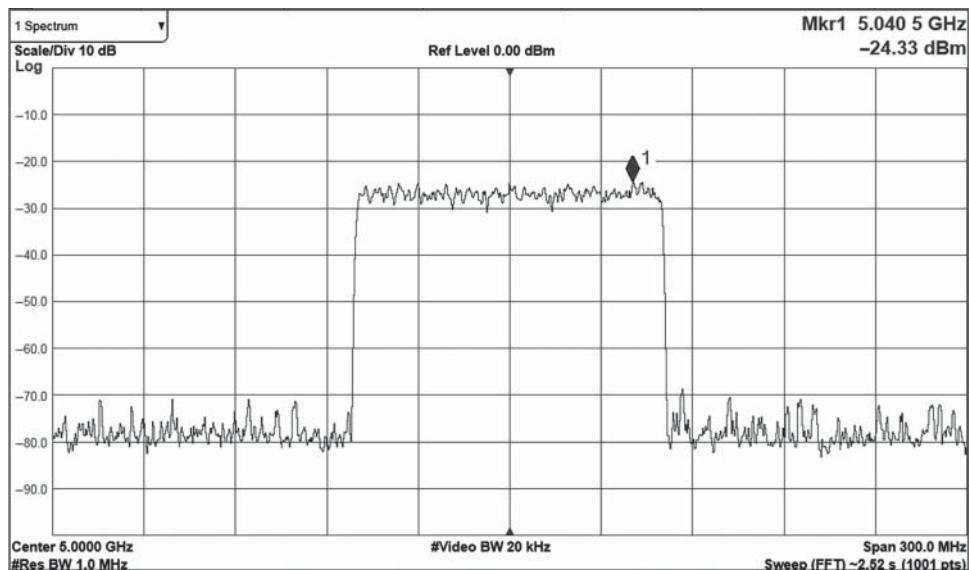


Figure 8.19 FFT mode measurement of AWGN with 30 kHz VBW.

The measurement can be repeated for the FFT mode spectrum analysis, and the results, shown in Figure 8.19, still show a small but distinct difference from the swept-mode above, where the two spectrums now look more similar. In the case of Figure 8.18, the VBW filter is implemented as a moving average filter of acquisitions. In the case of Figure 8.19, the averaging occurs as an integer number of FFT averages before the peak detection.

When comparing figures between the swept and stepped-FFT mode, we can still discern a difference, but with a narrower VBW, the sweep time is slowed and the signals are averaged so that the differences in acquisition appear less, and the signal more closely resembles the true spectrum, which we know must be a flat spectrum for an AWGN signal. But still we see the spectrum vary from sweep to sweep, but not as much as in Figure 8.11.

8.1.3.6 RBW Effects

While the VBW affects the visual look of the signal, it does not change the fundamental acquisition capability. Reducing the VBW will reduce the *variation* of the noise displayed, so it will tend to a mean value but will not reduce the mean value of the noise. The resolution bandwidth has a much different effect, reducing noise floor itself and improving the ability to resolve closely spaced signals. It is this second effect that becomes very important in understanding the RBW effect on modulated signals. Compare the appearance of an AWGN signal in Figure 8.16 to that in Figure 8.20, where the RBW is lowered to 75 kHz. We can see the apparent level has dropped, and the signal has much less variation across frequency. The drop in the signal level is usually understood as a consequence of measuring a noise-like signal with a narrower acquisition bandwidth. In Figure 8.16 the resolution was much wider, 1 MHz, so more noise power would be acquired in each FFT bucket, and thus a high peak would be measured. If a noise spectrum has a constant spectral density (in terms of dBm Hz^{-1}), then we expect the measured noise to drop when the RBW is lowered. Also, in FFT-based measurements the FFT resolution (which sets variation as a function of frequency across the x-axis of the display) is much narrower, and so we can see more variation (faster changes in Y-amplitude) across frequency. With the wider RBW of Figure 8.16, the

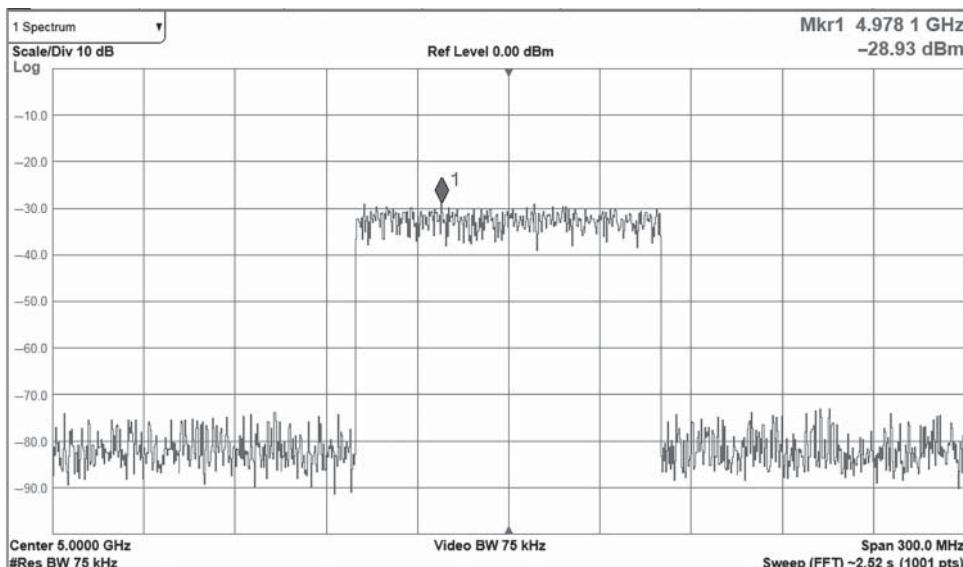


Figure 8.20 Measure of AWGN signal with 75 kHz RBW.

variation in the peaks of signal cannot exceed approximately 3 dB in any 2 MHz span, over an FFT acquisition, due to the 1 MHz RBW filter effects. Of course, at the edge of separate FFT acquisitions, there is no limit to the variation of the next segment as it is acquired at an entirely different time, and the signal may be completely different than during a previous FFT segment. This is especially noticeable in the noise floor region of the signal in Figure 8.16.

In contrast, the variation in the peaks of the signal in Figure 8.20 are much more closely spaced as the FFT resolution is much narrower. Further, the top of the signal envelop appears much more constant. Still, in this RBW setting the data acquisition time (about 20 μ s) is shorter than the signal period for the AWGN waveform (at 40 μ s), and we will see the trace response change on sweep-to-sweep basis, in a manner that looks noisy to an uninformed observer. But we know the spectrum must be flat, and uniform, for an AWGN signal, so still the observed display cannot be a proper representation of the AWGN signal.

8.1.3.7 The “Scales Fell from His Eyes” RBW < (1/T)

Finally, let's see what happens when the RBW is narrowed further, such that the data acquisition time becomes longer than the signal period. Because the RBW has windowing associated with it, the RBW must be quite a bit narrower to resolve fully the proper observation of a modulated signal waveform. If the RBW is narrowed further, to 30 kHz, the resulting display will appear as shown in Figure 8.21. Now we see a nearly flat display in the band of power representing the AWGN bandwidth, and some spectral regrowth, or out-of-band distortion due to the distortion of the amplifier through which this signal passes. Finally, we have true observation of the signal. In fact, if we switch the RBW lower still, there will no perceptible change in the signal.

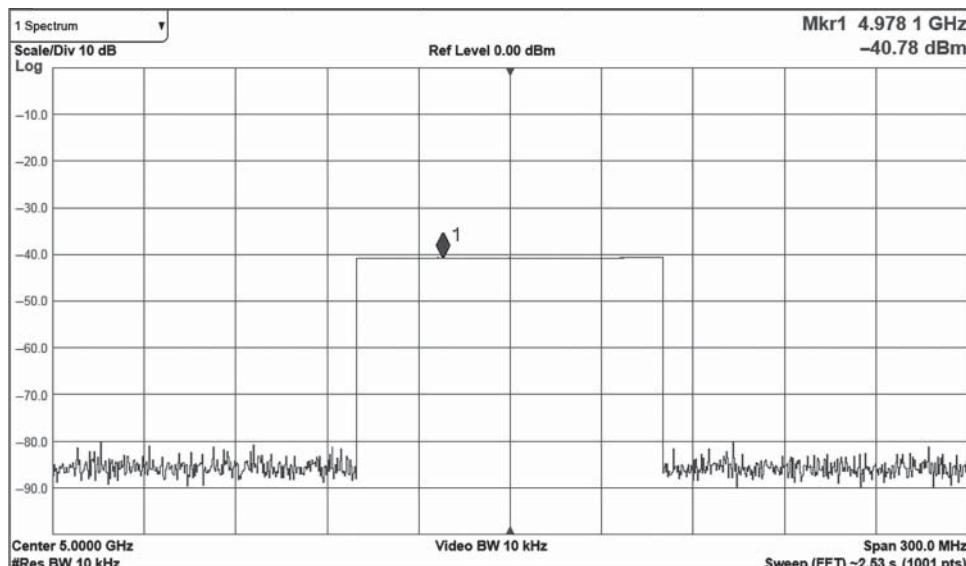


Figure 8.21 AWGN signal where RBW acquisition time is longer than the waveform period.

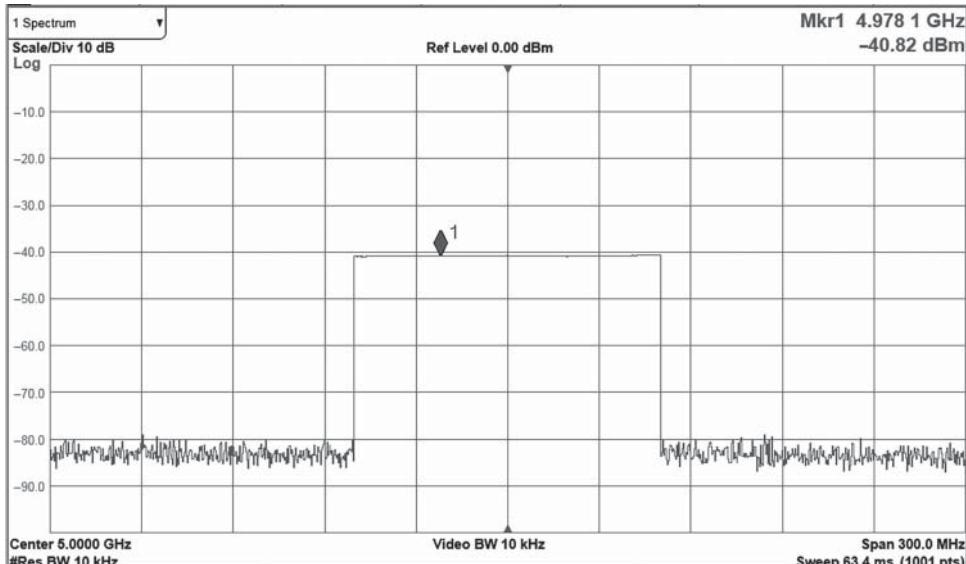


Figure 8.22 Swept-mode display with narrow RBW matches FFT mode.

Surprisingly, if we switch now from a stepped-FFT mode to swept mode, an identical signal is observed, as demonstrated in Figure 8.22. Here now, with the RBW narrow enough so that each acquisition is longer than the waveform period, each point of the swept-mode display results from an entire acquisition of the waveform period.

The conclusion of this section must be “for the proper measurement of the spectrum of a signal created from a waveform playback of an AWG, the RBW filter of the spectrum analyzer must be set so that the acquisition time of the filter is longer than the period of the waveform.” Of course, to set the RBW filter properly, one must have some foreknowledge of the signal, in particular the period of the waveform that creates the signal. While this is not practical in over-the-air testing of unknown signals (signal analysis), it is eminently practical in the area of component test, where the input signal must be created in the test equipment, as well as measured by the test equipment. In such a case we are not particularly interested in the signal, but in *what the DUT does to the signal*. We take this fact to a new level in the next section.

There is an alternative view that can be taken for what occurs when we reduce the RBW to have an acquisition time longer the modulated waveform period. We know a repetitive modulated waveform has a period or repetition, and a consequence of that periodic nature is that the underlying spectrum *must* be a multitone (sometimes called *multisine*) spectrum. Periodic signals have discrete spectrums, which have zero value between the spectral tones, and for an AWGN signal, each spectral tone has an amplitude that is $1/N$ times the total power of the AWGN signal. Figure 8.23 shows the spectrum of the same signal, but with the span set small enough to see the individual tones associated with the periodic waveform of an AWGN signal, in this case to 250 kHz span.

This also makes it clear why the spectrum now appears static and uniform, because we are resolving down to the individual multitone elements of the spectrum. Further reducing the RBW will not lower the spectral height or change the apparent composition of the

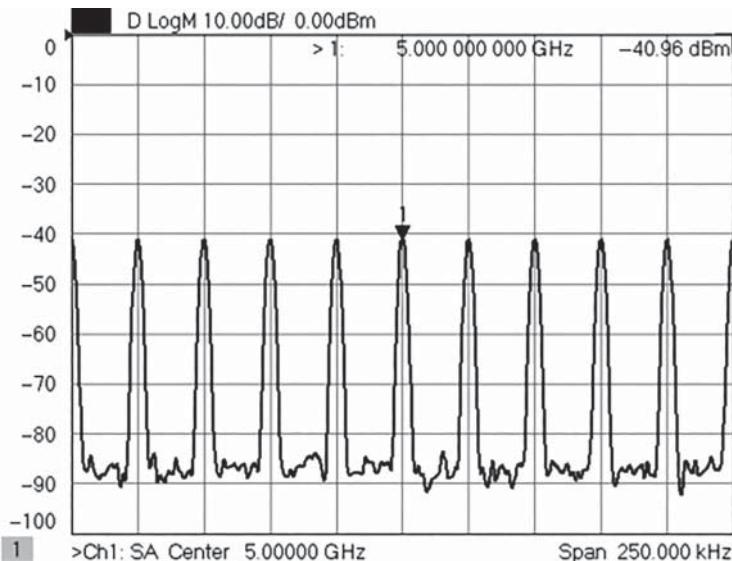


Figure 8.23 Close-in spectrum of AWGN waveform showing multitone components.

signal, although it will lower the noise floor between the multitone elements. In a wideband sweep, each display point may contain many multitone elements. If the detector type is set to Peak-Detection, then the value of the highest of the multitones will be shown. On a densely modulated signal, over the span of the signal, each display point contains many tones, and for a proper AWGN signal, they are all at the same height; thus, the spectral density appears constant. We can see clearly that lowering the RBW, such that the acquisition captures the entire period of the waveform, has the consequence of resolving the underlying nature of the signal. We take this one step further, in the next section, by forcing the data acquisition to be *exactly* N times the waveform period, where N is any integer. When we sample the signal at exactly its waveform period, we call that *coherent spectrum analysis*.

8.1.4 Coherent Spectrum Analysis

In traditional use of spectrum analyzers, virtually no information of the test signal is assumed before making measurements. The spectrum analyzer can search for and determine the frequency, in some cases the modulation type, and finally the modulation quality. In a harmonic test, a wideband sweep is used to find the largest signal (presumably the fundamental), and the harmonics are discerned from peaks in the spectrum at farther spaced frequencies. Usually a peak-maker search is used. But clearly, if one is also supplying the input signal, this is not necessary, and one might simply tune the spectrum analyzer – in a narrow span sweep – to each of the fundamental and harmonic frequencies. This is using foreknowledge of the input signal (e.g. its frequency) to simplify, speed up, and improve the measurement.

A similar example exists in the area of testing components with modulated signals, resulting in profound improvements in measurement capabilities. In the case of modulated signals,

there are two aspects that are important to know: the center frequency of the modulation *and* the period of the waveform used in the modulation. In the component test world, almost always these two values are available, and even if they are not known, they can be identified with additional measurements. We call this mode of operation *coherent spectrum analysis*, as the spectrum analyzer settings are adjusted to match those of the signal source. In particular, when using non-image-protected systems, such as a VNA, there are additional benefits from utilizing coherency.

8.1.4.1 Attributes of Coherency

We say a source/analyizer system is using coherency if we know the input frequencies of the source and set the analyzer's measurement attributes depending upon that knowledge. Fifty years ago, before the advent of VNAs, the RF world depended upon scalar network analyzers (SNAs), which used a power detector as a receiver. As such, there was no need to be concerned about the coherency between the source and receiver, as the receiver could measure signals at any frequency. But SNAs had poor dynamic range and were susceptible to interfering signals, among other faults. VNAs used tuned receivers with narrow bandwidths to improve dynamic range, and were insensitive to interference, but required the source and receiver operate coherently: the receiver must be tuned to the source frequency. The advantages of VNAs was so overwhelming that virtually all network analyzers today are VNAs. These same advantages accrue in modulated measurements but have the same prerequisite that the receiver be aware of the source frequencies.

In coherent spectrum analysis of a modulated signal, the receiver is tuned to the center frequency of the source, and the RBW filter is set so that its acquisition time is an exact integer multiple of the acquisition waveform; in practice, a minimum of three times the waveforms' period is required. We have already seen in Section 8.1.3 that if a modulated signal is measured with a short acquisition time, shorter than the modulation waveform time, the true nature of its spectrum will be distorted. Coherent spectrum analysis takes advantage of the foreknowledge of the stimulus signal period to ensure that exactly an entire waveform is captured with each data acquisition. A periodic waveform, played back from an AWG, must be a multitone signal, and those tones must land on a grid of frequencies that is exactly the reciprocal of the waveform period, normally centered about the source center frequency. In some cases, it is desired to create a waveform that is "single-sideband" in nature, which is where the modulated waveform is only above, or only below, the carrier frequency, but even in such a case, there remains tones, perhaps imperceptible, at each of the grid frequencies.

A particularly interesting consequence of using an exact integer multiple of the waveform is that it is no longer necessary to use windowing or shaping of the RBW filter. Since the FFT data buckets of an acquisition that exactly matches a multiple of the waveform period land exactly on grid with the multitone signal, there is no possibility of spectral leakage such as occurs on with asynchronous measurements. However, it is not necessary to provide any particular triggering to the acquisition (such as at the start of the waveform); all that is required is that the FFT acquisition time exactly matches the waveform time. It does not matter where in the waveform the acquisition starts, since the entire waveform is acquired.

8.1.4.2 Coherent-Time-Averaging (Vector Averaging)

An additional benefit occurs with coherency that is not possible in traditional spectrum analysis: we can utilize coherent time averaging, sometimes called *vector-averaging* or *stacking*. To understand coherent time averaging, consider Figure 8.24, which shows a modulated signal over several periods of the waveform, each offset to see the various acquisitions, where the noise is apparent on this two-tone modulated signal. If the digitized version of the signal of the first period is added to that of subsequent periods, the signal retains its inherent shape. Incoherent noise, which is often caused by the noise floor of the measurement equipment, is reduced in the time average. This is like stacking signals, one period at a time, and taking the average of the stack (hence the term *stacking* in some papers). The bottom trace is the coherent time average of the four waveform periods above it. It is readily apparent that the noise of the signal is greatly reduced, and an FFT of this signal will show the same amplitude for the modulated spectrum but a much lower noise floor.

An alternative frequency-domain view is that each period is comprised of a multitone signal, and each tone has a particular magnitude and phase, relative to the other tones. On top of this multitone signal is noise (often from the test receiver) that makes slight variations in the magnitude and phase. Each subsequent waveform has the same multitone signal but with a different noise contribution. If each tone is treated as a vector and averaged with the other acquisitions at the same frequency, the mean vector value of the noise will average to zero and the true value of the tone, without noise, will be revealed. This is particularly useful in measuring the signal power of low-level distortion products, which are caused by intermodulation of the signal tones and which must therefore also land exactly on the coherent grid frequencies. More about this in the Section 8.1.4.4.

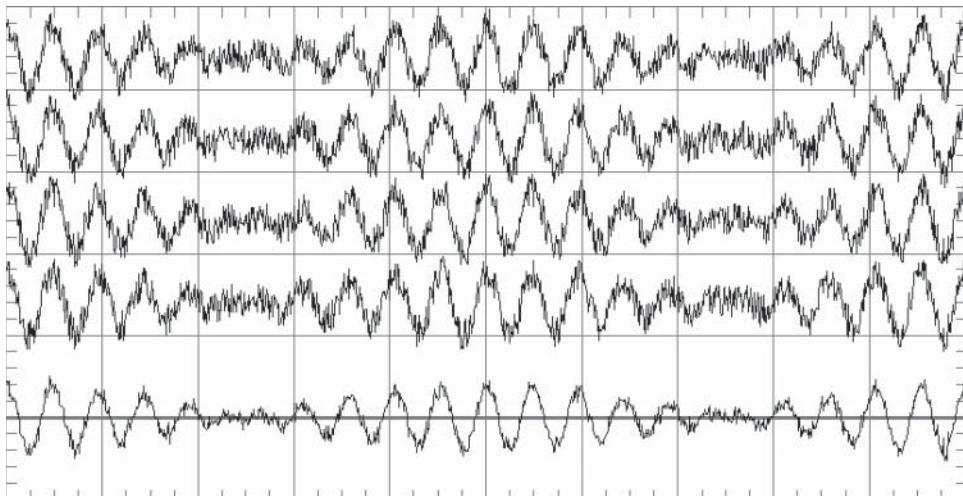


Figure 8.24 Coherent time averaging (or vector averaging) reduces noise on a coherent signal.

8.1.4.3 Coherent Image Rejection

In addition to improvements in measuring the spectrum of a signal from a DUT, coherency can also improve the rejection of unwanted images from the VNA or non-image-protected spectrum analyzer. The image-reject algorithms work remarkably well but occasionally can give spurious results if a dense multitone signal is not measured with a sufficiently narrow RBW.

When coherency is used, the LO can be selected to avoid known image problems such as aliased signals or higher-order mixing products. Thus, coherency mode adds an additional protection against failing to reject image signals or incorrectly rejecting true signals.

8.1.4.4 Modulated Power Detection for Coherent Signals

When modulated signals are coherent with the measurement analyzer, the ability to correctly measure the power of the modulated signals is greatly improved, particularly the low-level distortion of the signal in the form of adjacent power outside of the passband of the modulated signal. The power of a modulated signal is traditionally measured with either a power meter or a spectrum analyzer using a band-power measurement function.

The power meter measures the total power of the signal and is adequate if the signal level is large or if there are no other signals present. However, power meters have limited dynamic range (no lower than about -70 dBm) due to the kTB noise floor of the sensor. In essence, it is a wideband detector and broadband noise integrated over the entire bandwidth of the detector results in a minimum reading of about -70 dBm. Further, it cannot measure one signal in the presence of others. Thus it could not be used to measure a small signal in the presence of a large signal such as needed for intermodulation or ACPR measurements.

A spectrum analyzer measures the total power in a modulated signal by using band-power functions such as band-power markers. In traditional spectrum analyzers, the band power is integrated over the detected and displayed result, such that the detector type must be set to average type before applying the band power measurement; normally the spectrum analyzer will automatically switch the detector type to average when the band power markers are activated. However, after they are activated, the trace detector may be manually changed, for example to peak-detector type, which will result in an erroneous band power reading. It is often desirable to show the signal with normal or peak detection, while measuring the band power, and for a traditional spectrum analyzers two traces must be used: one in normal detection mode for the display, and the other in average detection for band power measurements. The displays of band power measurements using these two detector types is shown in Figure 8.25. The left display has peak detection mode enabled, and the incorrect band power is displayed; the right display is using average detector value, and the correct band power is displayed (the source was set for $+5$ dBm).

In some VNA-based spectrum analysis, such as the Keysight PNA, the band-power measurement is independent of the detector type (since band-power should always be done with a power-average detector), so the on-screen display may remain in peak-detect mode while the band function operates on the underlying average power.

With coherent spectrum analysis, we know that all energy associated with the signal must land on an FFT grid associated with the period (or tone spacing) of the signal. If multiple periods of the signal are acquired due to using a lower RBW (which extends the data acquisition time), then the width of each FFT bin is reduced, and the energy associated with the

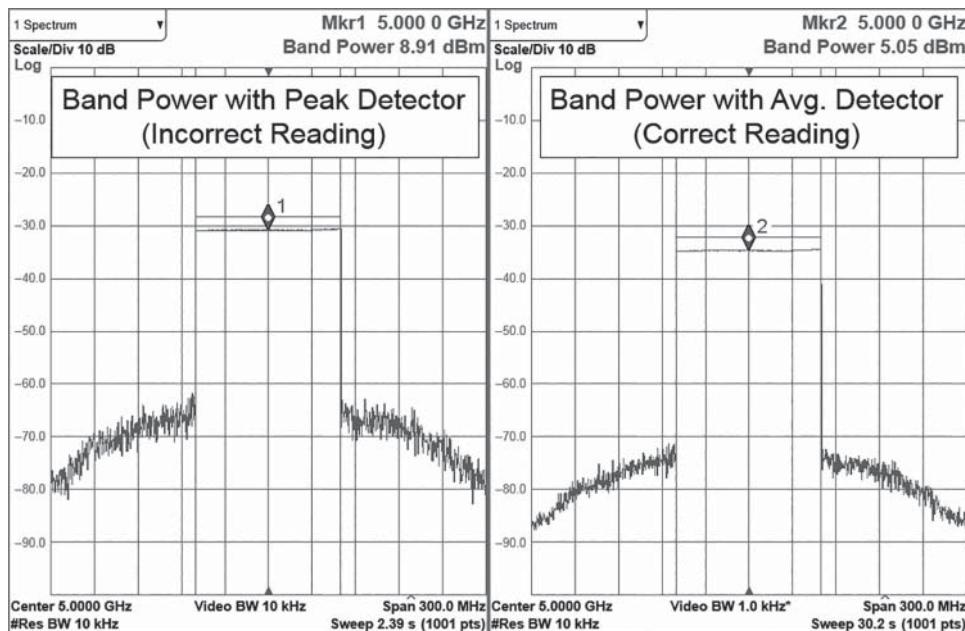


Figure 8.25 Band-power measurements show difference between using peak and average power.

signal, while still on the grid, falls more sparsely on the grid. For example, if the RBW is set so the acquisition time is 10 times the period of the waveform, the FFT tones associated with the signal will land in every tenth FFT bin. In measuring the power of the modulated signal, we need only sum the up the power in the tones that land on the known tone spacings, and we can ignore the power that lands on FFT bins between the tones associated with the modulated signal. We call this kind of detection *coherent power detection* or *multitone power detection*.

This kind of power detection ignores the power associated with noise or interfering signals that are not on the multitone grid. If one wants to measure the total power from a system (including added noise), then one cannot use coherency methods, as the noise from the DUT will not be coherent, nor the power from any interfering signals. But for component test, such as determining the output power of an amplifier or determining the adjacent power caused by distortion in the amplifier, the noise measured is typically due to limitations of the instrument and is not generally associated with the DUT. Consider an amplifier being measured for ACPR, producing a large signal. The measurement instrument must have sufficient attenuation such that the ACPR created in the test instrument is much lower than that of the DUT. To achieve this, a large amount of attenuation is used to reduce the signal level, which has the effect of degrading the noise figure of the instrumentation. For this reason, ACPR measurements are almost always limited by the noise floor of the test instrumentation. Using coherent power detection will allow the noise floor to be lowered, while maintaining sufficient attenuation to ensure the test instrument is not creating its own distortion. Examples of ACPR measurements are shown in Section 8.2.1.

In addition to ACPR measurements, coherent power detection can be used to measure very low-level signals in the presence of instrumentation noise. A key test associated with communications systems is receiver sensitivity level. In a receiver sensitivity test, the source signal is lowered until the receiver can no longer capture and demodulate the signal and is limited by the noise figure of the receivers. These receivers typically have quite low noise figure, often below 1–2 dB noise figure. The difficulty in the measurement is setting and verifying that the power level of the source signal is known with good accuracy at very low levels. The difficulty arises because the test instruments used to measure and verify the signal level often have higher noise figures than the DUT and are used in systems where there might be substantial loss from the DUT test interface to the test instrument used to calibrate the source signal. In these cases, the noise from the test system dominates the low power measurements, and traditional methods have no way to lower the band power noise. But, with coherent power detection techniques, there are two methods for lowering the incoherent noise and recovering the true power of a low-level signal.

The first method relies on lowering the RBW of the measured test signal. As the RBW of the SA is reduced, FFT detection grid becomes much smaller, so there are many more FFT bins, but the effective noise floor is lowered in each bin. This technique does not lower the total band power, which is integrated over all FFT bins and so remains constant with changes in RBW. Figure 8.26 shows a measurement of the band power of a modulated signal under the conditions of two different RBW values. In this case, the signal source was set to –60 dBm,

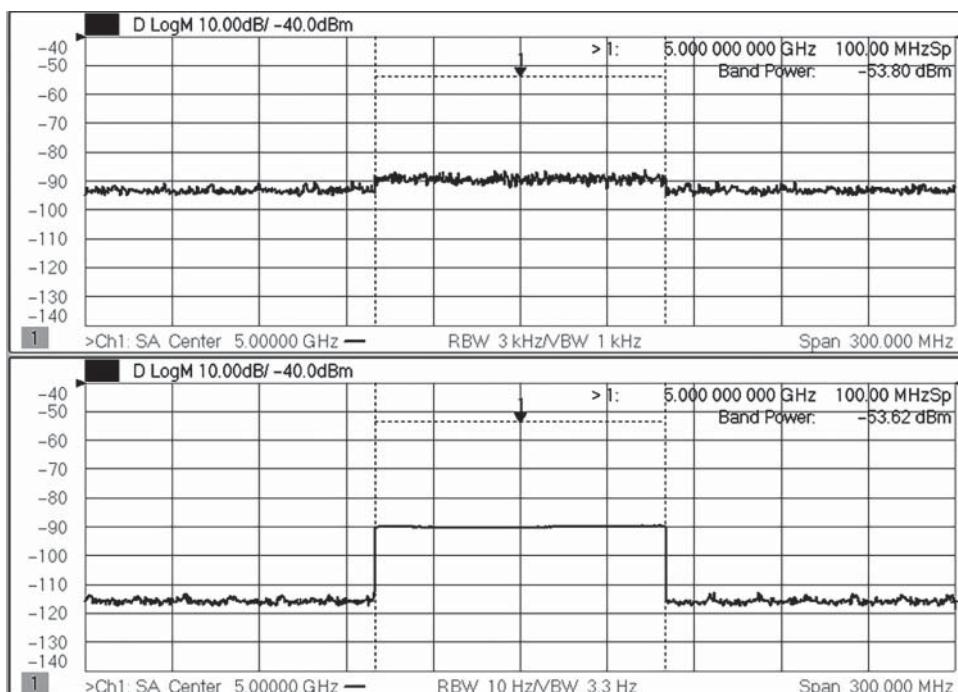


Figure 8.26 Band-power readings don't change with change in RBW.

but the band power reads significantly higher (-53.62 dBm) due to the integrated noise of the receiver. The upper plot shows the measurement for a 3 kHz RBW, with the signal just out of the noise floor. The lower trace shows the same signal, measured with a 10 Hz RBW. The total band power measured does not change even though this multitone signal is clearly far out of the noise floor in the case of the narrower RBW. This is why a band power measurement sensitivity cannot be improved using traditional noise reduction techniques.

But in the coherent detection mode, only the signals on FFT grids associated with the multitone period are used in detecting the band power. In effect, the FFT bins that are not on the multitone grid are set to zero. Thus, the power detection range is improved by the ratio of the RBW to the tone spacing of the modulation signal used. Figure 8.27 shows the band-power measurement of a signal near the noise floor, with a ratio of the modulation signal tone spacing to RBW of 30 (upper plot) and again with the ratio of 1000 (lower plot). The power level changes by about 0.5 dB due to power in the noise adding jitter to the reading of the signal in the upper plot. Here we see that in both cases, the power detected (about -60 dBm) is very close to the expected value, even though the actual band power (including noise from the VNA receiver) is on the order of -53 dBm .

To illustrate the effectiveness of this method, an additional 20 dB attenuation is added after the signal, and the expected power measured should drop by 20 dB. The results are shown

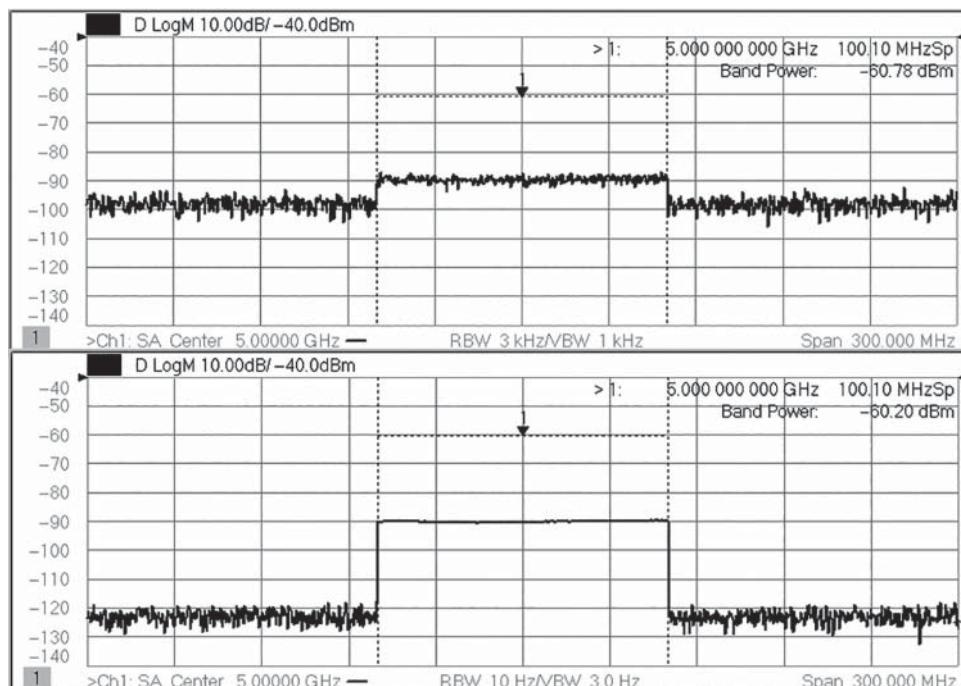


Figure 8.27 Lowering the RBW lowers the minimum detectable power with multitone coherent detection. Band power of a signal with tone spacing to RBW of ratio of 30 (upper), plot with ratio of 1000 (lower).

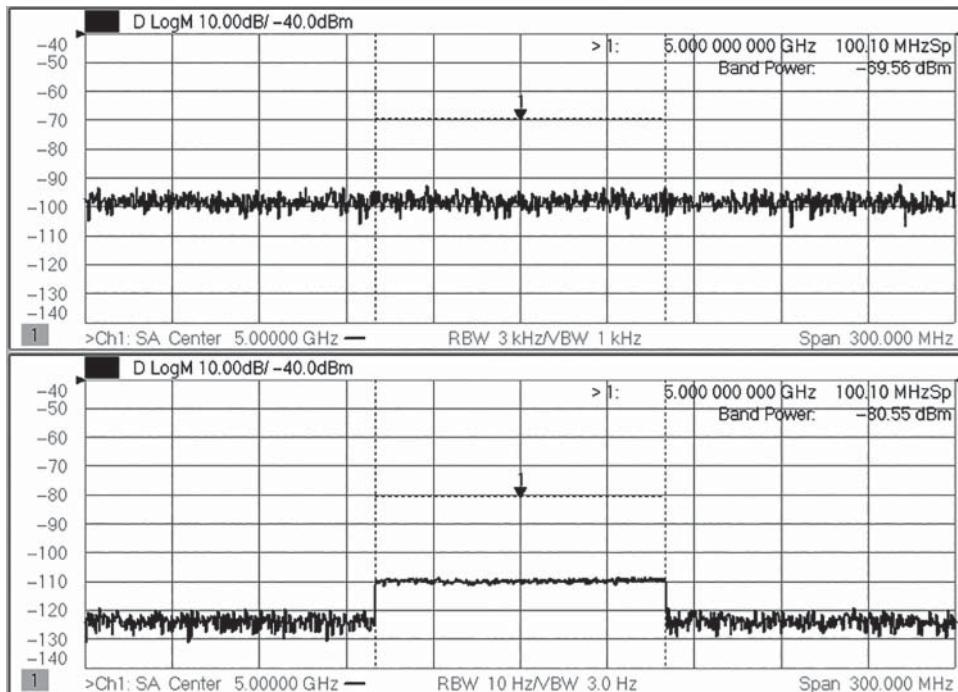


Figure 8.28 Same measurement as above with 20 dB lower signal power.

in Figure 8.28. Now the signal cannot be detected in the upper plot, and power displayed by the band-power marker shows only the power in the noise floor of the measurement receiver (which is less than the total noise power because only noise on the multitone grid is measured). The lower plot shows the power dropping by almost exactly 20 dB. It does not drop exactly by 20 dB because there is still some small contribution from noise power even in the lower RBW.

One could further lower the RBW and the expense of longer measurement time, but the measurement time will slow considerably for very low bandwidths as the FFT computation time increases with a factor of $N \cdot \text{Log}(N)$ where N is the number of sample points. But there is another method available to lower the noise floor when taking advantage of coherency: vector averaging.

The sweep-time difference between the upper and lower examples on the plots of Figures 8.27 and 8.28 is about 500 times (92 ms versus 45 s) when lowering the RBW. This is due to the longer FFT acquisition time (which scales linearly) and the FFT processing time (which does not scale linearly). Rather than using RBW reduction to lower noise floor, the vector averaging method of Section 8.1.4.2 can be used to lower the noise floor. Vector averaging occurs in the time domain, and processing can be done in such a way to reduce the total data need to go the FFT. In effect, the FFT has less samples to process and thus processes faster. Figure 8.29 shows the same condition as Figure 8.27 but instead of a lower RBW by a factor of 300, the vector averaging is increased by a factor of 300.

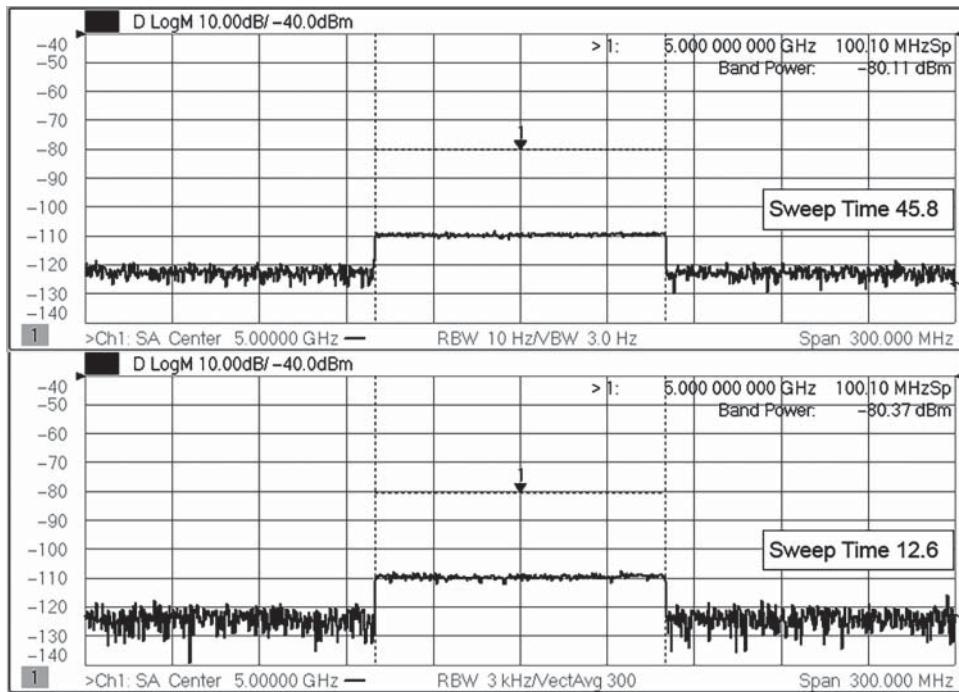


Figure 8.29 Vector averaging used to improve detection sensitivity.

Here the band-power measurements are improved in a similar way, but the measurement time does not slow down as much. Vector averaging in this case gives a similar noise floor but is nearly 4 times faster at 12.6 s versus 45.8 s for reducing IF BW by a factor of 300. There is another interesting consequence of using vector averaging: since the noise reduction occurs in the time domain (before the FFT), the traditional band-power measurement will also show a lowering of the noise floor. Thus, it is not necessary to go into the multitone power detection mode (that is, it is not necessary to zero the value of the FFT bins between the multitone spacings), as the effective noise power density is lowered by the vector averaging. The RBW remains the same, so there is the same number of FFT bins (as contrasted with an increase in FFT bins when the RBW is lowered), and since the noise power is lower, for the same bins, the integrated band power due to noise is lowered as well. This is demonstrated in Figure 8.30, where the zeroing of non-multitone bins is turned off; that is, all tones are detected. Here the source is set to -60 dBm, and we see the upper plot shows a much too large band-power value (as the integrated noise dominates the measurement), and the lower plot shows exactly to the proper band-power measurement. Since there remains some band power between tones, there may be a small offset due to the integrated instrument noise contributing to the band power; in this case, we read about 0.16 higher than in Figure 8.27.

These methods of coherent power measurement are generally applicable whenever the period of the multitone signal is known, and this knowledge is utilized to improve the measurement results.

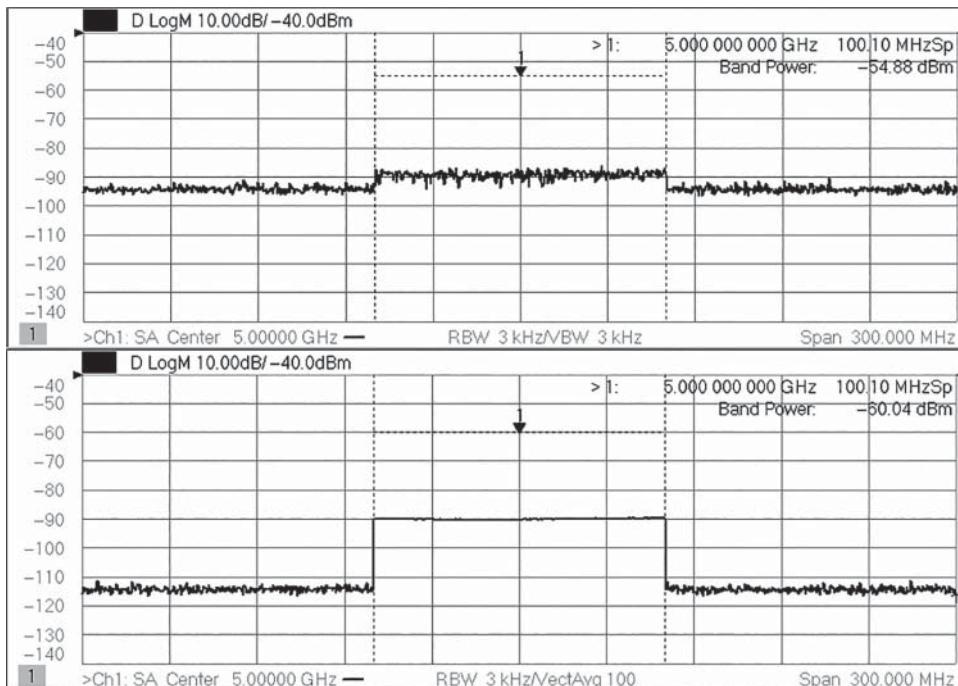


Figure 8.30 Vector averaging improves power detection in normal band-power measurements (all tones detected).

8.1.5 Calibration of SA Results

Traditional spectrum analyzers are calibrated at the factory for power measurements and have hard specifications on the power level accuracy. In the lower band of operation, where the spectrum analyzers utilize up/down-conversion for image rejection, the quality of the power measurement is usually quite good, on the order of ± 0.2 to ± 0.4 dB absolute. But in the region where a swept-tuned YIG preselector is used, the specified performance becomes much worse, on the order of ± 0.6 dB at the low end of the pre-selector band to ± 1.6 dB at the high end, near 50 GHz. And this ignores mismatch effects between the source and receiver (the specifications presume and ideally matched $50\ \Omega$ source). But, the spectrum analyzer is typically connected to a DUT through a cable. Most modern spectrum analyzers allow an amplitude correction for the cable loss but provide no means of correcting for the mismatch effect of combining the S_{22} of the cable to the S_{11} of the analyzer; this can add as much as ± 0.15 dB of ripple to the measurement at each connection (assuming an SA mismatch of -18 dB and a cable mismatch of -18 dB, quite typical at microwave frequencies), ignoring the effects of mismatch between the DUT and the analyzer. These accuracies are typically specified in swept-tuned mode, where the IF frequency is fixed. In FFT mode, where the wider IF is used to measure the signal, the IF frequency response of the spectrum analyzer will contribute an additional error; see Section 8.1.2.3. This IF frequency response is typically calibrated in place with an internal calibration, that often automatically refreshes if the

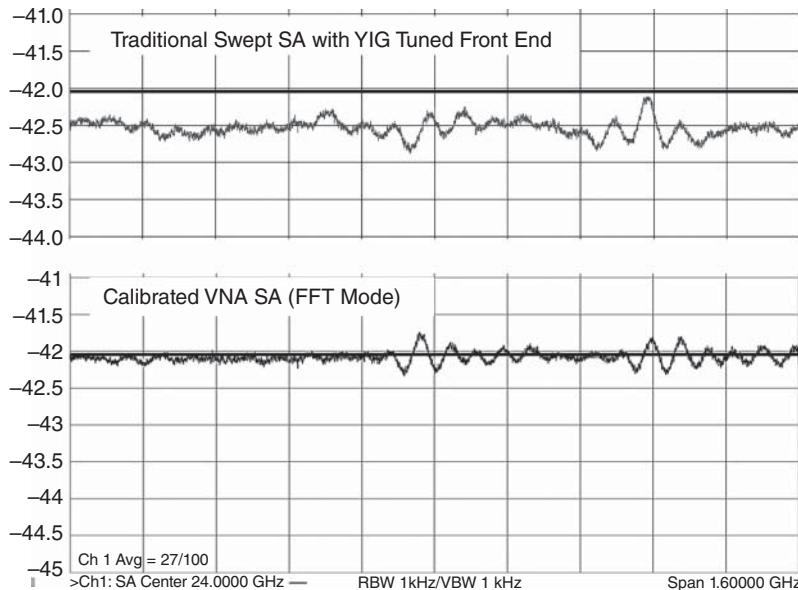


Figure 8.31 Amplitude accuracy in swept mode SA at 24 GHz center frequency: traditional SA (upper), VNA-based SA with VNA-based power cal (lower).

analyzer detects a temperature change, or too much time passes. Figure 8.31 shows the swept frequency response of a traditional SA compared to the response of a SA mode on a VNA with power calibration, at the end of the cable. This shows the flatness response of a multitone signal across 1600 MHz bandwidth, centered at 24 GHz. The signal was corrected for flatness measured on a well-matched, independently calibrated receiver.

The frequency response shown in upper trace demonstrates a relatively good flatness response (on the order of 0.2 dB) but an offset of about 0.5 dB. The dark line at -42.04 dBm represents the expected power of the calibrated source signal. The response of the lower trace, from a calibrated VNA, shows much less absolute offset and a little less relative ripple. Because there is a common ripple effect between the two traces, we might expect that to be an error contribution from the source flatness due to mismatch ripple during the source calibration.

The accuracy of a spectrum analyzer is best in swept-tune mode as the IF frequency response of does not affect the result. But for lower RBW measurements, it is often desirable to use the FFT mode. In the FFT mode, the quality of calibration of the IF response adds to the uncertainty of the measurement, but also the response of the pre-selector, which might change over frequency range and is a source of additional error. A prior figure, Figure 8.9, shows the response of a traditional spectrum analyzer, in the preselector band, with different FFT widths. The wider the FFT, the faster the data acquisition, but the more the pre-selector filter shaper will affect it. At sufficiently wide bandwidths (typically more than 40 MHz wide), the preselector must be bypassed to use a wide FFT. In such a case, the response is no longer image protected. Of course in the low-band region, where up/down-conversion is

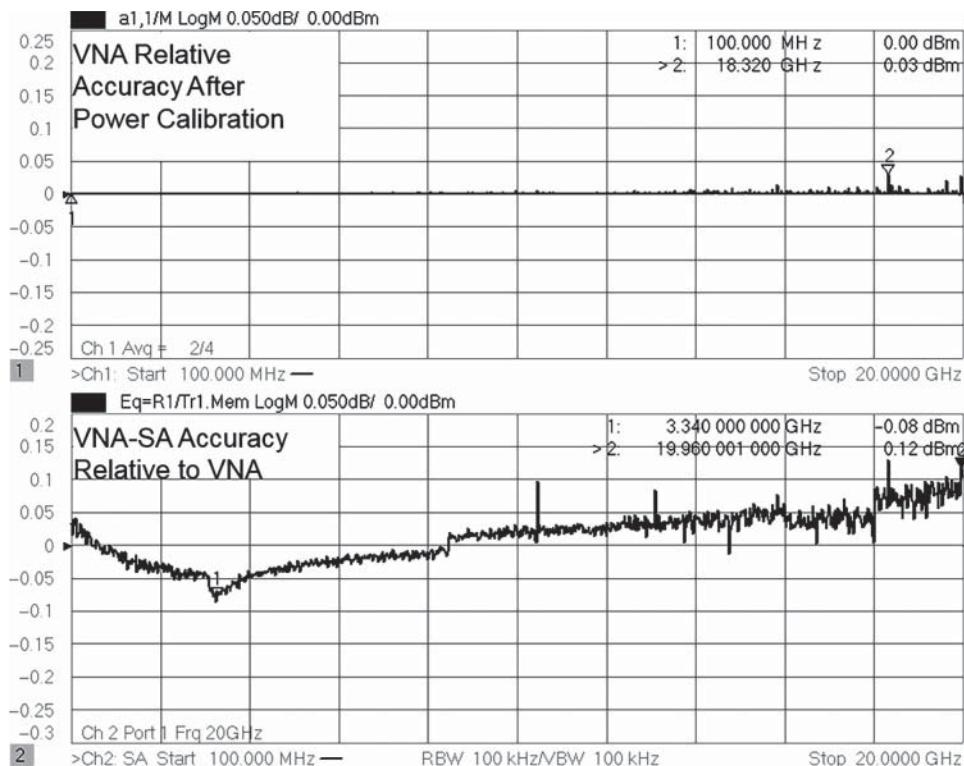


Figure 8.32 Wideband power accuracy of VNA spectrum analyzer mode.

used for image protection, there is no preselection filter, so only IF filter response (which is usually calibrated by an internal algorithm) affects the power accuracy.

A VNA-based spectrum analyzer typically does not have YIG-tuned preselector but uses digital image rejection. The RF receiver path is identical between VNA S-parameter measurements and SA power measurements; however, the IF path is not the same since the SA mode uses a wideband FFT processing step, where the VNA S-parameter channel uses a narrow-band digital filter. The difference between the IF response in the two cases can cause a power accuracy difference, but fortunately the IF response is calibrated in the factory, so the effect is minimal. Figure 8.32 shows a comparison of a VNA sweep with a power calibration compared with the same measurement using an SA mode sweep (in the SA-mode the source was stepped 1001 times over the same frequency range). The upper trace shows the effect of trace noise on repeated measurements of in the standard VNA power mode. The lower trace shows a “max-hold” trace of the stepped source, showing the effective response of the SA across frequency. The agreement here shows less than 0.12 dB offset. There are occasional spikes in the response on the order 0.1 dB, which occur at VNA receiver-band crossings.

In fact, the VNA SA mode power accuracy is specified relative to the VNA S-parameter mode power calibration, rather than independently. The S-parameter power calibration may be specified as described in Chapter 3, which depends upon the calibration accuracy

of a power meter, plus the residual mismatch between the source and the power meter (the raw mismatch removed by a mismatch calibration in the VNA channel) and power meter cal-factor uncertainty. To that, add the SA relative power accuracy (relative to the VNA-based calibration) to obtain the total uncertainty in the SA mode for power measurements, plus mismatch uncertainty between the DUT output S_{22} and the load match of the test port used for the VNA SA mode. One may add an additional trace that is a direct measurement of the source over the same frequency range, using a power meter independently. One small point: this last trace, while using a power meter directly, is not quite as accurate as the VNA-based S-parameter channel power measurement because the power meter reading is not compensated by the slight mismatch between the power meter and the source (on the order of 0.03 dB effect). The VNA source was set to deliver 0 dBm to a 50Ω load, but the power meter load is slightly different from 50Ω .

A VNA-based spectrum analyzer can utilize the built-in correction functions of the VNA to nearly perfectly compensate for the cable loss and mismatch effect, to the uncertainty of the power meter used to calibrate the VNA system. The VNA-based calibrations can also apply advanced functions like automatic fixture removal (see Chapter 11) or fixture de-embedding, to allow accurate spectrum analyzer measurements in cases of on-wafer or in-fixture devices.

8.1.6 Two-Tone Measurements, IMD, and TOI Definition

For many amplifiers, especially narrow-band amplifiers, the key specification for distortion is two-tone third-order intermodulation distortion, usually called IMD. While a two-tone measurement may be created through modulation methods, it is more common to create the two tones by combining two CW sources. As the name implies this distortion is a result of having two tones (usually of equal amplitude) of frequency F_L and F_U (for lower and upper tone) applied to the input of the amplifier. If the power in the two tones is sufficiently large to drive the amplifier into non-linear operation, the output spectrum will contain at least two other tones at frequencies of

$$\begin{aligned} F_{3U} &= 2F_U - F_L \\ F_{3L} &= 2F_L - F_U \end{aligned} \quad (8.8)$$

Higher levels of drive will create higher-order products such as the fifth- and seventh-order products. If we define the center and delta frequencies as

$$\begin{aligned} F_C &= F_U + F_L \\ F_\Delta &= F_U - F_L \end{aligned} \quad (8.9)$$

Then Nth order IMD product is defined as

$$\begin{aligned} F_{NU} &= F_C + \frac{(N-2)}{2} \cdot F_\Delta \\ F_{NU} &= F_C - \frac{(N-2)}{2} \cdot F_\Delta \end{aligned} \quad (8.10)$$

There is some confusion over the definition and description of terms for two-tone intermodulation distortion; for the purposes here, the following definitions will be used:

TOI: Third-order intermodulation product; the power in the third-order tone (also used to refer to this class of distortion measurements). Sometimes used for third-order intercept point (more properly called IP3).

IMD: Intermodulation distortion; a general reference to this class of distortion measurements.

Main Tone: One of the two drive tones, measured at the output PwrMain.

Upper Tone: Higher-frequency main tone; aka Hi Tone, PwrMainHi.

Lower Tone: Lower-frequency main tone; aka Lo Tone, PwrMainLo.

IM: IMD product in dBc relative to the nearest product.

IM3Hi: Upper third-order IM product in dBc.

IM3Lo: Lower third-order IM product in dBc.

Pwr3Hi: Upper third-order IM product in dBm.

Pwr3Lo: Lower third-order IM product in dBm.

IP3: Third-order intercept point.

OIP3: Output referred IP3, usually used with power amplifiers.

IIP3: Input referred IP3, usually used with a low-noise amplifier (LNA) and receiver amplifiers.

PwrMainHiIn: Upper frequency input power tone.

PwrMainLoIn: Lower frequency input power tone.

From Eq. (8.8) the IM3 products are created by modulation of one tone's second harmonic with the other tone's fundamental. This could be considered essentially a mixing product, and the power of the IMD product is related directly to the power of the two components of the mixing product. The power of the second harmonic changes 2 dB for every 1 dB of its fundamental, and the power of the other product changes dB for dB, so the power of the third-order IMD terms changes 3 dB for every 1 dB change in the pair of tone powers, and the dBc level of the IMD products changes 2 dB for every 1 dB of change in the fundamental tones.

Whenever dBc values are used for IM tones, it is always relative to the nearest main tone, if the two main tones are not exactly equal. Often the frequency response of an amplifier will cause the two main tones to be somewhat unequal. By using the definition of the nearest tone as the reference for dBc values, the value of the IM tone, in dBc, will be the same as if the two main tones were exactly equal. To understand this, consider the IMD products in Figure 8.33. This condition holds even if the tones are widely separated in power, for normal amplifiers.

Here the IM values are compared in two cases: in one, the IM tones are computed when the main power tones are equal, taking the lower IM value from the difference between the Pwr3Lo tone and the PowerMainLo tone. In the second case, the two tones are intentionally set different, with the upper tone raised by 5 dB and the lower tone reduced by 5 dB; this case is marked "Unequal." As indicated in the figure, the IM computed for this case is nearly identical to that for the "Equal" tone case, even though the output tone powers are substantially different. The average tone power (averaged in the dB sense) is the same. Raising the upper main tone raises the upper IM tone by two times the dB number (in this case 10 dB) because it is related to the nearest tone by a second-order term, as shown in Eq. (8.8), and raises the lower IM tone by one times, or 5 dB. In compensation, lowering the lower main tone by 5 dB lowers the lower IM tone by 10 dB and lowers the upper IM tone by 5 dB. As a result, the two main tones are 10 dB apart as are the two IM tones, and the dBc value relative to the nearest main tone is exactly the same as before the main tone powers were changed. In this way, one can accurately compute the IM products, and the OIP3 value, even if the measurements of the two main tones are not equal, by computing the value for the average of the two powers.

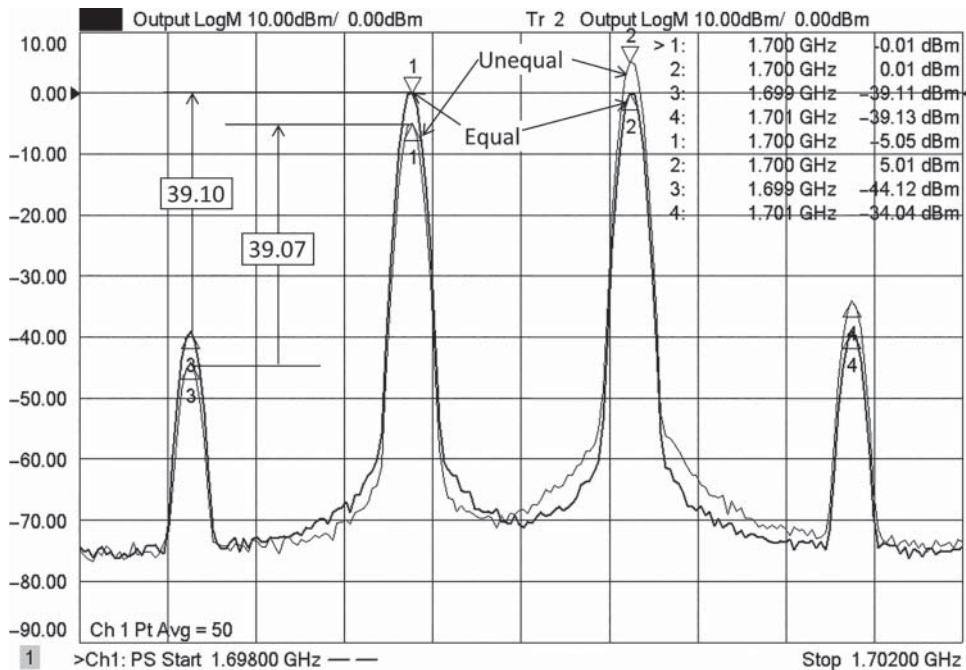


Figure 8.33 IMD measurements with offsets in the main tone power.

Of course, if only one main tone is changed in power, the average power is changed, and the dBc value will be changed as well, to the value that represents the new average power.

8.1.6.1 Intercept Points: OIP3

The output-referred intercept point is computed by projecting this 3:1 power response of the IM tones along a straight line projection with respect to an increased main tone power until they cross (or intercept) the projection of the main tone power, as illustrated in Figure 8.34. Note that this intercept point presumes a 3:1 slope of the IM tones versus main tone power so that only one IM reading is needed to compute the IP3 point. Since the lower the IM products are, the further they will have to increase to theoretically intercept the main tone power, and thus the higher the IP3 power will be, one can conclude that higher IP3 points are indicative of more linear amplifiers.

However, the IP3 point is simply a figure of merit; no actual intercept is measured. Its computation or value does not depend upon the actual linearity behavior of the amplifier's IM products. It is computed as

$$IP3 = PwrMain + \frac{|IM3|}{2} \quad (8.11)$$

For example, if at a power of 0 dBm output, the IM product is -38 dBc, then the IP3 will be +17 dBm. In fact, the IP3 value can and does change with drive level. Also shown in the figure is the computed OIP3 for each point of the power sweep. As expected, at lower

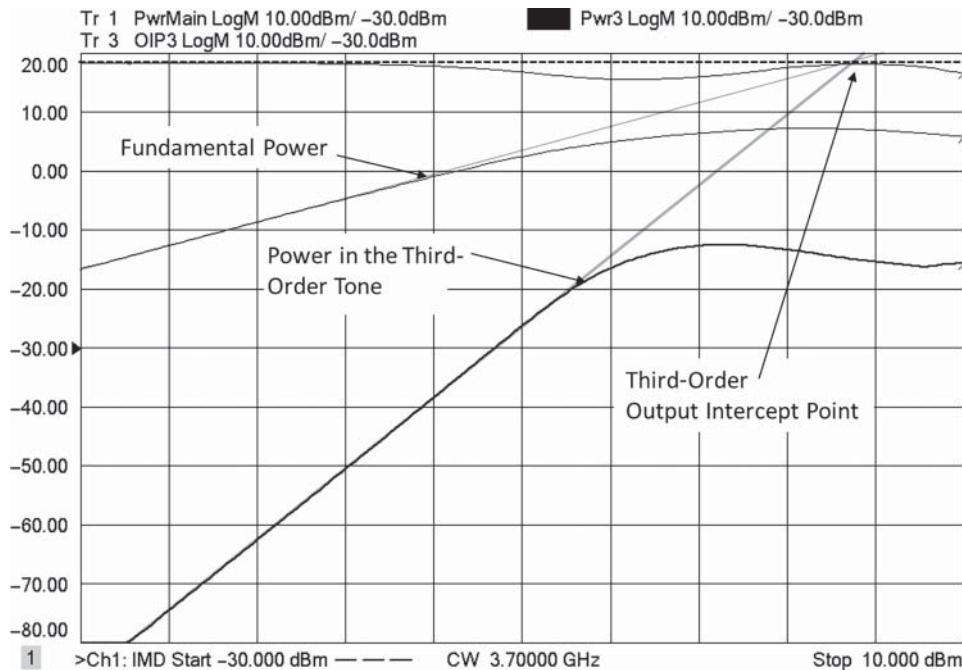


Figure 8.34 Projecting IM tones to obtain the IP3 point.

powers, the projected value for OIP3 and the computed value from the trace are the same. For simple amplifiers, the IP3 often decreases as the power increases. But some amplifiers are designed to enhance their IP3 value at some higher drive powers, and the IM tones may have some regions where they decrease before increasing again. In these areas the IP3 values can dramatically increase.

The IM products and intercept points for higher-order products follow a similar set of definitions; for example, the fifth-order products simply replace the 3 with a 5 in the definition of the products, and the intercept point is computed projecting the IM produce with a slope of 5:1 rather than 3:1.

8.1.7 Measurement Techniques for Two-Tone TOI

The measurement methods for IMD have historically used two sources and a spectrum analyzer. Some modern sources have built-in AWGs so that a single source can create a two-tone or even multitone signal, but sometimes the self-generated IMD of the modulator limits the quality of the two-tone signal. More recently, some modern VNAs have provided two independent sources that can be combined to create a two-tone signal; in at least one case the sources have built-in banded filters so that the self-generated IM products are very low, typically less than -90 dBc at max power.

For a receiver, a spectrum analyzer is usually used, but recently some VNAs have been developed with spectrum response modes and include integrated applications that

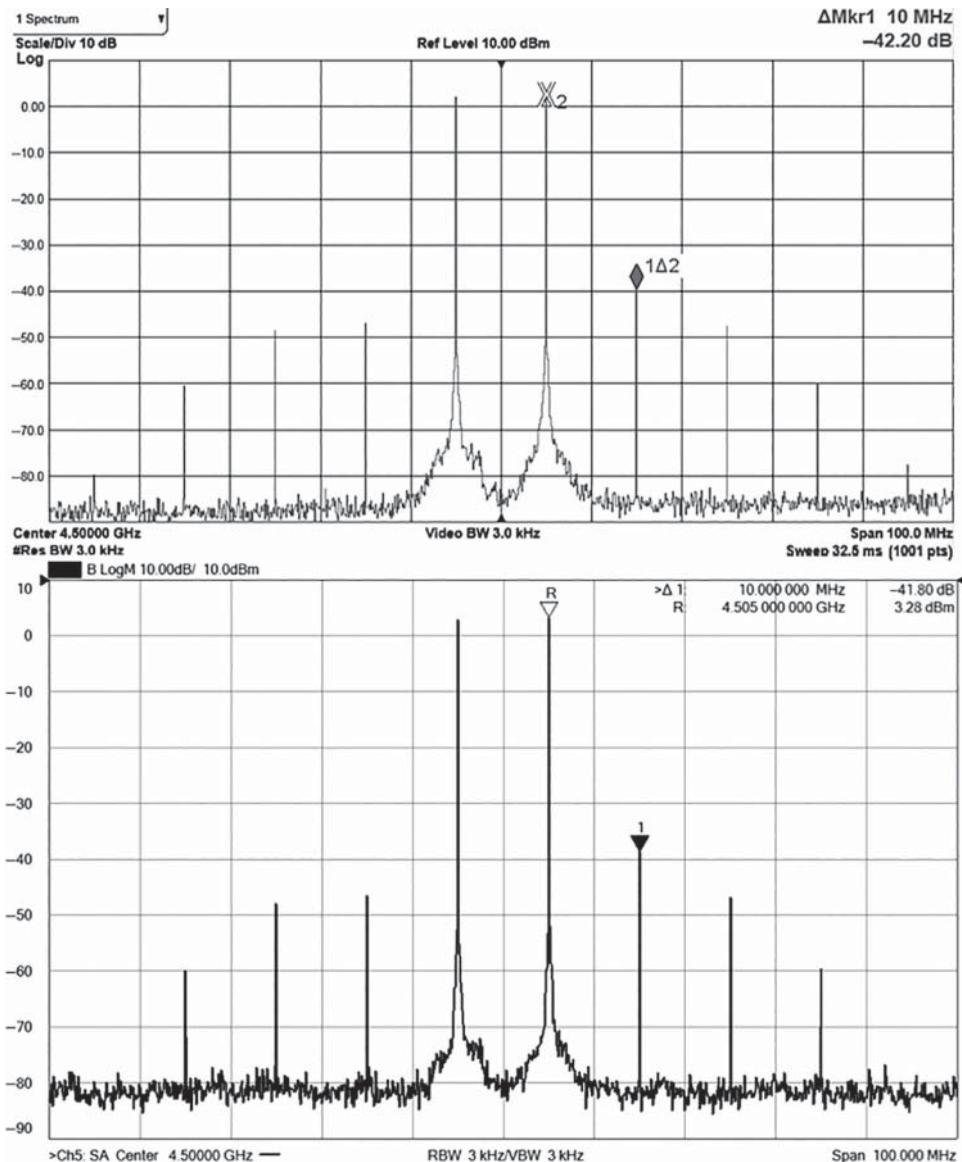


Figure 8.35 SA and VNA measurement of IM spectra.

automatically control the sources and receiver frequencies to tune to the F_C and set the span to accommodate the delta frequency. Typically, the best spectrum analyzers are better than the best VNAs in linearity for measuring IM products, on the order of 5 dB or more at lower frequencies to about equal at higher microwave frequencies. For single-frequency measurements, either may be used. Markers can be used to find the power in the tones and IM products, as shown in two examples of a SA and VNA measurement in Figure 8.35.

In this figure, the upper plot is from a Keysight UXA spectrum analyzer, and the lower plot, from a VNA SA, is a Keysight PNA-X model. This UXA spectrum analyzer has a noise reduction feature, called noise-floor extension (NFE), whereby the power of the noise floor of the SA receiver is subtracted from the measured signal to yield a lower displayed average noise level (DANL). Such a noise floor extension can be emulated on a VNA by measuring first the noise floor of the receiver, and then the IMD of the amplifier, and subtracting the noise power of the former from the latter, which is displayed in the right plot of the figure. This provides about 5 dB improvement in the displayed noise level for the VNA and improves the measurement accuracy of power for the IM tone. Here the SA and VNA measurements agree within a few tenths of dB. The signals shown were created with the VNA sources; using precision stand-alone signal sources will usually give lower phase noise. Also, since a similar synthesizer is used as the LO for the VNA, the phase noise shown in the VNA plot is about 3 dB higher than in the SA plot. Spectrum analyzers typically have much better close-in phase noise than a VNA.

For fixed single-frequency measurements, there may be an advantage to using the newer VNA spectrum analyzer mode, where the FFT-based capability makes the sweep faster for the high-dynamic range cases, when the IF is lowered to improve the noise floor, as compared to the traditional SA. Both SA and VNA in spectrum mode must sweep the frequency spectrum to search for the maximum signal level, even if the exact frequencies of the IM products are known. With external programming, the SA may be put in a zero-span mode and tuned to exactly the frequency of each tone. But many VNAs have a built-in mode to sweep receiver frequencies separately from source frequency; thus, VNAs can measure at only the tone and IM frequencies. This is known as a *frequency offset mode* (FOM) and can be used to support a swept IMD measurement, but that function has largely been replaced with custom applications that are specific to swept IMD measurements as described in the next section.

8.1.8 Swept IMD

In swept IMD measurements, either the center of the tone frequencies is swept, holding the power level and the tone separation (or delta frequency) constant; or the tone frequencies are held constant and the tone powers are swept; or the center frequency and tone powers are held constant and the delta frequency is swept. For any of these cases, the receiver is set to measure the main tones and IM products at each source setting. The swept variable is displayed on the x-axis, and the IMD results are shown as traces versus the swept variable. Many VNAs have frequency-offset mode, where the source frequencies can be swept independently from the receiver frequencies. If this mode is used for IMD measurements, at least four VNA measurement channels will be required, one for each tone. However, some modern VNAs have built-in applications that provide a complete solution for the swept IMD measurement, meaning a single channel can be used to measure and display all the tones and IM products, along with some precalculated functions for IM and IP3 values.

In this way, plots such as the IMD versus power plot shown in Figure 8.34 become easy to create. For many amplifiers, the swept power IMD measurements are the most interesting because various linearity enhancement techniques can yield vastly different IMD versus input-power traces, and the plot of higher-order terms can sometimes give the designer insight into the processes that create the IMD results.

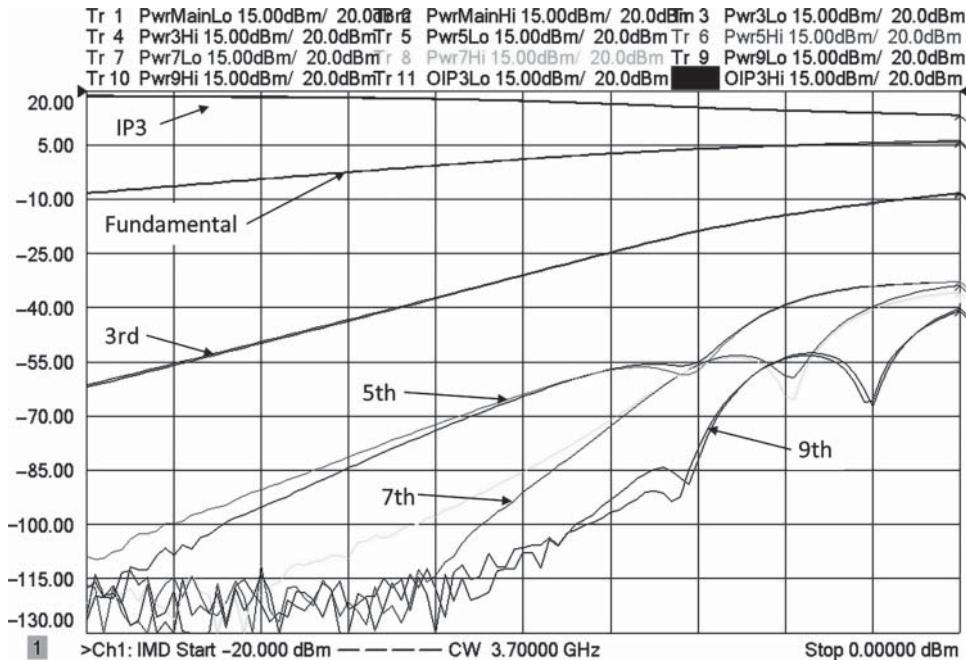


Figure 8.36 Swept-power IMD measurements.

Figure 8.36 shows the result of a swept power IMD measurement. On the rather crowded screen are the power of the upper and lower main tones (PwrMainHi and PwrMainLo), the power in the IM tones up to ninth order (Pw3Hi, Pw3Lo, Pw5Hi, Pw5Lo, Pw7Hi, Pwr7Lo, Pwr9Hi, Pwr9Lo), and the output-referred IP3 point for the high and low tones. The inflections in the higher-order tones are likely caused by the mixing of different order products (which generate higher-order tones) adding and subtracting in phase.

For the same amplifier, Figure 8.37 shows a plot of IMD measurements versus sweeping the center frequency of the IM tones, holding the tone spacing and power constant. The shape of the input filter is clearly shown; for a constant input power, the IM tones are greatly reduced in the out-of-band area. It is interesting to note that the higher-order tones show significant variation in signal across the band and peak at the filter band edges.

Finally, again for the same amplifier, Figure 8.38 shows the result of sweeping the delta frequency or tone separation frequency while keeping the center frequency and power constant. For this amplifier, there is some change in response versus delta-frequency in the fifth-order IM product. This might be due to some effect in the amplifier such as response of the bias network to the tone separation frequency that remodulates the output in a way that causes some products to have variation with tone spacing and may be the reason that the upper and lower tones generate different responses.

The unique advantage of using a VNA in IMD measurements is that both the sources and receivers are internally controlled and derived from the same reference. As such, very narrow IF bandwidths (the equivalent of a resolution bandwidth in an SA) can be used, as only one frequency point for each tone power is measured to compute the overall result. Also, the

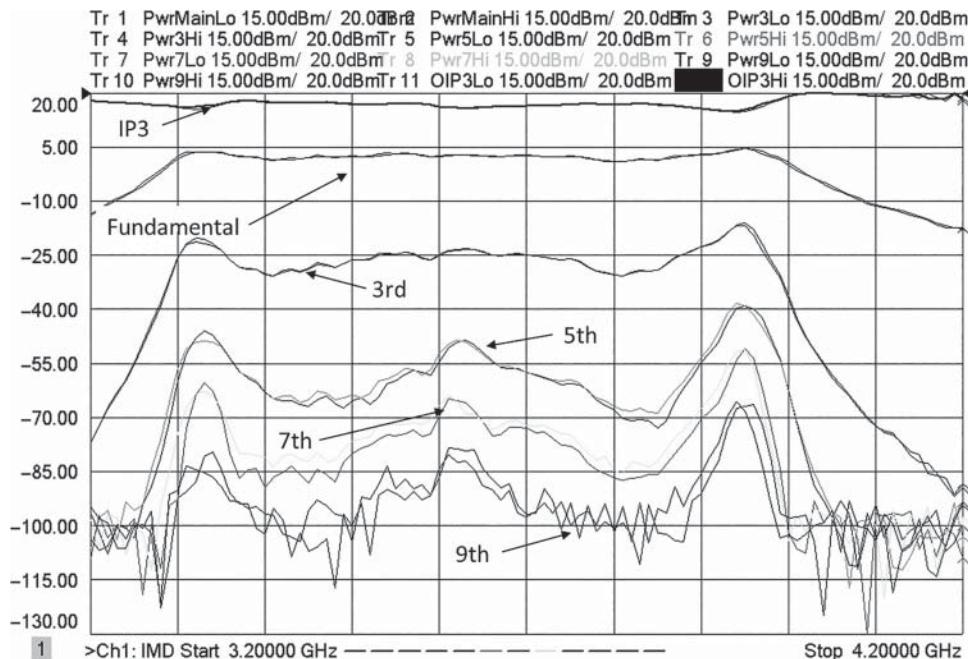


Figure 8.37 Swept-center frequency IMD measurements.

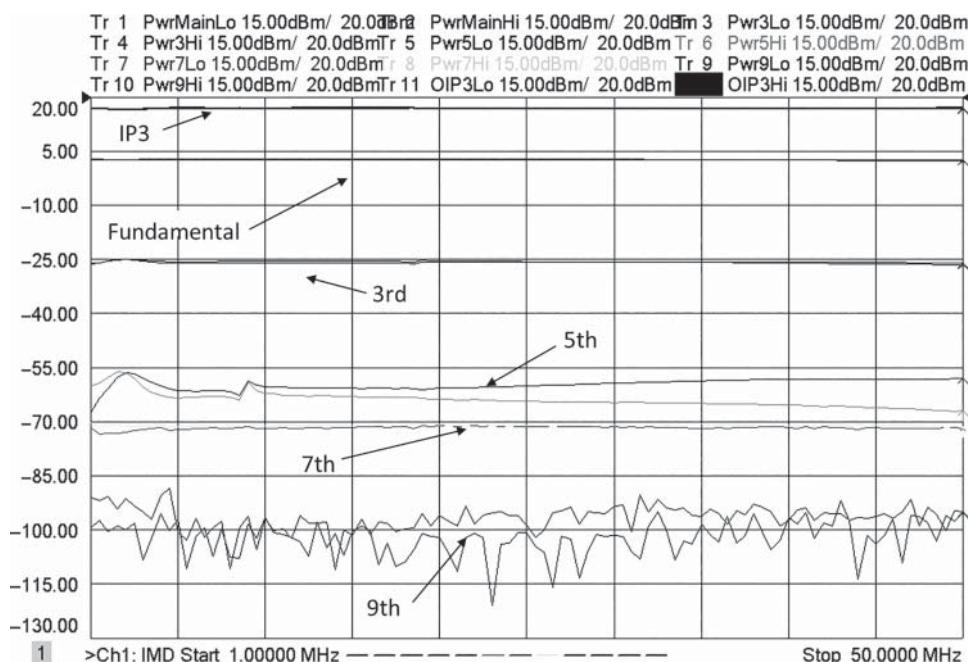


Figure 8.38 Swept-delta frequency IMD measurements.

resolution bandwidth can be set wider for the main tones and narrower for the IM tones to optimize the measurement speed.

8.1.9 Optimizing Results

The measurement of IMD products may be limited by the linearity of the receiver, as well as by the IMD content and signal quality of the source. Optimization of the measurement setup can greatly improve the measurement results.

8.1.9.1 Source Optimization

Measurement optimization includes choosing the tone spacing, when such a choice is possible. Wider tone spacing helps eliminate masking of the IM signals due to the phase noise of the source. If the tone spacing is sufficiently close, the phase noise of the source at the IM product spacing will swamp the IM product signal, unless very narrow IF bandwidths are used. If possible, selecting a wider tone spacing will reduce or eliminate the effect of phase noise on the measurement result. In Figure 8.35, one can see the phase noise just dropping down into the noise floor at about 150 kHz off the tone frequency, for the VNA spectrum. Closer tone spacing would cause the IM product to be lost in the phase noise, especially in the VNA-based measurement.

Another area of optimization consists of ensuring that the sources self-generated IM products are below the level of the DUT products.

A possible cause of source-generated IM is due to harmonics (particularly second harmonic) of one-source mixing with the output of another. If two sources are combined to create the two-tone signal, adding a filter that suppresses the second harmonic can substantially reduce the self-generated IMD. Some VNAs and signal sources include many narrow-band switched filters that provide for very low harmonic levels. Sources without integrated filters often must be followed with external narrow-band or low-pass filters to remove the harmonics, and these filters will limit the frequencies that can be measured.

Even with filtering, non-linear effects in the sources output amplifiers can cause source-generated IMD. Using a coupler instead of a simple combiner can provide dramatically increased isolation at the expense of power in one of the tones. Some modern VNAs either use a switched internal coupler or use test port coupler from an unused port, to combine the two sources together. This does result in a loss of maximum power that can be applied with two equal tones to the DUT. If the power is too limited in this way, the internal combiner may be bypassed by an external equal-loss combiner. This will increase the available tone power by approximately the loss of the coupler less the loss of the combiner, or 16 dB minus 6 dB for around 10 dB improvement in most cases, as illustrated in Figure 8.39.

The main cause of source-generated IM products is direct cross-modulation between the two sources, as well as IM generation in the combining network. When a coupler is used, the isolation of the coupler prevents this source of IM generation, but when higher power is needed and an external combiner is used, the cross-modulation of sources will likely occur. One way to improve the source direct cross-modulation (from one source to the other) is to add isolators between each source and the combiner, also illustrated in Figure 8.39. The combiner and isolator can be used in cases where higher power is needed, for example if

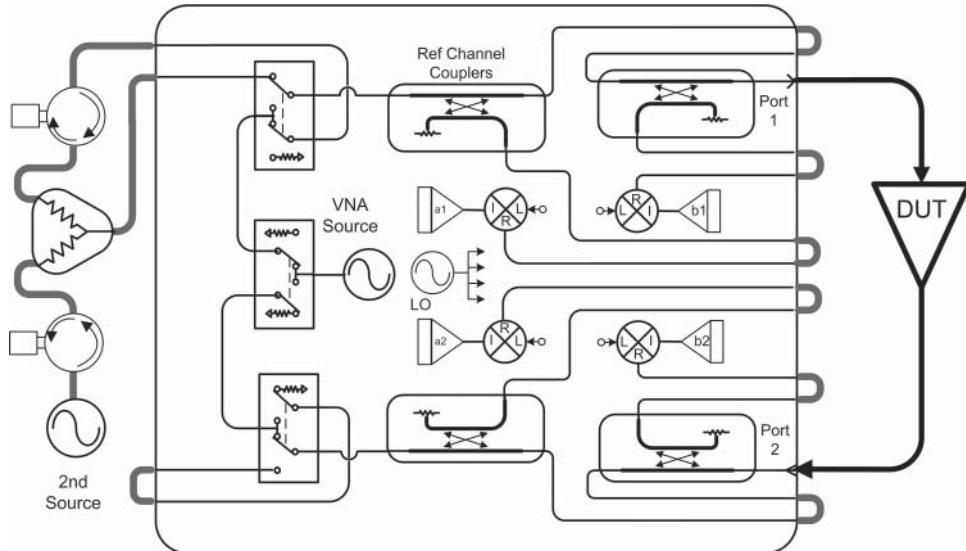


Figure 8.39 Block diagram of combined source with isolators.

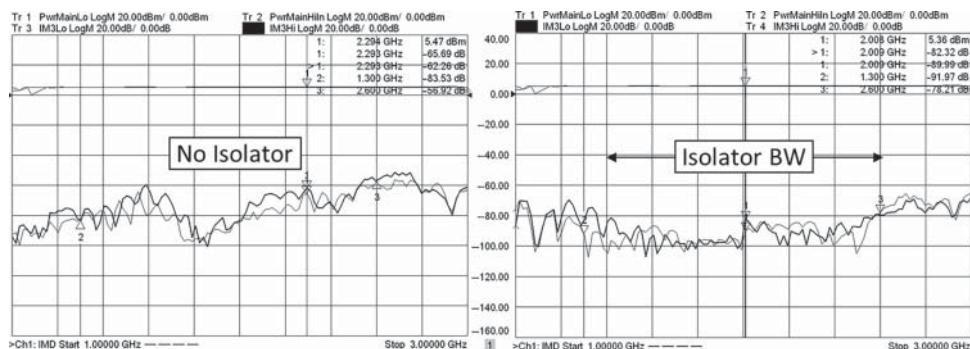


Figure 8.40 Source-generated IMD due to direct cross-modulation with and without isolation.

a booster amplifier is applied to the either source. With very high powers, one must ensure that the combiner and couplers following the combiner themselves don't generate passive intermodulation. In the diagram, an external source is used as the source for the second tone, but some VNAs provide two or more internal sources that can be used as well.

Figure 8.40 shows the result of combining two sources with a low-loss combiner. In one case (left plot), the sources are directly connected, and in the other (right), the sources have an isolator placed between each source and the combiner network. The isolator doesn't cover the full band, so at the edges, the source-generated IM products are not improved. Here, over the bandwidth of the combiner, the worst-case source-generated IM products are improved by about 20 dB. In this measurement, the receiver is padded down by 25 dB, and an IF BW of 1 Hz is used to obtain the best noise floor.

In some cases, including the case illustrated previously, the source IM products are generated by some mechanism that is extremely sensitive to power so that lowering the power by even 5 dB will cause the source IM products to drop to the noise floor. This can occur in circuits that have some limiting aspect to them; such limiting is sometimes placed internal to instrumentation sources to protect against reverse power from a DUT causing damage to the source. Thus, there is a trade-off between maximum source power and IM generation.

Finally, one should avoid using tone spacings that match identically with internally generated clock signals of the measurement instrument, with the worst-case example being 10 MHz spacing. Almost all instruments provide a 10 MHz reference oscillator and derive their own frequencies from this reference. Any leakage of this 10 MHz signal may show up in very low-level IMD measurements if a 10 MHz tone spacing is used. Slightly changing the tone spacing, even by a few kilohertz, will eliminate this issue.

8.1.9.2 Receiver Optimization

For the most part, source effects can be removed to a negligible level with isolation and filtering between the sources, but receiver effects generally are more difficult to account for. The non-linear behavior of the receiver will generate IM products from the main tone powers hitting the receiver. In general, it is difficult to remove the main tone products as very narrow notch filters, or a pair of band pass filters, are required. Thus, the linearity of the receiver often sets a limit to the level of IMD products that can be measured. The main method for reducing receiver-generated IM products is to add attenuation to the receiver channel.

Adding attenuation reduces the main tone values, and for every 5 dB of attenuation added, the IM products go down 15 dB, for a 10 dBc improvement. However, the noise floor comes up by 5 dB and so further IF bandwidth reduction is needed to maintain the same noise floor.

The IMD properties of the receiver can be evaluated by changing the receiver attenuation value and looking for any change in the IMD product level, up to the point that the noise floor dominates. In a spectrum mode, the noise floor is set by the RBW, and narrow RBWs will dramatically slow the sweep. However, in swept IMD mode, the narrow RBW is applied only to the IM tone, rather than sweeping the entire frequency range around the main tones and IM products, as the spectrum mode does, so the speed will be much faster.

For many systems, the source IM level can be improved sufficiently so that the receiver dominates. In such a case, a large source signal into the receiver, with low or no attenuation, will show the receiver-generated IM level directly. Adding attenuation should show the expected drop until the source generated level is reached.

Figure 8.41 shows the result of driving the receiver with a pair of +5 dBm tones, with 0 dB and 15 dB and 30 dB attenuation. The receiver-generated IMD is clearly visible. In this case, a 1 Hz IF BW filter was used to set a noise floor of approximately -120 dBm. At the highest receiver attenuation, the IM tones are limited by the source IM generation. At any particular frequency, the combination of source and receiver IM products can either add or subtract, depending upon phasing, so there will be ripple in the IM tone if its contributions from both the source and receiver take a part in generating it.

It is typical for an SA to have a default minimum attenuation of 10 dB; this is to protect the input mixer. VNAs have a 13–16 dB input coupler and so generally don't have any extra

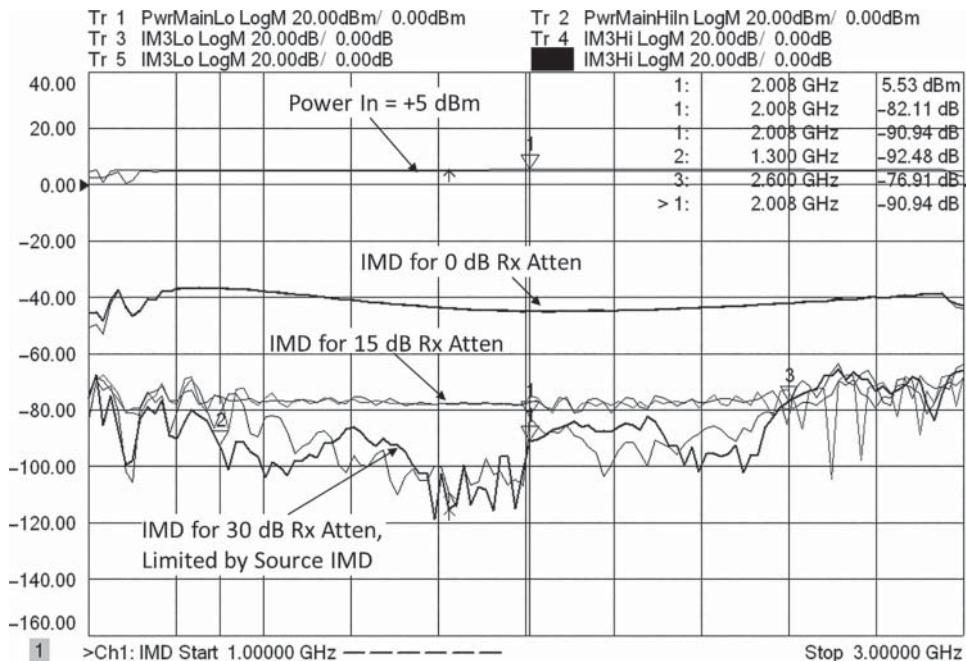


Figure 8.41 Receiver-generated IMD in swept IMD mode, at +5 dBm input, with 0, 15, and 30 dB of receiver attenuation.

default receiver attenuation. For cases where the signal being measured is so small that the receiver will not generate any IMD products, the SA attenuator may be set to 0 dB using a manual override. On a VNA, the test port coupler may be reversed (see Section 5.2.6) to improve the sensitivity of the receiver. For microwave VNAs, whose couplers often have corner frequencies in the 500 MHz range, it may always be necessary to reverse the coupler for lower-frequency IM measurements to remove the loss effect of the coupler roll-off. In some case, the coupler could roll off at very low frequencies (10 MHz) by as much as 35 dB. In these cases, the coupler may be reversed, but it might then be necessary to add some receiver attenuation to avoid distortion effects due to direct receiver access by the thru arm of the coupler.

8.1.10 Error Correction

Error correction of the IMD measurements is typically provided by a source power calibration and a receiver response calibration. For a two-tone stimulus, the source power calibration may need to be separated into two steps, one for the lower tone and one for the upper tone. If using sources and an SA, each source will need to be calibrated independently for cable and fixture loss, typically with a power sensor. The combiner should be in place to account for its loss. After one source is calibrated (over all frequencies needed for all tones), then a

thru connection can be made to the SA, and it may be calibrated. Finally, the second source may be turned on and calibrated by the SA, leaving the first source on (and offset by the delta frequency for each cal point) to account for its mismatch effect on the output of the second source. If the match of the second source changes when it is turned on, the first source will need to be re-calibrated as well. The SA is used as a frequency discriminating receiver to calibrate each source independently.

Unfortunately, the mismatch between the SA and the sources is a main cause of error, and cannot be removed, so making the match very good is important for accuracy. The SA response calibration will typically be around ± 0.3 dB due to source mismatch interactions, given a 15 dB source power match and a 15 dB match of the test system (including cables and connectors).

When using VNA, very similar calibrations can be made, but because the VNA contains a built-in reflectometer, the effects of source and load match on the receiver response can be nearly eliminated. A source power cal is performed on the first tone, along with a reference receiver response calibration. Then a 2-port cal is performed between the input and output. From these two calibrations the exact b_2 receiver response can be computed removing the source and load match effects to a residual level. For a typical setup as described, with calibration kits having 40 dB or better residual directivity and match, the error in the match-corrected response tracking will be less than 0.03 dB, or 10 times better than the simple response cal. Finally, a source power calibration can be performed on the second source for the upper tone. The modern VNAs that have built-in swept IMD measurements also have built-in calibration processes that essentially automate all the steps described previously, requiring only a power meter connection and an Ecal thru connection to give fully match corrected source and receiver power corrections.

Experience shows that the source and receiver tracking terms are quite flat over narrow bandwidths, on the order of 0.01 dB per 10 MHz, so if the delta frequency is less than 10 MHz, it is not required to calibrate at every tone frequency, but only at the center frequency of the tones, and then interpolate the tracking term for each of the tone frequencies. For higher orders of tones, this can dramatically shorten the calibration at virtually no degradation in quality of measurement.

8.2 Distortion Measurement of Complex Modulated Signals

Complex modulated signals are used in almost all communications systems and are distinguished from other modulations, such as AM, FM, and pulse modulation, in that the information in the signal is contained in both the magnitude and phase of the signal or, more precisely, in the complex representation of the signal. Often this is referred to as the I/Q information, for in-phase and quadrature-phase, which is another way of saying the real and imaginary parts of the signal. The distortion measurement of two-tone IMD provides a means to characterize the linearity of an amplifier or mixer, but the response to complex signals may be difficult to predict from the two-tone IMD, and further the distortion caused by the in-channel part of the signal may produce products outside the signal band interfering with the operation of adjacent channel services. Some common figures of merit for distortion of modulated signals are described in this section.

8.2.1 Adjacent Power Measurements

For many amplifiers or systems driven with a modulated signal, a key measure of distortion performance is the adjacent channel power level (ACPL), adjacent channel power ratio (ACPR), or adjacent channel leakage ratio (ACLR). ACPL is a measure of the power in an adjacent communications channel caused by distortion in the desired channel and is essentially a measure of the third-order-intermodulation products caused by intermodulation distortion due to non-linear amplifier behavior in the main channel. Consider a modulated signal with flat spectral distribution: this is similar to what is seen for an OFDM signal and can be created using a band-limited AWGN signal, as shown in Figure 8.42. This signal is driven into an amplifier that is starting to approach compression, and the distortion will appear as a triangle shape on a linear watts scale, or a pair of rounded shoulders in dBm format. The reason for the shape of the distortion is related to the particular power spectrum density of the signal. This main signal has a uniform power spectral density. It is an AWGN signal generated from a repetitive waveform, so it has a discrete spectrum of uniform amplitude, random-phase multitone signals with a spacing that is the reciprocal of the waveform period. The triangle nature of the linear (watts) response is due to the fact that only signals that are spaced exactly one signal-bandwidth apart can intermodulate to generate a signal that is a bandwidth-and-a-half away from the center frequency. There are only two such multitones in this signal, so at the extremes of the adjacent channel, the distortion is lower. But in the edge of the adjacent channel near the signal, any two multitone signals that are one-half a bandwidth apart can intermodulate to generate the distortion near the band edge. There are many more multitone signals so situated, so the amount of distortion near the band edge is maximum. In fact, the distortion also occurs in-channel, with uniform distortion across the signal bandwidth, but

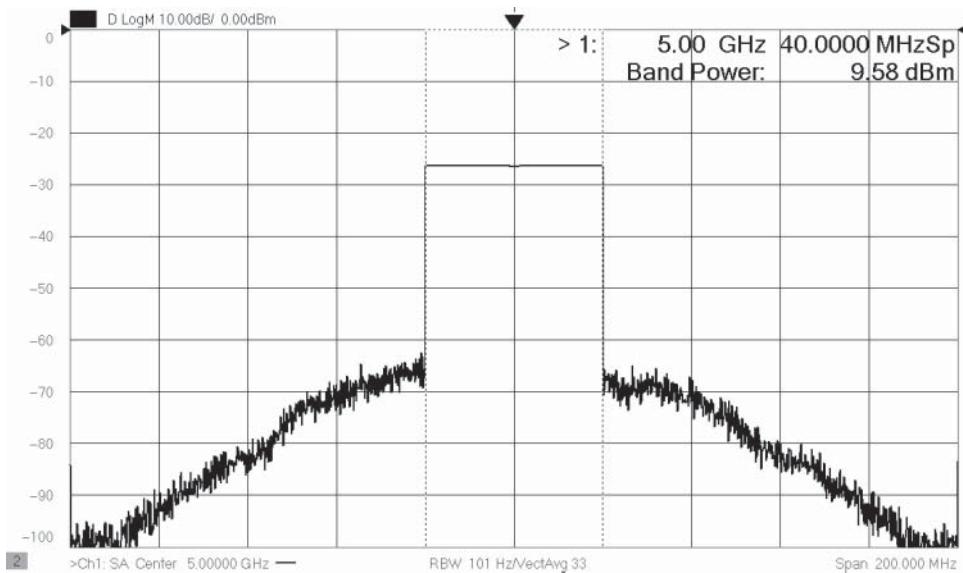


Figure 8.42 Spectrum of an AWGN driven into an amplifier that is non-linear.

it is not easily discerned because it lies on top of the main signal. This in-channel distortion is often measured in another way, using an NPR measurement; see Section 8.2.2.

8.2.1.1 Adjacent Channel Power Ratio (ACPR)

ACPR is a similar measure of distortion, where the ratio of the integrated power (also called the *band power*) in the adjacent channels to the total power in the main channel is displayed. It is typically measured using a spectrum analyzer with the band-power marker functions. Some spectrum analyzers have predefined applications that allow one to enter just the values for center frequency, main band power span, adjacent channel offset frequency, and adjacent channel bandwidth, and it will automatically compute the band-power values.

In Figure 8.43, the band-power markers are set up to measure the adjacent power ratio. The reference marker is used to measure the main channel band power, and Markers 1 and 2 are used in a “delta-marker” mode to measure directly the ACPR in dBc.

Some test scenarios require measurements of adjacent channel power where there are multiple carriers present in the DUT. In this type of measurement, several carriers are generated across the radio spectrum, with one or more carriers turned off, and the adjacent channel power that occurs at the frequency where a carrier is turned off is measured, which is demonstrated in Figure 8.44.

In this measurement the entire radio spectrum may be loaded with carriers, and the test scenarios may include various power levels for individual carriers, but the key measurement remains the power level caused by intermodulation or other distortions in the ACP channel that has had the carrier removed.

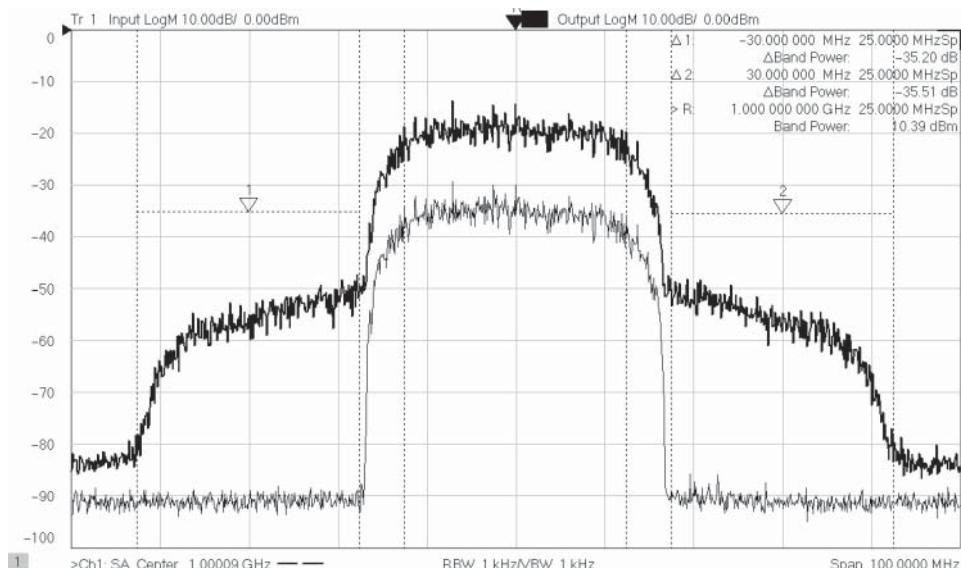


Figure 8.43 ACPR measure of a 64QAM signal measured through an amplifier; lower trace is the input, and upper trace is the output with band-power markers showing ACPR.

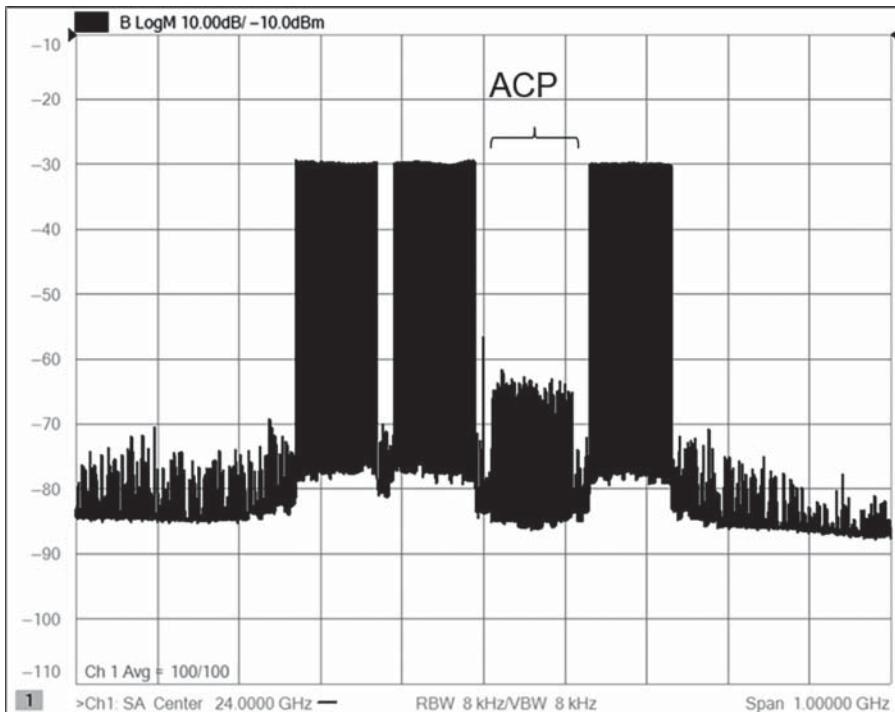


Figure 8.44 Multicarrier adjacent channel level measurements with one channel turned off.

8.2.1.2 Adjacent Channel Power and EVM

Adjacent channel power can affect the quality of a communications system in several ways: the power in the adjacent channel from a transmitter may overwhelm a near-by receiver, which is tuned to a farther transmitter thereby blocking the weaker signal, or it may add an interference signal to the same transmitter output on an adjacent carrier. In the first case, it is simple to understand how this channel blocking can affect an adjacent receiver. But it might be less clear how such a low signal level, typically -55 dBc, would affect the adjacent channel signal from the same transmitter, as the adjacent signal may be quite large. In most systems, the power level of each channel, transmitting to each subscriber, is controlled to provide just enough power to main a high-quality signal, based on a two-way feedback from the subscriber. In some cases, the power from the adjacent channel may be higher than from the subscriber channel, and the ACPL in the subscriber channel can be significant. It adds an interference signal to the subscriber channel's signal, and this interference will affect another measure of distortion: EVM. For a first estimate, the EVM level (typically described in percent) will closely associated with the ACPL (after conversion from dBc to percent).

Additionally, the ACPL from a transmitter, being a measure of the distortion of the transmitter, can also cause the transmitting channel to have in-channel distortion, as the same mechanism that causes adjacent channel distortion also causes in-channel distortion that causes

EVM. In-channel distortion is difficult to measure directly, and EVM measurements normally require demodulation to determine the EVM, but EVM levels are closely related to the distortion seen in the adjacent channel, and the in-channel distortion can be measured directly, using a specialized signal described in the next section.

8.2.2 Noise Power Ratio (NPR) Measurements

While adjacent-channel power is the most common way of measuring distortion of a modulated signal, it does so only in the out-of-channel region. In some communications systems, for example, satellite systems, the adjacent channel response is strongly filtered using a multiplexing filter so that adjacent channel power does not appear at the output of a system. However, the in-channel distortion remains, and it is difficult to estimate its level, especially in a generalized sense. While it is possible to use a modulated signal of some particular format, the wideband channels of a satellite system can be reconfigured for different signaling types, so it is desired to have a method to measure in-channel distortion without depending on demodulation to determine the distortion level. The generalized method that has been developed for many years is the NPR measurement.

8.2.2.1 Generating NPR Signals

In the past, it was difficult and expensive to generate wideband modulated signals (which in early satellite business was often generated in practice by combining many narrow-band signals), so an alternative method to generate a wideband signal was used, utilizing a broadband noise source, followed by a bandpass filter (which is set to the width of the channel) followed by an amplifier to achieve the desired power level, finally followed by a narrow-band band-stop filter in the center of the pass band. This created a signal with uniform power spectral density (thus uniform loading of the amplifier across its operating width). A block diagram of the setup is shown in Figure 8.45. However, with the advent of modern wideband modulated RF sources, most manufacturers have moved to a different method of generating NPR signals, utilizing a AWG playing back a waveform that emulates an ideal broadband noise signal with a deep notch.

As described in Section 8.1.3, a signal played back from a waveform is periodic in nature and may be represented by a multitone signal. Further, we know that an AWGN signal must have a uniform spectrum, so one can easily represent a noise-like signal by defining a set of multitone signals across the desired bandwidth, but assigning a random phase to each tone. The notch is created by setting the amplitude of the tones in the notch region to zero.

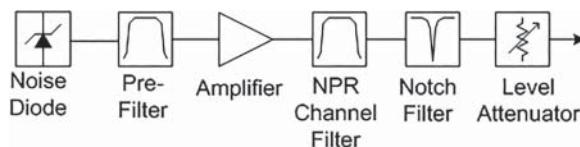


Figure 8.45 Traditional setup for generating an NPR signal.

The duration of the waveform is set by the tone spacing, so for a given NPR bandwidth, the number of tones is determined from the overall bandwidth and the tone spacing.

One figure of merit for modulated signals is the peak-to-average power ratio. For a true noise-like signal, the peak-to-average power would be infinite, over an infinitely long interval, as a true noise-like signal can take on any value, including an arbitrarily large value (though very rarely). For any finite waveform, played back from an AWG, the peak-to-average power is constrained by the number of sample points (in the time-domain waveform) or the number of tones (in the multitone spectrum, which is equivalent). The largest peak-to-average ratio, in dB, that can be achieved is $10 \log(N)$ and would occur if one of the time samples was 1 and the rest were 0, that is, an impulse signal. But in practice a random signal would not have such a waveform, so the peak-to-average signal would be much lower.

For noise-like signals the peak-to-average ratio goes roughly as $10 \log(\ln(N))$, where N is the number of sample points in a waveform. Table 8.1 shows the computed peak-to-average ratio versus number of tones (or number of samples) for a randomly generated noise-like waveform where for each example number of tones, 1000 different random phase seeds were used, and the mean and standard deviation of the peak-to-average was computed. This table makes it clear that the more tones used, the less the variation in the peak to average, but also that the mean peak-to-average ratio grows slowly with the increasing number of tones.

One interesting outcome of the considering noise-like signals waveforms is the recognition that the signal is only noise-like (in terms of amplitude) in the time domain. A true-noise signal will have a uniform spectrum (that is, flat response looking at the frequency domain) when measured over a long time. To create a noise like waveform of some duration (or period), a multitone signal with tone spacing of 1/period is generated with constant amplitude and random phase. Making the phase random over 0–360° ensures the signal follows a Gaussian noise-like CCDF curve. Figure 8.46 shows the CCDF curve of a 1001-tone noise-waveform and a 1 000 001-tone waveform. These curves were generated by the method of creating a flat-spectrum multitone with random phase. In each case, the curve follows closely the ideal Gaussian noise curve (which follows the formula)

$$10 \log(CCDF) = \frac{10^{(-P^2)}}{\ln(10)} \quad (8.12)$$

where P is the peak-to-average power ratio (in dB) on the x-axis and the dB value of CCDF is on the y-axis. It is clear that when the CCDF percentage approaches $1/N$ (where N is the number of time samples in the waveform, or tones in the spectrum), the CCDF curve appears discrete valued. Of course, this is because the actual waveform has only discrete values.

Table 8.1 Peak-to-average statistics for noise-like waveforms

Number of Tones	Peak-to-Average Ratio (Mean)	Standard Deviation	Peak-to-Average Ratio Max	Peak-to-Average Ratio Min
101	7.58	0.87 dB	11.14	5.37
1 001	9.04	0.65	11.76	7.55
10 001	10.16	0.49	12.2	8.99
100 001	11.03	0.40	12.54	10.03

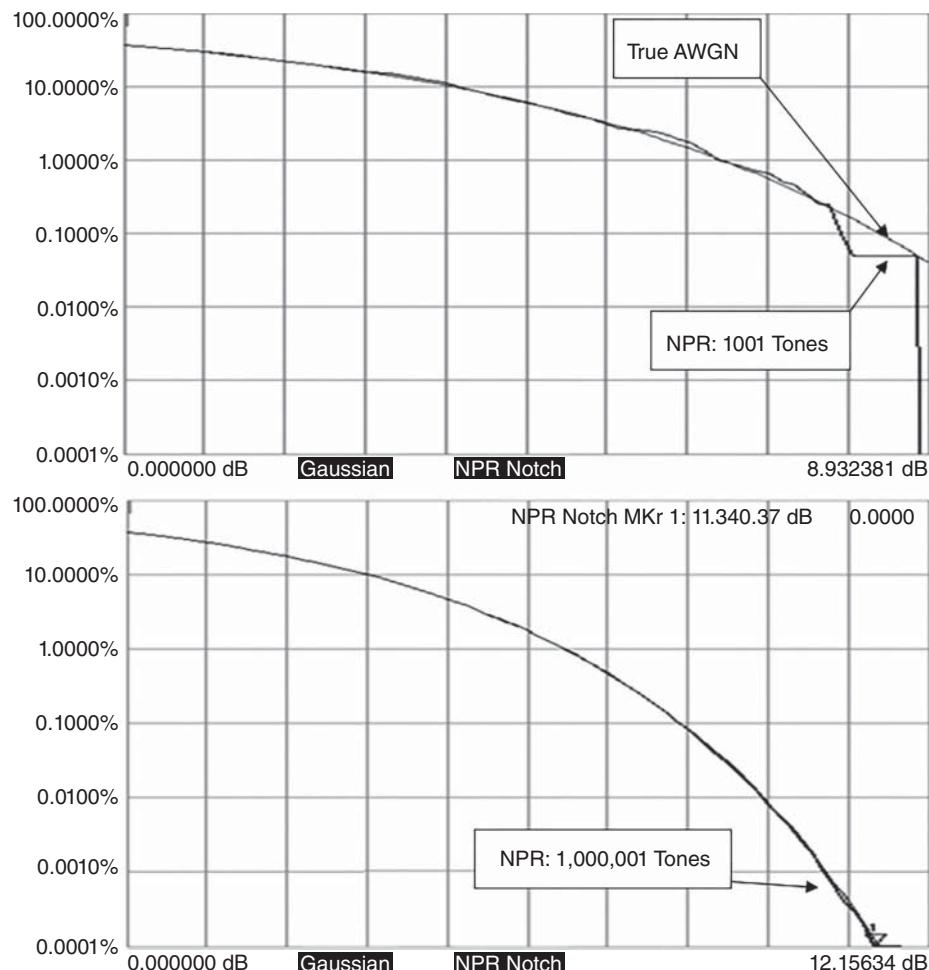


Figure 8.46 CCDF curve of a noise-like waveform generated from a random phase multitone signal.

Once the CCDF curve falls below $1/N$ in percent, the CCDF must go to zero. Note that on the 1001-point NPR waveform, only one point of the 1001 points does not fall closely to the true AWGN curve, and for nearly all purposes, this can be ignored. On the 1 000 001-point waveform, nearly all points fall on the AWGN curve. Finally, note that even true-noise signal, if sampled over a finite time, will show samples that do not fall on the ideal AWGN CCDF curve.

8.2.2.2 NPR Measurement: Band Power

An NPR signal is created as an AWGN waveform, but with one or more notches in desired locations. The notch is typically in the center of the spectrum to be analyzed, but it is

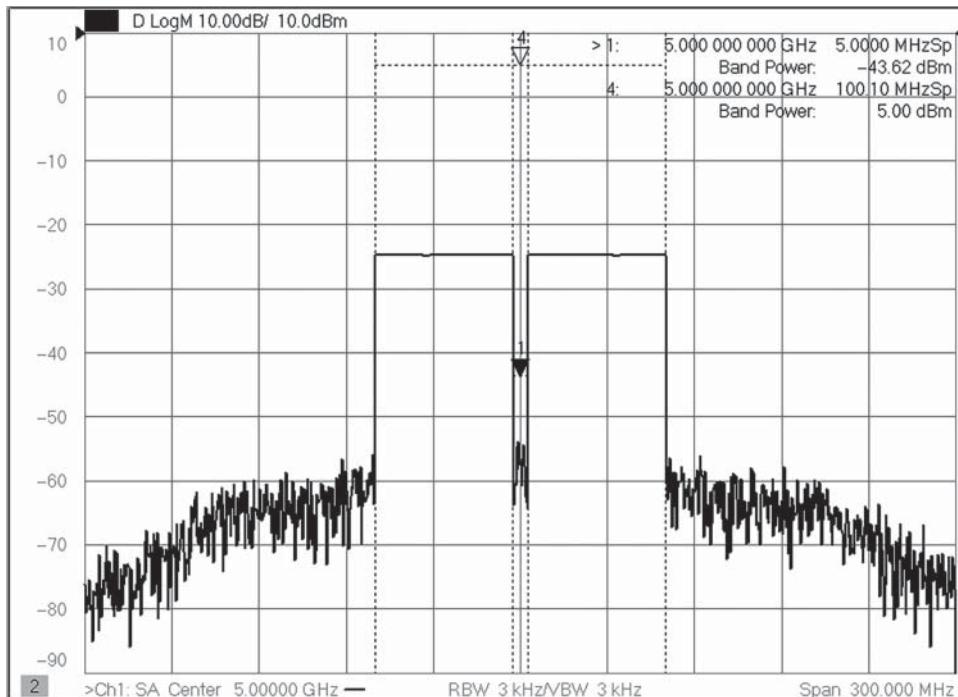


Figure 8.47 NPR signal with band-power markers.

sometimes offset by half the notch width to avoid the LO leakage that can occur with some methods of signal generation. An NPR signal is shown in Figure 8.47, generated using 1001 points across 100 MHz, with a 5 MHz (5%) notch. The traditional band-power markers are shown that measure the band power in the total spectrum (100 MHz) and the band power in the notch. From this, the NPR can be computed as $-43 - (+5) = -48$ dBc, but it is more common to refer to band power density (in dB MHz^{-1} , or dB Hz^{-1}) when measuring NPR. This is because the traditional way to make the measurement was to estimate a line through the signal region, on a spectrum analyzer analog display, as well as through the noise region, and take the vertical distance as the NPR. Computed that way, the NPR is found to be larger, with the power density of the main signal being

$$\begin{aligned}NPR_{\text{dBc}} &= 10 \log \left(\frac{\text{Signal_Density_W/MHz}}{\text{Notch_Density_W/MHz}} \right) = 10 \log \frac{\left(10^{(\text{SigPwr_dBm}/10)} \right) / (\text{BW} - \text{Notch})}{\left(10^{(\text{NotchPwr}/10)} \right) / \text{Notch}} \\NPR_{\text{dBc}} &= -10 \log \frac{\left(10^{(5.0/10)} \right) / (100 - 5)}{\left(10^{(-43.62/10)} \right) / 5} = -35.83 \text{ dBc}\end{aligned}\quad (8.13)$$

This definition matches more closely the traditional measurement of visually identifying the dBc difference and provides a means to compute NPR based on definitively computed values of band-power density.

8.2.2.3 NPR Measurement: Band-Power Density

The band-power method described previously requires additional computation on the results to obtain the generally accepted definition of NPR. However, if one used band-power density (which is power per unit Hz) instead of band power, one can measure the NPR more directly.

Many traditional spectrum analyzers, and VNA-based SA mode, include a “band-noise” marker function that displays power in a band in terms of dBm Hz⁻¹. This is essentially the band-power density and is often used with noise measurements, since noise measurements must always be made over some band (unlike a CW signal, the probability of noise power at a zero bandwidth tends to zero). Figure 8.48 shows the result of using Band-Noise markers directly. The Reference Marker is used on one-half the signal to capture the signal density, marker 4 is used in band-power mode to determine the total signal power, and marker 1 is used with both Band-Noise and Delta Marker functions to directly show the NPR level in dBc (only dB is displayed). This results in a direct measurement of NPR. For this reading to be accurate, the noise power density across the span must be constant. If the Reference Marker covers the entire span of the signal, the NPR will be degraded slightly by the ratio of

$$\text{Offset} = 10 \log \left(\frac{\text{Signal_BW}}{\text{Signal_BW} - \text{Notch_BW}} \right) = 10 \log \left(\frac{100}{100 - 5} \right) = 0.22 \text{ dB} \quad (8.14)$$

as the power density in the main signal is computed slightly lower than the true value.

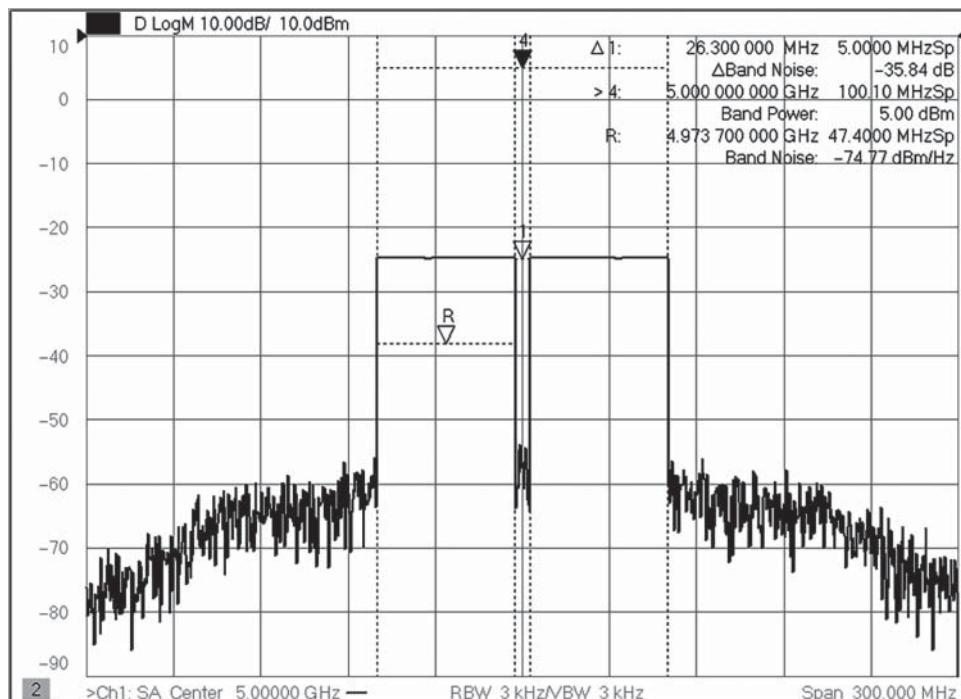


Figure 8.48 NPR using Band Noise markers to get a direct reading of NPR.

8.2.3 NPR Signal Quality and Correction

Since NPR signals are commonly created using an AWG upconverted in an RF or microwave source, the quality of the microwave source will affect the NPR reading from the device under test. There are three common issues that affect the NPR signal quality, listed here:

1. Flatness and power correction of the signal: The NPR signal should have a uniform (flat) spectrum to match the uniform spectrum density of a noise signal. Loss, ripple, and other compression in the source (or associated power amplifiers used to boost the source power) will affect the quality of the signal. It is common to do some type of flatness correction, as well as offset the nominal power to make up for nominal loss
2. LO Leakage: Often an IQ modulation method is used to create the NPR signal, and if the desired notch is in the exact center of the signal, the NPR signal may have LO feed-through that dominates the power in the notch. There are three methods generally employed to avoid the LO leakage, listed here:
 - a. Offset the notch by one-half the notch width, so the LO leakage is in the large-signal portion of the signal, and not in the notch.
 - b. Use single-sideband modulation to create an NPR signal only above (or below) the LO signal. For this to be effective, the NPR signal width must be less than one-half the modulation bandwidth of the source. Further, IQ imbalance can give an image of the NPR signal offset on the opposite sideband of the LO.
 - c. Use LO suppression methods to lower the LO signal level. Since the LO leakage falls on the DC element on the IQ signal, it is possible to somewhat suppress the LO by generating a DC element that will exactly cancel the LO signal. Usually this is used only with advanced software delivered by instrumentation manufacturers. Some signal generators have an automated LO suppression that will suppress the LO through an automated DC offset process.
3. Signal distortion caused by the signal source or booster amplifier that creates its own undesirable level of notch distortion. Some instrumentation manufacturers have developed software that can determine the level of distortion (using a spectrum analyzer) and produce a compensating signal in the IQ data that will predistort the signal resulting in significant reduction in the self-generated distortion, even through a pre-amplifier. In the past, this process was quite slow, but recent improvements in VNA-based spectrum analyzers along with utilizing coherent techniques has sped up the process by 100 to 1000 times, making it practical for large multitone signals.

Figure 8.49 shows the result of programming an NPR signal over the full bandwidth of the AWG modulation source. This signal shows all three defects mentioned previously.

This signal was set to create a 160 MHz wide NPR signal with a 16 MHz (10%) notch, at 5 dBm power. The output power measured with Marker 4 shows we are about 0.4 dB below the target power, and we also see about 2 dB of slope across the spectrum with some ripple. The ripple is likely due to mismatch in the system; this illustrates issue 1 from earlier.

We also see the LO feed-through quite clearly in this signal. The nominal carrier power is 5 dBm, and the LO feed-through is around -33 dBm, so this source has 43 dB of LO suppression, but still the LO level is so high it completely dominates the NPR notch power.

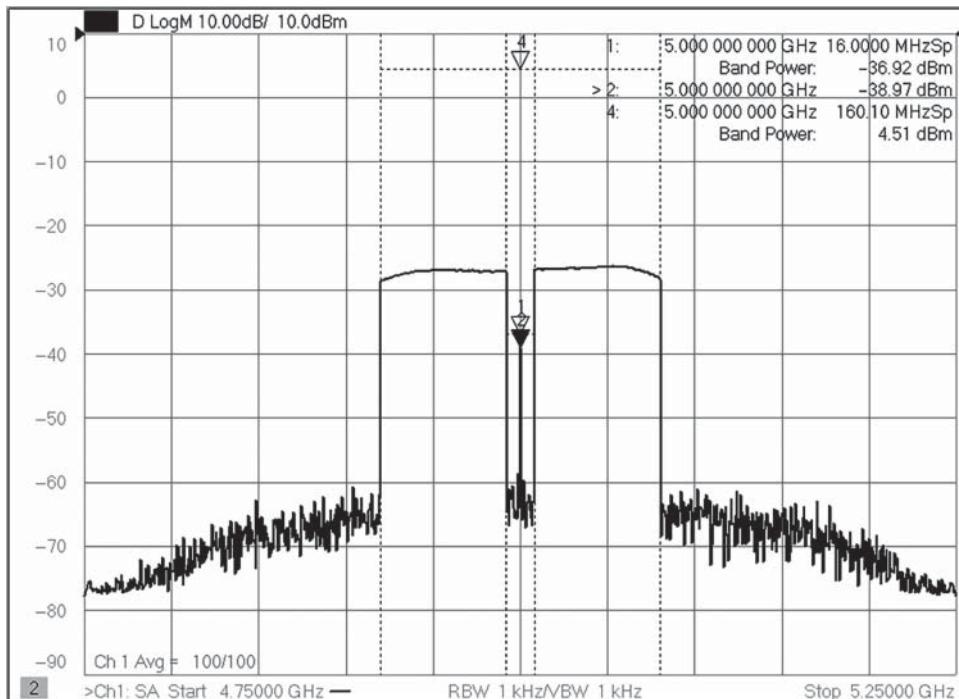


Figure 8.49 NPR signal with no corrections.

Finally, from the shape of the out-of-band distortion, we can estimate that this source is creating at least 20 dB greater NPR than the noise floor due to distortion in the source.

The NPR of this signal (measured using the band power method) from Eq. (8.13) is computed as -28 dBc . This value of NPR is limited primarily from the carrier feed-through, as the band power of the notch is nearly the same as the LO feed-through (measured by Marker 2 in the figure).

These attributes may be corrected by applying advanced correction methods, and offsetting the notch to remove the LO feed-through effect, as demonstrated in Figure 8.50. Here the roll-off of the signal is corrected to be flat within 0.1 dB across the bandwidth of the signal, and the total power is corrected to be exactly the desired power. Further, by offsetting the notch by one-half the notch width, the signal generator carrier feed-through is removed from the notch. Finally, the notch distortion correction is applied (using the instrumentation manufacturers' software tool) to lower the distortion by at least 10 dBm. And the marker band-power density function is used to compute the NPR directly.

With these corrections, the NPR signal level measures more than 10 dB better than in the uncorrected case, here shown as -38 dBc , but the NPR signal here is limited by the noise floor of the VNA. Marker 1 shows the NPR in the notch. If the RF signal is turned completely off, the band power in the notch drops less than 2 dB, indicating that it is dominated by the VNA receiver noise floor (if it dropped by 3 dB, it would indicate that exactly half the power was from the VNA receiver).

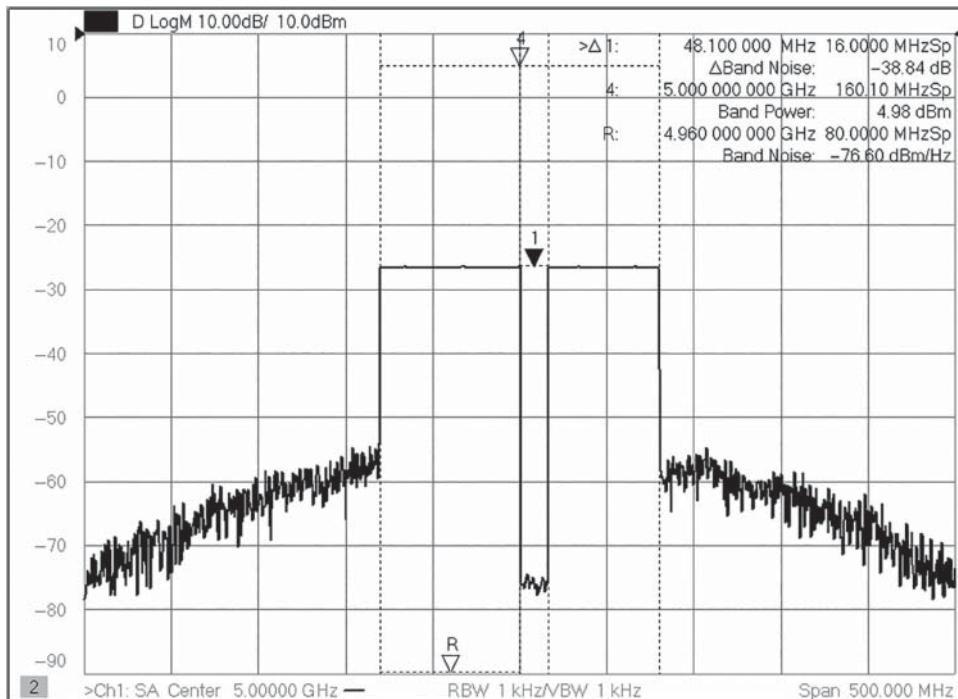


Figure 8.50 Corrected NPR signal.

8.2.3.1 NPR and Coherent Power Measurements

The final improvement to be made in the NPR measurement is to remove the effect the test receiver noise floor. The noise floor dominates because the signal must be lowered enough, using the receiver attenuators, to ensure that distortion in the test receiver does not contribute to the NPR value. But, using the techniques of Sections 8.1.4.2 and 8.1.4.4 the receiver noise effects can be removed from the NPR measurement.

Figure 8.51 shows the result of applying the multitone detection method described in Section 8.1.4.4. In this method, the noise power contained in FFT bins, which do not lie on the grid of multitone frequencies where the NPR signal intermodulation will appear, is set to zero. As such, only the intermodulation signals contribute to the band power, and we see the NPR level is reduced by 20 dB to -58 dBc.

This demonstrates the effectiveness of using the multitone power detection method. An alternative method of noise floor reduction, vector averaging (i.e. coherent time averaging), can be applied as well, with or without multitone power detection. Figure 8.52 shows the result of applying only vector averaging to the signal. To obtain the same level of noise reduction as the multitone power detection (by looking at the notch power), a vector averaging factor of about 100 is needed. This shows a similar NPR result, but vector averaging does take a great deal more time, compared to zeroing the noise between NPR tones.

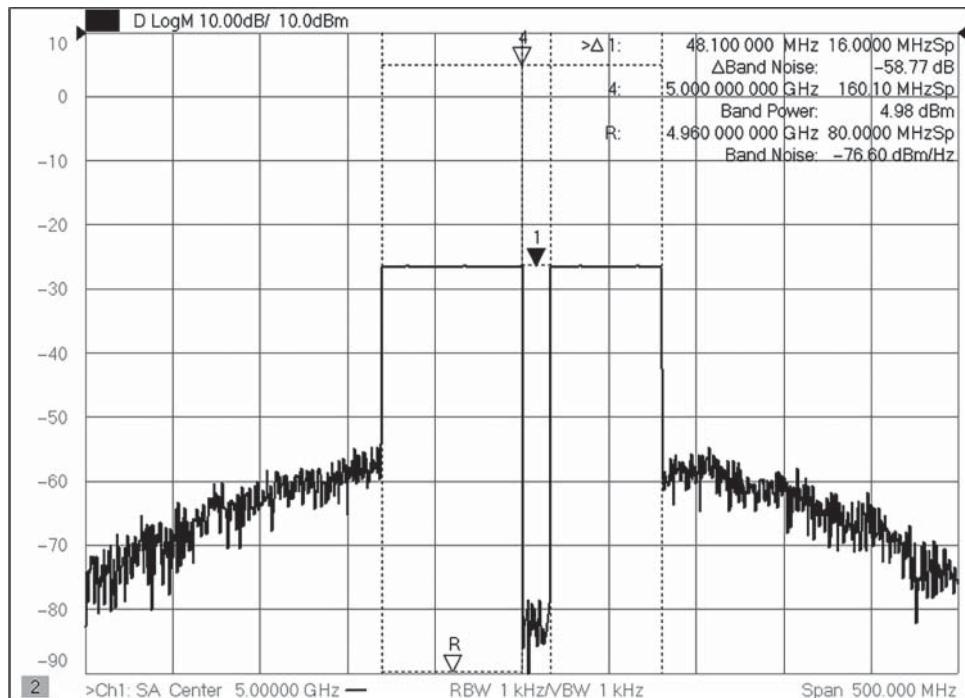


Figure 8.51 NPR measurement using mutitone detection methods.

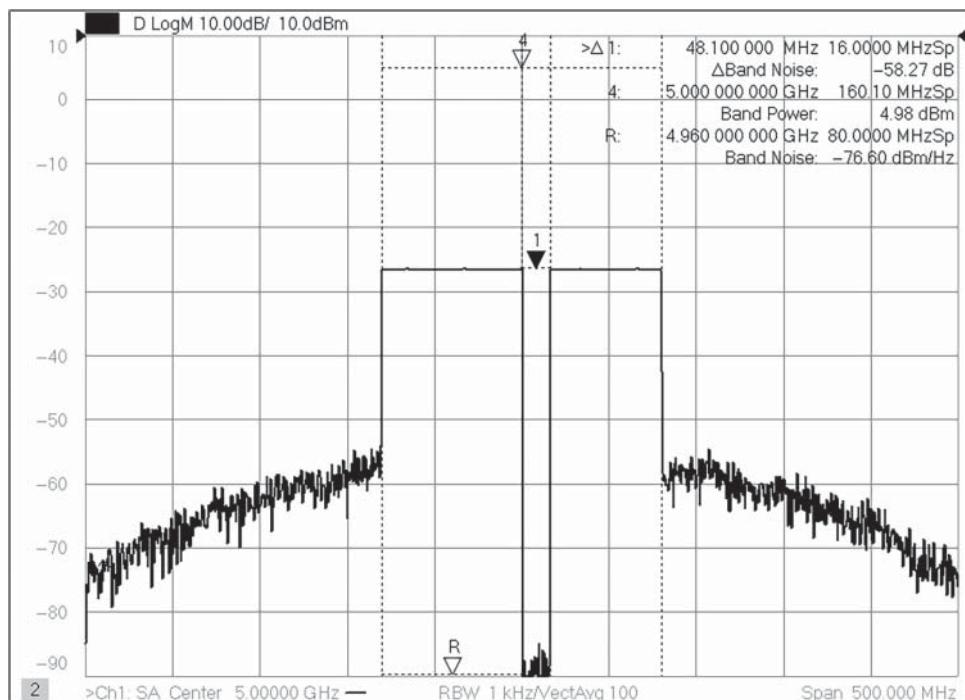


Figure 8.52 NPR with vector averaging.

8.2.4 EVM Derived from Distortion Measurements

Recent work has been developed to directly derive EVM measurements from a two-channel measurement of the spectrum of a modulated waveform, without the need to demodulate the signal. To understand how this is possible, consider how a modulated signal is created from a stream of data, as illustrated in Figure 8.53. Here the bit stream is broken into sets of 4 bits and mapped to an IQ pattern based on its hex evaluation. The mapping can be viewed as a trajectory across the IQ space. Each IQ position is called a *symbol*, and this particular mapping is known as 16 QAM, as there are 16 symbols used to encode the data stream, and they are encoded using both amplitude (distance from the center) and phase (angle from the origin), if considered as a complex number. The real part is typically referred as the *in-phase* or *I signal*, and the imaginary is called the *quadrature-phase* or *Q part*.

These signals are used to modulate a carrier wave to transmit the data. For wired communications, this is what is transmitted along the data bus. But for wireless communications, the IQ signals thus formed have very sharp edges, and this sharp transition cause them to have very wide frequency spans. To preserve bandwidth and limit adjacent-channel interference, the IQ signals are filtered to produce a band-limited response, as illustrated in Figure 8.54. Special filters, typically based on a root-raised-cosine response are used that have zero inter-symbol interference. This means the ringing of the filter crosses exactly zero at the time of the next symbol. In the figure, the symbol time is defined as the middle of the period. One can see that the value of the waveform is correct at exactly the center of the period but can vary considerably, even overshooting the original amplitude, between symbol points, as illustrated by the waveform appearance at the edges of each segment.

If this is mapped to the complex plane, it forms a pattern called a *constellation diagram*, as illustrated in Figure 8.55. Here, if the data is sampled at exactly the symbol time, the data will be seen to land exactly on one of the symbol points.

But consider what happens if this signal is passed through a non-linear device, such as an amplifier operating near its compression point. This is illustrated in Figure 8.56. The signal is amplified, of course, but also shifted in time (due to the delay of the amplifier), and the peaks of the signal may be clipped due to amplifier compression.

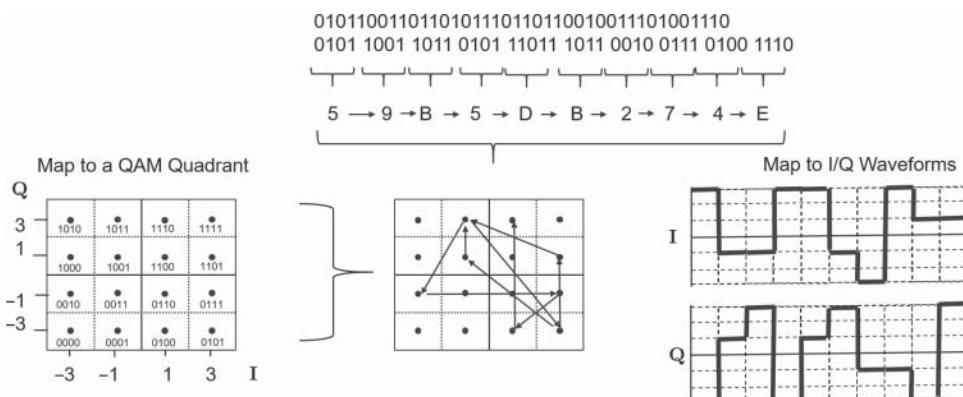


Figure 8.53 Mapping data to IQ space.

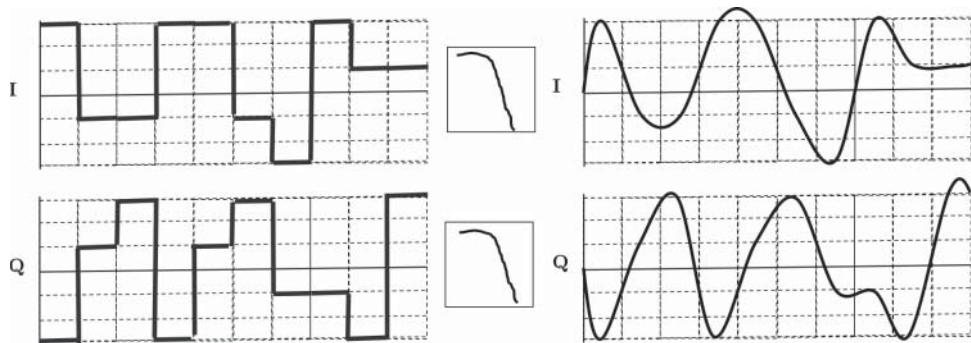


Figure 8.54 Filtering the IQ waveform.

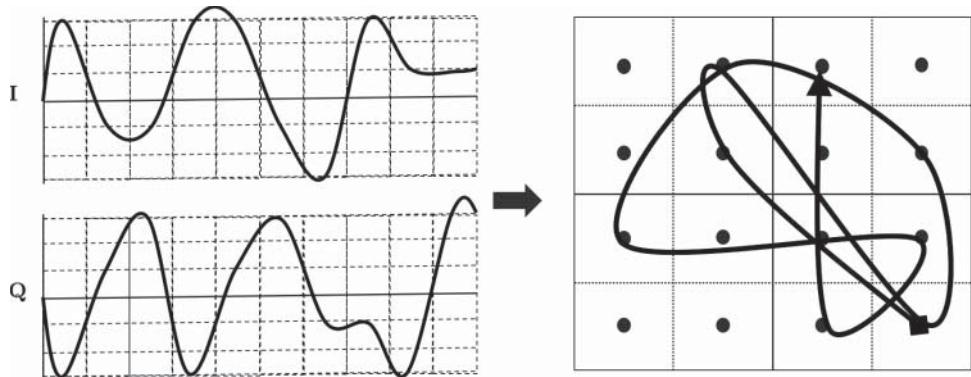


Figure 8.55 Mapping IQ to a constellation diagram.

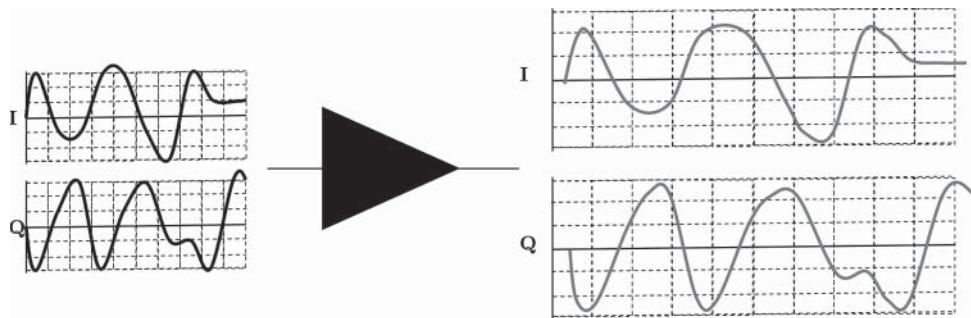


Figure 8.56 Gain, delay, and clipping on a signal.

Further, the frequency response (and delay) of the amplifier may not be flat, so it can affect the response of the modulated waveform (which is wideband) differently over different portions of the spectrum of the waveform. The frequency response ripple of an amplifier is a mostly linear function, and so a linear filter, properly chosen, can compensate for the non-ideal frequency response of the amplifier. This process is called *equalization* and is typically used in most communication systems. The signal from the amplifier is transmitted over the air and captured by a receiver. While it is possible to directly digitize an RF signal, it is most common to down-convert this signal, either at an IF or, more commonly now, using a synchronous converter or IQ demodulator that directly produces the IQ waveform. This waveform is processed by sampling at a rate similar to the symbol rate. It may not be sampled at exactly the symbol rate, because the receiver's clock is not synchronized to the transmitter's clock. The sampling of the signal is illustrated in Figure 8.57.

Further, since the distortion can also cause adjacent channel distortion, the signal is filtered (in the digital domain) as illustrated in Figure 8.58. It's important to point out that this is accomplished in the digital domain, so only the sample points have meaning. But the notional waveform is shown, and the peak of one portion is circled to illustrate that the filtering has the effect of removing sharp transitions associated with the clipping. This filtering ensures that processes that follow are not affected by out-of-channel signals, including the ACP distortion.

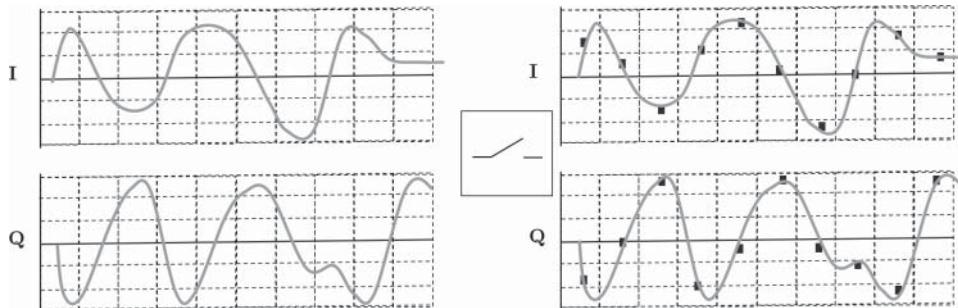


Figure 8.57 Sampling the signal at the symbol rate.

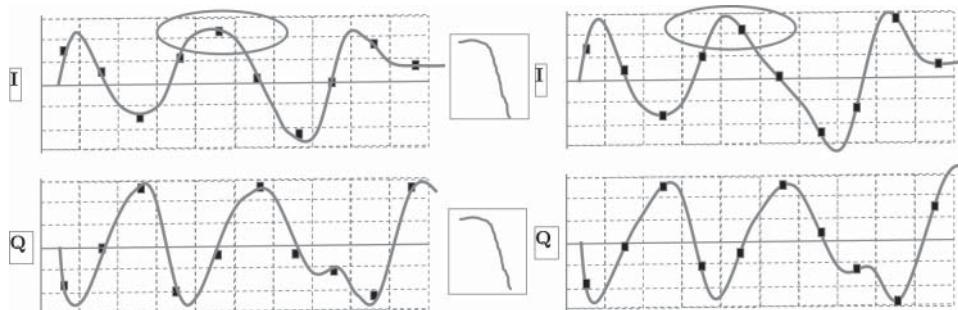


Figure 8.58 Filtering the received signal.

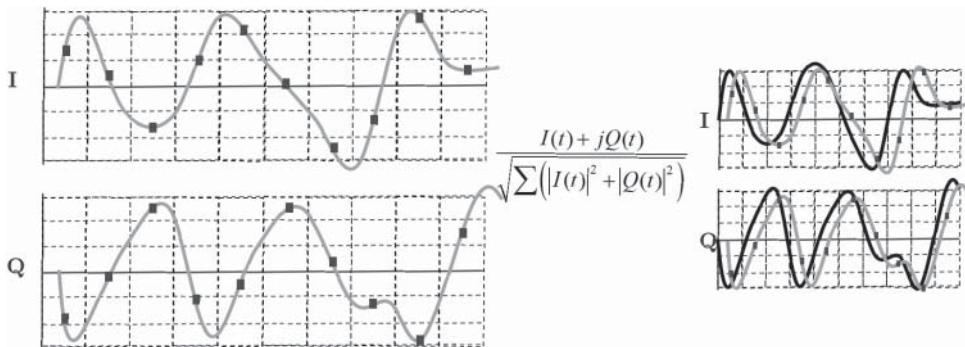


Figure 8.59 Renormalizing the received signal.

If one compares the original signal to the received signal, it is clear that the amplitude has changed, the response is delayed, and the shape is somewhat changed through distortion or clipping, even after filtering. The first step in recovering the data is to renormalize the signal to the unit circle, as illustrated in Figure 8.59. Here it is plotted with the original waveform from which we can see it is both time-shifted (so sample points don't land on the symbol points) and somewhat distorted.

The renormalization can be done in one of two ways: it can be renormalized to the peak of the signal (so the peaks line up), or it can be renormalized to the RMS value of the signal (which is more typical). Here RMS renormalization is illustrated. (Note: Some spectrum analyzers default to peak re-normalization, and others default to RMS; be sure to check the renormalization function when evaluating the EVM results.)

The next step in finding the EVM is to time-align the signals. Time alignment can be done in a number of ways, for example, through a correlation with the reference signal. In over-the-air testing, the reference signal is not necessarily known, so a multistep process is needed to acquire an estimate of the reference signal by sampling the waveform and extracting the data, then forming an idealized signal, and then re-aligning to the reference signal and re-sampling. The details can be quite complex and may vary from one receiver to another even for the same signalizing method. Other systematic defects such as frequency offset, IQ imbalance, and phase drift can also be determined and compensated for. The time alignment step is shown in Figure 8.60.

After time alignment, the sample points of the received signal are not at the proper time in reference to the symbol points of the reference signal. The final step before determining the EVM is to resample the received signal to obtain a sample point at the symbol time (the middle of each interval in the figure). This is a crucial step, and the reader is reminded that the data is known only at the sample points; the illustrated waveform is not available to the resampling process. But, the filtering at the transmitter and receiver, and the sampling in the receiver ensure that the waveform is sampled sufficiently to satisfy the Nyquist criteria. In practice, the receiver may sample faster than the symbol rate to make resampling and filtering more robust. Figure 8.61 illustrates the resampled points as circles on the waveform.

Also highlighted is an example of one delta-I and delta-Q point. The difference between the reference waveform and the measured waveform, expressed as a vector of the I and Q

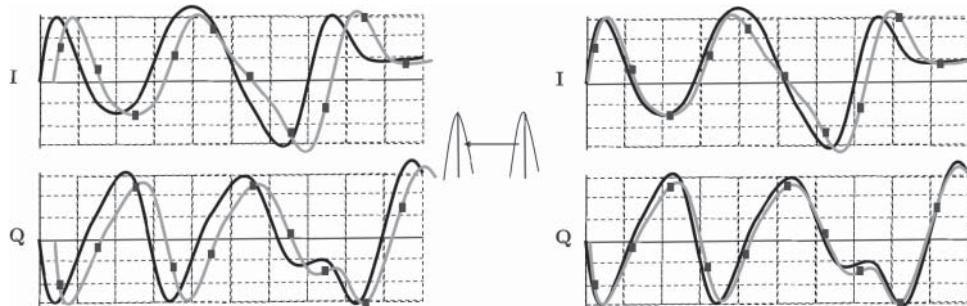


Figure 8.60 Time aligning the signal.

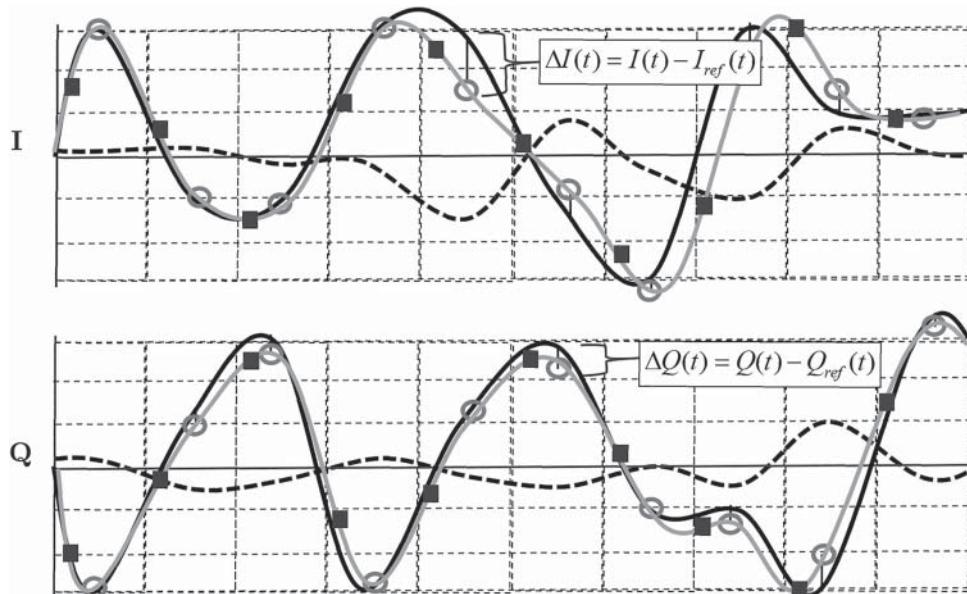


Figure 8.61 Resampling and the delta IQ waveforms.

difference, is called the *error vector*. A key figure of merit is the sum of all the error vectors, relative to the RMS value of the waveform, and this is the definition of EVM, and can be expressed as

$$EVM = \sqrt{\frac{\sum |\Delta I(t_N)|^2 + \sum |\Delta Q(t_N)|^2}{\sum |I(t_N)|^2 + \sum |Q(t_N)|^2}} \quad (8.15)$$

Also shown in the figure, in a dotted line, is the notional delta-I and delta-Q waveforms. We say notional here as we are in the digital domain, and so the waveform may be reconstructed for all time by taking the difference between the I and Q reference and measured samples and resampling that difference at any time to obtain this IQ waveform. It is important to note that this is possible due to the sampling meeting the Nyquist criteria.

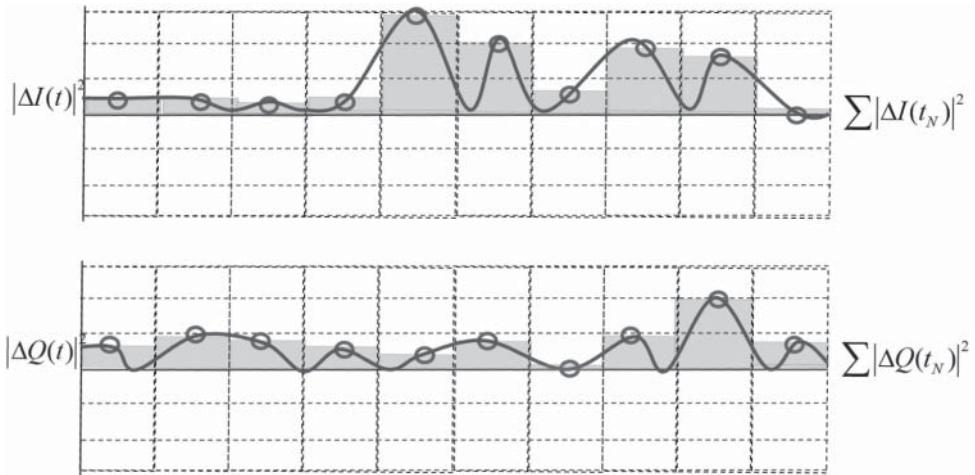


Figure 8.62 I and Q magnitude-squared waveforms.

This is the essential process (of course many real-world details have been simplified) for computing EVM through a demodulation process. However, one need not go through the demodulation process if these two conditions exist:

1. The signals are sampled sufficiently to meet the Nyquist criteria.
2. The test signal is repetitive and measured over at least one (or more) entire periods of the waveform.

The first condition tells us that magnitude of the error vector is essentially the area under the curve of $|\Delta I(t_N)|^2$ and $|\Delta Q(t_N)|^2$, as illustrated in Figure 8.62.

Nyquist sampling ensures the error under the curve is the same as the sum of the samples, so we can say, for k periods of the signal

$$\sum (|\Delta I(t_N)|^2 + |\Delta Q(t_N)|^2) = \int_{kT} \Delta I(t)^2 + \Delta Q(t)^2 dt, \quad (8.16)$$

Nyquist sampling also implies for a periodic signal that waveform has been band-limited in the frequency domain, which is the case for these signals due to the transmit and receive filters.

The second condition, that the test signal is periodic, and that the time-domain measurement is performed over exactly k periods of the waveform allow us to use Parseval's theorem to state

$$\int_{kT} \Delta I(t)^2 + \Delta Q(t)^2 dt = \int_{FreqSpan} |\Delta I(\omega)|^2 + |\Delta Q(\omega)|^2 d\omega \quad (8.17)$$

where the frequency span of integration is over the band-limited signal. Parseval's theorem holds in general only if one integrates over all time and all frequency, but our band-limited periodic signal needs to be measured only over an exact period and an exact span. Finally, note that for a periodic signal, the frequency spectrum is discrete (essentially a set of tones

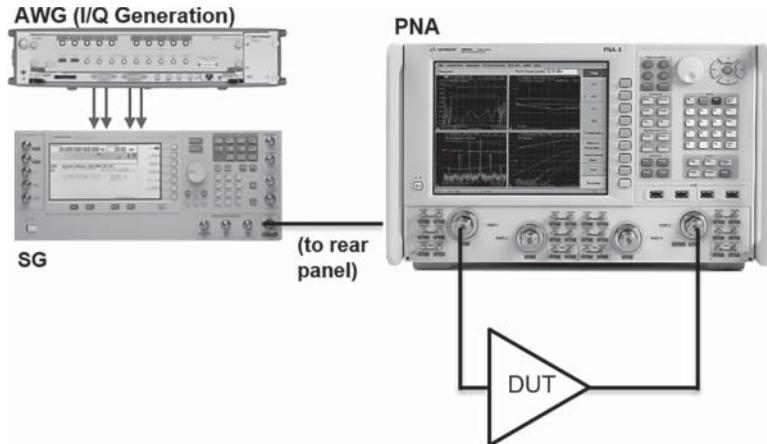


Figure 8.63 Measuring EVM with a VNA.

separated by the reciprocal of k times the waveform period), and so to compute EVM we only need to sum the magnitude of the vector difference of the spectrum of the measured signal relative to the reference signal, to get EVM as

$$EVM = \sqrt{\frac{\sum_{FreqSpan} |\Delta I(\omega_n)|^2 + |\Delta Q(\omega_n)|^2}{\sum_{FreqSpan} |I(\omega_n)|^2 + |Q(\omega_n)|^2}} \quad (8.18)$$

where we recognize here that these are after normalization. What this means in practice is that to measure the EVM contribution of a device under test, such as an amplifier, one can simply measure the output spectrum relative to the input spectrum (after accounting for gain and delay of the amplifier) and compute the EVM without any demodulation. Modern VNAs with multiple channels can capture the vector spectrum of output versus input easily, and the processing, based on spectral correlation, is used to determine the EVM. This has been detailed in a recent paper (Verspecht et al. 2019). A typical system is shown in Figure 8.63. A modulated signal is created (here shown with a wideband AWG and a signal generator as an up-converter) and inserted before the reference channel of a VNA. The port 1 reflectometer allows the measurement of the input signal (on the $a1$ or Reference channel), and port 2 allows measurement of the output signal (on the $b2$ or Test channel).

Unique to this system is the delta between input to output signal can be computed on a tone-by-tone basis, so even a narrowband VNA receiver can compute the EVM of a device driven with a wideband signal, measured over multiple acquisitions. This is called *coherent acquisition*.

Measuring a DUT EVM in this way provides several advantages, listed here:

1. The errors associated with the input and output response of cables and connectors can be removed using normal VNA calibration techniques.
2. The RBW can be made very small, and vector averaging can be added to lower the noise floor of the system. The consequence of this is measuring over more periods of the waveform.
3. The comparison between input and output is exact: in normal demodulation, for high QAM rates, the reconstruction of the reference signal can be faulty, if the peak EVM exceeds the distance between symbols. This results in an artificially low EVM number. As such high EVM results cannot be properly obtained on high QAM (e.g. 64, 256, 1024) signals using normal methods.
4. Errors in the source signal, which result in non-zero EVM at the input, do not affect the result of the EVM measurement of the amplifier, as the output is not compared to an idealized reference signal but to the actual reference signal applied to the DUT input.

A plot of the EVM of a power amplifier versus output drive power is shown in Figure 8.64 for a traditional demodulation method and for this new frequency-domain method, for two

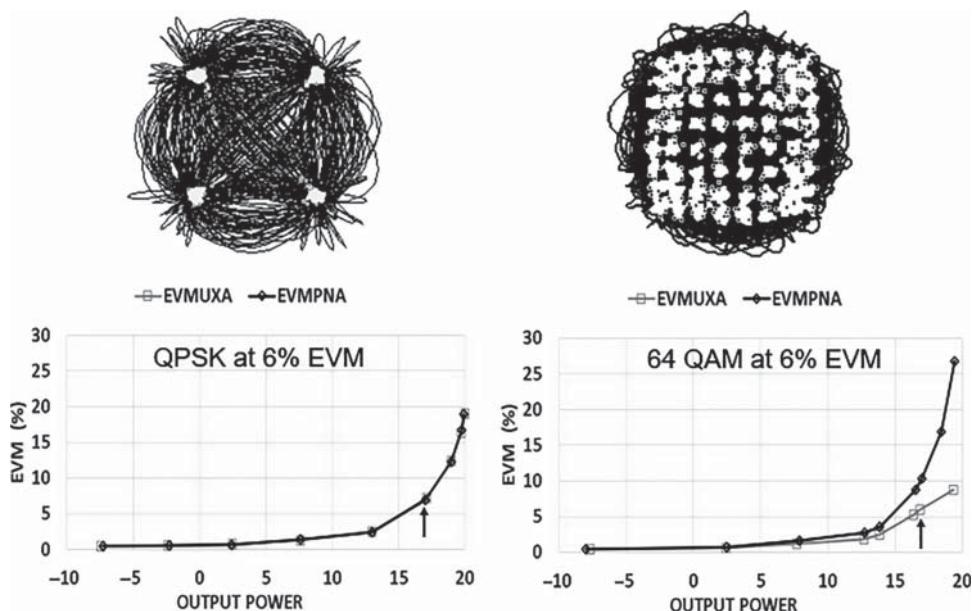


Figure 8.64 QPSK and QAM signals; upper plot is constellation diagrams, and lower is EVM versus power.

modulation types: QPSK and 64 QAM. The upper portion shows the constellation diagrams as measured on a Keysight UXA spectrum analyzer. The lower portion shows EVM versus output power for the cases of measurement using the UXA and traditional demodulation (marked EVMUXA) and using the new distortion computation based on frequency-domain methods using a Keysight PNA-X (marked EVMPNA). In the low-power region, the demodulation method shows almost identical results. The frequency domain method can apply coherent averaging to substantially reduce the noise floor. In the high-power region, using QPSK modulation, the results are also almost identical between VNA method and traditional demodulation. But for 64 QAM, the results are quite different. Traditional demodulation derives the reference signal from the demodulated symbols, and if the peak EVM exceeds one-half the symbol spacing, the reported symbol is not correct (but closer to the wrong symbol point). This under reports EVM, as “symbol skipping” causes the wrong reference waveform to be used in comparison. The VNA method compares input to output waves such that even with high distortion at the output, the VNA method does not cause any errors in the reference waveform; further, if the source itself has some errors in the waveform, producing an input EVM that is not perfect, it will be corrected for using the VNA method. The constellation diagrams are shown for 16.8 dBm output power (marked by an arrow on the graphs), and one can see in the upper constellation diagrams that while the spread of the “dots” representing symbols, around any particular symbol point, is about the same in each case, in the 64 QAM case the spread is large enough to cross a symbol boundary and will give under-reported EVM values. In fact, at even higher power, the constellation is unrecognizable (the demodulation fails to show a 64 QAM picture), but still the max EVM reported is only 8%.

An example of the display from the PNA-X Modulation Distortion application, for the QPSK signal of Figure 8.64, is shown in Figure 8.65.

In addition to equalized EVM, the un-equalized EVM is also shown, along with the input and output powers, the signal bandwidth, and the ACPR for the input signal and the output response, for the lower adjacent channel and the upper adjacent channel.

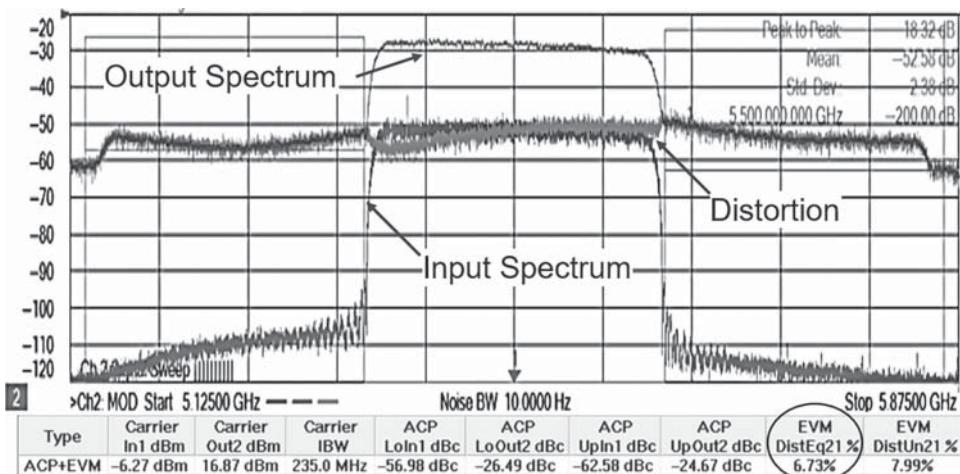


Figure 8.65 Display from PNA-X Modulation Distortion application, with equalized EVM highlighted.

8.3 Measurements of Spurious Signals with VNA Spectrum Analyzer

One of the most common uses of a spectrum analyzer is looking for spurious (unwanted) signals from a DUT. Spurious signals come in two forms: known and unknown spurious. Known spurious refers to signals that are at a known frequency and maybe have some specification on the DUT that they are below some target level. They are most common in mixer measurements and are found at the high-order mixing product frequencies. Unknown spurious are found in many devices, but amplifiers have some notable attributes that make them susceptible to particular unknown spurious signals. The first attribute is sometimes called the *amplifier stability factor* (see Section 6.1.5) and is a measure of the susceptibility of an amplifier to have spurious oscillation. A second attribute of amplifiers and other active devices is that the DC circuitry driving the DUT might have low-level signals on it, such as a signal associated with the switching frequency of the DC supply, and these low-level signals are modulated onto the carrier frequencies; this causes what is sometimes called *close-in* spurious and are apparent only when a signal is present.

8.3.1 Spurious at Predictable Frequencies

8.3.1.1 Multistage Converter Example

With mixer measurements, particularly multistage converter measurements, the design of the converter is dictated by the need to suppress higher-order mixing products. By using multiple stages of conversion, the lower-order products (such as two times the LO minus the RF, or three times the LO minus two times the RF) are pushed out of band where they may be filtered, but higher-order products may still remain in-band. It is possible to compute all higher order products, which for a single conversion mixer is

$$F_{\text{Products}} = m \cdot F_{RF} \pm n \cdot F_{LO} \quad (8.19)$$

For a double conversion mixer, the order of products can expand to

$$F_{\text{Products}} = m \cdot F_{RF} \pm n \cdot F_{LO1} \pm k \cdot F_{LO2} \quad (8.20)$$

Thus, for every first converter spur there are $2 \cdot k$ spurs from the second conversion. This very large number of spurs on multiple conversion mixers makes the direct measurement of spurs as described in Section 7.9.3 impractical. It is easier to simply sweep the output spectrum and look for any higher-order products that exceed a particular threshold. Figure 8.66 shows the output spectrum result from a two-stage frequency converter. The first stage up-converts the RF signal from 650 MHz to 3.65 GHz, using a 3 GHz LO; the image is of course 650 MHz lower than the LO at 2.35 GHz. The desired signal is filtered through a 3.65 GHz bandpass filter, before going to a second-stage filter. The second conversion up-converts the signal from 3.65 to 17.65 GHz using 14 GHz LO. This allows a wider filter at the higher output frequency to remove the image. If one tried to direct convert, the LO would be only 650 MHz away at 17 GHz and would be proportionally harder to reject. But it is difficult to predict which high-order products will cause signals in the IF, so a sweep is performed looking for any spurious above the desired level. The FFT-based VNA SA is uniquely qualified to perform this high dynamic-range spur search as the sweep speed

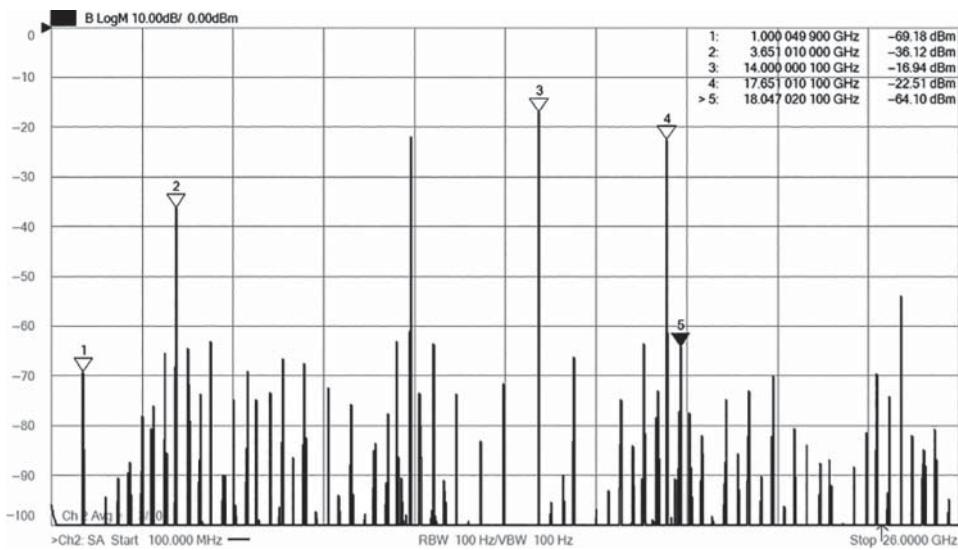


Figure 8.66 High-order products from a two-stage frequency converter.

versus RBW setting allows much faster sweeps at low RBW settings than the traditional swept-frequency spectrum analyzer. The sweep time for Figure 8.66 with a 26 GHz span at 300 Hz resolution bandwidth is about 82 seconds; if a traditional swept-frequency spectrum analyzer had similar settings, to get the same dynamic range and noise floor, the sweep time would be 30 000 seconds! Even in the stepped FFT mode, the sweep time would be 4000 seconds, limited by the FFT width and the tuning time of the YIG tuned preselector.

8.3.1.2 Determining Spur Order

With the high dynamic range available in the VNA-based SA, we can see the larger order of spurious signals coming out of the final stage of the two-stage mixer. Considering the first stage was filtered using a 3.65 GHz bandpass, it might be surprising to see such a high number of spurs at the output, but this is typical when the first stage filter utilized quarter-wave transmission-line construction. Such filters often have secondary passbands at the odd-order multiples ($3\times$, $5\times$, ...) of the passband. Identifying each of these causes (mixing products) of these spurs can be complicated, but a simple trick can help to determine the spur order as illustrated in Figure 8.66.

To accomplish the determination, one first sets each of the RF and LO signals slightly off frequency. In this case, the input, nominally at 650 MHz, is set to 651, 1 MHz above nominal. The first LO, nominally at 3.65 GHz, is set to 3.650 01 GHz, or 10 kHz above the nominal frequency. The second LO is set to 14.000 000 1 GHz or 100 Hz above the nominal frequency. With these settings, peak search is used with markers to establish the exact frequency of the signals in the spectrum.

In the previous figure, Marker 2 shows the first LO feedthrough (which was not completely filtered out by the first IF filter), and Marker 3 shows the second LO feedthrough. Marker 4

shows the desired output signal at 17 051 010 100 Hz. Note that we can see the offset of each of the RF, first-LO and second-LO in this figure: 1 in the MHz place shows it is one times the RF, 1 in the 10 kHz place shows it is one times the first LO, and 1 in the 100's Hz place shows it is one times the second LO.

Now consider Marker 5, which has an output signal at 18 047 020 100 Hz. If we reset the values of the RF, and LOs, this would be at 18.050 GHz. To determine the order of spur, we can look at the MHz position and see that it is -3 MHz lower than the nominal value, indicating it is a negative third-order product of the IF input. The 20 in the kHz position indicates that this is a second-order product of the first LO. The 100 in the units position indicates it is first order of the second order. So, the computation of this spur is

$$\begin{aligned} Spur_{18.050\text{GHz}} &= 2 \cdot LO_1 - 3 \cdot RF + LO_2 \\ &= (2 \times 3\text{GHz} - 3 \times 0.65\text{GHz}) + 14\text{GHz} \\ &= (6\text{GHz} - 1.95\text{GHz}) + 14\text{GHz} = 4.05\text{GHz} + 14\text{GHz} = 18.05\text{GHz} \end{aligned} \quad (8.21)$$

This indicates that the 2LO-3RF spur from the first mixer likely passed through the 3.65 GHz band-pass filter (at 4.05 GHz offset) and converted with the second LO. Thus, this is not a good frequency plan to avoid such a spur, or a better band pass filter is needed. Let's consider another spur, the one indicated by Marker 1 at 1 000 049 900 Hz. It has a nominal value of 1 GHz, but we see it is off by 49.9 kHz with the offsets given. To understand the order, note that there is nothing in the MHz column, so it is not a product of the input IF. Round to the nearest 10 kHz and it becomes 50 kHz higher than nominal indicating it is a fifth order of the first LO. But it is actually 100 Hz lower, indicating a minus-one order of the second LO, from the units column. So, we can conclude (as might be obvious) that this spur can be computed as

$$\begin{aligned} Spur_{1\text{GHz}} &= 5 \cdot LO_1 + 0 \cdot RF - LO_2 \\ &= (5 \times 3\text{GHz} - 0 \times 0.65\text{GHz}) + 14\text{GHz} \\ &= (15\text{GHz}) - 14\text{GHz} = 1\text{GHz} \end{aligned} \quad (8.22)$$

8.3.2 Multiport Mixer Spurious Measurements

Beyond just looking at the output of a mixer to discern the higher-order products, it is sometimes quite useful to see what products are emanating from the RF input or even from the LO port. Because a VNA-based spectrum analyzer has multiple receivers, along with directional-couplers on each port, one can detect the spectrum emanating from any port of the mixer, as well as viewing the spectrum of the RF and LO applied signals incident on the ports of the mixer. This might be useful to verify if any harmonics or spurious at the output are caused by harmonics or spurious on the source input signals. Often, ripple in a mixer's conversion loss can be traced to an image signal emitting from the RF or LO port, reflecting off a high refection at that port (often due to filtering applied at that port to removed unwanted signals from outside the mixer) and re-entering the mixer to reconver to the desired output frequency. If this is the case with a mixer, one solution is to use so-called non-reflective filters (which essentially have diplexer at the port to shunt the stopband

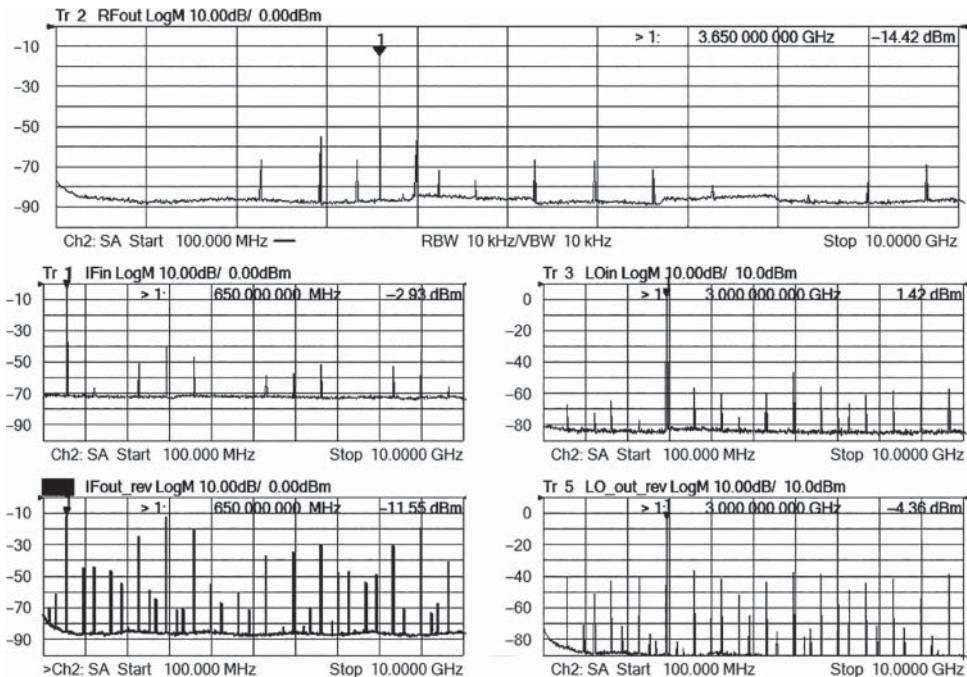


Figure 8.67 Measurement of the three ports (RF – input, LO, and IF – output) of a mixer.

frequencies to an internal load). One can see the spectral elements emanating from all three ports of a single-stage mixer (without filters) in Figure 8.67.

The RF output signal is shown in the Tr2 plot, labeled (appropriately enough) RFout. Also apparent are some other higher-order products. Tr1 in the smaller, left-middle plot shows the spectrum of the IF input signal, and it might appear to have spurious as well, but viewing the signal coming out of the IF input port of the mixer, the smaller lower-left plot labeled IFout_rev, shows a large number of spurious signals coming out of the IF port of a mixer. This might come as a surprise to an inexperienced engineer, but often the IF port can have a very dense spectrum coming out. The reflections of these signals off port 1 of the VNA are what cause the apparent spurs in the IFin trace.

Also note the LO drive signal can be displayed as well (middle-right, labeled LOin). The spectral content of this drive signals also appears to be not completely pure, but one can see the spurious product level is much smaller than the level coming out of the LO port of the mixer: thus we can conclude that this spurious content appearing on the LO drive signal is really a result of the mixer creating these products, emanating out the LO drive port, and then reflecting off the internal source in the VNA.

8.3.3 Spurious Oscillations

An amplifier can oscillate in or out of band, and a spectrum analyzer is used to look for these unwanted oscillations. For this measurement, it is most common to turn off the input signal

and look for any signals above the noise floor. Spurious oscillations may not be very stable and can move in frequency and in amplitude depending on the mechanism used to measure the spurious signal.

With a traditional swept-frequency spectrum analyzer, spurious signals must be stable and apparent when the analyzer sweeps past the signal. If a spurious signal is time varying, the analyzer must be slowed down so that over any particular RBW span, if the signal is unstable (for example, oscillating up and down in power), it goes through its variation cycle while the SA is on frequency. There is a similar limitation for a VNA using digital image rejection, in an FFT analysis mode. The signal must be stable during each acquisition of the FFTs used to perform the image rejection. If the signal is not stable, the image rejection code will erroneously treat it as an image (as the level changes with acquisitions of the different overlapped LOs used for the image detection). However, FFT detection schemes have modes that can improve detection of time-varying signals. One mode is to use averaging before the FFT image rejection; this is accomplished by lowering the VBW, which in FFT mode means there are more acquisitions averaged together before the image reject algorithm is applied. In such a case, the spurious signal's average power will be displayed, provided the time window over which the averaging occurs (essentially $1/\text{VBW}$) encompasses the time variation of the spurious signal. But often spurious must be specified to be below some threshold level even instantaneously. In such a case, the VBW averaging function can be changed from averaging the acquisitions to a peak-mode detection in the VBW averaging method (note that this is different than peak detection in the FFT detector type). With this mode, the peak of a signal during the VBW acquisition time is acquired, and as long as the acquisition time is longer than the signal-variation time, the peak acquired will be properly detected and retained by the image rejection code. However, this will also give a little higher reading than the actual spur as broadband noise power may be added to the spur and so the peak reading will include the time varying spurious signal plus the noise of the measurement.

Figure 8.68 shows the result of measuring an unstable spurious signal with a traditional spectrum analyzer (upper) and with an FFT-based VNA spectrum analyzer mode with various settings for the VBW. The signal was created using a signal generator with 400 Hz FM (100 kHz deviation) to emulate frequency instability and 10 dB of AM at a 1 kHz rate. This is a broadband sweep over 1 GHz span (spurious tests are often measured over wide spans), with the RBW set to have a noise floor of about -80 dBm with a nominal spur level of -70 dBm . The upper plot (labeled A) shows a trace with traditional SA sweep mode, with trace peak-detection turned on. The default sweeps do not show the full peak value of the spur, which is expected to be -70 dBm .

The lower part of the figure shows three examples of the FFT VNA-based measurements. Similarly, the default sweep (labeled B) does not show the full extent of the spur; this channel is set also set to default except for changing the RBW to achieve a -80 dB noise floor. The next trace (labeled C) has the VBW set to be 300 Hz, which has the effect of creating a video averaging factor of 5. The noise floor deviation is reduced (due to the averaging of the noise floor before the peak detector), and the signal variation from sweep to sweep is reduced, but the peak level is below the true peak level. The final trace has the video averaging method set to voltage max, instead of average power, and the peak level of the spurious is now seen. Note that this is different than the display peak detector, as the max-voltage detection occurs before the image-rejection function. As long as the time acquisition associated with VBW averaging is longer than the variation cycle of the spurious signal, the peak value will be captured. The

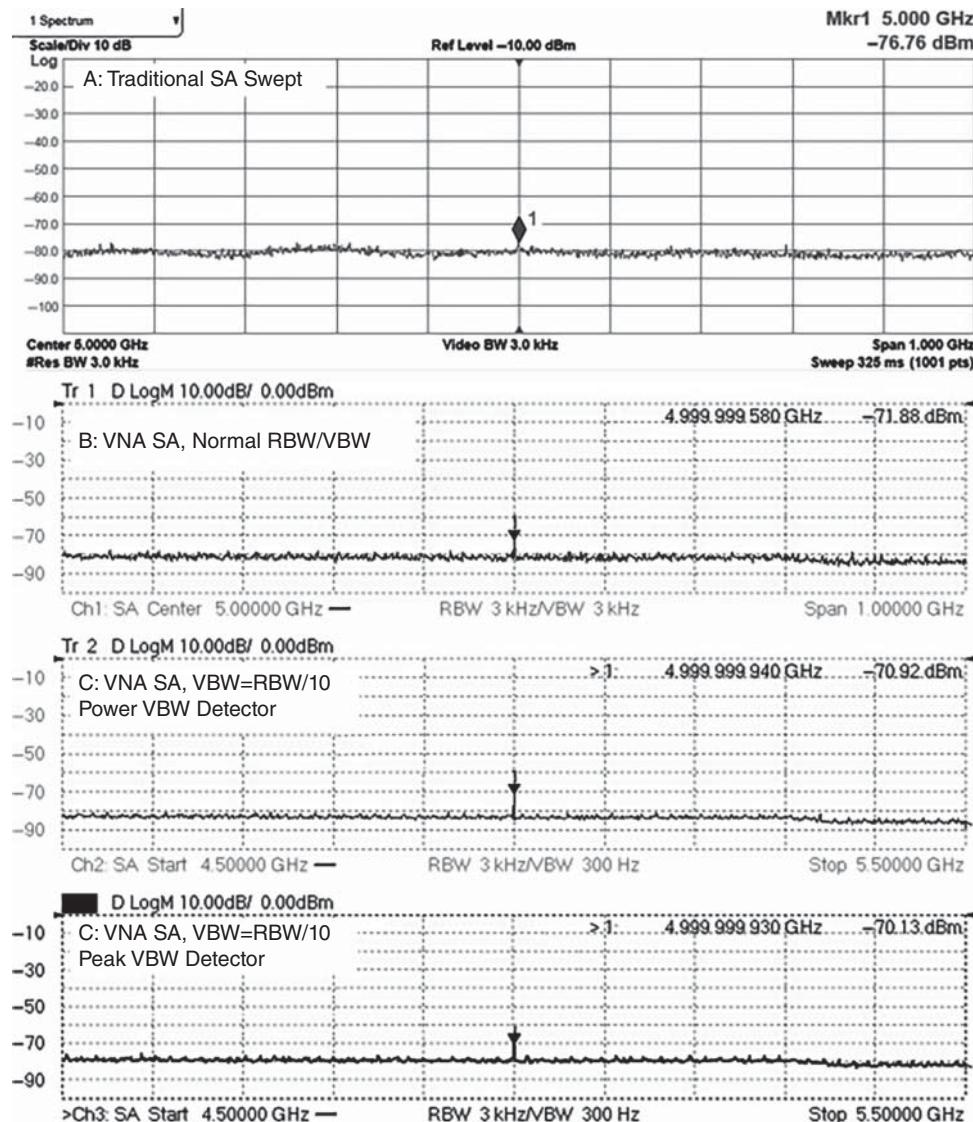


Figure 8.68 Measurements of an unstable spur using traditional swept SA (upper) and VNA-based FFT SA with various VBW settings.

noise value is also raised a bit, but because both the display detector type and the VBW averaging time are set to peak detection, the peak of several acquisition at each frequency is higher due to finding the largest noise signal over the several acquisitions. But this method ensures that the proper height of time-varying signals is captured.

8.4 Measurements of Pulsed Signals and Time-Gated Spectrum Analysis

In many applications the DUT cannot support a constant, CW signal. For example, in on-wafer testing, proper heat sinking of the device is not possible, and so the duty cycle of the RF signal (or similarly the DC bias) must be reduced to a fraction of the CW case, often below 1% duty cycle. For high-power amplifiers such as those used in radar systems, the system is designed to switch between Rx and Tx, and the desired operation of the amplifier must be measured in the same condition, which is typically 10%–50% duty cycle. In these cases, measuring the spectral content of the DUT, especially looking for spurious signals close to the carrier, becomes very difficult, because the pulsing of the RF source implies the spectrum from the DUT will be very broad due to the square-wave modulation of the source.

8.4.1 Understanding Pulsed Spectrum

A CW sinusoidal signal, which is pulse modulated, is essentially a sine wave multiplied by a pulse in the time domain, and has the spectral content of a sine wave, convolved with a sinc envelope where the pulse width is T.

$$f(t) = P_T(t) = \begin{cases} 0 & |t| > \frac{T}{2} \\ 1 & |t| < \frac{T}{2} \end{cases}$$

$$F_T(\omega) = \frac{\sin\left(\frac{\omega T}{2}\right)}{\omega} \quad (8.23)$$

But since the pulse is repetitive at a rate

$$T_R = \frac{T}{DutyCycle} \quad (8.24)$$

the spectral content will be a multitone signal with tone spacing of

$$F_{P_MT} = \frac{1}{T_R} \quad (8.25)$$

and an example of a spectrum measurement of a pulsed signal is shown in Figure 8.69. Here a 0 dBm, 3.3 GHz, CW signal is pulsed with a 1 μ s pulse width, and a 10% duty cycle, resulting in a 10 μ s period. This is equivalent to a 100 kHz multitone signal, due to the repetition rate. It is clear from Marker 1 in the figure that the peak value of the center tone is 20 dB below the CW power. Marker 2 shows the occupied bandwidth (99%) of the signal, and its power is nearly exactly the duty cycle less than the CW power. Similarly, the spectral spreading of the signal also reduces the main lobe power by a similar amount, so that we can say the power in the center tone is reduced by $20 \log(DutyCycle)$.

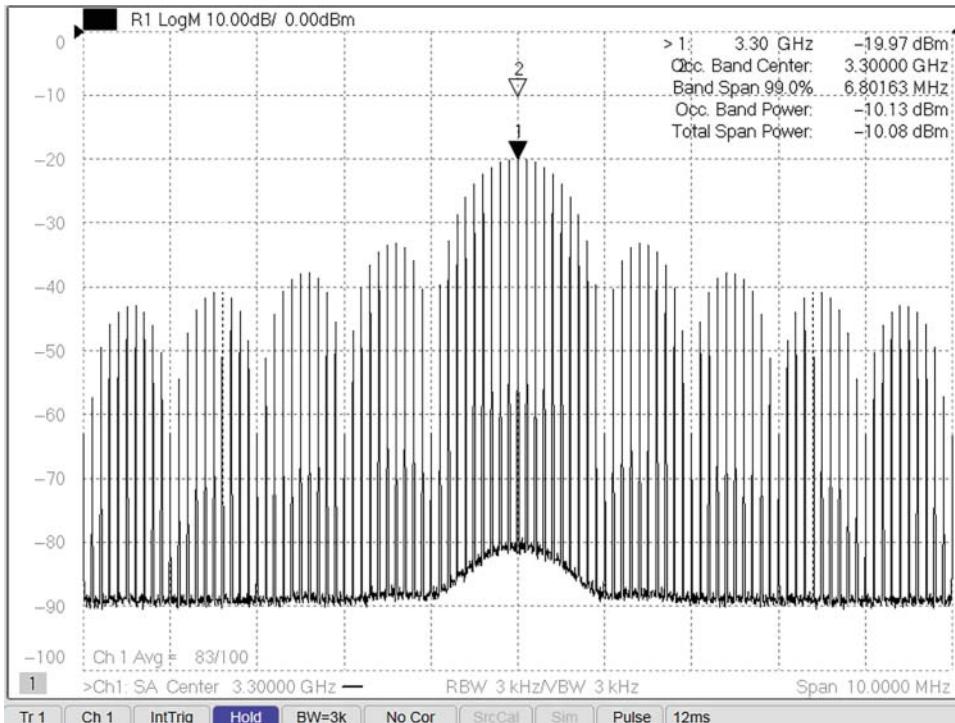


Figure 8.69 Spectrum of a signal pulsed with $1\ \mu\text{s}$ with 10% duty cycle.

But for the most part, measuring the true spectrum of a pulsed signal is unsatisfying if one is looking for spurious signals during the pulse-on time. This is because the spectral spreading covers such a wide range that any lower-level spurious signals are not visible. Consider the signal shown in Figure 8.70. Here the signal is a pulsed sine at 3.3 GHz, with a 10 dBm output power and 43% duty cycle and $1\ \mu\text{s}$ pulse. The spectrum is spread across 200 MHz, and it would be very hard to discern any spurious or other signals below $-40\ \text{dBm}$ (close-in) or $-50\ \text{dBm}$ (far out). But it is often desired to see the close-in spectrum for these pulsed situations, and it cannot be done with normal spectrum analysis. The solution is time-gated spectrum analysis.

8.4.2 Time-Gated Spectrum Analysis

Time gating of the spectrum analysis essentially means taking the data acquisition only when a pulse is on, and not taking data when the pulse is off. For a traditional spectrum analyzer, this is accomplished by adding a gate-trigger, which holds off the SA acquisition until the gate is high. The gate trigger can be generated externally (for example, from the DUT pulse gate) or may be internally generated from the IF envelope. In a VNA-based spectrum analyzer, the same principle applies where the data acquisition for the FFT is used only during the pulse-on time of the gate signal. Some VNAs include internal multichannel pulse generators and

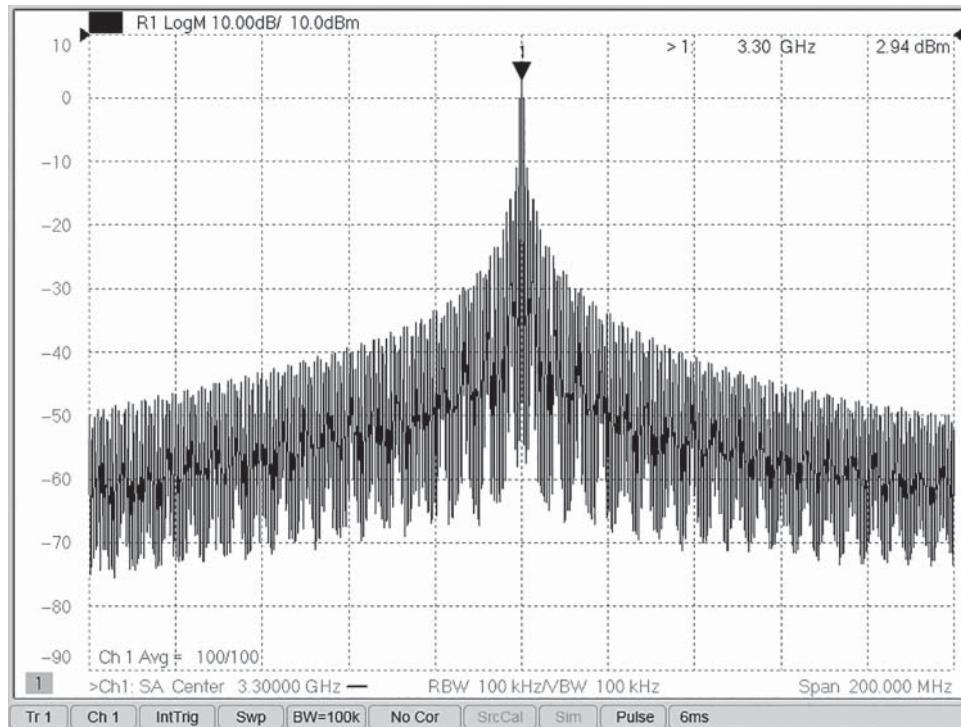


Figure 8.70 Wideband spectrum of a pulsed signal.

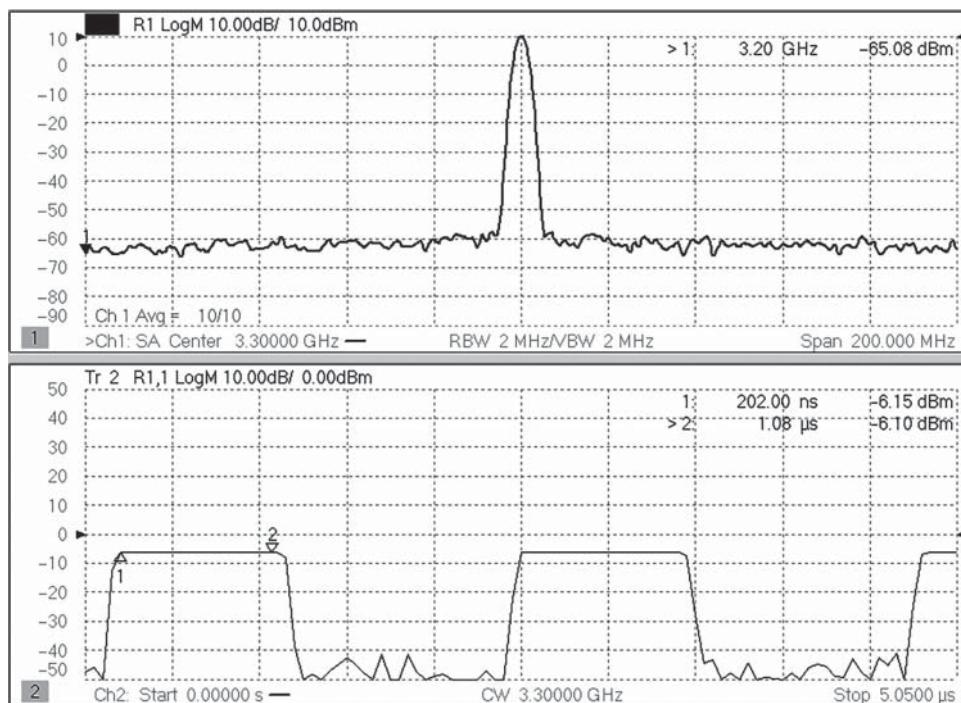


Figure 8.71 Time-gated spectrum analysis (upper) and pulse profile (lower).

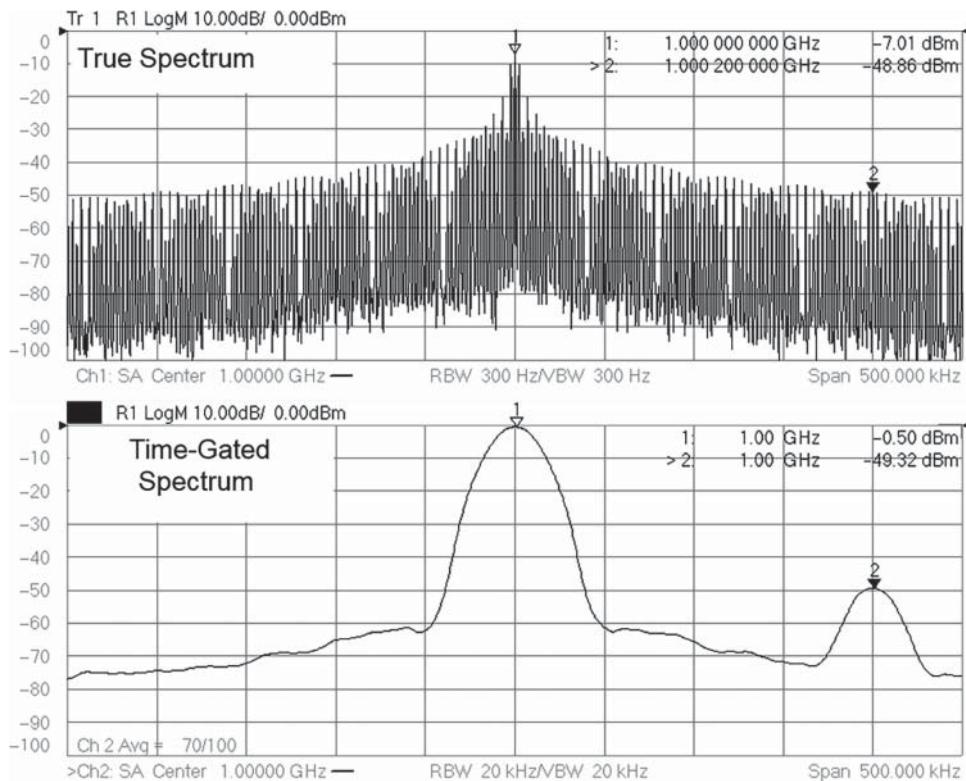


Figure 8.72 Measuring a close-in spur in a pulsed signal. The upper plot shows the true spectrum (narrow RBW), and the lower plot shows the time-gated spectrum (wider RBW).

modulators, so these may be internally reconfigured to supply the pulsed-gated signal. Others require an external gate signal to be used. A key aspect of the pulse-gated signal is that the resolution bandwidth must be set wide enough so that the FFT acquisition time is less than the pulse-on time. In VNAs where the pulse generator is internal, the settings for RBW and pulse on-time can be coupled to automatically set the data acquisition for measurement of the pulse. But the settings for an externally generated pulse signal must be set manually. Figure 8.71 shows a time-gated spectrum response of the signal from Figure 8.70, in the upper window and the time-domain pulse-profile response in the lower window. For setting the time gating we should set the gate delay to 200 ns and set the gate width to about 900 ns. Even though the pulse was defined to have a pulse width of 1 μ s, the rising and falling edges mean we should set the ADC acquisition time to something like 10% less than the pulse width to ensure we don't capture a rising or falling edge. Wide acquisition means we can have a narrower RBW and lower noise-floor, but if the acquisition is wider than the pulse, the sharp edges in the pulse rise and fall will spread the spectrum and make it harder to see close-in signal.

The advantage of time-gated spectrum analysis is clear when looking at a close-in spurious signal, as shown in Figure 8.72. The upper plot shows the true-spectrum, measured with a narrow RBW filter; the filter is narrow enough so that its acquisition time is longer than the

pulse-repetition time so that more than an entire pulse is captured. This gives the ability to resolve the underlying multitone spacing that is inherent in a repetitive signal. This example was created using a 1 GHz pulsed signal at 0 dBm, with 200 μ s on-time and a 425 μ s pulse period. A second signal at 1.000 2 GHz was coupled in at -50 dBm. Marker 2 in the upper trace shows the location of the second signal, but it is very difficult to discern it as a spurious signal from all the spectral content of the true spectrum of the pulse.

In comparison, the lower plot is a time-gated version of the same signal, where the time gate is set just shorter than the pulse on-time, and the RBW is set much wider, so that it can complete an entire acquisition during the duration of the pulsed signal. Here we can clearly distinguish the spurious signal from the main signal. Of course, the limit here is one cannot measure close-in spurs with a frequency offset of less than the reciprocal of the pulse width.

8.5 Summary

While in the past most components were tested with CW signals, or at most two-tone modulation, in modern systems the concern over the effect of broadband modulated signals on the performance, particularly non-linear performance, of amplifiers and systems has led to utilizing modulated signals as stimulus for testing these devices. However, most methods evaluate the signal in terms of the total modulation quality and not in terms of the effect of the device on the modulated signals. Further, the methods currently in use take small time samples of the modulated signals evaluating them for distortion (such as ACPR or NPR) and signal quality (such as EVM), as though they were making an over-the-air signal measurement.

In this chapter, the concept of coherent spectrum analysis was introduced as well as coherent power detection. In the area of component test, there is no reason to not take advantage of the knowledge of the input stimulus signal, particularly its waveform period in the case of a modulated signal, to enhance the measurement of the DUT. Utilizing a VNA-based spectrum analyzer, including the advantages of error correction, port extension, and automatic fixture removal, will likely become the gold standard for accurate, mathematically correct measurements of the response of amplifiers and systems to these modulated signals.

Reference

- J. Verspecht, A. Stav, J. Teyssier and S. Kusano. (2019) "Characterizing Amplifier Modulation Distortion Using a Vector Network Analyzer," *2019 93rd ARFTG Microwave Measurement Conference (ARFTG)*, Boston, MA, USA, pp. 1–4.

9

Measuring Noise Figure and Noise Power

When considering noise-figure measurements, most engineers naturally think of testing low noise amplifiers. Noise figure measurements can be especially challenging as a state-of-the-art low-noise amplifier (LNA) can have a noise figure substantially lower than the available equipment to test it with. These problems are compounded at microwave and mm-wave frequencies.

A further challenge exists in determining the noise figure of systems that integrate the LNA stage into a single-chip down-converter, where the output is not at the same frequency as the input and the LNA cannot be separated from the rest of the system. When the individual system elements such as LNAs, mixers, filters, and local oscillators are assembled into a complete converter assembly, the simple assumption is the first-stage LNA sets the noise figure of the system, but the other components often have a large, or even dominant, effect on the overall system noise performance. This chapter illuminates the details of noise figure and noise-power measurements, including their uncertainty, verification, and methods to improve them.

9.1 Noise-Figure Measurements for Amplifiers

Noise figure, though seldom measured on power amplifiers, is a key measurement on amplifiers used for receivers and other low-noise applications. Usually referred to as LNAs, these amplifiers typically have gain in the 10–20 dB range, although low-noise modules can have much higher gains as they are comprised of multiple stages of LNA.

Because of the design constraints to create low-noise amplifiers, the other parameters such as S_{11} or input match, stability, compression, and intermodulation distortion (IMD) are sometimes sacrificed for better low-noise performance; these other parameters must still be measured, and occasionally there is a tuning step to trade off lower noise for better performance in one of the other areas.

The subject of noise-figure measurements could support an entire text book or more, but here the key concepts used in modern measurement methods are introduced, along with the key formulas for measurement and correction.

9.1.1 Definition of Noise Figure

The definition of noise figure is at the same time simple, but with some subtleties. In Chapter 1, noise figure is defined as (Eq. 1.55)

$$N_{Figure} = 10 \log_{10} \left(\frac{\text{Signal}_{Input}/\text{Noise}_{Input}}{\text{Signal}_{Output}/\text{Noise}_{Output}} \right) \quad (9.1)$$

From this, the more useful definition can be derived as

$$N_{Figure} = 10 \log_{10} \left(\frac{DUTRNP}{G_A} \right) \quad (9.2)$$

DUTRNP is the available relative noise power from the device-under-test (DUT), where relative means relative to thermal noise floor, kT_0B , where k is Boltzmann's constant, T_0 is defined as 290 K, and B is the measurement bandwidth and is computed by dividing the available noise temperature from the DUT by T_0 , as

$$DUTRNP = \frac{T_{A_DUT}}{T_0} \quad (9.3)$$

Since available power does not depend upon the load impedance, and available gain does not depend upon load impedance, this definition of noise figure also does not depend upon load impedance.

Noise figure is always expressed in dB, but a related linear form of noise figure is called the *noise factor* and is often shown with the symbol F . This is a common source of confusion, and here the distinction will not be made between noise figure and noise factor; rather, the two terms are used interchangeably with the understanding that, unless explicitly indicated otherwise, the linear form noise factor is used in all computations and equations, even if it is called noise figure.

A related value, DUTRNPI is the incident relative noise power from the DUT. Here, *incident* means the noise power delivered into a non-radiating non-reflecting load, or an ideal cold Z_0 load; this is the power measured by an ideal Z_0 noise receiver. This is distinguished from the total power measured in the receiver (which includes the receiver's own noise), defined as the system-relative noise power incident (SYSRNPI), and which represents the total raw noise power measured at the receiver output, including the noise from the DUT input, amplified by the DUT gain, and added to the input noise of the receiver, provided the receiver has a Z_0 impedance, relative to kT_0B . Finally, for completeness, the total noise power measured at the receiver, available from the DUT, including the noise power generated in the receiver, is called the *available system-relative noise power* and is represented by the acronym SYSRNP. In newer instruments, where the receiver noise power can be calibrated out, the DUTRNPI values are usually displayed. Older instruments did not have a means to automatically remove their own noise contribution, so they displayed only SYSRNPI.

For most measurement purposes, the noise figure is *defined* from a 50Ω source and into a 50Ω load, so incident noise power makes the most intuitive sense to measure if the system is Z_0 . In most measurement cases, the source and load impedances are not 50Ω . Using the available powers in calculations removed the load impedance mismatch effect, to the first order, but did not remove the source mismatch effect, which remains a substantial source of error. In the case that the system is not Z_0 , the DUTRNPI can still be computed by applying the appropriate match correction factors, as described in the following discussion.

9.1.2 Noise-Power Measurements

Noise power measurements are usually performed on a noise figure analyzer (NFA), spectrum analyzer (SA), or more recently, on a vector network analyzer (VNA) with a noise measurement application. The NFA is really a specialized version of an SA, with more flexibility in the RF and intermediate frequency (IF) gain path to optimize the noise measurement to the receiver's most linear region, including a built-in LNA at the input. NFAs and SAs have an advantage in measuring noise in that the input is image-free, so the SA only measures the noise power at the selected frequency and is insensitive to noise at other frequencies. Newer SAs have very similar capabilities to NFAs and can provide very similar noise-figure measurements, provided they also have an LNA in front of the first converter. Calibration and measurement of noise power with either is essentially the same. The measured noise power in a receiver is a function of the receiver gain and bandwidth. For any given passive resistance, the noise voltage generated by that resistance is equal to

$$V_N = 2\sqrt{kTBR} \quad (9.4)$$

where k is Boltzmann's constant (1.38×10^{-23}), T is the temperature in Kelvin, B is the bandwidth over which the voltage is observed, and R is the value of the resistance. The available power from this resistor is the power that would be dissipated in a matched noiseless resistor (essentially an ideal noise receiver); since the load resistor is defined to be matched, half the noise voltage appears across each resistor, so power delivered to the load resistor is

$$P_N = \frac{(V_N/2)^2}{R} = kTB \quad (9.5)$$

Thus, the available noise power from any passive source depends only on the temperature of the source, and not on its impedance.

9.1.2.1 Noise Receiver Bandwidth

A VNA can also be used as a noise receiver, under some special conditions. In general, the effective noise bandwidth of a VNA is twice as wide as the IF BW, since in general the VNA does not have an image-protected mixer. That is, noise from one IF above and one IF below the VNA local oscillator (LO) frequency will be mixed in the final IF and measured. If the IF frequency is high (e.g. 10 MHz), then the noise will come from the displayed frequency (LO + IF) and a frequency 20 MHz away (LO - IF), and leads to uncertainty in the measurement due to not knowing which frequency is the major source of noise. Recently, some

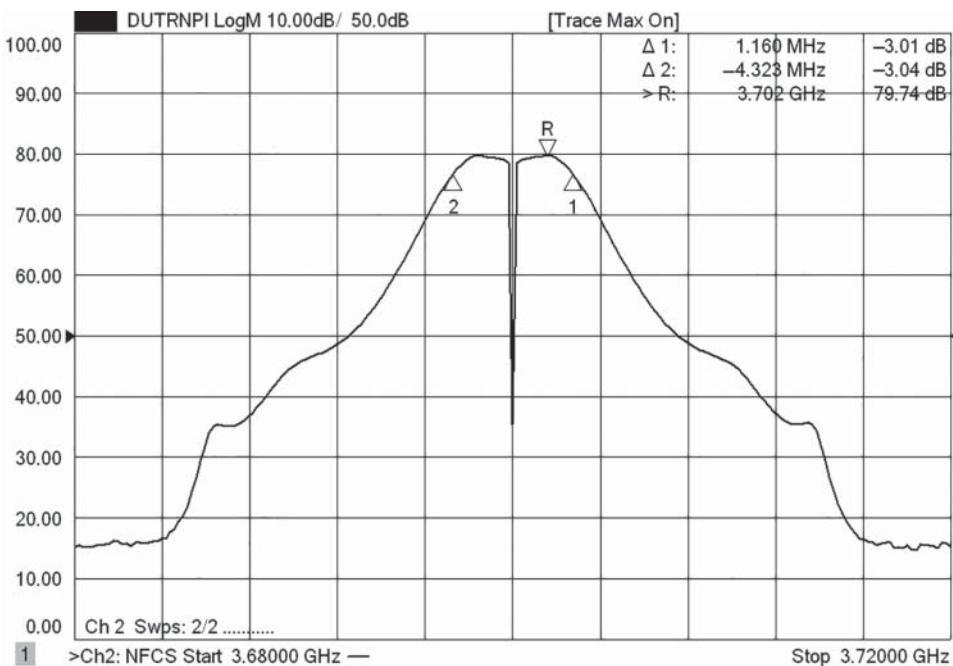


Figure 9.1 Noise bandwidth of a VNA with zero-IF receiver.

VNA have provided flexible IF structures, where the IF frequency can be arbitrarily set, even to 0 Hz. In such case, the effective bandwidth is still twice the IF bandwidth, but centered at the LO frequency (which is the same as the RF frequency for zero IF), thus having no uncertainty on which frequency the noise comes from. Often with zero-IF detection, there is an intentional notch in the IF response exactly at DC to avoid DC offset errors. One example of the IF response in a VNA used for measuring noise power is shown in Figure 9.1. To create this measurement, the VNA source is turned on, fixed at the center frequency. The noise receiver sweeps across the source power and traces out the effective noise bandwidth. The markers show the 3 dB down points from the reference, with an apparent bandwidth of about 5.5 MHz; however, with the notch in the center, the effective noise bandwidth is about 4 MHz.

There can be additional sources of error in VNA measurements due to noise conversion on higher-order LO products such as the third or fifth harmonic of the LO. In such a case, a band-limited filter must be placed in front of the noise source or DUT to ensure out-of-band noise is not converted into the IF. At least one manufacturer, Keysight, provides a special low-noise measuring receiver in a VNA as a hardware option, with a built-in switched-gain LNA and some built-in filtering to protect against images from LO harmonics.

For all the noise-measuring instruments, the bandwidth sets the resolution of the measurement, meaning the ability to distinguish the noise figure at different frequencies. The most common bandwidth used is 4 MHz (historically, due to the original HP 8970 NF meter), with narrower bandwidths available for narrow channel measurements, and wider bandwidths available for better speed and lower jitter.

9.1.2.2 Understanding Jitter

Since the NF receivers are measuring random noise, there will be a substantial variation in the measurement results from sample to sample. The variation in the noise goes down as the square root of the noise samples; the number of samples per unit time goes up as the bandwidth goes up. Wider bandwidths therefore give more samples in the same time and thus have lower jitter for the same measurement time, sometimes called the *integration time*, as the noise power is integrated over the samples.

The effect of the integration, or number of samples averaged, to generate a noise reading is illustrated in Figure 9.2. Here a system is set up with a noise receiver that is defined to be -164 dBm Hz^{-1} noise floor, measuring a noise signal with a noise power density or -154 dBm Hz^{-1} . Since the noise of both the signal and the receiver add, with these levels the expected noise power is $-153.6 \text{ dBm Hz}^{-1}$. In the upper plot, the integration factor is set to 1, and we see a tremendous variation in the noise reading. But in the lower plot the integration-factor is set to 100, and the variation, while still large, is now only about 1.5 dB. In commercial system, a minimum factor of 10 000 might be typical, resulting in a variation of 0.15 dB, when the noise power is 10 dB above the receiver's internal noise power.

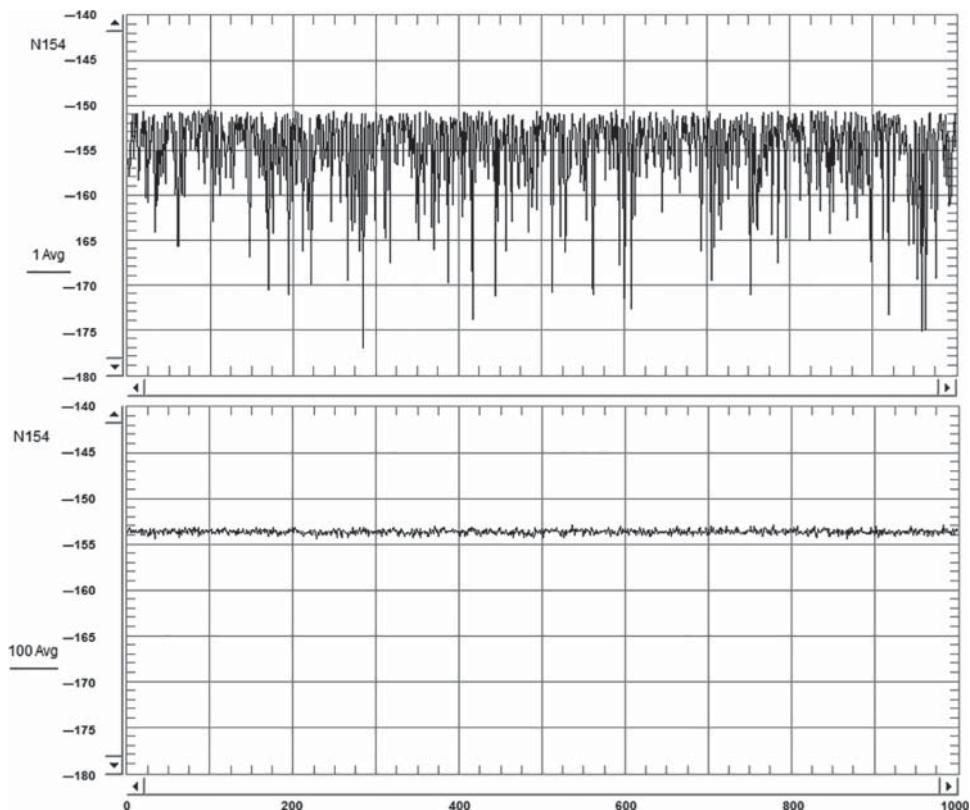


Figure 9.2 Noise variation over 1000 samples, with a 1 or 100 averaging.

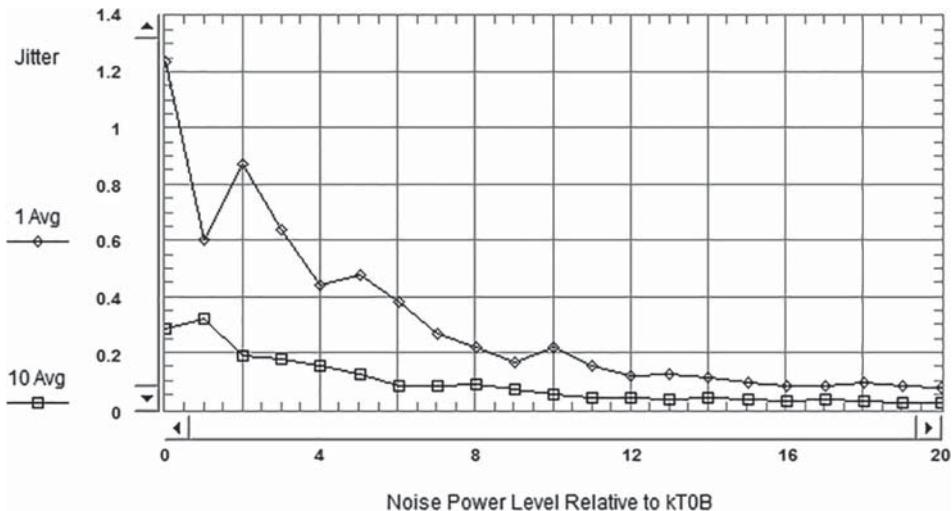


Figure 9.3 Measured error versus input noise power.

But, the noise variation changes dramatically as the noise power approaches or falls below the noise receiver's own noise power. That does not mean that one cannot measure noise lower than the internal noise power of the receiver, because the noise signal applied will always add to the internal noise power producing a small increase in the overall noise reading. If the internal noise power is known very accurately, it may be subtracted from the measured noise power to find the applied noise power. However, the measurement of the internal noise power is subject to the same jitter as the measurement of any external signal, and so the effect of averaging power on the estimate of the internal noise power will be similar.

Figure 9.3 shows the result from two different experiments, where the internal integration factor is set to 1000, and the noise measurement is averaged either 1 time or 10 times. Here an external signal is swept on the x-axis from -174 to -154 dBm Hz^{-1} , and the internal receiver noise power is -164 dBm . Note that there are two noise measurements, one to determine the noise power of the internal receiver (which will be subtracted from the total noise measurement of the external signal) and one for the external signal. Each is averaged the same number of times. It is clear from the figure that when the externally applied signal is much higher than the internal (at 20 dB above kT_0B , the input signal is -154 dBm Hz^{-1} and internal signal is -164 dBm Hz^{-1}), the error is small regardless of averaging factor. But when the external signal is smaller (at 0 dB relative noise power, the input signal is -174 dBm Hz^{-1}), the error is much larger with lower averaging factor.

For most cases, commercial instruments restrict the minimum integration factor to a large number and then allow additional noise averaging on top of that. For example, if the internal integration factor is 10 000 (typical for a commercial instrument) and the noise averaging is set to 100, then 1 million noise samples are actually acquired for the noise power reading. Jitter cannot be eliminated from noise-power measurements; after all, the signal that is being measured changes from one acquisition to another, because it is noisy. It's in the name. But with this understanding of how the estimate of noise power is affected by the noise power

level relative to the instrumentation noise level, and the averaging factor, we can now define some noise-figure measurements.

9.1.3 Computing Noise Figure from Noise Powers

9.1.3.1 Y-Factor Correction and Noise Receiver Calibration

A noise source is used in the calibration and measurement of a DUT from an SA or NFA by providing either a cold state (kT_0B noise) or a hot state, which produces a known amount of excess noise above the cold state. The total noise power detected depends upon the gain and bandwidth of the receiver. The noise source excess noise is defined as an *excess noise ratio* (ENR) and is related to the noise power (or noise temperature) above the kT_0B noise by

$$ENR_{dB} = 10\log_{10}(ENR) = 10\log_{10}\left(\frac{T_H - T_C}{T_0}\right) \quad (9.6)$$

This definition is slightly unusual in that if the hot noise equals the cold noise, the excess noise ratio is not 0 dB (as one might intuitively expect), but instead is the log of zero, or negative infinite dB. If the hot noise is two times the temperature of the cold noise, the ENR is 1 or 0 dB. So, excess noise ratio should not be interpreted as simply the noise power above kT_0B . The definition of (9.6) is useful, however, in simplifying the computation of the noise figure of a system measured with at hot and cold sources. Normally, ENR is defined for T_C equal to 290 K, but if the noise source is not at T_0 , Eq. (9.6) accounts for the difference. (This presumes that the hot noise temperature, relative to the cold noise, changes with a change in the ambient temperatures; hot/cold noise resistors follow this rule, but some solid-state noise sources do not; see the references for details.)

The noise figure of any system can be related to the pair of hot/cold measurements directly as

$$N_{F_Sys} = \frac{ENR - Y}{Y - 1}, \quad \text{where } Y = \frac{P_H}{P_C} \quad (9.7)$$

And P_H , P_C are the noise powers measured for the hot state and the cold state of the noise source, respectively. In the case where the noise source is not at 290 K temperature, a slight modification accounts for the difference as

$$N_{F_Sys} = \frac{\frac{ENR - Y}{Y - 1} \cdot \left(\frac{T_C}{T_0} - 1\right)}{Y - 1} \quad (9.8)$$

where T_C is the temperature in Kelvin of the noise source in the cold state, which is the ambient temperature, and T_0 is the reference noise temperature, defined as 290 K.

To calibrate an NFA, SA, or VNA, it is necessary to determine the contribution of the receiver noise to the overall system noise measured. The receiver noise figure is computed by applying the noise source directly to the input of the receiver and computing the system noise figure as shown previously. Because the powers are used as a ratio, they can be expressed as either watts/Hz (noise power density) or simply noise temperature. The power conversion for the hot power to noise temperature is

$$T_H = \frac{P_H}{kB} = \frac{P_{H_density}}{k} \quad (9.9)$$

where k is Boltzmann's constant, B is the receiver bandwidth, and noise power density is defined as noise power relative to a 1 Hz noise bandwidth. The cold temperature is simply the ambient temperature of the noise source. Most noise measurement systems give noise power in terms of noise power density in dBm Hz⁻¹, or noise temperature, and noise temperature is sometimes more convenient. Noise figure may also be expressed as a temperature (called *excess noise temperature*) and is defined as

$$T_{E_Rcvr} = T_0 \cdot (F_{Rcvr} - 1) \quad (9.10)$$

where F_{Rcvr} is the receiver noise factor. Once the noise figure of the system is known, the gain-bandwidth of the system can be computed by knowing the ENR of the noise source and the hot temperature measurement, as

$$GB_{Rcvr} = \frac{T_{H_Rcvr}}{(ENR \cdot T_0 + T_C) + T_{E_Rcvr}} = \frac{T_{H_Rcvr}}{(ENR + 1) \cdot T_0 + T_{E_Rcvr}} \Big|_{T_C=T_0} \quad (9.11)$$

With these factors, the noise receiver can be calibrated to read noise temperatures directly using the receiver correction

$$T_A = \frac{T_{M_Rcvr}}{GB_{Rcvr}} - T_{E_Rcvr} \quad (9.12)$$

where T_A is the actual available noise temperature, and T_{M_Rcvr} is the raw measured noise temperature on the receiver, and the measured noise power density can be computed as

$$P_{A_density} = \frac{kT_{M_Rcvr}}{GB_{Rcvr}} - kT_{E_Rcvr} \quad (9.13)$$

With this correction, the second-stage noise contribution of the receiver is removed.

9.1.4 Computing DUT Noise Figure from Y-Factor Measurements

After the calibration, the noise source is applied to the input of the DUT, and the output of the DUT is connected to the noise receiver. The hot and cold states of the DUT driven from the noise source are measured and recorded, and the noise figure of the overall system is computed as

$$N_{F_Sys}^{DUT} = \frac{ENR - Y \cdot \left(\frac{T_C}{T_0} - 1 \right)}{Y - 1}, \quad (9.14)$$

$$\text{where } Y = \frac{P_{H_DUT}}{P_{C_DUT}}$$

This represents the combined noise figure of the DUT and the noise receiver, not just the DUT by itself. If the noise power is corrected as in Eq. (9.13), then this is the corrected noise figure. However, some systems measure the overall noise power (SYSRNP versus DUTRNP) instead. If the DUT has high gain, the noise contribution of the receiver is not significant, and the noise figure is nearly the same as the DUT noise figure. For cascaded devices, the Friis equation can be used to compute the overall noise figure and states that for N cascaded, the total noise factor is

$$F_{Total} = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 \cdot G_2} + \cdots + \frac{F_N - 1}{G_1 \cdot G_2 \cdots G_{N-1}} \quad (9.15)$$

which can be used to compute the noise factor of just the DUT separate from the noise receiver as

$$N_{F_DUT} = N_{F_Sys}^{DUT} - \frac{(N_{F_Rcvr} - 1)}{G_{DUT}} \quad (9.16)$$

where gain is the power gain, such as $|S_{21}|^2$. This can also be expressed in terms of excess noise temperature as

$$T_{E_DUT} = T_{E_Sys}^{DUT} - \frac{T_{E_Rcvr}}{G_{DUT}} \quad (9.17)$$

To find the noise figure of the DUT, all that is needed is to find the gain of the DUT. The gain of the DUT can be computed from the two sets of noise-power measurements as

$$G_{DUT} = \frac{P_{H_DUT} - P_{C_DUT}}{P_{H_Rcvr} - P_{C_Rcvr}} = \frac{T_{H_DUT} - T_{C_DUT}}{T_{H_Rcvr} - T_{C_Rcvr}} \quad (9.18)$$

Much of this is automated when used with an NFA or SA with noise figure personality. Most VNAs do not rely on the Y-factor method, as they can measure gain independently. An example results of the Y-factor technique is shown in Figure 9.4.

The Y-factor result here is a direct computation from two noise-power measurements of hot and cold power using Eq. (9.14), after applying the correction of (9.13). A particularly interesting observation is that the noise figure has a ripple in the middle of the response and the peak and valley correspond at the frequencies exactly at the extremes of impedance as

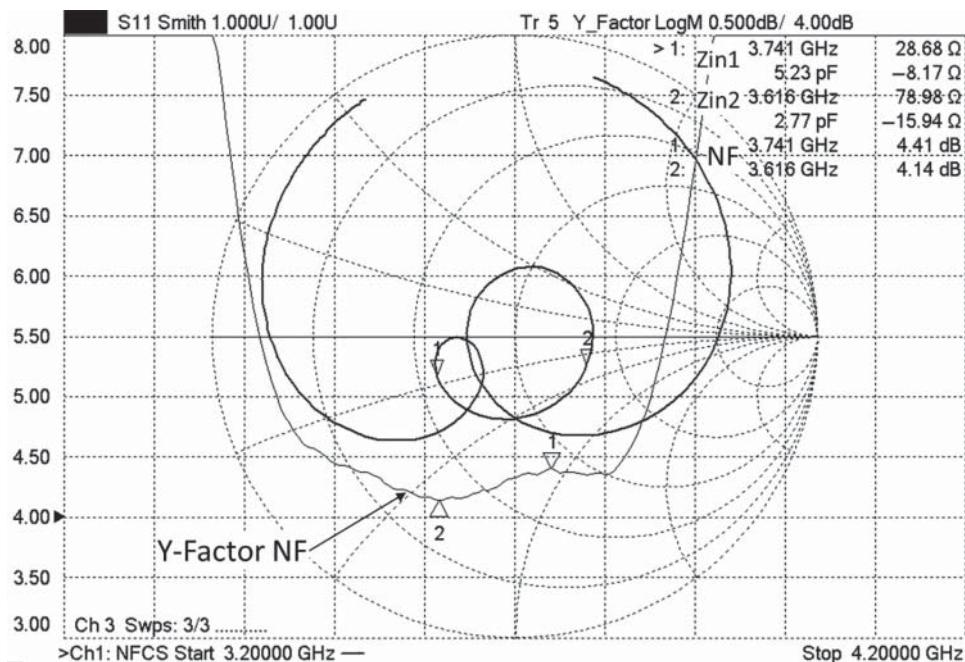


Figure 9.4 Y-factor computation based on hot and cold sources, and effects of S_{11} mismatch.

shown with Z_{in1} and Z_{in2} on the Smith chart trace. This indicates a likely error due to the noise source not being a perfect 50Ω source. Thus, the measurement displays a lower noise figure for the Z_{in2} than it does for the Z_{in1} , likely due to the noise parameters of this particular amplifier.

A few notes on the Y-factor technique:

First, the gain measurement is not the normal S_{21} gain, but the insertion gain of the DUT. This differs from the S_{21} gain due to the mismatch at the input and output of the DUT during the measurement.

The Y-factor measures the noise figure of the DUT in the impedance of the noise source, and (for the most part) it is the cold impedance that is important. The noise figure of a DUT depends upon the source impedance (see next section), so if the noise source is not exactly 50Ω , it is as if the noise figure has been changed or pulled by the noise source off of its 50Ω noise figure. The hot state of the noise source is designed to measure the gain, and it usually overwhelms the DUT noise figure, so pulling of the DUT during the hot state measurement is not so important.

For Y-factor measurements, it is usually recommended to use a low-ENR noise source. This has two benefits: the lower ENR sources are manufactured by adding high-quality attenuators to the high-ENR sources, thereby improving both hot and cold matches. Further, one source of error in Y-factor measurements is the linearity of the receiver, so a lower ENR causes less of a change in the power measured than a high ENR source. However, the DUT noise figure should not be significantly bigger than the ENR source or there will be too little change in the measured noise power for the on-state for the receiver to measure.

The Y-factor is very effective for very high gain devices, as only the change in noise is measured, not the absolute noise power, so drift in output cables or adding attenuation at the output has almost no effect on the measured noise figure (although it will have an effect on measured gain).

9.1.5 Cold-Source Methods

The principal advantage of the Y-factor method is that neither the gain of the receiver nor the gain of the DUT needs to be known; they are computed as part of the measurement, although the computed value is susceptible to error due to system mismatch. However, when using a VNA in a noise measurement system, the determination of the DUT gain is easily and precisely given by the S-parameter calibration and measurement. Since the principal reason for using a hot noise source is in the determination of the gain the DUT, such a hot source is not needed when measuring noise figure using a VNA. Rather, the so-called cold-source method is used which can result in a simpler and faster measurement scenario.

In the cold-source method, the noise figure of the DUT is computed essentially from first principles. The definition of noise figure can written as Eq. (9.2), and if the receiver can be calibrated to measure the relative noise power from the DUT (DUTRNP) and the available gain can be computed from the S-parameters of the DUT and knowledge of the source impedance, then the noise figure computation can be completed without resorting to hot noise measurements. Since noise measurements are usually quite slow compared with S-parameter measurements, due to the averaging needed to reduce jitter, the cold-source method is theoretically nearly two times faster than the Y-factor method; in practice it is much faster as most

SA and NFA implementations utilize a swept-tuned yttrium-iron-garnet (YIG) filter for image rejection, and the filter's sweep response time is quite slow. It is typical for a VNA-based cold-source measurement to be 5–10 times faster than an SA- or NFA-based approach, for the same noise bandwidth and jitter.

In the cold-source method, the gain-bandwidth product of the receiver must be characterized. This can be characterized using a noise source and a pair of hot/cold measurements as described in Eq. (9.11). If the noise-power measured by the receiver is converted to an equivalent temperature, then the effective noise temperature can be computed as

$$T_E = \frac{T_A}{G_A} - T_0 \quad (9.19)$$

where T_A is the available noise and G_A is the available gain. This is clearly recognized as the excess noise temperature at the input above T_0 . The noise factor is similarly computed, as

$$NF = \frac{T_A}{T_0 \cdot G_A} \quad (9.20)$$

This can be converted to S-parameters and incident noise power as

$$NF = \frac{T_{Inc}}{T_0} \cdot \frac{|1 - \Gamma_S S_{11}|^2}{(1 - |\Gamma_S|^2)|S_{21}|^2} \quad (9.21)$$

where $DUTNPI = \frac{T_{Inc}}{T_0}$, $DUTRNPI = DUTRNPI \cdot (1 - |\Gamma_2|^2)$, and Γ_2 is the output match of the device connected to the noise receiver.

If the VNA source impedance is matched (Z_0), the noise figure is simply

$$NF_{50} = \frac{DUTRNPI}{|S_{21}|^2}, \quad NF_{50_dB} = DUTRNPI_{dB} - S_{21_dB} \quad (9.22)$$

Thus, for a cold-source measurement, the noise figure given a 50Ω source impedance is simply the excess incident noise power (in dB above kT_0B) minus the S_{21} gain in dB. For example, if a DUT has 22 dB of excess noise and 20 dB of gain, it has a 2 dB noise figure.

Figure 9.5 shows the cold source measurement of an amplifier, computed as in Eq. (9.21), as well as the Y-factor measurement and the measurement presuming a matched source, from Eq. (9.22). From the results, it appears that the Y-factor measurement has the most ripple (~0.25 dB), as it does not apply any correction for mismatch. The trace labeled “Matched-Source Approximation” takes a simple ratio of DUTRNPI and $|S_{21}|$, which is exactly correct if the source-match is 50Ω , though here it also shows some effect due to source mismatch. In this case, it is considerably less than the Y-factor (~0.1 dB). The smallest ripple, as well as the lowest consistent noise figure, is from the fully corrected cold-source method of Eq. (9.21). The ripple here is on the order of 0.05 dB, and the trace shows a response consistent with the inverse of the gain of the amplifier, which is to be expected. In this case, the cold source measurement is improved by adding an external 6 dB attenuator pad to the test port cable, to improve the raw source-match. This is a good practice to reduce the residual error from the non-ideal source impedance.

Equation (9.21) gives the noise figure for a source that is not 50Ω , but it is *not* the noise figure as commonly understood, N_{F_50} . It is instead the noise figure for the DUT in the reference impedance of the source, just as the Y-factor gives the noise figure of the DUT at the

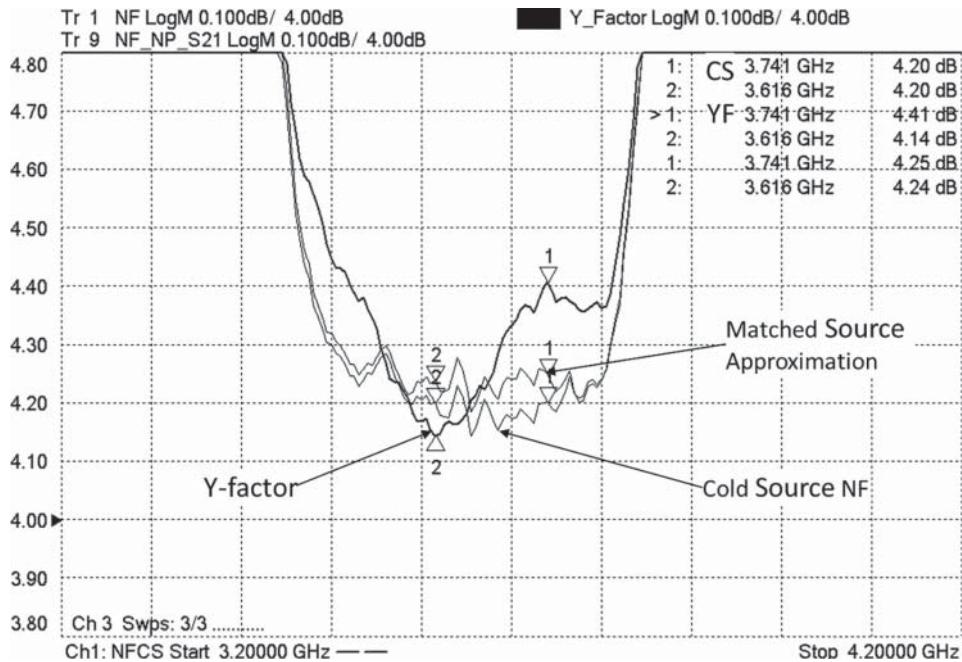


Figure 9.5 Noise figure computed from cold source; also shown is the Y-factor NF and NF assuming a matched source.

impedance of the cold state of the noise source. And, with the information so far discussed, it is not possible to predict or compute the $50\ \Omega$ noise figure from this value. However, methods based on the cold-source measurement can be used to find the noise parameters of a DUT, and from those, the exact $50\ \Omega$ noise figure may be accurately computed, as discussed in the next section.

9.1.6 Noise Parameters

The noise figure for amplifiers varies as the source impedance changes. For most amplifiers, the noise figure is specified at Z_0 , usually $50\ \Omega$. However, many amplifiers have a lower noise figure at some different impedance, and a key design task is to create a matching network that transforms $50\ \Omega$ to the optimum impedance for minimum noise figure.

Noise parameters provide the basis for understanding how the noise figure of an amplifier changes with the source impedance. The noise figure of a device at any reflection coefficient is described in Eq. (1.90) using four values as

$$N_F = N_{F\min} + \frac{4R_n}{Z_0} \frac{|\Gamma_{opt} - \Gamma_S|^2}{|1 + \Gamma_{opt}|^2(1 - |\Gamma_S|^2)} \quad (9.23)$$

This function can be described intuitively as a parabolic shaped surface above a Smith chart, where the distance from the Smith chart represents the noise figure, the lowest point of the

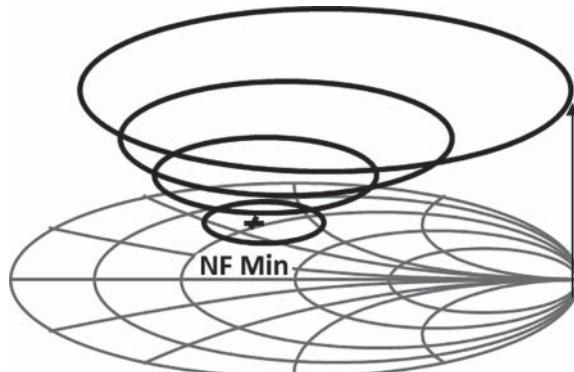


Figure 9.6 Noise parameters describe the noise figure as a function of source impedance.

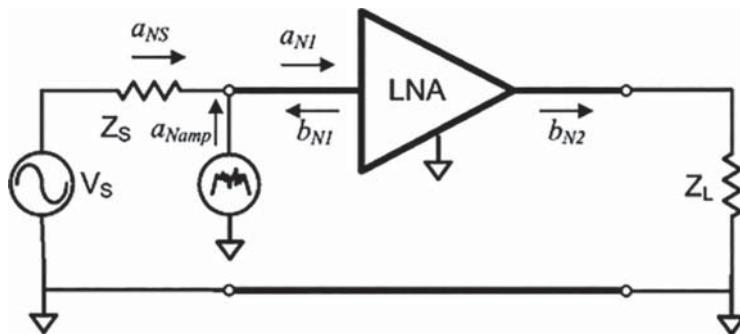


Figure 9.7 An amplifier with internal noise sources.

parabola lies above Γ_{Opt} and its height represents the minimum noise figure, as illustrated in Figure 9.6.

An amplifier with higher R_n will see the noise figure increase more quickly as the source reflection coefficient moves further from Γ_{Opt} . In the illustration, the arrow shows the noise figure of an amplifier with source impedance near the edge of the Smith chart, which represents the condition of an amplifier with an open circuit at the input. The circles are typically set to represent 1 dB steps in the degradation of the minimum noise figure and are typically plotted directly on the Smith chart for a single-frequency noise circle plot.

To understand the cause of the noise-parameter effects, consider the amplifier in Figure 9.7, with internal noise sources; the effect of the noise source is to produce noise power waves that may be treated similarly to normalized power waves, a and b .

The source termination produces an incident noise wave a_{NS} , simply due to the temperature of the source impedance, which adds to the internal noise created in the amplifier, and is represented as an input noise source a_{Namp} . From this figure, one can make a direct comparison to the S-parameters and see that reflected noise power might add to or subtract from the incident noise power and affect the total noise power. However, at the input of the amplifier, the noise generated inside the amplifier is in general not correlated with the noise coming from the source termination so that they don't add together in a simple way. Due to this, the

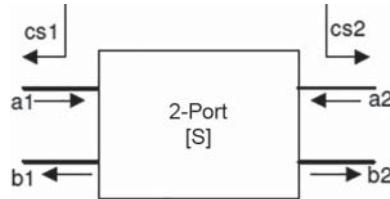


Figure 9.8 Noise representation of a 2-port network using s-parameter representation.

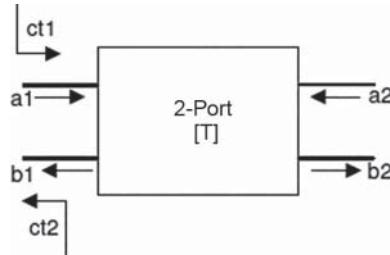


Figure 9.9 Noise parameter represented as T-matrix.

noise power at the output of the amplifier, and therefore the noise figure, depends upon the source impedance in a complex way.

There are scattered noise waves represented by the noise emitted from the input of the amplifier, b_{N1} , and the noise incident on the load is b_{N2} . Another representation can be shown as in Figure 9.8 (Keysight Technologies 2018; Randa et al. 2011).

The noise waves out of the amplifier can be represented as

$$\begin{bmatrix} b_{1N} \\ b_{2N} \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \cdot \begin{bmatrix} a_{1N} \\ a_{2N} \end{bmatrix} + \begin{bmatrix} cs_1 \\ cs_2 \end{bmatrix} \quad (9.24)$$

or alternatively, as shown in Figure 9.9, where the noise waves can be computed as

$$\begin{bmatrix} a_{1N} \\ b_{1N} \end{bmatrix} = \begin{bmatrix} tn_{11} & tn_{12} \\ tn_{21} & tn_{12} \end{bmatrix} \cdot \begin{bmatrix} b_{2N} \\ a_{2N} \end{bmatrix} + \begin{bmatrix} ct_1 \\ ct_2 \end{bmatrix} \quad (9.25)$$

where the t -matrix uses the alternative definition as described in Chapter 1, and the lower case tn is used to distinguish the Tn -matrix as being used for noise measurement purposes, distinct from the first form of T-matrix described in Eq. (2.20).

For this representation, the transformation from S-parameters to t -parameters for noise figure (tn) and back are defined as

$$\begin{bmatrix} tn_{11} & tn_{12} \\ tn_{21} & tn_{22} \end{bmatrix} = \frac{1}{S_{21}} \cdot \begin{bmatrix} 1 & -S_{22} \\ S_{11} & -(S_{11}S_{22} - S_{21}S_{12}) \end{bmatrix}$$

$$\begin{bmatrix} S_{11} & S_{12} \\ S_{12} & S_{22} \end{bmatrix} = \frac{1}{tn_{11}} \begin{bmatrix} tn_{21} & (tn_{11}tn_{22} - tn_{21}tn_{12}) \\ 1 & -tn_{12} \end{bmatrix} \quad (9.26)$$

The noise correlation matrix is defined as

$$C_t = \begin{bmatrix} ct_1 \cdot ct_1^* & -ct_1 \cdot ct_2^* \\ -ct_2 \cdot ct_1^* & ct_2 \cdot ct_2^* \end{bmatrix} = \begin{bmatrix} |ct_1|^2 & -ct_1 \cdot ct_2^* \\ -ct_2 \cdot ct_1^* & |ct_2|^2 \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \quad (9.27)$$

The noise correlation matrix can be found from the noise parameters, as (Dunsmore 2017)

$$C_t = \begin{bmatrix} F_{\min} - 1 + \frac{4R_n \cdot |\Gamma_{opt}|^2}{Z_0 \cdot |1 + \Gamma_{opt}|^2} & \frac{-4R_n \cdot \Gamma_{opt}}{Z_0 \cdot |1 + \Gamma_{opt}|^2} \\ \left(\frac{-4R_n \cdot \Gamma_{opt}}{Z_0 \cdot |1 + \Gamma_{opt}|^2} \right)^* & \frac{4R_n}{Z_0 \cdot |1 + \Gamma_{opt}|^2} - F_{\min} + 1 \end{bmatrix} \quad (9.28)$$

Alternatively, the noise parameters can be found from the noise correlation matrix, as

$$\begin{aligned} F_{\min} &= \frac{ct_{11} - ct_{22} + \sqrt{(ct_{11} + ct_{22})^2 - 4 \cdot |ct_{12}|^2}}{2} + 1 \\ |\Gamma_{opt}| &= \frac{ct_{11} + ct_{22}}{2 \cdot |ct_{12}|} - \sqrt{\left(\frac{ct_{11} + ct_{22}}{2 \cdot |ct_{12}|} \right)^2 - 1} \\ \theta_{opt} &= \arctan \left(\frac{\text{Im}(ct_{12})}{\text{Re}(ct_{12})} \right) \\ R_N &= \frac{Z_0(ct_{11} + ct_{22} + \text{Re}(ct_{12}))}{4T_0} \end{aligned} \quad (9.29)$$

The measurement of noise parameters often uses the noise wave method, resulting in a noise correlation matrix, which is converted to the noise parameters. An interesting point is while the noise correlation matrix can always be computed from the noise parameters, the noise parameters, in some cases, cannot be computed from the noise correlation matrix due to the difference under the square root sign in Eq. (9.29). Physically realizable devices must always have these values be positive, but errors and noise in the measurement can sometimes cause these values to be negative, and in such cases the values for noise parameters become imaginary, which is not possible. Thus, in these cases, the values for F_{\min} are typically forced to NF50, and the value for Γ_{opt} is set to zero. Methods for measuring noise parameters are discussed in the next section.

9.1.6.1 Noise Parameter Measurement Systems

Noise parameter measurements require, in addition to normal noise-figure measurements, some method for changing the impedance at the source of the amplifier under test. Traditional noise parameter measurement systems were comprised of a set of equipment including a VNA, NFA, noise impedance tuner, noise source, and switches to connect the various pieces of equipment for various measurements, as illustrated in Figure 9.10.

The source-impedance tuners are typically constructed by creating a well-controlled “slab-line” transmission line with a slot in one of the ground planes that allows a capacitive probe to come close to the line. The height of the probe sets the magnitude of the reflection

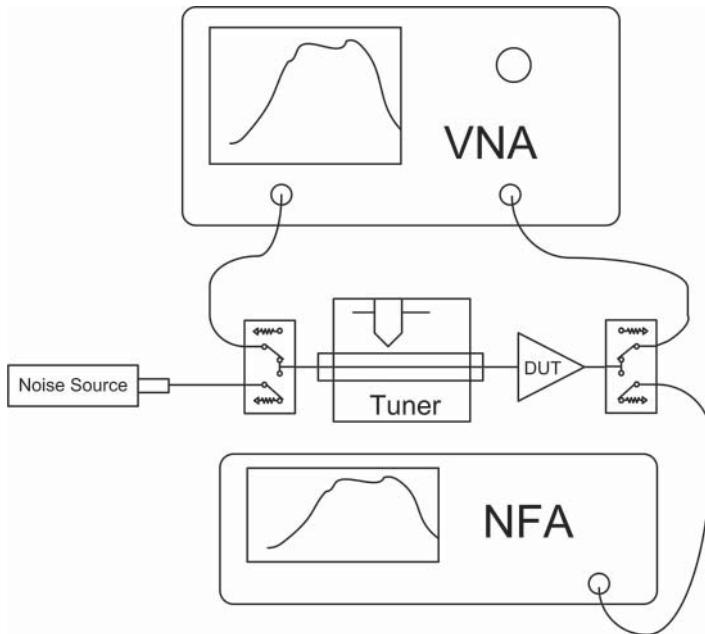


Figure 9.10 Traditional noise parameter measurement system.

coefficient, and the distance along the line sets the phase. These systems worked by premeasuring the impedance of the tuner at many different positions to find the desired impedance at each frequency, generating hundreds or thousands of measurement points. Then, with the amplifier inserted, the noise figure for each impedance is measured at many preselected impedance points, from which the over-determined sets of measurements can be used to solve Eq. (9.23). Unfortunately, older systems operated on a point-by-point basis, going from one CW frequency to the next and moving the impedance tuner to a variety of impedance states for each point. This process was very slow, and a typical system could take up to 20 minutes per frequency point including the time to calibrate the tuner.

With the introduction of noise-figure measurements in the modern VNA, the process was greatly simplified by removing the need for external switching and noise sources. The measurement algorithms were optimized to select tuner position rather than impedance points, with sufficient spread of the positions so that they would not overlap over the frequency range of interest. In this way, the tuner could be positioned one time for an entire sweep of frequency for both S-parameters and noise figure. The tuner was then repositioned and data taken again. Since the tuner moved only to a few positions (in one manufacturer's case, 21), the precharacterization was much faster, as was the measurement, resulting in measurement times as short as a few seconds per point, or several hundreds of times faster than old techniques.

The noise parameters don't require a mechanically based tuner, and at least one manufacturer has integrated an electronic tuner (similar to an Ecal module) to provide a single connection measurement of noise parameters, from which the exact 50Ω noise figure is derived by using noise parameters to correct for the error caused by the non-ideal VNA source-match. An example of such a system is shown in Figure 9.11, where a noise source

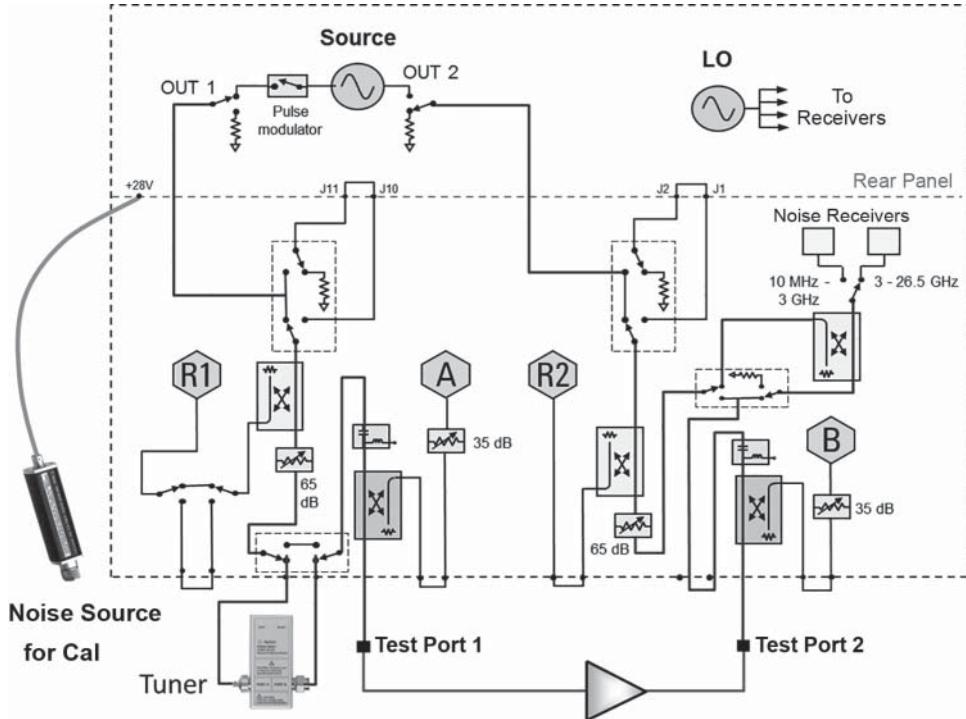


Figure 9.11 VNA system for making vector-error-corrected noise-figure measurements.

is used only during the calibration portion, and the tuner is realized with an external Ecal module.

The system shown uses the electronic tuner to produce just a few states, on the order of seven distinct impedances, to characterize the noise parameters of the DUT. These states surround the 50Ω impedance point but do not extend to the edge of the Smith chart. In such a case, there can be some substantial error in determining the values for N_{Fmin} and Γ_{Opt} , if Γ_{Opt} is far from 50Ω , but these errors are correlated in such a way that the computation of the corrected 50Ω noise figure has very little error, due to the impedance states surrounding the 50Ω point. Since the magnitude and phase of the source impedance affects the noise figure, this correction method is sometimes called a *vector-corrected* noise-figure measurement, in contrast to the *scalar* (magnitude-only) correction from Section 9.1.5.

A noise parameter data acquisition requires multiple noise measurements, each a different impedance tuner state, in addition to the S_{21} measurement required for the gain measurement. The relationship between the noise parameters and these measured values is given by

$$T_{RN} \cdot |1 - \Gamma_S S_{11}|^2 = B \cdot |S_{21}|^2 \left(c t_{11} 2 \operatorname{Re}(\Gamma_S) + 2 \operatorname{Im}(c t_{21}) + |\Gamma_S|^2 c t_{22} + (1 - |\Gamma_S|^2) \frac{T_{\Gamma_S}}{T_0} \right) \quad (9.30)$$

where

$$T_{RN} = \frac{T_{\text{Meas}}}{T_0}, \quad T_{\Gamma_S} = \text{Temperature of } \Gamma_S \quad (9.31)$$

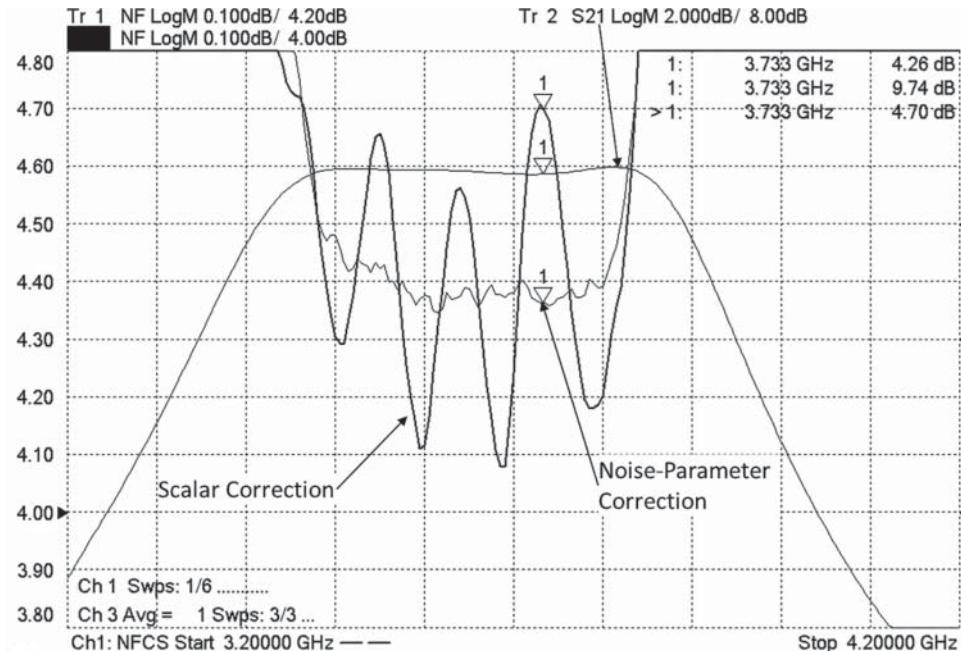


Figure 9.12 Vector-corrected noise-figure measurement, compared to Y-factor and cold noise.

and B is noise bandwidth. The ct terms are the unknown terms, and by making at least four noise temperature readings, T_{meas} , with a four different source impedance states Γ_S , at known source impedance temperatures, T_{Γ_S} , they maybe be solved for using a variety of methods. Normally the number of source impedance states is much larger to produce an overdetermined set of equations from which the best-fit values for the noise correlation matrix are found.

A measurement of the 50Ω vector-corrected noise-figure measurement is shown in Figure 9.12, along with a cold-source scalar-corrected measurement of the same amplifier.

The ripples in the measurements are an indication of the source pulling effect of non-ideal source impedance. It is much larger in this figure as no 6 dB attenuator was used on the port 1 for the scalar noise correction, so the full mismatch of the VNA source impedance pulls the noise figure of the DUT. This mismatch is only properly corrected when using the noise parameter correction method, the details of which are discussed next. In this case, the source-match, at the end of the VNA test port cable, is about 15 dB. The full vector calibration improves the effective match to better than 40 dB.

9.1.7 Noise Parameter Measurement Results

The noise measurement system described in Figure 9.11 can be utilized to measure the noise parameters directly. Although the limited number of states associated with using an Ecal as a tuner gives rise to in a lot of variation in the noise-parameter results, there is still value in viewing these noise parameters to understand the nature of the behavior of the DUT.

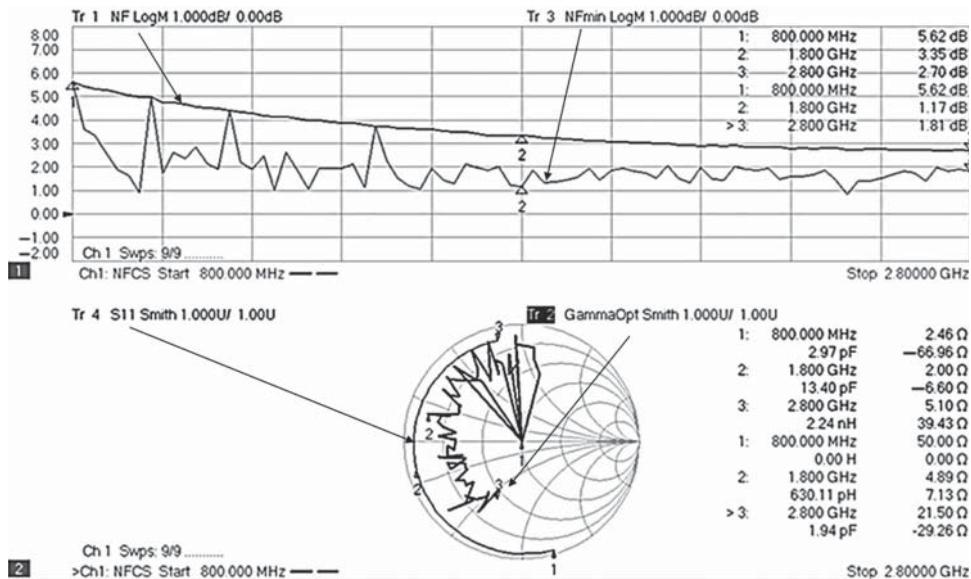


Figure 9.13 Noise parameter using Ecal as a tuner.

Consider the noise parameters shown in Figure 9.13. It is clear from the figure that the NF_{min} is lower, on average, than the NF50 by a considerable amount, especially at low frequencies, and for the most part, the GammaOpt is far from 50Ω . This is a single-transistor amplifier with a poor input match, so it is expected the GammaOpt might be quite far out the Smith chart as well. Notice there are four points where the NF_{min} is the same as the NF50, and at these same four frequencies, the GammaOpt is zero (50Ω). This is the result of a poorly formed noise-correlation matrix as used in Eq. (9.29). Whether due to poor separation of the impedance tuner states, or other uncertainties, the values under the square-root sign have become negative at these four points. Thus, it is not possible to compute the noise parameters, and the analyzer simply returns NF50 for NF_{min} and zero for GammaOpt.

Recently, a new capability has been added to some VNA-based noise-measurement methods, which allows a mechanical tuner to be substituted for the Ecal. It may be placed either in the same location as the Ecal or even directly in-line with the DUT between port 1 of the VNA and the DUT input, as shown in Figure 9.14. With a mechanical tuner, many more reflection states can be used in determining the value of noise correlation, giving a larger over-determined matrix of Eq. (9.30).

Figure 9.15 shows the resulting noise parameters on the same device as Figure 9.13, but this time using a mechanical tuner instead of an Ecal, and extracting the noise parameters with 21 impedance states instead of only the seven states available from the Ecal. The mean value of the Ecal roughly matches the values taken with the mechanical tuner, but clearly the larger number of impedance states yields a much cleaner noise parameter result. The NF_{min} is essentially flat, and any increase in 50Ω noise figure at low frequencies is strictly due to mismatch effects. Likewise, the GammaOpt is much smoother and clearly shows that GammaOpt is following a curve of nearly constant resistance. Interestingly, the S_{11} of the

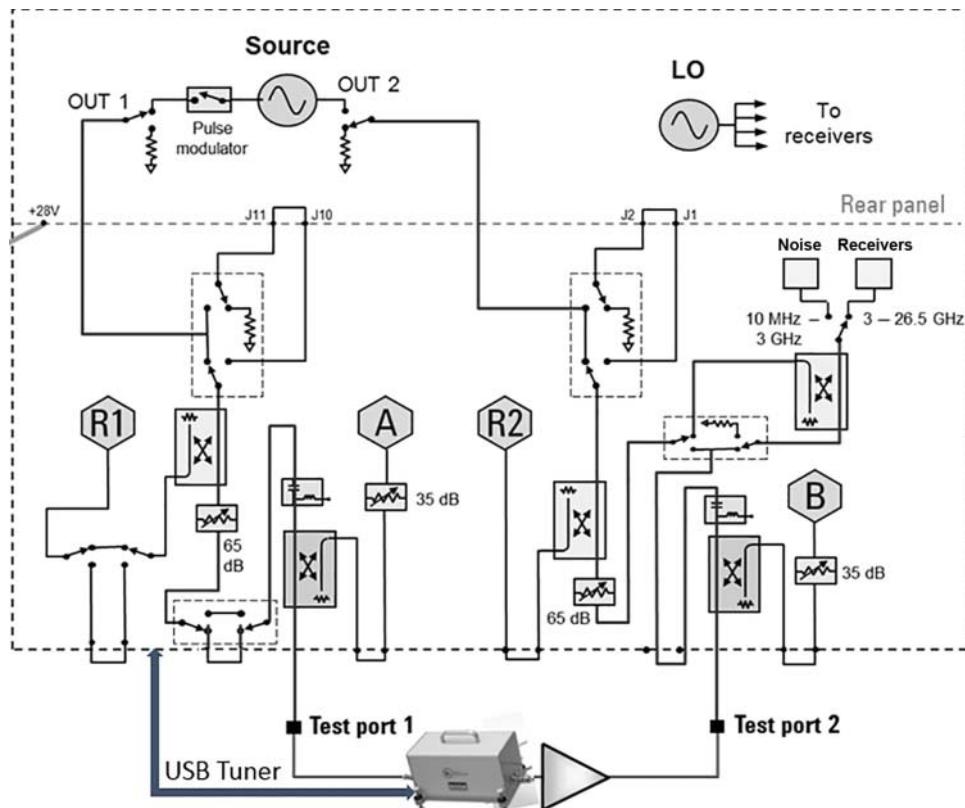


Figure 9.14 Tuner-based noise parameters.

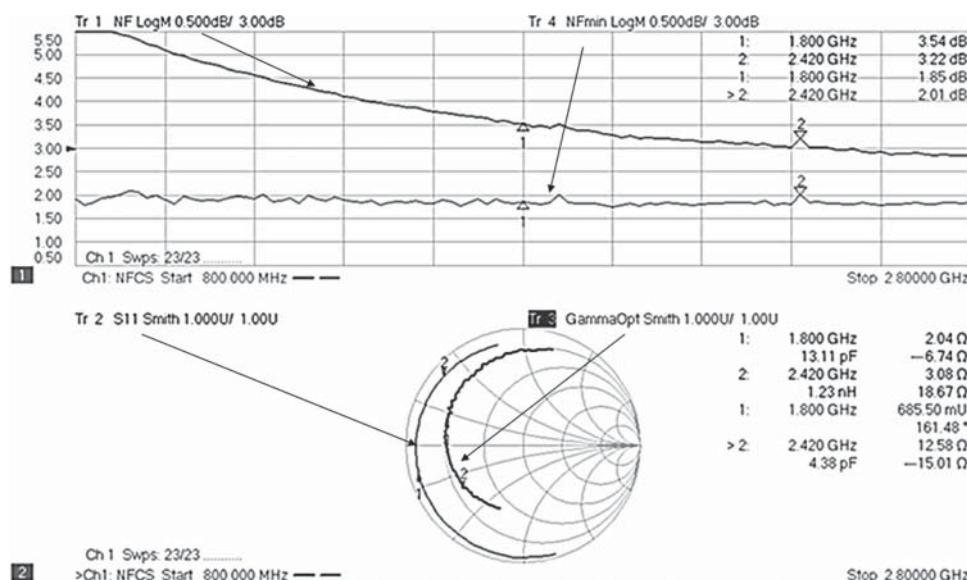


Figure 9.15 Noise parameters using a mechanical tuner.

amplifier is even further out on the Smith chart than the GammaOpt but follows a similar curve (accounting for the fact that GammaOpt rotates in the “counter-Foster” direction).

With such a system, it is clear that even for a poorly matched device, the noise parameters are easily determined. This system requires about 60 seconds to make a 101-point swept-frequency noise parameter characterization. Finally, the data may be saved in an S2P file, which includes a definition for noise parameter data.

9.1.8 Error Correction in Noise Figure Measurements

Error correction for noise-figure measurements consists of a few parts. In Y-factor measurements, the calibration measures noise at the NFA receiver input, and the correction consists of removing the effect of the NFA noise from the overall noise figure. However, the receiver noise figure can be “pulled” by the output match of the DUT so that the correction has some residual error if the DUT output match is not near 50Ω , or rather, if it is not near the impedance of the noise source used to characterize the NFA receiver. The Y-factor also is susceptible to the errors due to the impedance of the noise-source driving the DUT being different from the reference impedance. The errors in impedance will also reflect in errors in the gain as measured by the Y-factor technique. Thus, systematic raw errors can cause significant overall errors in measurement; Y-factor measurements rely on having good, well-matched source and load impedances. This is one reason why isolators are often used in precision Y-factor test systems.

For the scalar-corrected cold-source method, the gain and noise power are measured separately, and error correction is also applied separately. Normal S-parameter error correction is used for the gain computation, and so the errors from gain are the same as for S-parameters, which can be reduced to very small errors using good techniques. For the noise measurement, the receiver noise can be measured using the hot/cold method, but this does not remove the noise pulling effects of the receiver. However, if a tuner is used, the noise parameters of the VNA receiver can be characterized, and the noise-power measurements can be corrected exactly for the noise of the second stage as it is pulled by the output match of the DUT. For this correction, an additional S_{22} measurement is required to ascertain the DUT output impedance and provide an exact determination of the second stage receiver noise. In this way, for low-gain devices, the cold-source method has less error than the Y-factor method. Of course, for high-gain devices, the second stage noise figure has a much smaller detrimental effect.

For vector-corrected cold source measurements, that is, measurements that include a tuner on the input, the pulling effects of the test system source impedance are also corrected, in addition to the gain and noise receiver correction. Thus, all systematic errors in a noise-figure measurement are corrected for in this method. This makes the vector-corrected noise-figure measurement very suitable in cases where the impedance at the input or output of the test system cannot be well controlled. Some examples include on-wafer measurements, systems with multiport switching matrixes before or after the DUT, or in-fixture measurements of DUTs.

Recently, new methods for VNA based noise-figure measurement systems have replaced the noise source as a reference for receiver gain-bandwidth determination with the use of a power sensor. A guided match-corrected power calibration is used to transfer the accuracy of the power sensor frequency response to the VNA receiver. This provides the gain portion of the receiver characterization. The bandwidth is measured independently by sweeping the

VNA source while holding the VNA receiver constant (much like the measurement that shows the noise receiver bandwidth in Figure 9.1), to trace out the shape of the IF response; the shape is integrated to find the total noise bandwidth. This is combined with the receiver gain to produce the total gain-bandwidth product.

9.2 Active Antenna Noise-Figure Measurements (G/T)

The recent implementation of massive multiple-input multiple output (MIMO), along with the push to move communications bands higher in frequency with the new 5G rollout, has caused engineers to develop large-scale antennas with active circuitry integrated into the antenna. Some antenna are intended for point-to-multipoint applications, and these will include phase-array steering to overcome the loss of the channel at the new 5G millimeter frequencies. Other uses are for very large arrays (more than 1000 elements) to be utilized in the “new-space” applications, which include large constellations of low-Earth-orbit (LEO) satellites numbering in the thousands. The gain required to overcome the path loss is accomplished by utilizing these large-element phased-array antenna, and in these large-element antenna, every element has an LNA behind the antenna, usually before the phase shifter used to form the antenna beam, on the receiver side of the antenna.

Unlike traditional phased-array antenna where the LNA is typically a separate module that can be individually tested, these new arrays are intended for low-cost applications and often use a large PC board with patch antennas for the radiators, along with chip and small-packaged LNAs mounted directly behind the patch network. Sometimes the phase shifter is included in with the LNA in a monolithic microwave integrated circuit (MMIC), and the combining network follows. For these kinds of structures, it is difficult to measure the noise figure beforehand, so some means of over-the-air (OTA) noise-figure measurement are required. For antennas, this measurement of the noise characteristics is called G/T.

G/T, as a figure of merit, is a measure of the gain of the antenna divided by its equivalent input noise temperature. The value increase as the antenna gain goes up and also increases as in internal noise generation goes down. The higher the G/T, the better the antenna. Resistive losses in the combining structure degrade the G/T, so adding an LNA directly behind the radiator can improve the G/T, even if though the active device does add excess noise. The mathematical definition is

$$G/T|_{T_0} = \frac{G_{Ant}}{T_0 + T_E}$$

G_{Ant} relative to an isotropic antenna (9.32)

at the output of the antenna. The antenna is always measured referring to T_0 even though in applications such as space the environmental equivalent noise temperature can be much colder. The G/T of an antenna can be computed from knowledge of the gain of the antenna and the noise figure and gain of the LNA. For example, consider the antenna shown in Figure 9.16.

The excess noise T_e plus the ambient noise T_0 are both amplified by the LNA gain, which also amplifies the passive antenna gain. The G/T can be computed by considering just the antenna gain and LNA noise figure. From the definition of noise figure

$$T_e = T_0 \cdot (NF - 1) = T_0 \cdot NF - T_0$$

(9.33)

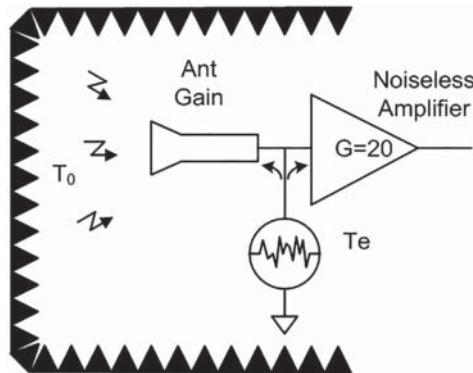


Figure 9.16 Defining the G/T of an active antenna.

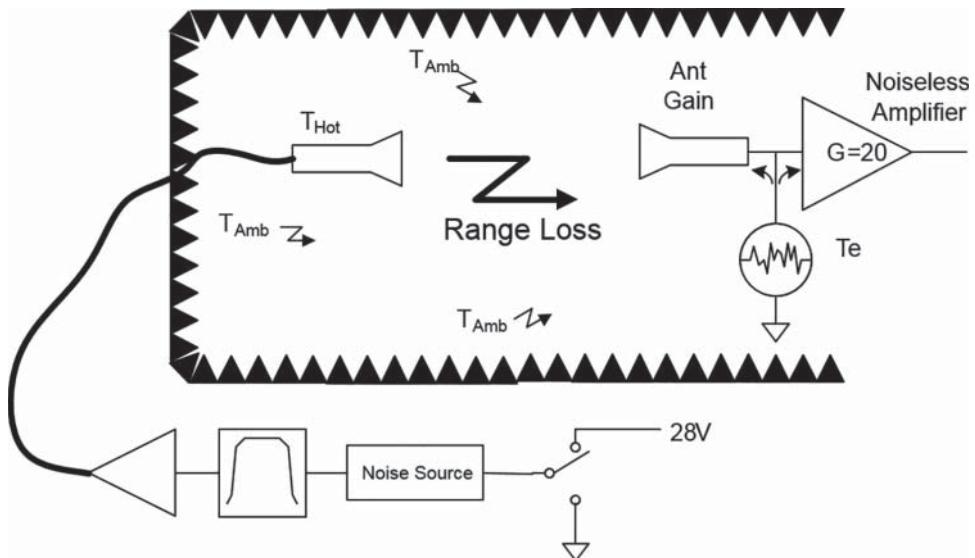


Figure 9.17 Using Y-factor to measure G/T.

Combining with Eq. (9.32),

$$G/T = \frac{G_{Ant} \cdot G_{LNA}}{T_0 \cdot NF \cdot G_{LNA}} = \frac{G_{Ant}}{T_0 \cdot NF} \quad (9.34)$$

where G_{Ant} , the antenna, is the passive or directional gain of the antenna; this means if we know each of these gains and the noise figure of the LNA, we can define the G/T. The gain of the antenna is the power gain, T_0 is in Kelvin, and NF is the noise factor (linear). The traditional way of measuring G/T is to measure the noise figure of the system, using a Y-factor method, as shown in Figure 9.17. One can measure the noise figure of the overall system, but to know the G/T, the range loss must be accounted for.

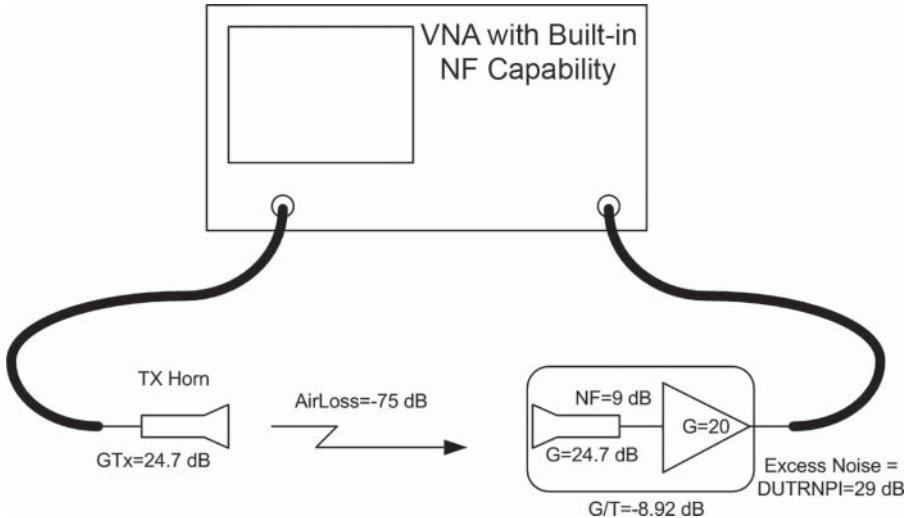


Figure 9.18 VNA using cold-source method to measure G/T.

This method works fine for a traditional active antenna, but it fails when trying to measure an active phased-array antenna, because the noise from the ambient chamber is applied to each radiator and the gain of each is quite small. The hot noise will be lost in the range loss, and the overall gain of the antenna is not achieved until the beam is formed electronically; while the hot noise is coherent and will experience gain, the ambient noise at each radiator may not be coherent and so will not experience the same gain. To overcome some of these issues, very large amplifiers are used to increase the hot noise, but this adds a host of other problems and uncertainties. Recently, VNA-based cold-noise methods have been introduced to solve this problem, as illustrated in Figure 9.18.

The G/T for the antenna illustrated can be determined from the given noise figure, antenna gain and LNA gain, by converting each to their linear form, as

$$NF_{Amp} = 10^{(9/10)} = 7.943$$

$$G_{Ant} = 10^{(24.7/10)} = 295.12$$

$$10 \cdot \log_{10}(290) = 24.62 \text{ dB}$$

$$G/T = \frac{295.12}{290 \cdot (7.942)} = 0.128$$

Or, in dB format :

$$G/T_{dB} = 10 \cdot \log_{10}(G/T) = 10 \cdot \log_{10}(0.128) = -8.92 \text{ dB}$$

$$G/T_{dB} = G_{Ant_dB} - NF_{dB} - 24.62 = (24.7 - 9 - 24.62) = -8.92 \quad (9.35)$$

It's interesting to note that the gain of the LNA does not affect the G/T of the antenna, as it adds to the overall antenna gain and noise at the output in equal amounts and so cancels out. It is only the noise figure of the amplifier that is important. Of course, if the antenna is followed

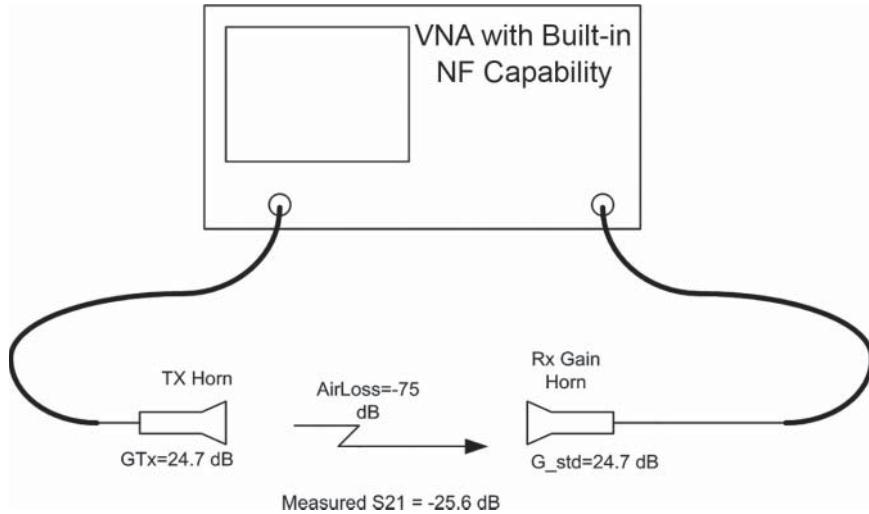


Figure 9.19 Range loss measurement.

by a second stage LNA, then the gain does become significant as the overall noise figure is affected by the gain of the first stage.

From the previous figure, the noise figure of the overall system (from port 1 to port 2) can be found from the antenna gain noise figure and range loss (which is the combination of the Tx horn gain and the air loss of the range) as

$$NF_{Sys} = \frac{NF_{Amp}}{G_{Ant} \cdot RangeLoss} \quad (9.36)$$

The NF_{Sys} is the noise figure measured by the VNA from port 1 to port 2. The range loss can be determined by measuring what is known as a *standard gain horn*, as illustrated in Figure 9.19. The standard gain horn has a calibrated gain (it is just a coincidence that it has the same gain as the Tx horn, although this is common as a similar style is often used for the Tx horn).

Here we have two values that can be determined: the measured range loss ($S21_{RL_meas}$) and the true range loss (RL). The true range loss is computed from the measured as

$$\begin{aligned} |S21_{RL_meas}|^2 &= RL \cdot G_{std} \\ RL &= |S21_{RL_meas}|^2 / G_{std} \\ RL &= \frac{(10^{(-25.6/20)})^2}{10^{(24.7/10)}} = \frac{2.754 \times 10^{-3}}{295.12} = 9.331 \times 10^{-6} \\ RL_{dB} &= 10 \cdot \log_{10}(9.331 \times 10^{-6}) = -50.3dB \\ RL_{dB} &= S21_{RL_meas} - G_{std_{dB}} = -25.6 - 24.7 = -50.3dB \end{aligned} \quad (9.37)$$

Here note that the gain (as with all elements related to noise measurements) is power gain, and thus one must use the square of the magnitude of S_{21} in the equation. From the noise

figure and the range loss, the G/T can be computed directly as

$$\begin{aligned} G/T &= \left(\frac{1}{NF_{sys}} \right) \cdot \frac{1}{T_0 \cdot |S21_{RL}|^2} = \left(\frac{1}{NF_{sys}} \right) \cdot \frac{G_std}{T_0 \cdot (S21_{RL_meas})^2} \\ &= \left(\frac{G_{Ant} \cdot \cancel{(S21_{RL_meas})^2}}{NF_{Amp} \cdot \cancel{G_std}} \right) \cdot \frac{\cancel{G_std}}{290 \cdot \cancel{(S21_{RL_meas})^2}} = \frac{G_{Ant}}{T_0 NF_{Amp}} \end{aligned} \quad (9.38)$$

Equation (9.38) shows that the G/T measurement only needs the measured noise figure of the overall system, comprised of the transmitter horn, the air loss, and the target antenna, as well as a previously measured value for the range loss.

Alternatively, one can separate the noise figure in two measurements, the gain of the overall antenna and the noise power out of the antenna, as

$$\begin{aligned} G/T &= \left(\frac{1}{NF_{sys}} \right) \cdot \frac{|S21_{std_gain}|^2}{T_0 \cdot |S21_{RL_meas}|^2} = \left(\frac{|S21_{sys}|^2}{DUTRNPI} \right) \cdot \frac{|S21_{std_gain}|^2}{T_0 \cdot |S21_{RL_meas}|^2} \\ &= \left(\frac{|S21_{sys}|^2}{DUTRNPI} \right) \cdot \frac{|S21_{std_gain}|^2}{|S21_{RL_meas}|^2} \end{aligned} \quad (9.39)$$

This does not require a noise-figure measurement, but simply an incident relative noise-power (that is the noise power above kTB level), an overall gain, and a measure of the range loss with knowledge of the gain of the range-loss measurement antenna. G/T measurements of this type have been used on phased array antenna with greater than 1000 elements, to a level better than 0.5 dB.

9.3 Noise Figure in Mixers and Converters

Mixers and frequency converters are often the first component after the antenna in a receiver system, and as such the noise figure of the first converter predominantly sets the noise figure of the system. Thus, noise-figure measurements on frequency converters are an essential aspect of their full characterization.

Most frequency converters have filters that keep unwanted images out of the response and have amplifiers either before or after the embedded mixer as well and so have overall gain and excess noise at the output above the kT_0B noise floor of a passive device. Thus, characterizing the noise figure a frequency converter is quite similar to that of an amplifier (see Section 9.1), where one measures the excess noise, then determines the gain of the mixer under test (MUT), and essentially computes the noise figure from these two values. Y-factor or cold-source techniques both work reasonably well on frequency converters. For the next sections, the aspects where mixer characterization differ from amplifier characterization will be highlighted.

9.3.1 Y-Factor Measurements on Mixers

Characterizing the noise figure of stand-alone mixers, in contrast to that of frequency converters, is a substantial challenge because many passive mixers have no excess noise and so their

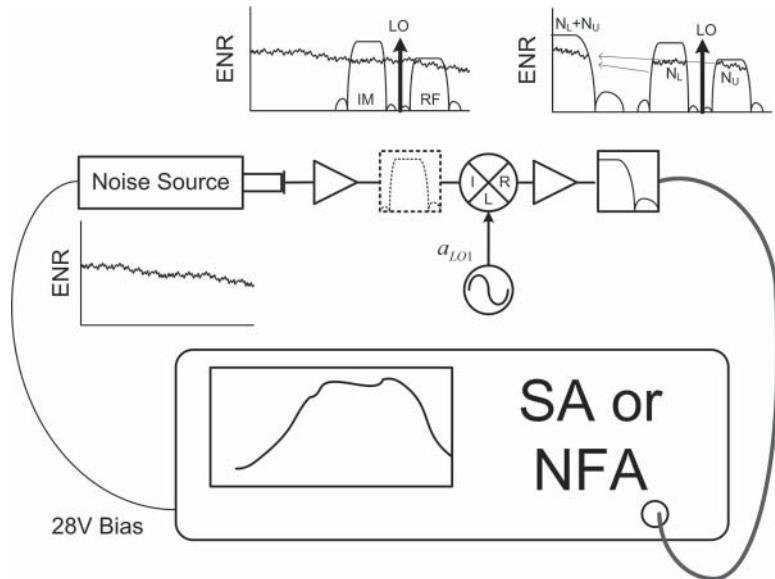


Figure 9.20 Y-factor mixer measurement illustration of double-sideband noise.

effective noise figure is simply the inverse of their conversion gain. If the mixer has no frequency selectivity at the input, the traditional Y-factor technique gives the wrong answer as the gain estimate has a substantial error; for many mixers the Y-factor measurement error for gain, and thus the noise figure error, is around 3 dB! Consider the comparison of measurement of a mixer or frequency converter illustrated in Figure 9.20.

For the mixer (or a converter without image protection, that is, without an input filter), the Y-factor technique applies the noise source at the input with the noise source turned on and measures the output noise; then the noise source is turned off, and again the output noise is measured. In this case, a down-converter, the noise source provides excess noise at both the desired RF input (one IF frequency above the LO) and the image input (one IF frequency below the LO); both noise powers are converted to the IF, where they are added together and are measured by the NF receiver. To illustrate the point, consider a noise-free mixer ($T_E = 0$) with a conversion loss of 6 dB, equally, at the RF and image (IM) frequencies; this would be the case of a broadband mixer. If the ENR noise source has a hot noise temperature of T_h that is constant with frequency and a cold noise temperature is T_0 , then the noise temperature measured at the output consists of the kT_0B noise plus the hot noise of the noise source at the RF times the conversion gain at the RF, and the hot noise at the IM plus the conversion at the IM to yield

$$T_{H_Rcvr} = T_{h_RF} \cdot |SC_{2IF,1RF}|^2 + T_{h_IM} \cdot |SC_{2IF,1IM}|^2 + T_0 = T_h \cdot (0.5) + T_0 \quad (9.40)$$

The cold measurement simply measures the cold temperature, as the mixer has no excess noise, so

$$T_{C_Rcvr} = T_0 \quad (9.41)$$

The noise figure computed from the Y-factor becomes

$$\begin{aligned}
 Y &= \frac{T_{H_Rcvr}}{T_{C_Rcvr}} \\
 NF_{DSB} &= 10\log_{10} \left(\frac{ENR}{Y - 1} \right) = 10\log_{10} \left(\frac{\frac{T_h}{T_0} - 1}{\frac{T_h(0.5) + T_0}{T_0} - 1} \right) \\
 NF_{DSB} &\approx 10\log_{10} \left(\frac{\frac{T_h}{T_0}}{\frac{T_h(0.5)}{T_0}} \right) = 3dB \\
 NF_{SSB} &= NF_{DBS} + 3dB = 6dB
 \end{aligned} \tag{9.42}$$

In fact, from first principles, the noise figure should be identical to the conversion loss, or 6 dB, but the implied conversion loss from the measurement is 3 dB too high, due to double-sideband conversion of the hot noise power. In most Y-factor measurement instruments, there is a mode to add 3 dB to the measured noise figure for mixers to compensate for double-sideband noise effects. If the mixer to be measured is not a bare mixer but instead has a pre-amplifier, as in the case shown in the previous figure, the amplifier at the input of the mixer may have some significant gain slope, so the noise power from the unwanted image (which is at a lower frequency) may be quite a bit higher than from the desired RF band above the LO. In this case, the error can be substantially more than 3 dB. The only way to know for certain how to correct for the gain in the undesired sideband is to measure the conversion loss of both sides, and compute the noise contribution from each to know how to correct the overall Y-factor.

In short, the cold-noise-power measurement in a Y-factor method is the correct cold noise, but the hot noise depends upon the sum of the upper and lower sideband conversion. If the mixer is image protected so that hot noise power from the noise source at the image frequency is blocked from the mixer, then the proper implied conversion gain is measured, and the Y-factor will give a reasonable result. In fact, the error could be greater than 3 dB if the sideband of interest has *lower* conversion gain than the undesired sideband. This might happen while evaluating the noise figure of a mixer at the edge of its specified range.

Another measurement technique, the cold-source method described next, doesn't have an issue with images in the mixer as the gain is computed independently; the image noise is a true source of noise at the output, during the cold-power measurement so it must be included. The image noise in a hot measurement is a *source of error* in the Y-factor method but does not occur in cold source measurements. A comparison is shown in the next section.

9.3.2 Cold-Source Measurements on Mixers

For noise-figure measurements using a modern VNA, the cold-source method can be applied, which may give better results than Y-factor for mixers without image protection.

The cold-source method follows the same in mixers as in amplifiers, with two measurements required to compute the noise figure. The first measurement is the gain of the mixer, SC_{21} , and is measured as described in Chapter 7, for standard mixers and for image mixers. The noise power is measured on a noise receiver, which has been calibrated in accordance to the method in Eq. (9.12). It can be converted to incident noise power, and the converter noise figure is computed as

$$NF = \frac{T_{Inc}}{T_0} \cdot \frac{|1 - \Gamma_{S_RF} S_{11_RF}|^2}{(1 - |\Gamma_{S_RF}|^2) |SC_{21}|^2} = \frac{|1 - \Gamma_{S_RF} S_{11_RF}|^2}{(1 - |\Gamma_{S_RF}|^2)} \cdot \frac{DURRNPI_{IF}}{|SC_{21}|^2} \quad (9.43)$$

For a well-matched system, the noise figure is simply

$$NF = \frac{DUTRNPI_{IF}}{|SC_{21}|^2} \quad (9.44)$$

That is, the mixer noise figure is the measured excess noise power in the IF divided by the square of the conversion gain. In dB terms, it is simply the excess noise power in dB minus the gain in dB.

Figure 9.21 shows a comparison of a Y-factor measurement, and a cold-source measurement, for the gain, relative noise power (DUTRNPI), and noise figure. In this case, the Y-Factor and cold-source gain and DUTRNPI match almost exactly. The noise figure of the Y-factor is about 0.2 dB higher than for the cold source. This is most likely attributable to errors in the source-match of the noise source. Since the DUTRNPI and gain are nearly identical, the noise figure should match with one computed by Eq. (9.44). However, the cold

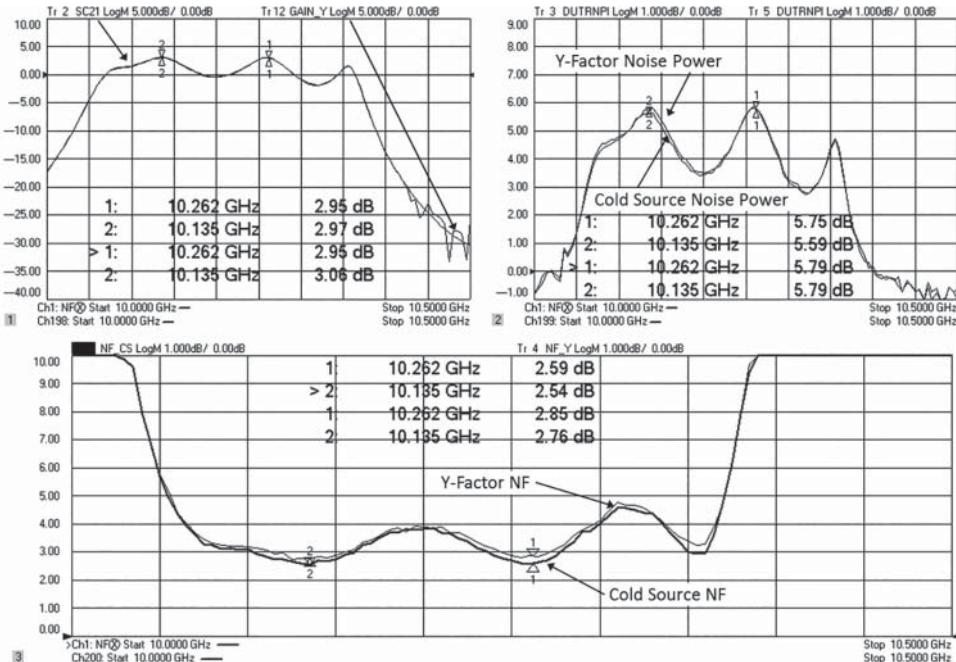


Figure 9.21 Comparison of Y-factor NF, cold-source NF, and SC_{21} .

source also does a mismatch correction for the effective source-match of the system, as in Eq. (9.43), which is why it displays a slightly lower noise figure. In the upper-left plot, the gain computed as SC_{21} and the gain computed as Y-factor are compared. In the passband of the mixer, the two compare nearly identically. In the stopband, however, the Y-factor gain shows significant error.

For this mixer, which has an image filter, both the Y-factor and the cold source give reasonable results. However, if the RF image-reject filter is removed from the converter's input, an entirely different result occurs.

The design of this converter has first an LNA and then an image filter (this configuration ensures the loss of the image filter does not hurt the noise figure and removes the excess LNA noise in the IM band from the mixer), then the mixer followed by another IF low-pass filter to remove any LO leakage. If the IF low-pass filter (LPF) is not used, the LO feed-through from the mixer can overdrive the NF receiver's LNA.

To understand the importance of the image filter, a second set of measurements on the same LNA/mixer/IF-filter set was performed, but in this case the RF image filter was removed. One would expect substantially higher noise figure as the LNA noise at the IM frequency will not convert to the IF. In this case, the LNA has a large gain slope, so the excess noise from the LNA is even higher in the IM range than the RF range. The results of making both cold source and Y-factor measurements are shown in Figure 9.22.

The measurement of the mixer in Figure 9.22 actually shows more gain variation in the SC_{21} trace. This is likely because the image filter at the input enhanced the conversion gain over some frequencies, so one sees about ± 3 dB change in gain without the filter. However, the Y-factor gain goes up by more than 6 dB! And it is no longer equal to the SC_{21} gain. Since

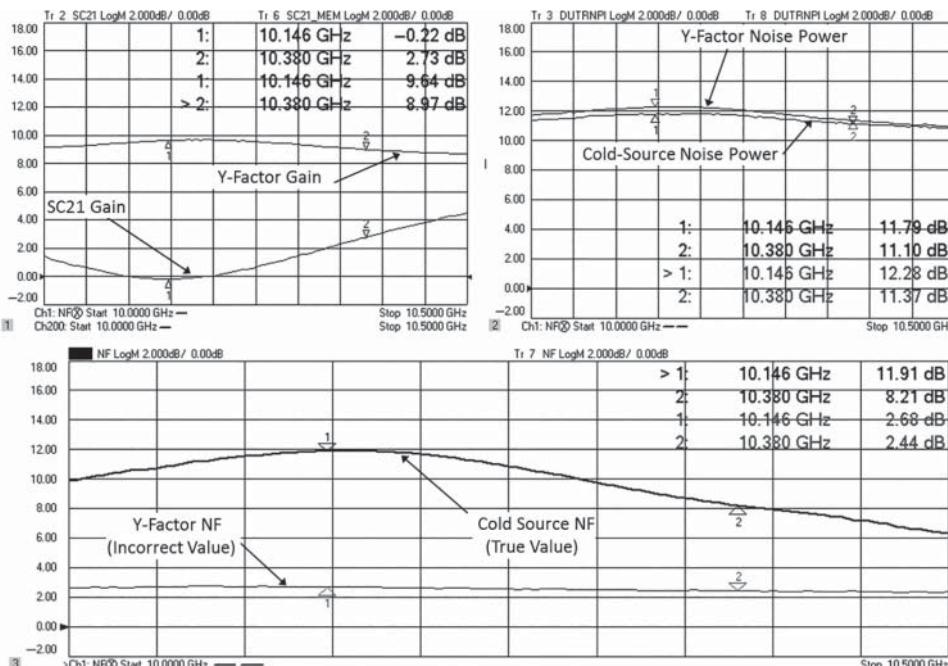


Figure 9.22 Comparison of Y-factor and cold-source on a mixer with no image filter.

the SC_{21} gain is simply computed by input and output power, there is very little uncertainty in this result, indicating that the Y factor gain has very large errors. Any time the Y-factor gain does not match the SC_{21} gain, the Y-factor noise figure must be suspect.

It's clear from this comparison that the Y-factor NF is in error by even more than the expected 3 dB associated with the double-sideband conversion. In this case, the image frequency is much lower than the RF frequency, so the gain of the LNA is much higher at the image frequency. Thus, the noise power from the unintended sideband, for both hot and cold measurements, is much too high. The DUTRNPI measurement should normally show similar powers for both cases, but this LNA is sensitive to match, and since the mismatch of the noise source is not identical to the mismatch of the VNA source, there can be some small noise power differences in the upper-right plot.

While the noise figure shown for Y-factor of less than 3 dB is appealing and many engineers would happily accept this result, it is most certainly also incorrect. The SC_{21} gain (performed using standard mixer measurements as described earlier, Section 7.5) shows the true gain of the MUT. The fact that the Y-factor gain is so much higher is the smoking gun for an invalid measurement. Unfortunately, ignoring the effects of the image gain is common mistake that inexperienced users make when using a Y-factor mixer measurement on an SA or NFA. In this case, only the cold-source technique can be used to show the proper noise figure for this mixer, which does not have image protection, and it shows it to have degraded by up to 8 dB due to the image noise.

9.3.2.1 Low-Gain Mixer NF Measurements

For bare mixers (or any device with loss), the measurement of noise figure becomes simply a measurement of the gain of the device, or very nearly so. If there is little or no excess noise in a device, the noise power density measured in the NF receiver is very near kT_0B and is usually swamped by the internal noise of the NF receiver. As part of the error correction, the noise of the NF receiver is subtracted from the overall noise measurement to try to obtain just the DUT noise. However, if the DUT is poorly matched, the NF receiver may be pulled by the match so the receiver noise figure measured during calibration is not the same as during the DUT measurement. Some advanced VNA systems do perform noise parameter extraction on the noise receiver, and so this effect is reduced. Also, because of the jitter in a noise-figure measurement, the characterized value for the receiver noise figure is only an estimate; substantial noise averaging (on the order of 100 or more) must be used if one wants to make measurements at kT_0B with low error. In most cases, the error due to jitter swamps other errors, and the noise figure of bare mixer is in error by the jitter in the receiver noise characterization, and the jitter in the MUT cold noise measurement. In these cases, a measurement of loss is probably a better estimate of the true noise figure than an actual noise-figure measurement. This is quite related to the effort to verify noise-figure measurement by measuring passive devices (see Section 9.5.2).

9.3.2.2 Excess LO Noise in Mixers

For measurements of mixers or frequency converters with an LO input, it is assumed that a perfect local oscillator is used to drive the LO port. In some cases, the local oscillator may have

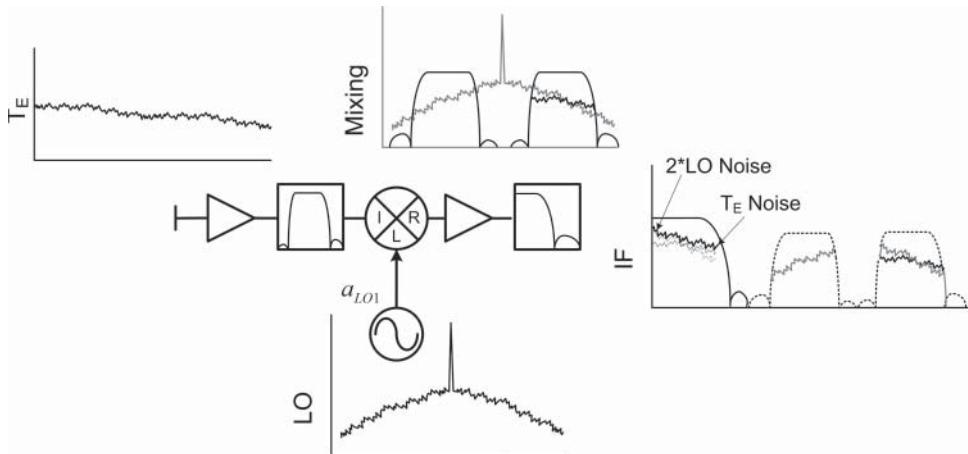


Figure 9.23 Errors due to excess converted noise of the LO.

significant amplification in the signal path before the DUT, and this noise can add substantial error to the measurement of noise figure of a mixer. Any excess noise from the local oscillator that is present at the LO port can be converted to the IF, as determined by the $SC_{2IF,3RF}$ and $SC_{2IF,3IM}$ terms (see Chapter 7). Figure 9.23 illustrates the source of this excess noise. Consider this converter, which is image protected by a filter after the LNA. The filter removes any excess noise from the amplifier out of band, so only the amplifier's noise in the RF band is converted to the IF. However, if there is excess noise on the LO port, it can be converted to the IF output from either above or below the LO signal, in either the RF or the image band. Thus, from the output one can see approximately twice the LO broadband noise in addition to the noise signal from the converter input. If the amplifier gain is sufficient, the RF noise will overcome the LO noise rendering it negligible. Also, if the conversion of LO noise to IF output is small (the $SC_{2IF,3RF}$ and $SC_{2IF,3IM}$ terms), then the LO noise contribution will be similarly reduced.

To fully understand the effect, consider a numerical example where the external LO has an amplification with 30 dB of excess noise (5 dB noise figure and 25 dB gain), which is called ENP_{LO_RF} and ENP_{LO_IM} as it appears at both the RF and the Image frequency, and the conversion terms of RF signal on the LO port to IF output is $SC_{2IF,3RF} = SC_{2IF,3IM} = -16\text{dB}$, so that the total excess noise power at the output of the mixer, ENP_{Tot_Hot} (presuming 3 dB conversion gain, and 15 dB excess noise of the noise source) is

$$ENP_{Tot_Hot} = |SC_{2IF,1RF}|^2(ENR_{RF} + NF_{Mix}) + |SC_{2IF,1IM}|^2(ENR_{IM} + 1) \\ + ENP_{LO_RF}|SC_{2IF,3RF}|^2 + ENP_{LO_IM}|SC_{2IF,3IM}|^2 \quad (9.45)$$

which for this example, makes the hot excess noise

$$SC_{2IF,1RF_dB} = 3\text{ dB}, NF_{Mix} = 3\text{ dB}, \\ |SC_{2IF,1RF}|^2 = 2, |SC_{2IF,1IM}|^2 = 0 \text{ (due to image protection);}$$

$$\begin{aligned}
NF_{Mix} &= 2, ENR = 10^{15/10} = 31.6; ENP_{LO} = 10^{30/10} = 1000; \\
|SC_{2IF,3RF}|^2 &= |SC_{2IF,3IM}|^2 = 0.00251 \\
ENP_{Tot_Hot} &= (2) \cdot 33.6 + 2 \cdot (0.0251) \cdot 1000 = 67.2 + 50 = 117.4 \\
ENP_{Tot_dB_Hot} &= 10\log_{10}(117.4) = 20.6 \text{ dB}
\end{aligned} \tag{9.46}$$

Here the excess noise from the LO adds about 2.4 dB to the hot noise. And the cold excess noise is

$$ENP_{Tot_Cold} = |SC_{2IF,1RF}|^2(NF_{Mix}) + ENP_{LO_RF}|SC_{2IF,3RF}|^2 + ENP_{LO_IM}|SC_{2IF,3IM}| \tag{9.47}$$

which becomes

$$\begin{aligned}
ENP_{Tot_Cold} &= (2) \cdot (2) + 2 \cdot (0.025) \cdot 1000 = 4 + 50 = 54 \\
ENP_{Tot_dB_Hot} &= 10\log_{10}(54) = 17.3 \text{ dB}
\end{aligned} \tag{9.48}$$

Here, the excess LO noise adds 11.3 dB to the cold noise power! From which noise figure for this example can be computed as

$$N_F = 10\log_{10} \left(\frac{ENR}{Y - 1} \right) = 10\log_{10} \left(\frac{31.6}{\left(\frac{117.4}{54} - 1 \right)} \right) = 14.3 \text{ dB} \tag{9.49}$$

Thus, in this example, the excess noise on the LO adds about 11.3 dB to the measured noise figure for this mixer. On the other hand, if the excess noise of the converter were not 3 dB but the converter had more gain, say 20 dB total with 3 dB noise figure, then the excess noise from the mixer would be 23 dB, and the error would be less than 1 dB in the measured noise figure.

This error, excess noise from the LO, also affects the cold-source method, as it will also measure a similar error since the added LO noise is unrelated to the single-sideband or double-sideband nature of the RF to IF conversion of the mixer, but only depends upon the RF and IM conversion of the noise impinging on the LO port. Figure 9.24 demonstrates this effect with two measurements of a mixer, one where the LO is applied directly to the MUT and another where a narrowband filter is add to the LO path to remove the RF and IM noise frequencies from the LO signal. The second condition is the same as used for Figure 9.21.

This measurement is for a mixer with about 11 dB gain at the input, for an overall gain of 3 dB, and a noise figure of about 3 dB, just as in the example computation of Eq. (9.49). The degradation due to excess noise on the LO port is lessened if there is more gain before the mixer or if the MUT is an active mixer (as is common in an IC frequency converter). In such a case, the added noise from the gain and mixer noise figure swamps the contribution of the excess LO noise. Without gain before the mixer, it is *imperative* that the LO be filtered to remove noise that could reconvert to the IF. In most converters, the LO is often embedded, so any excess noise it adds should be included in the converter noise figure. In some cases, using higher-performance signal synthesizers, instead of the built-in VNA sources, may lower the excess noise contribution from the LO, but this depends entirely on the particular synthesizer design.

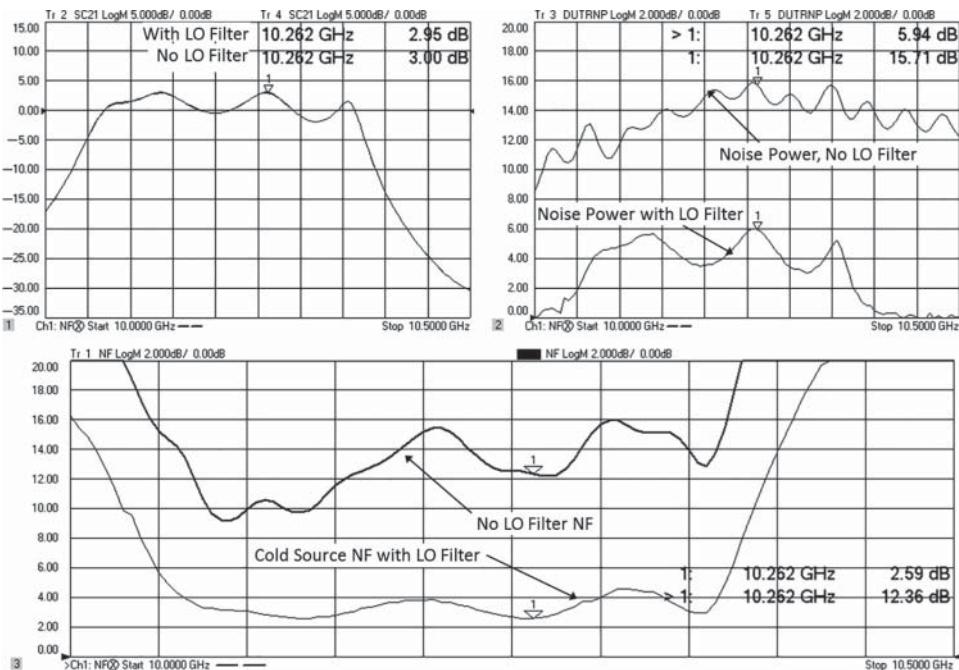


Figure 9.24 Error in noise figure due to excess LO noise.

9.4 Other Noise-Related Measurements

9.4.1 Noise Power Measurements with a VNA Spectrum Analyzer

Traditional spectrum analyzers have also been used to measure noise power, which is often used to determine the self-generated noise of a DUT. Unlike noise-power-ratio measurements, which are in fact a measure of distortion and not noise, measuring the intrinsic noise of a DUT is a true noise-power measurement. There are a few noise-power measurements that are significant and can be accomplished using traditional spectrum analyzers. These same measurements can also be performed with VNA-based network analyzers, but some specialized settings must be used to accommodate the effects of digital image rejection used in a VNA SA mode. These measurements are described next.

9.4.2 Noise-Power Measurements

Spectrum analyzers can directly measure noise power, if the noise power exceeds the noise figure of the SA. The noise-power measurements must be performed using the average power detection mode. It is typical to use either a noise marker (with display in dBm/Hz) or a band-power function or a band-noise function when measuring noise power. Since traditional spectrum analyzers are image protected, the only issue in measuring noise power is the calibration accuracy of the receiver. However, in the YIG-tuned microwave region, the absolute power uncertainty is rather high, on the order of ± 1 dB. A VNA-based SA using a

power-meter-corrected receiver calibration can have better uncertainty on the received power, but a noise-like signal will cause difficulties with the image rejection code, which may add additional error to the noise power reading. However, properly setting the image-rejection parameters, the video bandwidth (VBW) settings, and the analog-to-digital converter (ADC) filter settings can greatly reduce this error providing a good measure of noise power.

9.4.2.1 Traditional SA Considerations

For both traditional and VNA-based noise measurements using a spectrum analyzer, the key measures are noise power (in a band) and noise power density. Traditional spectrum analyzers are image-free by hardware design and so they may be simply used to detect the noise power using the marker readouts or band-power markers, but it is important to always use average detection and to set the detection method to power (or RMS) detection. In the past, the detection method for spectrum analyzers used logging amplifiers that produced an output signal based on the log of the input signal, and averaging was done after logging. Since the log function is non-linear, the average of a log is less, about 1.1 dB less, than the log of an average. For this reason, noise measurements using analog detection hardware had a negative bias. Modern spectrum analyzers utilize digital hardware to detect the signal voltage and convert to power or dBm using data processing, so any order is possible. To get the correct reading, one must use the average detector. Further, it is often common to add video bandwidth reduction to reduce the jitter. Many spectrum analyzers automatically set the VBW to one-tenth the resolution bandwidth (RBW) when placed in band-power or band-noise measurements, which slows the measurement but reduces the jitter resulting in a better estimate of the noise power.

9.4.2.2 VNA-Based SA Considerations

Because the receiver of a VNA-based SA is not image-free, its response to measuring noise power is more complicated than the traditional SA with hardware image protection. If the noise signal is quite broadband (wider than ADC filter in the IF path) and uniform, then no image rejection is needed, and it may be turned off for noise measurements. However, the measured noise will include contributions from both images of the noise signal, above and below the LO, and so the noise power will be two times (3 dB higher) than the true noise. One other issue is that if the LO is “hi-side” (meaning the SA RF frequency is below the LO), the measurement of noise power is skewed in frequency by the LO offset to a higher frequency than the displayed center frequency (since the measured power is from noise signal both above and below the LO). And if it is “low-side,” the RF measured is skewed lower than the displayed center frequency.

This effect can be avoided by using the image rejection function in conjunction with using a narrower ADC filter (many SAs including VNA-based SA have choices of fast Fourier transform [FFT] ranges set by the ADC filter) and using a lower VBW.

Using a lower VBW means that the signal will be averaged in power, before image rejection occurs. This will cause both images to be nearly the same value in power, and if they are nearly the same power, the resulting detected signal will be the average of the two measurements. This will give the equivalent of the power at the RF frequency exactly, without an offset.

At each RF frequency, four measurements of power (if two LO acquisitions are used for the image rejection): noise signal above (at the RF center frequency) and below the first LO; and noise signal below (also at the RF center frequency) and above the second LO. The average of these readings will be the displayed power. If the VBW is about 30 times smaller than the RBW, the effect of image rejection (which chooses the minimum image if the four readings are not close enough) is eliminated. Further, if the ADC filter is wide (say 40 MHz) the effective power is from as wide as 160 MHz; but if using a narrower ADC filter will restrict the measured noise power to a smaller range, a 10 MHz ADC filter restricts the range to 40 MHz.

Figure 9.25 shows measurement examples with various settings and RBW/VBW settings and for settings of image reject on and off. The signal source for this was an amplified precision noise source so the noise power density is very well known, at +35.74 dB above kTB or $-138.26 \text{ dBm Hz}^{-1}$. This is computed from the ENR of the noise source, the NF of the amplifier used, and the gain of the amplifier. The lowest trace shows a measure of noise power density (in dBm Hz^{-1}) using the VNA integrated low-noise receiver. The value shown is $-138.31 \text{ dBm Hz}^{-1}$, extremely close to the computed value, as one would expect in a precision noise receiver. The rest of the traces represent various settings of the SA in noise power (not density), and the noise power will change with the integrated bandwidth. Each trace has marker 1 showing the noise power density (in dBm Hz^{-1}), while marker 2 shows the noise power integrated over the RBW.

The upper two traces are both measured in a 1 MHz RBW, with the image reject mode on, but with different VBW.

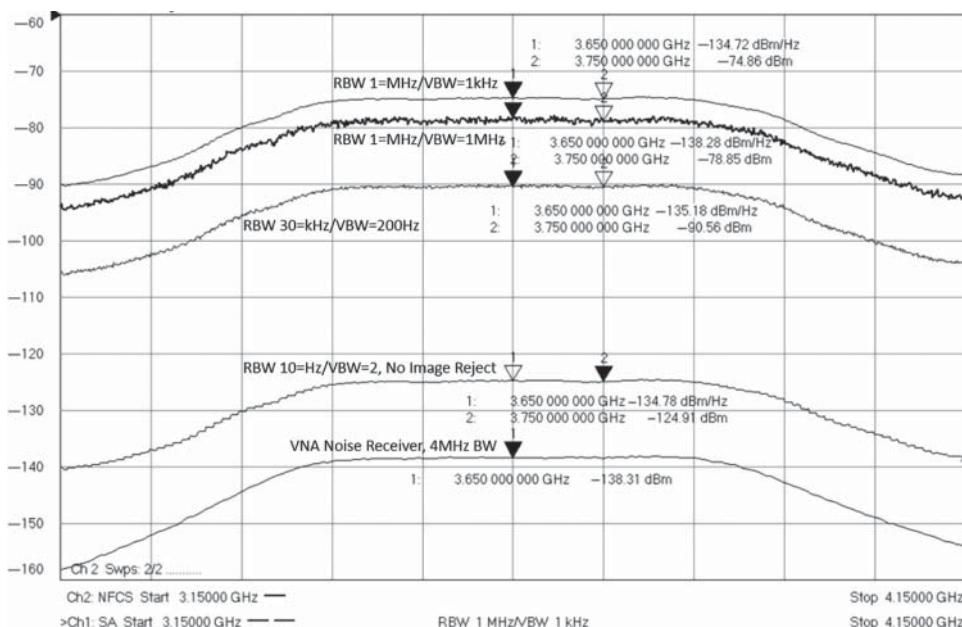


Figure 9.25 SA noise measurements with various settings image reject and RBW/VBW ratio, compared with VNA noise receiver.

9.4.3 Noise Figure Measurements Using Spectrum Analysis

Noise figure measurements are normally performed using the low-noise receiver of the VNA, but if that option is not available, a direct Y-factor measurement may be performed as well. The details of Y-factor measurements are included Section 9.1.3. The only measurement needed is the ratio of hot noise power (when the noise source is on) to cold noise power (when the noise source is off). The error in noise reading due to image rejection should the ratio between hot and cold measurements and so should not introduce an error when using the SA application to measure Y-factor noise figure. An example of the hot and cold measurement is shown in Figure 9.26, in the upper plot. A first set is of hot and cold noise powers are measured with the noise source connected directly to the SA port. These are labeled “Hot Cal” and “Cold Cal.” A second set of hot/cold measurements are performed with the LNA connected to the SA, and the noise sources at the input, labeled “Hot LNA” and “Cold LNA.” The middle plot shows the computed gain from the Y-factor measurements compared with the measured S_{21} (with full 2-port correction). As expected, the S_{21} trace is much cleaner and shows no ripple associated with imperfect test system match. The Y-factor gain is, however, quite close in value.

Finally, the lower plot shows the computed noise figure along with the measured noise figure using the integrated VNA noise measurement receiver. To compute NF using Y-factor, the ENR is needed, and this is read from the calibration data for the noise source.

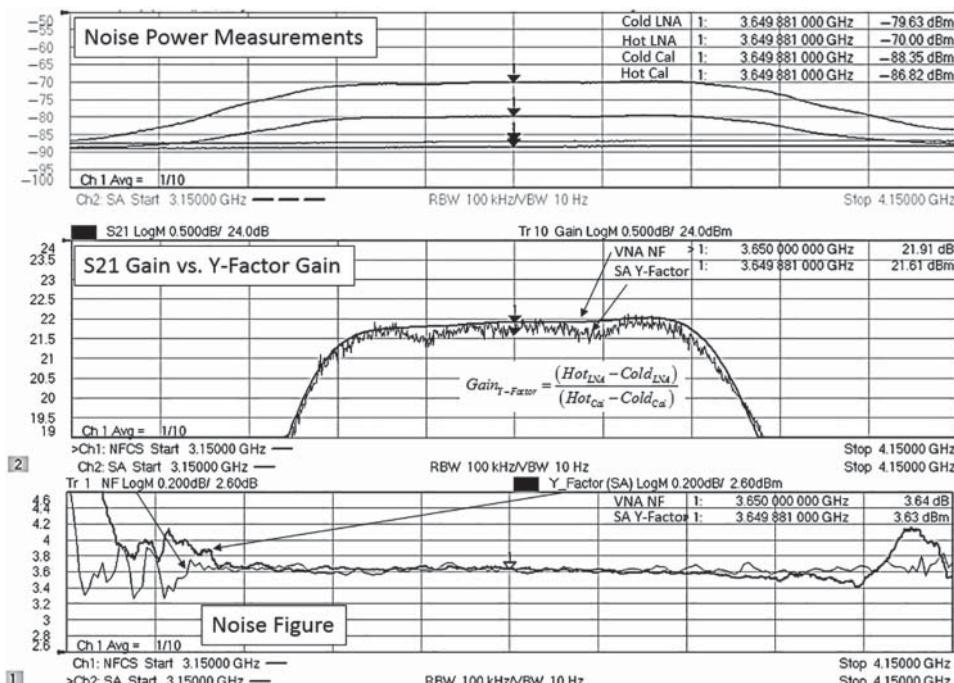


Figure 9.26 Hot and cold noise measurements and Y-factor computation, compared to norm NF mode.

The SA-measured NF matches very closely with the VNA noise-figure measurement, illustrating that Y-factor is an effective way of making noise-figure measurements. One note here is the VNA receiver was configured in high-sensitivity mode (reversed test port coupler) to improve its dynamic range. Finally, the noise figure for the Y-factor method also includes computing the second-stage noise figure and removing its effect, as described in Eq. (9.16).

9.4.4 Carrier-to-Noise Measurements

One measurement that requires the ability to measure both a CW signal power and a noise signal power is the carrier-to-noise measurement. This measurement is sometimes used to confirm the overall performance of a transmitter system. Noise figure is seldom measured in a transmitter, and if the transmitter is operated near saturation, traditional noise-figure measurement methods don't work well (the large carrier signal overdrives the noise receiver).

In a carrier-to-noise measurement, the carrier power is measured at the output, and the noise power, usually expressed as a density but not always in a 1 Hz BW, is also measured. If the input power is known, then the noise figure can be computed from the fundamental noise (Eq. 9.1), where signal to noise at the output is the carrier to noise measurement, and signal to noise at the input is the input power divided by kTB noise, expressed in the

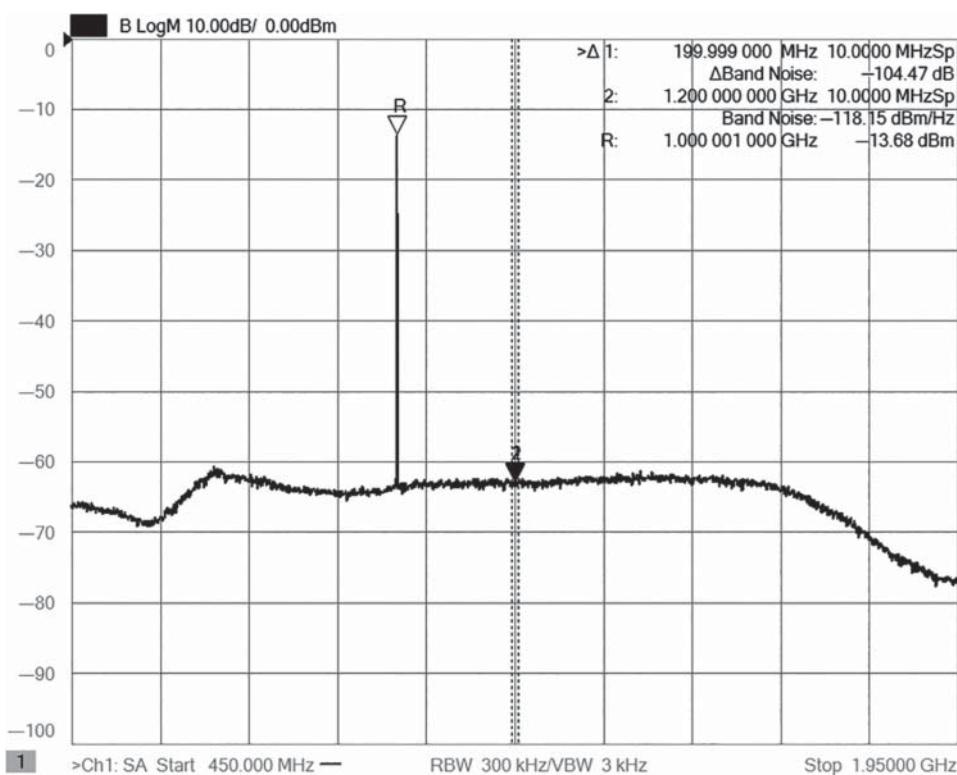


Figure 9.27 Carrier to noise measurements as well as a measure of input signal.

proper noise bandwidth. One cannot measure the carrier to noise at the input as the kTB level is always lower than the input receiver noise figure. And one must always ensure the noise power from the DUT is well above the kTB noise floor of the spectrum analyzers. Figure 9.27 shows the carrier-to-noise measurement of a low-noise block down-converter operating at a 12 GHz input frequency. Marker 2 shows the band-noise density at a specified offset. The reference marker shows the carrier power, and Marker 1 shows the carrier-to-noise measurement directly.

9.5 Uncertainty, Verification, and Improvement of Noise-Figure Measurements

9.5.1 Uncertainty of Noise-Figure Measurements

The accuracy of noise-figure measurements is probably the least understood and most optimistic in RF and microwave measurements. Instrument manufacturers and users have been a bit complicit in estimating the error in noise measurements, with the desire to have uncertainties below 0.1 dB. To achieve this value, the instrumentation errors are included, but many sources of error are often not described except in supplemental documentation. To be fair, the instrumentation manufacturers cannot control some of these external effects, but they really must be included in any assessment of overall uncertainty. Instrumentation errors are primarily linearity and mismatch pulling of the receiver used to measure noise figure, which are the most often quoted error. However, in many cases, the instrumentation error is perhaps the smallest contributor to the overall error. Fortunately, most manufacturers now supply a spreadsheet uncertainty calculator that does include many important effects; since uncertainty depends upon the actual characteristics for the DUT, one must compute a new uncertainty for each new device based on noise figure, gain, input and output mismatch, noise source used, and receiver used. In fact, one cannot really predict the uncertainty of a particular noise-figure measurement unless all the noise parameters of the DUT are measured, as the noise pulling effect of the input match cannot otherwise be properly accounted for.

A principal error in the uncertainty computation is the error associated with the ENR of the noise source, which translates almost directly to noise-figure error. This is sometimes ignored in a published specification of an instruments noise uncertainty, with the rationale that an arbitrarily good calibration can be obtained for a noise source, and the noise source is not part of an NFA instrument. But in practice, most users buy off-the-shelf noise sources without special calibrations. This error affects all methods of noise-figure measurement, except VNA systems calibrated with power sensors. In that case, the power sensor calibration factor becomes the primary uncertainty.

The mismatch error at the output of the DUT, between the DUT and the receiver, primarily affects the gain measurement in the Y-factor method but does not have a large effect on the noise figure. This is because the error due to mismatch affects both the gain and noise power in a similar way and is effectively canceled. The mismatch of at the output does affect the noise of the NFA receiver and can cause some pulling effects that may be significant for a low-gain, low-noise-figure device.

In many practical systems, the largest error is caused by input mismatch. This error can be fully understood only if the noise parameters of the DUT are also known.

Unlike S-parameters, it is rather difficult to verify noise-figure measurements, and no widely agreed upon method exists for such verification. Verification normally entails measuring a device with a known value of the attribute to be verified, and then one can determine the quality of measurement from the difference between the measured and expected value. Unfortunately, noise figure is somewhat complex, and the errors associated with the measurement come in two forms that cannot be easily distinguished. Recently, a new verification standard has been proposed that overcomes many of the issues with previous methods.

9.5.2 Existing Methodologies

A common verification standard for noise figure is to measure a known passive attenuator, whose noise figure will be $1/S_{21}$, and compare the measured result with the expected S_{21} . This method does not really validate noise-figure measurements, as it only tests noise-power measurement at exactly noise power of the kT_0B level, and as such really only verifies the S_{21} measurement. If the noise receiver has some uncompensated added noise (for example, the noise figure of the NF receiver depends upon the match of the DUT), then the noise power will not display the proper kT_0B value, but something either higher or lower and the error can be quite substantial depending upon the averaging used during the noise figure calibration and measurement. In this way measuring a passive device can give some indication of the sensitivity of the receiver noise power to S_{22} of the DUT. To see this effect clearly, a mismatched device must be used.

A measurement where the noise power of the DUT is below that of the measuring receiver, as is always the case with a passive device, is almost completely unrelated to the proper reading of noise power when testing an active device, where the DUT excess noise exceeds the noise generated in the NF receiver. If the DUT produces no excess noise, then the noise power reading is simply a remeasurement of the NFA self-generated noise power, and testing a passive noise figure is simply testing the stability of this noise power reading and is almost unrelated to the accuracy if the noise power from a real DUT is higher. For example, if the residual noise floor after calibration is 10 dB below the DUT noise power, the NF error is around 0.5 dB due to jitter; if the DUT noise is 15 dB above the residual system noise, the error is only 0.17 dB. Thus, unless one knows the actual total system noise figure, one cannot predict the accuracy of noise-figure measurements; and the use of a passive device is of almost no value.

A better technique to validate noise-figure measurements, especially cold-source measurements, is to understand the individual contribution errors to the overall result. In a cold-source measurement ,there are two factors: error in the gain (S_{21}) measurements, and error in the noise-power measurement. The S_{21} measurement error can be validated using well-known techniques to understand the S_{21} uncertainty. Noise-power measurements require measuring a known noise source. In fact, one of the best verification methods is to measure the noise power of a different noise source from the one used for calibration; the error in its measurement is a combination of the calibration error and the error in the verification noise-source ENR uncertainty. In some VNAs or NFAs, there is a direct ENR measurement parameter, so errors in the noise calibration can be readily discerned by this measurement.

The other key source of error in a NF measurement is due to input mismatch. (Because output mismatch-error affects noise power and output power in the same way, the composite

error in the NF measurement is negligible due to output impedance mismatch.) One way to characterize the quality of a NF measurement with respect to input mismatch is to add a small, low-loss, high-quality airline to the input of the DUT. Since the airline adds no mismatch and almost no loss, any change in the NF measurement after adding an airline can be associated with an error in the effective source-match and the measurement of NF of the DUT.

Full vector noise correction (using noise parameters) can minimize this error. Figure 9.12 provides a clear illustration of the improvement that a full-vector noise-parameter correction provides versus the scalar correction. Clearly, the full correction gives the best results.

9.5.2.1 Improved Traceable Verification Devices for Noise Figure

Until recently, the only ways to create a device with a traceable noise figure were to use a passive device as described earlier or to have an active device (e.g. an LNA) characterized at a certified calibration lab and use it as a transfer standard, as described in (Randa et al. 2011), where the authors use a mismatch line to verify that the changes in noise figure but admit that the nominal level of the noise figure is not verified due to issues with maintaining a constant temperature, and voltage on the amplifier across different calibration laboratories.

In 2017 this author proposed a new method to create a verification standard that is traceable to standards labs, that can be used to independently verify the noise-figure measurements of both Y-factor and cold-source methods, that has known noise parameters, and that produces substantial excess noise above the kTB level (Dunsmore 2017). In this paper, the proposed verification device consists of a noise source, with a traceable ENR, connected to a hybrid combiner with a traceable measurement of its S-parameters, as shown in Figure 9.28. The computation of the noise figure of this device can proceed by computing the incident noise at the output from knowledge of the loss in the hybrid, the temperature of the hybrid, and the available noise power from the noise source. Once the output noise is known, the loss through the input/output path can be computed (also accounting for mismatch of the noise source), and from these two pieces of information the noise figure can be directly computed.

Figure 9.29 shows an actual implementation of the noise verification device with comprised of a Keysight 346C-K01 noise source and an Anaran 90° hybrid.

In breaking down the problem into two pieces (noise computation and loss computation), it is useful to create a signal flow diagram for the device, as shown in Figure 9.30.

Note that there are two sources of noise represented in the signal flow diagram: one from the external noise source applied at port 3 (labeled $kT_{inc}B$), and one that represents the noise temperature of the hybrid (a_{n_hyb}), which is determined by the ambient temperature, $T_{Ambient}$.

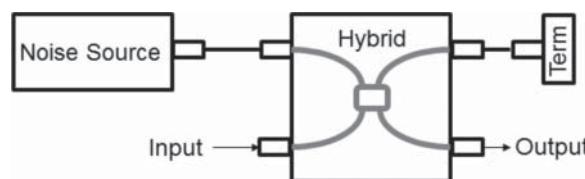


Figure 9.28 Noise-figure verification device.

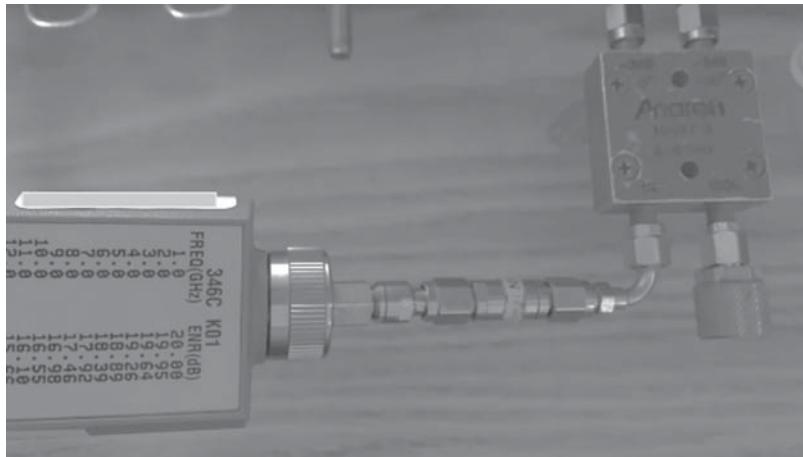


Figure 9.29 Photo of the actual verification device.

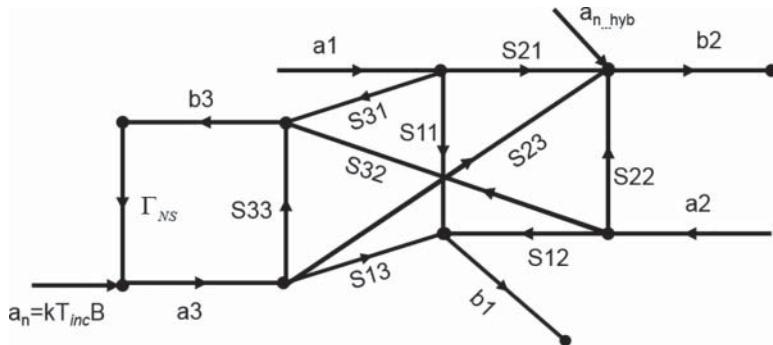


Figure 9.30 Signal flow diagram of a noise verification device.

The available noise from the external noise sources is found from the ENR as

$$\begin{aligned} ENR &= (T_{H_avail} - T_0)/T_0 \\ T_{H_avail} &= ENR \cdot T_0 + T_0 \end{aligned} \quad (9.50)$$

where $T_0 = 290K$

The incident noise can be computed if the return loss of the source, Γ_{NS} , is measured, as

$$T_{H_inc} = (1 - |\Gamma_{NS}|^2) \cdot T_{H_avail} \quad (9.51)$$

The noise source-match can be determined in a traceable way using standard VNA measurements, along with the uncertainties of the match determination.

The next step is to compute the S_{21} gain of the verification device, which can be derived from the simplified signal flow diagram of Figure 9.31.

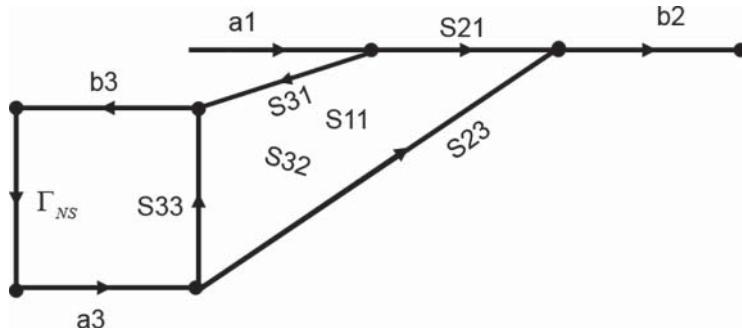


Figure 9.31 Signal flow diagram from computing the overall gain of the verification device.

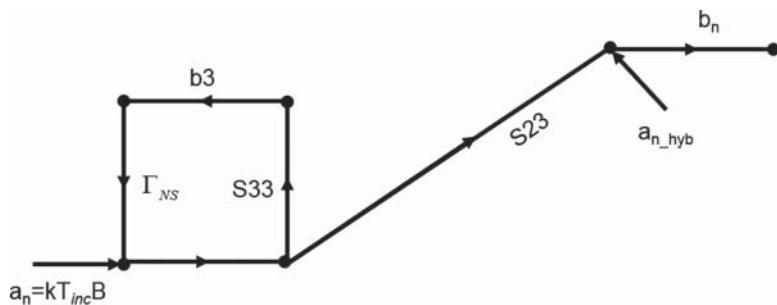


Figure 9.32 Signal low diagram for noise power out under matched conditions.

The overall transmission of the verification device is computed as

$$S_{21}^{Ver} = S_{21} + \frac{S_{31} \cdot \Gamma_{NS} \cdot S_{23}}{(1 - \Gamma_{NS} \cdot S_{33})} \quad (9.52)$$

Noise figure is almost always defined in a matched (50Ω) condition, and this simplifies the signal flow diagram, as shown in Figure 9.32.

The loss from port 3 (the port that to which the noise source is attached) is computed as

$$Loss_NS = \frac{S_{23}}{(1 - \Gamma_{NS} \cdot S_{33})} \quad (9.53)$$

The noise power out of the overall verification device can be computed as

$$T_{Hot_inc}^{Ver} = \left. \frac{(1 - |\Gamma_{NS}|^2) \cdot (ENR \cdot T_0 + T_0) \cdot |S_{23}|^2}{|1 - \Gamma_{NS} \cdot S_{33}|^2} \right|_{T_{Amb}=T_0} \quad (9.54)$$

presuming $a1 = 0$, $a2 = 0$ (matched condition). This also presumes the temperature of the hybrid is at T_0 , which is not usually the case. Equation (9.54) comes from the effect of loss on the noise power. The loss of the hybrid lowers the noise power, but it cannot be lower than the ambient temperature of the hybrid. For a simple case, the following the formula may be

used for computing the effective incident hot noise temperature at the end of a lossy device, such as a cable, from a hot noise source:

$$T_{Hot_Inc} = T_{Ambient}^{Cable}(1 - |S_{22}|^2) + (T_{Avail}^{Hot} - T_{Ambient}^{Cable})(1 - |\Gamma_{NS}|^2) \cdot |S_{21}|^2 \quad (9.55)$$

The hybrid, however, has two paths through it and can have two temperatures applied to the input; in one case it is the hot temperature of the verification noise source applied to port 3, and for the other it is the noise of the reference termination applied at the input, port 1. This reference termination is normally defined at T0, so the equivalent incident noise power at the output of the hybrid is

$$\begin{aligned} T_{Hot_Inc}^{Ver} = & \left\{ T_{Amb} \cdot \left(1 - \left| S_{22} + \frac{S_{23} \cdot S_{32} \cdot \Gamma_{NS}}{(1 - \Gamma_{NS} \cdot S_{33})} \right|^2 \right) \right\} \\ & + \left\{ (T_{Hot_avail} - T_{Amb}) \cdot (1 - |\Gamma_{NS}|^2) \cdot \left| \frac{S_{23}}{(1 - \Gamma_{NS} \cdot S_{33})} \right|^2 \right\} \\ & + \left\{ (T_0 - T_{Amb}) \cdot \left| S_{21} + \frac{S_{31} \cdot \Gamma_{NS} \cdot S_{23}}{(1 - \Gamma_{NS} \cdot S_{33})} \right|^2 \right\} \end{aligned} \quad (9.56)$$

where the S-parameters are from the hybrid coupler. The first term represents the incident noise power from the hybrid, as terminated with the noise-source impedance on port 3 and an ideal load on port 1, both at ambient temperature. The second term is the incident noise power from the noise source at port 3, above the ambient temperature, and passing through the loss of the hybrid coupler. The final term is the noise associated with the termination at port 1, at a temperature different than ambient (in this case negative as the input is defined to be at T0 which is typically lower than Tambient), and passing through the loss of the hybrid path from port 1 to the output.

The Incident power, relative to kT_0B , is then computed from

$$P_{Hot_Inc}^{Ver} = \frac{T_{Hot_Inc}^{Ver}}{T_0} \quad (9.57)$$

for the verification device and is shown in Figure 9.33.

From Eqs. (9.52) and (9.56), we can compute the overall 50 Ω noise figure of the verification device as Eq. (9.58).

$$NF_{50}^{Ver} = \frac{(T_{Hot_Inc}^{Ver})}{T_0 \cdot \left(\left| S_{21} + \frac{S_{31} \cdot \Gamma_{NS} \cdot S_{23}}{(1 - \Gamma_{NS} \cdot S_{33})} \right|^2 \right)} \quad (9.58)$$

This analysis of this verification is rather complicated and can be prone to errors in the proper handling of the various temperature and loss effects of the hybrid, so a complete model can be created using the Keysight Advanced Design System (ADS) simulator, with the circuit shown in Figure 9.34.

The ADS simulation is created by using an S1P model for the noise source utilizing the measured match of the noise source and by setting the temperature of the S1P model to the

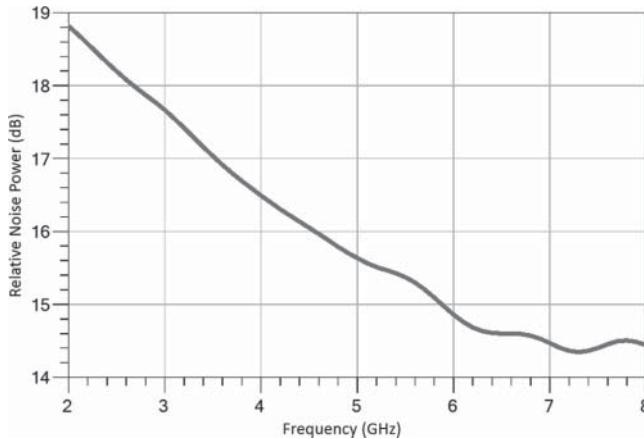


Figure 9.33 Relative noise power from the verification device.

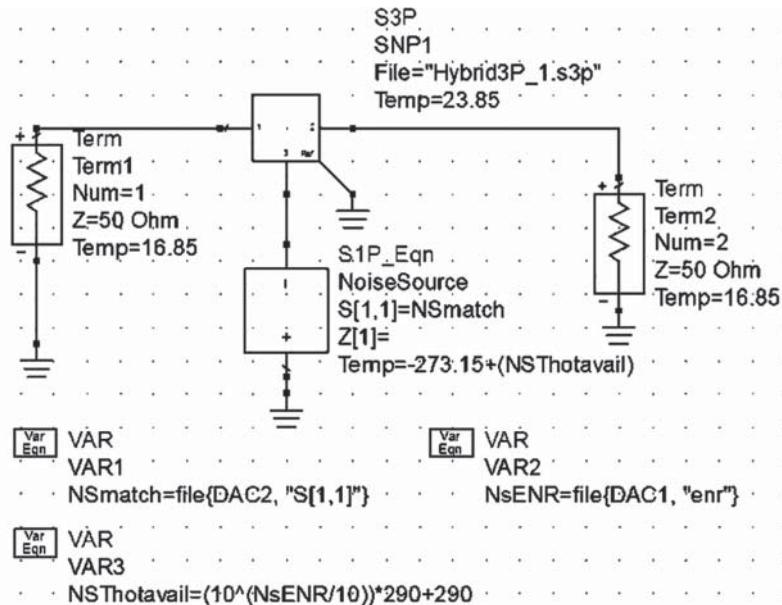


Figure 9.34 ADS simulation of the verification device.

noise source hot available temperature, as defined in Eq. (9.50). The simulation uses Celsius as the unit of temperature, so we must subtract 273.15° to convert the kelvin temperature to Celsius. This simulation can be compared with the analytical solution of Eq. (9.58). The resulting comparison between the ADS simulated noise figure and the analytically computed noise figure is shown in Figure 9.35. The results are nearly, but not quite, identical.

There is a very slight difference as shown in Figure 9.36. This slight difference is nearly uniform across the frequency range and is less than $1 \mu\text{dB}$. Most likely this difference is

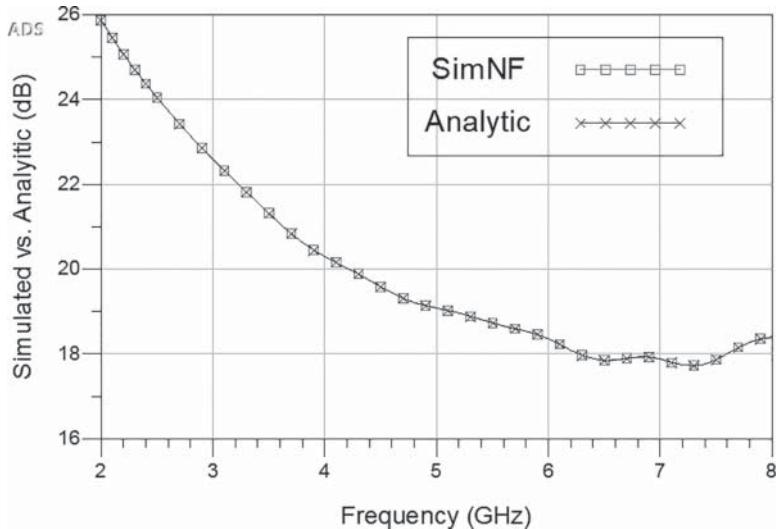


Figure 9.35 Comparing ADS simulation with analytic computation.

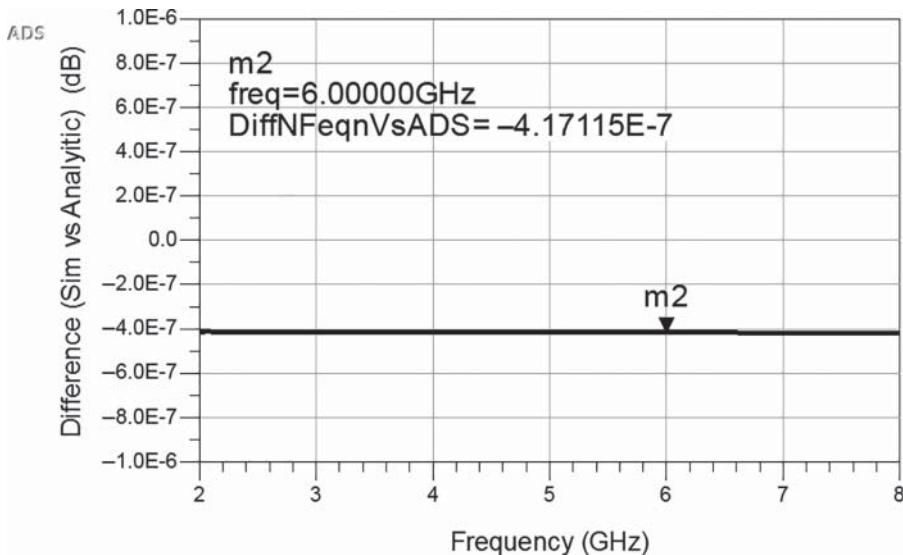


Figure 9.36 Difference between ADS simulation and analytic computation.

caused by the how the numeric data is handled between the simulator and data processing of the analytic equation. Thus, the analytic solution is verified as correct.

From this analysis, we can see that we now have a device whose noise figure can be computed from two sets of traceable measurements: the ENR of the noise source and the S-parameters of the hybrid. The traceable measurements mean the uncertainty of the noise

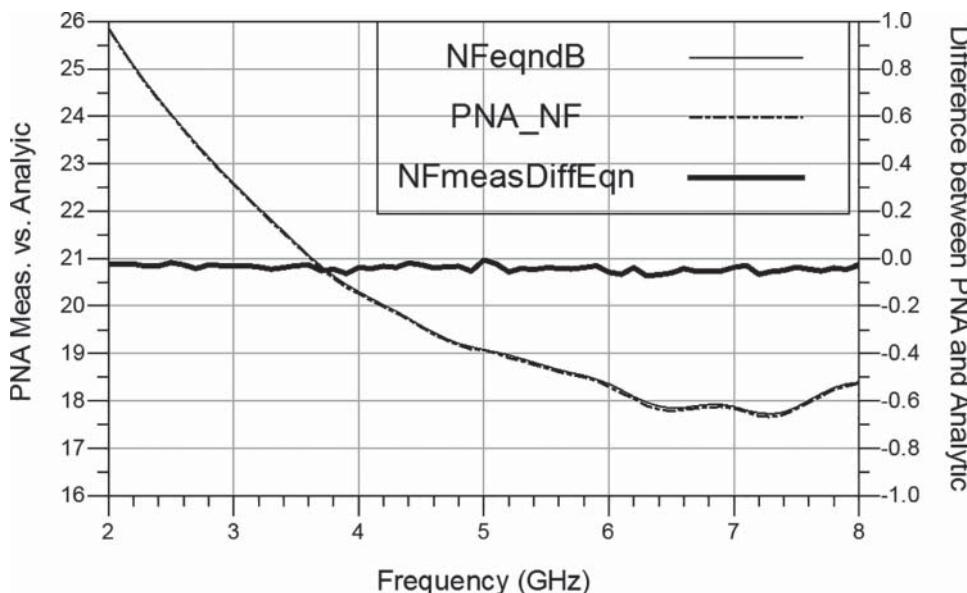
Table 9.1 NF verification uncertainty

Error Source	Specification or Expanded Uncertainty	Standard Uncertainty	Uncertainty Squared
ENR	0.2 dB	0.1	0.01
S21^2	0.03	0.015	0.0002
S31^2	0.03	0.015	0.0002
1-GammaNS^2	0.008	0.004	0.000016
1-GammaS11^2	0.008	0.004	0.000016
Root-sum-square (RSS)		0.102	0.0104
95% Confidence	2.77*RSS = 0.28 dB		

figure can be directly computed as well. The uncertainty of each portion of Eq. (9.58) can be determined, and the overall uncertainty can be computed, as illustrated in Table 9.1.

This verification standard can now be used to determine the quality of noise-figure measurement test equipment. The principal error is in the ENR of the noise source, but to test the actual contributions of the hybrid, the temperature effects, and mismatches, a first test is performed where the same noise source is used to calibrate the noise receiver (in this case a Keysight PNA-X with option 029 noise receiver) so that any error in the noise source is common to the measurement of the verification device and the calibration of the noise receiver. This will validate if the rest of the mathematical treatment is correct, to the limits of the uncertainty of the measurement methods.

The plot in Figure 9.37 shows the result of measuring the noise figure verification device on a PNA-X, which is calibrated with the same noise source as the verification device (which

**Figure 9.37** Measurement of NF verification device versus analytic computation.

removes the uncertainty of the ENR of the noise source from the results). The offset seen between the measured value and the computed value is shown on the right scale and amounts to a small (0.04 dB) offset with a slight ripple. This offset is well within the uncertainty of the S-parameter measurements, and the fact that these are nearly identical is a further indication that the analytic solution is properly derived.

From here this noise verification device can be used to evaluate other noise figure systems. Further, one could do an independent calibration of the PNA, using either an independent noise source or a power-meter based calibration, to compare the quality of calibration with a common device.

Figure 9.38 shows the measurement of the verification device on an N8975A noise-figure analyzer from Agilent. The noise source is a high-ENR noise source, which implies a relatively poor match. The overall response has a mean offset from the computed response of +0.22 dB, as well as about a 0.2 dB of ripple. This is most likely due to mismatch effects (noise pulling) due to the poor match of the noise source. The offset value is within the expected deviation as each noise source (the one used in the verification device and the one used for measurement on the NFA) has an uncertainty on the order of 0.2 dB.

To further illustrate the critical importance of the noise source used in the measurement to the overall uncertainty, consider the measurement in Figure 9.39. It shows a mean offset of -0.19 dB. This is in exact opposite direction of the previous measurement. Further, there is much more jitter (or noise) on the measurement because the ENR from this noise source is much lower than the previous one, but there appears to be less mismatch error.

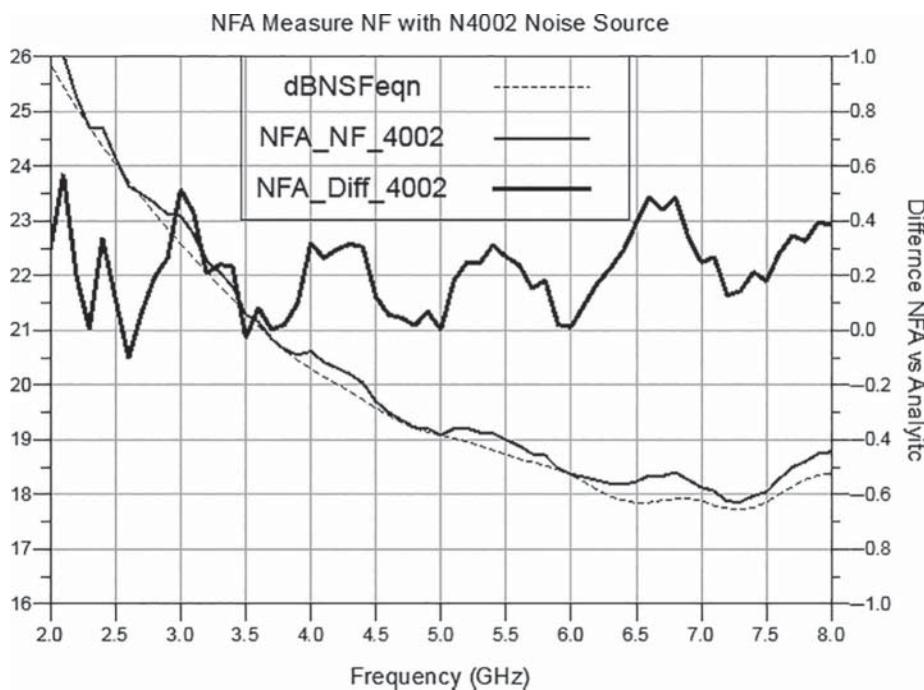


Figure 9.38 NFA 4002 measurement compared to analytic computation.

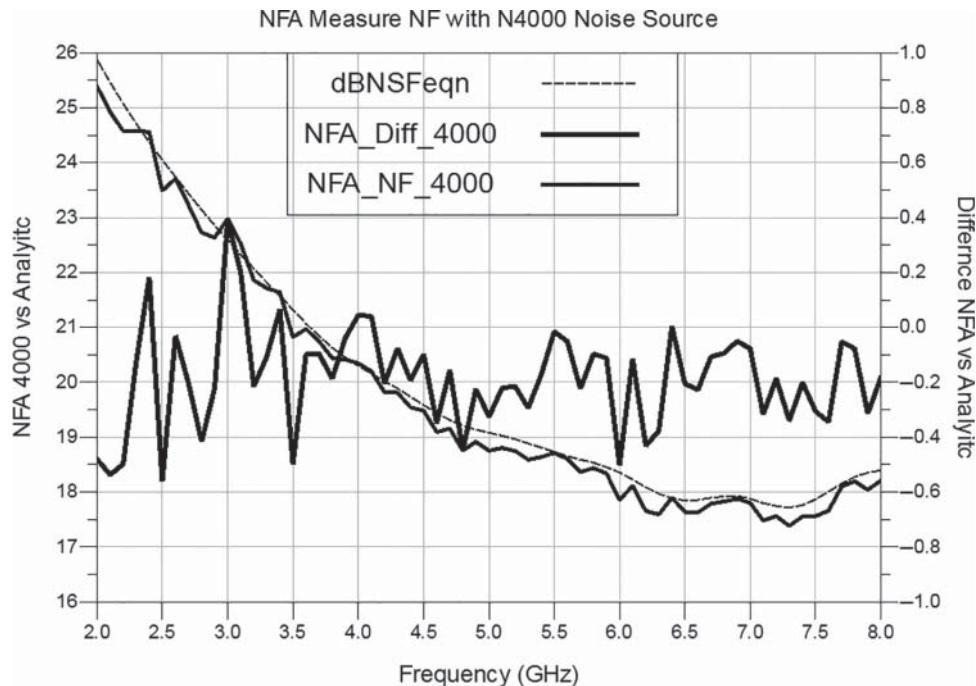


Figure 9.39 NFA 4000 measurement compared to analytic computation.

This measurement was done under the same conditions, but considering the lower ENR, the averaging factor could have been higher.

The variation of the noise-figure measurement here, compared to the PNA measurement in Figure 9.37, shows much more variation due to jitter, and even some underlying mismatch effect. The PNA-based measurement has virtually no mismatch effect due to the vector-noise-figure calibration, which essentially determines the noise parameters for the device and then generates the $50\ \Omega$ noise figure from those noise parameters.

This demonstrates both the relative quality of the measurement as well as the offsets due to ENR measurement accuracy. What is the best ENR accuracy? The higher ENR noise sources have more uncertainty because they have poorer match, but the mean value might be better as the higher excess noise means there is less error contribution due to the noise generated by the calibration system receiver. It is interesting to note that the verification device (made from a Keysight 346C-K01 noise source) shows a response in between the two other noise sources.

9.5.3 Techniques for Improving Noise-Figure Measurements

9.5.3.1 Improving Y-Factor Measurements

The overall error is dominated by the input mismatch error, and this can be minimized by using an attenuator or isolator at the input. Isolators are narrowband by nature, but attenuators are broadband, and adding them at the end of the port 1 connector improves overall accuracy of the measurement system with a few drawbacks. If an isolator or attenuator is used before

the DUT, its loss must be accounted for, and since it is resistive, the temperature must be accounted for as well. For a Y-factor measurement, loss before the DUT can be accounted for by modifying the ENR from the noise source to account for the loss. The modification can be computed simply as new hot temperature from

$$T_{H_Loss} = T_H \cdot L + (1 - L) \cdot T_L \quad (9.59)$$

Here, the loss is expressed as a linear power loss, and T_L is the temperature of the lossy element, provided that the loss is resistive. If the loss is purely reactive, just the first term is used.

Because mismatch cannot be corrected in Y-factor measurements, the use of cables or adapters should be minimized. Loss between the DUT and noise source should be compensated in the same manner as in Eq. (9.59).

Loss after the DUT must be compensated in a similar manner and can be accounted for in the computation of the effective temperature of the DUT, but typically, the loss after the DUT can be included in the calibration of the noise system when measuring the noise figure of the second stage. In fact, it adds directly to the second-stage noise figure.

For Y-factor measurements, using a low ENR noise source improves the accuracy of the measurement in two ways. First, it typically provides a better match at the input of the DUT, and so reduces the mismatch error. Second, it provides a smaller change from hot to cold state, so the receiver is measuring smaller differences in noise and is theoretically more linear. However, many modern receivers with digital IF receivers have very good linearity over wide ranges, and this error is relatively small. Using a small ENR requires more averaging to reduce jitter in the noise measurement and will give a less accurate measure of the DUT gain. Since Y-factor relies on measuring the slope of a line, the wider the spacing between hot and cold, the less fixed measurement error affects the result. On the other hand, for very high-gain devices, the hot noise should be low enough so that the total noise power from the noise source plus the gain does not put the DUT into compression.

9.5.3.2 Improving Cold-Source Measurements

The cold-source method can provide accurate noise-figure measurements, but some careful attention is needed to avoid common problems.

Since gain is measured using the VNA S-parameter method, one must be sure that the settings for S-parameter measurements do not affect the accuracy of the noise figure. In particular, for high-gain or low-power amplifiers, one must be sure the input power does not compress the amplifier under test. For most devices, this means using a power at least 20 dB below the 1 dB compression point. In setting the power, one must be sure that the calibration for S-parameter has low trace noise, so using techniques described in Chapter 6 for high-gain amplifiers is recommended. For example, for high-gain amplifiers, one must be sure to set the port 2 power higher than the power 1 power, by approximately the gain of the amplifier.

For the noise-source specific calibration of the VNA receiver, one must be sure that the ENR of the noise source is sufficiently high to overcome the noise figure of the VNA and any

loss between the DUT reference plane and the VNA. For most cases, using a noise source with *High* ENR will yield the best results. This is exactly opposite of the recommendation for Y-factor. In Y-factor, the difference in power between the hot and cold levels can cause some error due to the linearity of the noise receiver. In the cold-source method, the excess noise from the DUT will be measured in one region of the receiver linearity curve. For optimal results, the ENR from the noise source should match the DUT excess noise to reference the same point. And unlike Y-factor, the cold-source method measures the output match of the noise source and corrects for its effect so that there is no reason to use a low ENR source to improve its match.

For very high-gain devices, with excess noise exceeding 60 dB, or very wide bandwidth devices, the noise from the DUT can overwhelm the noise receiver. In such a case, an attenuator on the output of the DUT can be used to reduce the overall gain. Cold-source systems typically support de-embedding of loss from either the input or the output of the test system, provided the S-parameters of the loss are known.

If the test system has high loss at port 2, such as testing a high-power DUT or testing at the end of a very long cable, the loss from the reference plane to the VNA will be so great that the ENR from the noise source used for calibration will be absorbed in the loss, and there will be no excess noise left at the end of the loss. In such cases, the noise source should be connected directly to the VNA for calibration and the external path loss separately measured and de-embedded after calibration.

Alternatively, new methods have been introduced that allow using a power meter to calibrate the VNA receiver for gain-bandwidth product. This new method is not sensitive to loss at port 2. As part of the advanced power calibration techniques, loss either before or after the reference port planes can be removed using a multistep calibration and de-embedding process, which is built into the calibration method, making it suitable for on-waver, in-fixture, and waveguide calibrations.

The cold-source method (with scalar correction) relies on the VNA source providing a good matched Z_0 (50Ω) reference impedance to get good noise figure results. Adding an attenuator to the end of the port 1 test cable can substantially improve the match, allowing very good noise-figure measurements. Unlike the Y-factor method, the calibration process naturally removes the effect of the attenuator loss, and no special care is needed. A value of 6–10 dB is very reasonable; higher loss attenuators give better match but make the S_{11} more susceptible to cable drift and degrade the raw directivity.

For the full vector-corrected noise-figure measurements, having a good constellation of impedance states is important. Often the impedance tuner is set behind the test port coupler, so if there is significant loss in the port 1 cable, the impedance states will collapse to the center of the chart and not provide good separation for extraction of the noise parameters. However, the tuner or Ecal may be placed directly in front of the DUT or anywhere in the port 1 path. The only drawback to this is that the loss and linearity of the Ecal or tuner may degrade the maximum power or create distortion if other measurements such as gain compression or IMD are desired for the same setup. These issues are avoided in at least one VNA implementation as the manufacturer provides a bypass switch to remove the tuner from the path during non-noise-figure measurements.

9.6 Summary: Noise and Noise-Figure Measurements

Noise and noise-figure attributes are fundamental to all communications systems, and the measurement of these attributes is one of the more confusing areas of RF and microwave engineering. While a fundamental issue, the fact that when one is measuring noise, it means measuring an attribute that does not have a simple, single value, but rather one that can be expressed only in a statistical way, and we can never say what a true-value is, but only what the expected value (in a mean or averaged sense) is with some additional statement of the variance or standard deviation of the result. In short, noise is noisy.

Further, in many cases, the test instrumentation may produce a greater noise signal than the attribute trying to be measured, and so some correction is needed to accommodate the error associated with the instrumentation.

This chapter has attempted to describe noise power and noise-figure measurements in a correct mathematical way, while maintaining a practical engineering approach.

The author would like to acknowledge that much of this content is informed by the compilation of the work of Roger Pollard, PhD, as well as the implementation of Jim Kerr, Lynn Rymes, Bob Shoulders, and Ken Wong, all of Keysight Technologies.

References

- Dunsmore, J. (2017). Noise figure verification of Y-factor and cold-source methods. In: *2017 International Conference on Noise and Fluctuations (ICNF)*, 1–4. Vilnius: <https://doi.org/10.1109/ICNF.2017.7985945>.
- Keysight Technologies (2018) Fundamentals of RF and Microwave Noise Figure Measurements, Available at: <https://literature.cdn.keysight.com/litweb/pdf/5952-8255E.pdf> (Accessed 3 Nov 2018)
- Randa, J., Dunsmore, J., Gu, D. et al. (2011). Verification of noise-parameter measurements and uncertainties. *IEEE Transactions on Instrumentation and Measurement* **60** (11): 3685–3693. <https://doi.org/10.1109/TIM.2011.2138270>.

10

VNA Balanced Measurements

10.1 Differential and Balanced S-Parameters

Traditionally, common RF structures involved single-input and single-output devices, with a common ground reference. But with the advent of advanced microwave monolithic integrated circuits (MMICs) and the move to using CMOS technology to higher frequencies, more and more RF circuits are being designed using differential devices. Even computer backplanes and clock rates are reaching such speeds that RF and microwave considerations must be taken into account. Because of this, differential or balanced S-parameters have become an important area of RF and microwave research and application. Fortunately, the theoretical foundation for differential S-parameter theory has been well founded, and a consistent definition has been accepted (Bockelman and Eisenstadt 1995).

Consider the network described in Figure 10.1. It is a 4-port network, with an associated 16-term S-parameter matrix. The input and output waves are defined by the matrix formulation

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} & S_{13} & S_{14} \\ S_{21} & S_{22} & S_{23} & S_{24} \\ S_{31} & S_{32} & S_{33} & S_{34} \\ S_{41} & S_{42} & S_{43} & S_{44} \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} \quad (10.1)$$

But for a differential amplifier, the input differential port is considered to comprise both ports 1 and 3, and the output differential-port is comprised of both ports 2 and 4. Note that port numbering here is arbitrary, and another choice for defining the input and output differential ports is ports 1 and 2 at the input and ports 3 and 4 at the output. The original reference paper used this definition, but conventional test equipment often puts the input ports as ports 1 and 3, and the output ports as ports 2 and 4, which has become commonly accepted. The literature is about equally split on the definition, so following common industry practice, the odd ports (1 and 3) will be defined here as the inputs, and the even ports (2 and 4) will be defined as the outputs. Also, it must be recognized that whenever a port pair is described as a differential port and there exists a common ground node, then there can also exist a common mode signal on the port pair. Thus, all 4-port devices that also have a common ground node

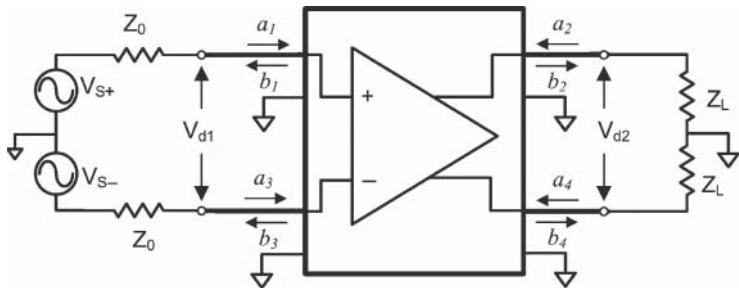


Figure 10.1 A 4-port network used as a balanced amplifier.

must be properly described as having both a differential mode and a common mode on each port pair, which is commonly called *mixed mode*.

From this description, it must be recognized that a new definition for differential input waves and differential output waves must be created. In this case, one can define the differential incident (or forward) waves and scattered (or reflected) waves from those shown in Figure 10.1 based on the differential input voltage and differential input current

$$V_d^F = \frac{1}{2}(V_d + I_d Z_d), \quad V_d^R = \frac{1}{2}(V_d - I_d Z_d) \quad (10.2)$$

where V_d and I_d are defined as the difference in input voltage and the difference of input currents between the ports 1 and 3, as shown in the figure. If the node voltages are V_1 , V_3 and the node currents are I_1 , I_2 then the differential input voltage and current can be defined as

$$V_d = V_1 - V_3, \quad I_d = \frac{1}{2}(I_1 - I_3) \quad (10.3)$$

The definition of V_d is intuitive, but the factor of $\frac{1}{2}$ in the definition of I_d is a little less intuitive and needs further explanation. In a single-ended ground-referenced circuit, the input voltage is the difference between the voltage on the node and ground, and the current is the current flowing into just the input node; the ground current isn't considered when computing the input current. Similarly, the differential input voltage is the difference between the two inputs, with port 3 playing the role of ground; if the circuit were purely differential, the current into port 1 would equal the current out of port 3. However, the internal connections of the network can allow different current to enter port 1 and exit port 3; for example, port 1 could be connected to ground through a resistor and port 3 could be open circuit. In such a case, there would be no "differential ground" current flowing out port 3, so how does one compute a differential current? This is done by presuming an averaging of the current flowing into port 1 and out of port 3 as the differential current. The "leftover" is due to common mode current flowing into the network and out of the actual ground node. Thus, one-half of the current flowing into port 1 is differential, and one-half is common; one-half the current flowing out of port 3 is differential, and one-half is common. Therefore, the average differential current is one-half of port 1 current plus one-half of port 3 current. Finally, recognize that differential

current out of port 3 is the same as $-I_3$, so the average differential current should be described as in Eq. (10.3).

There is a corresponding definition for common mode voltage and current

$$I_c = I_1 + I_3, \quad V_c = \frac{1}{2}(V_1 + V_3) \quad (10.4)$$

where the common-mode current into the network is the sum of the currents going into ports 1 and 3 (the return current flows through the ground node), and the common mode voltage into the network is the average of the voltage on ports 1 and 3. Thus, take one-half of each of the voltages, in much the same manner as the differential current was defined previously. The common mode forward and reverse voltages can be similarly defined to be

$$V_c^F = \frac{1}{2}(V_c + I_c Z_c), \quad V_c^R = \frac{1}{2}(V_c - I_c Z_c) \quad (10.5)$$

The final aspect of the definition of differential and common mode circuits is in the definition of the differential-mode reference impedance as it appears in Eq. (10.2), and the common mode reference impedance as it appears in Eq. (10.5). The impedance of a single-ended termination is simply the node voltage divided by the current into the node. A differential termination can be defined to say it is the differential voltage divided by the differential current into the node, whereas the common mode impedance is the common mode voltage divided by the common mode current, and thus one finds

$$Z_d = 2Z_0, \quad Z_c = \frac{Z_0}{2} \quad (10.6)$$

if the source impedance for both sources at the port pair is Z_0 , and the sources are independent, that is, there is no coupling elements between the sources.

The last thing needed now is a definition of differential and common mode S-parameters, and they are derived from our earlier definitions, and a new definition of mixed mode a and b waves

$$a_d = \frac{V_d^F}{\sqrt{Z_d}}, \quad b_d = \frac{V_d^R}{\sqrt{Z_d}}, \quad a_c = \frac{V_c^F}{\sqrt{Z_c}}, \quad b_c = \frac{V_c^R}{\sqrt{Z_c}} \quad (10.7)$$

This is presuming Z_d and Z_c are real. From this, the S-parameters can be defined as

$$\begin{bmatrix} b_{d1} \\ b_{d2} \end{bmatrix} = \begin{bmatrix} S_{dd11} & S_{dd12} \\ S_{dd21} & S_{dd22} \end{bmatrix} \cdot \begin{bmatrix} a_{d1} \\ a_{d2} \end{bmatrix}, \quad \begin{bmatrix} b_{c1} \\ b_{c2} \end{bmatrix} = \begin{bmatrix} S_{cc11} & S_{cc12} \\ S_{cc21} & S_{cc22} \end{bmatrix} \cdot \begin{bmatrix} a_{c1} \\ a_{c2} \end{bmatrix} \quad (10.8)$$

Here, there is also a port numbering associated with the mixed-mode waves, where port $d1$ and $c1$ are the differential and common input ports, respectively, comprised of ports 1 and 3, and $d2$ and $c2$ the differential and common output ports, respectively, comprised of ports 2 and 4. In addition to the differential-mode S-parameters and the common-mode S-parameters of (10.8), it is entirely reasonable to consider a network that is driven in one mode but outputs a signal in another mode. The cross-mode parameters come in two versions: drive a differential

incident wave and measure a common-mode scattered wave; or drive a common-mode incident wave and measure a differential-mode scattered wave. They are defined as

$$\begin{bmatrix} b_{c1} \\ b_{c2} \end{bmatrix} = \begin{bmatrix} S_{cd11} & S_{cd12} \\ S_{cd21} & S_{cd22} \end{bmatrix} \cdot \begin{bmatrix} a_{d1} \\ a_{d2} \end{bmatrix}, \quad \begin{bmatrix} b_{d1} \\ b_{d2} \end{bmatrix} = \begin{bmatrix} S_{dc11} & S_{dc12} \\ S_{dc21} & S_{dc22} \end{bmatrix} \cdot \begin{bmatrix} a_{c1} \\ a_{c2} \end{bmatrix} \quad (10.9)$$

The 16 mixed-mode S-parameters are all needed to fully describe a 4-port network, and are often shown in matrix form as

$$\begin{bmatrix} b_{d1} \\ b_{d2} \\ b_{c1} \\ b_{c2} \end{bmatrix} = \begin{bmatrix} S_{dd11} & S_{dd12} & S_{dc11} & S_{dc12} \\ S_{dd21} & S_{dd22} & S_{dc21} & S_{dc22} \\ S_{cd11} & S_{cd12} & S_{cc11} & S_{cc12} \\ S_{cd21} & S_{cd22} & S_{cc21} & S_{cc22} \end{bmatrix} \cdot \begin{bmatrix} a_{d1} \\ a_{c2} \\ a_{c1} \\ a_{c2} \end{bmatrix} \quad (10.10)$$

With this, the definition of mixed more S-parameters is nearly complete, and in fact many references stop here. However, it would be convenient to be able to describe the mixed-mode S-parameters in terms of the single-ended S-parameters. One can use the previously defined waves to make the following observations; from Eqs. (1.8) and (1.9), and (10.2) and (10.5), one can derive

$$a_{d1} = \frac{(a_1 - a_3)}{\sqrt{2}}, \quad a_{c1} = \frac{(a_1 + a_3)}{\sqrt{2}}, \quad b_{d1} = \frac{(b_1 - b_3)}{\sqrt{2}}, \quad b_{c1} = \frac{(b_1 + b_3)}{\sqrt{2}} \quad (10.11)$$

and

$$S_{dd11} = \left. \frac{b_{d1}}{a_{d1}} \right|_{a_{c1}=a_{d2}=a_{c2}=0} \quad (10.12)$$

And recognizing that $a_{c1} = 0$ implies that $a_3 = -a_1$ so that $a_{d1} = 2a_1$, then

$$S_{dd11} = \left(\frac{b_1 - b_3}{a_1 - a_3} \right) = \left(\frac{b_1 - b_3}{a_1 - (-a_1)} \right) = \left(\frac{b_1 - b_3}{2a_1} \right) \quad (10.13)$$

This is combined with a version of Eq. (1.17) considered for ports 1 and 3 to obtain

$$\begin{aligned} S_{dd11} &= \left(\frac{b_1 - b_3}{2a_1} \right) = \left(\frac{[S_{11}a_1 + S_{13}a_3] - [S_{31}a_1 + S_{33}a_3]}{2a_1} \right) \\ &= \left(\frac{[S_{11}a_1 - S_{13}a_1] - [S_{31}a_1 - S_{33}a_1]}{2a_1} \right) \end{aligned} \quad (10.14)$$

Factoring out the common a_1 yields

$$S_{dd11} = \frac{1}{2}(S_{11} - S_{13} - S_{31} + S_{33}) \quad (10.15)$$

A similar computation can be performed for each of the mixed-mode parameters to find their equivalence in the single-ended S-parameters. S_{dd21} is the most important attribute of a balanced amplifier, and one can compute its value as a function of single-ended parameters by recognizing that

$$S_{dd21} = \left(\frac{b_2 - b_4}{a_1 - a_3} \right) = \left(\frac{b_2 - b_4}{a_1 - (-a_1)} \right) = \left(\frac{b_2 - b_4}{2a_1} \right) \quad (10.16)$$

And following the same logic as Eq. (10.14),

$$\begin{aligned} S_{dd21} &= \left(\frac{b_2 - b_4}{2a_1} \right) = \left(\frac{[S_{21}a_1 + S_{23}a_3] - [S_{41}a_1 + S_{43}a_3]}{2a_1} \right) \\ &= \left(\frac{[S_{21}a_1 - S_{23}a_1] - [S_{41}a_1 - S_{43}a_1]}{2a_1} \right) \end{aligned} \quad (10.17)$$

which, when factored, yields

$$S_{dd21} = \frac{1}{2}(S_{21} - S_{23} - S_{41} + S_{43}) \quad (10.18)$$

The mixed-mode conversions are sometimes expressed in the form of a matrix transformation, where

$$\begin{aligned} [S_{MM}] &= [M] \cdot [S] \cdot [M]^{-1}, \\ [M] &= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}, \quad [M]^{-1} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix}, \end{aligned} \quad (10.19)$$

A complete list is shown in Table 10.1.

For the most common of the mixed mode S-parameters, the interpretation is straightforward: S_{dd21} is the differential gain, S_{cc21} is the common mode gain, but for the other, cross-mode parameters, the practical meaning is not so obvious. When a device is driven differentially, it is essentially self-shielded, but if it produces common mode signals at its output terminals, it implies significant ground current is flowing through the common ground. This condition may lead to radiated emissions from the device. Thus, the S_{cd} parameters are

Table 10.1 Mixed-mode S-parameters expressed as single-ended S-parameters

Differential-mode parameter	Cross-mode parameter: common to differential	Common-mode parameter: differential to common	Common-mode parameter
S_{dd11}	$(S_{11}-S_{13}-S_{31}+S_{33})/2$	S_{dc11}	$(S_{11}+S_{13}-S_{31}-S_{33})/2$
S_{dd12}	$(S_{12}-S_{14}-S_{32}+S_{34})/2$	S_{dc12}	$(S_{12}-S_{32}+S_{14}-S_{34})/2$
S_{dd21}	$(S_{21}-S_{41}-S_{23}+S_{43})/2$	S_{dc21}	$(S_{21}-S_{41}+S_{23}-S_{43})/2$
S_{dd22}	$(S_{22}-S_{42}-S_{24}+S_{44})/2$	S_{dc22}	$(S_{22}-S_{42}+S_{24}-S_{44})/2$
Cross-mode parameter: differential to common		Common-mode parameter	
S_{cd11}	$(S_{11}+S_{31}-S_{13}-S_{33})/2$	S_{cc11}	$(S_{11}+S_{31}+S_{13}+S_{33})/2$
S_{cd12}	$(S_{12}+S_{32}-S_{14}-S_{34})/2$	S_{cc12}	$(S_{12}+S_{32}+S_{14}+S_{34})/2$
S_{cd21}	$(S_{21}+S_{41}-S_{23}-S_{43})/2$	S_{cc21}	$(S_{21}+S_{41}+S_{23}+S_{43})/2$
S_{cd22}	$(S_{22}+S_{42}-S_{24}-S_{44})/2$	S_{cc22}	$(S_{22}+S_{42}+S_{24}+S_{44})/2$

sometimes associated with a measure of potential radiated emissions of the device. Similarly, if a device that is intended to have differential outputs produces such an output signal when there is a common-mode signal applied, then it implies it is susceptible to ground currents, and the S_{dc} parameters are associated with a measure of its potential external-signal immunity.

10.2 3-Port Balanced Devices

The mixed-mode parameters may also be defined for a 3-port network, comprising one single-ended port and one mixed-mode port, which is usually defined as the differential or balanced port. But again, it must be recognized that for a device referenced to ground, a balanced port always has the possibility to support a common mode signal as well. A schematic of such a device is shown in Figure 10.2.

The mixed-mode parameters for a 3-port network are defined by

$$\begin{bmatrix} b_s \\ b_d \\ b_c \end{bmatrix} = \begin{bmatrix} S_{ss} & S_{sd} & S_{sc} \\ S_{ds} & S_{dd} & S_{dc} \\ S_{cs} & S_{cd} & S_{cc} \end{bmatrix} \cdot \begin{bmatrix} a_s \\ a_d \\ a_c \end{bmatrix} \quad (10.20)$$

Since for the 3-port case, the ports are unambiguous, there is no need to use port numbers, but one can refer only to the three port modes; however, it is also commonly found in the literature to include the port numbers, especially if defining the 3-port, single-ended (SE) to balanced properties of a 4-port network, where it may have more than one SE input. The values of the 3-port mixed-mode parameters can also be computed in terms of the single-ended parameters.

The same definitions hold as in Eq. (10.7) for differential and common waves so that the 3-port mixed-mode parameters can be computed in terms similarly to Eq. (10.19) but in the 3-port case

$$[M] = \frac{1}{\sqrt{2}} \begin{bmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 1 & 1 \end{bmatrix}, \quad [M]^{-1} = \frac{1}{\sqrt{2}} \begin{bmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 1 \\ 0 & -1 & 1 \end{bmatrix} \quad (10.21)$$

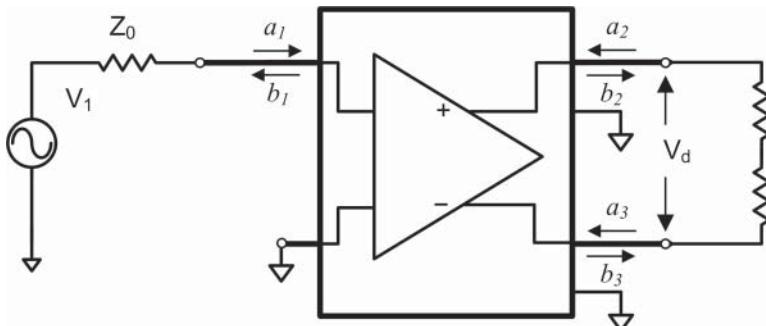


Figure 10.2 A 3-port device used as a single-ended to balanced device.

which yields the following conversions for mixed-mode parameters

$$\begin{bmatrix} S_{ss} & S_{sd} & S_{sc} \\ S_{ds} & S_{dd} & S_{dc} \\ S_{cs} & S_{cd} & S_{cc} \end{bmatrix} = \begin{bmatrix} S_{11} & \frac{1}{\sqrt{2}}(S_{12} - S_{13}) & \frac{1}{\sqrt{2}}(S_{12} + S_{13}) \\ \frac{1}{\sqrt{2}}(S_{21} - S_{31}) & \frac{1}{2}(S_{22} - S_{23} - S_{32} + S_{33}) & \frac{1}{2}(S_{22} + S_{23} - S_{32} - S_{33}) \\ \frac{1}{\sqrt{2}}(S_{21} + S_{31}) & \frac{1}{2}(S_{22} - S_{23} + S_{32} - S_{33}) & \frac{1}{2}(S_{22} + S_{23} + S_{32} + S_{33}) \end{bmatrix} \quad (10.22)$$

The most common use of a single-ended to differential device is as a balun (which comes from **BAL**anced-to-**UN**balanced transformer), used to drive a differential device from a single-ended measurement instrument. In the past, when vector network analyzers (VNAs) were limited to only two ports, baluns were used extensively, and sometimes incorrectly, in making differential measurements. With that advent of 4-port VNAs, the need for baluns in characterization of linear, passive measurements was essentially eliminated. However, even today they remain a key component for testing other more complicated characteristics such as compression, distortion, and noise figure.

The concept of mixed-mode parameters is not limited to 3- or 4-port devices. In fact, they can be extended to an arbitrary number of ports. The convenience of mixed-mode parameters as developed is that the familiar formulations for computing with single-ended S-parameter applies just as well to mixed-mode S-parameters including concepts for maximum power transfer, stability, and the effects of cascading networks and de-embedding networks.

10.3 Measurement Examples for Mixed-Mode Devices

10.3.1 Passive Differential Devices: Balanced Transmission Lines

Perhaps the simplest differential device is a balanced transmission line. In fact, one of the oldest styles of transmission lines, called *twin-lead*, used extensively with early television receivers, was essentially a balanced transmission line. The modern implementation of balanced or differential transmission lines is found in the high-speed low-voltage differential-signaling (LVDS) communication paths found in high-speed digital and communications circuits. These lines now tend to be dual PC board traces that inter-connect pairs of differential drivers to differential receivers. One interesting aspect of these signaling lines is that making them differential can often lead to much better signal-to-noise performance, reduced interference, and improved frequency response compared to single-ended transmission lines. Of this, the self-shielding aspect may be most important.

Figure 10.3 shows an example of a test board used for evaluating PC board transmission lines characteristics. With current speeds of digital devices approaching 3 GHz clock rates and data transmission at 10 Gbps, the integrity of these signaling lines as they pass between board layers, around PC board vias, and transition through connectors is becoming critical to understand. This is often called the *physical layer* of a communications stack, which includes modulation, formatting, paging, and framing all of which are higher orders of processing the raw signals to improve reliability of signaling. This area of study has become known as



Figure 10.3 A PC test board for characterizing differential lines.

signal integrity and is a separate but closely related field to that of the RF and microwave communications engineer.

This test board has a design artifact where the traces are narrowed in one region, perhaps to represent an area where a smaller footprint on the PCB is required. A full 4-port S-parameter characterization was performed, with the 16 S-parameters shown in Figure 10.4. In the test boards, the input and output lines are narrowed as the line form a differential pair so that the differential-mode impedance is maintained at a constant value.

While very busy, this display is useful in noting that the one may quickly see if there are any unusual aspects of the measurement. One apparent issue is the high loss in the transmission paths, S_{21} and S_{43} , on trace 5 (high-lighted) and 15. Further, there is very strong coupling between the lines as evinced by S_{41} and S_{32} (traces 13 and 10). From a cursory look at the single-ended transmission lines, one might conclude that these traces can be used only up to about 3 GHz, where S_{21} goes below -3 dB.

However, when the mixed-mode S-parameters are examined, a different story unfolds. Figure 10.5 shows a comparison of S_{dd21} with the two of the transmission parameters associated with port 2, S_{21} and S_{23} . From this result, one can see that the differential transmission is very good, less than -3 dB, beyond 5 GHz. In fact, because of the particular phasing of the signals, the coupling of S_{23} exactly compensates for the loss of S_{21} so that the differential mode of transmission sees much less loss than the single-ended mode.

There is still some ripple and loss in the differential S_{dd21} (likely due to the stepped impedance test feature), and evaluation of other differential parameters can shed light on the attributes of the differential line that contribute to the loss in S_{dd21} .

Figure 10.6 shows a display of all the 16 mixed-mode parameters, with each mode in its own window. The cross-mode terms show very little signal, indicating that the lines are quite well balanced, which mean they are well shielded, even if they do have significant mismatch in the differential mode as exhibited by the S_{dd11} , and in the common mode, as shown by the S_{cc11} traces.

Mixed-mode parameters allow analysis in exactly the same manner that single-ended S-parameters allow, so one may investigate attributes of differential lines using all the tools of S-parameter analysis, including time-domain transforms.

In Figure 10.7, the upper window shows S_{dd11} and S_{cc11} , and the lower window shows the time-domain response of each. From the time-domain response, it is very clear that the ripple in the transmission, and the mismatch in the reflection is due to the discrete change in line impedance that occurs at about 250 mm down the line (as indicated by marker 2). This plot illustrates several interesting points. First, note that the marker readout displays have been

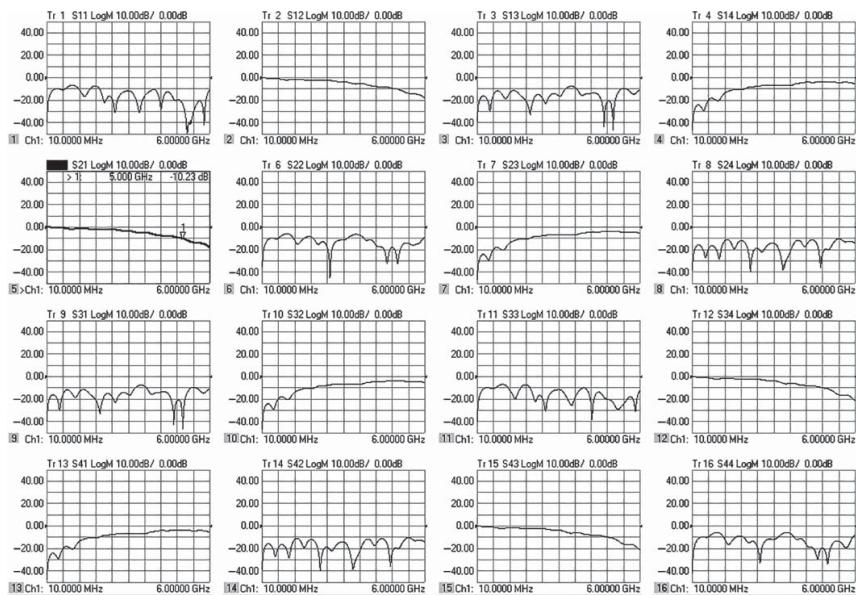


Figure 10.4 16 S-parameters of a balanced transmission line.

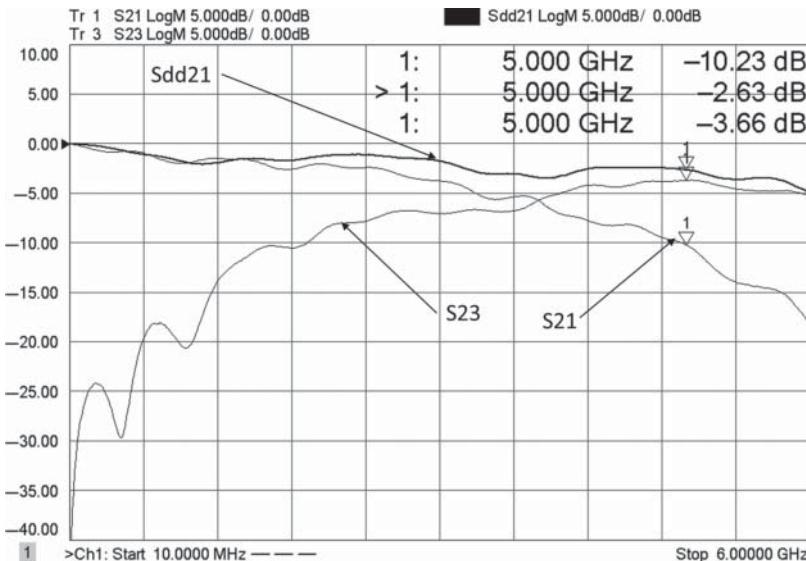


Figure 10.5 Comparing S_{21} , S_{23} , and S_{dd21} .

changed from default to read the equivalent impedance (essentially a Smith chart marker) and is a convenient marker feature found on some modern VNAs. This provides for directly reading the impedance of the lines. One can see that at 0 time (the first graticule line), the differential impedance is nearly matched, but the common mode impedance is *not* matched. This also implies that the single-ended result will not be matched as well. Because the lines are tightly coupled (the S_{23} trace of Figure 10.5 demonstrates this), the SE line impedance must be modified to maintain a constant differential impedance. From Figure 10.3, one can tell that the line widths have been narrowed where the differential portion starts (the PC board was measured with differential port 1 on the right).

The step in impedance due to the thin sections of line are clearly shown in both time-domain reflectometer (TDR) plots, but it is interesting to see that the step is slightly delayed in the common-mode transmission, when compared to the differential-mode transmission. That is because the field distribution in the differential line has substantial fields in the air above the lines. The common-mode transmission, from line to ground, has the fields substantially more in the PC board, with its higher dielectric constant. Thus, the velocity factor in the common-mode is slower than in the differential mode, and this is reflected in the time-domain response.

For linear, passive devices, making balanced or differential measurements is very straightforward using the newest 4-port VNAs. All of the fixturing, de-embedding, and impedance transformation that is possible with standard S-parameters is also available for mixed-mode parameters. Since devices are linear, making single-ended fully corrected S-parameter measurements and converting to mixed-mode parameters gives exactly correct results. Whether this also holds true for active devices has been a topic of a great deal of speculation, but recently new studies have set very good guidelines for applying mixed-mode analysis to active devices, as discussed in the next section.

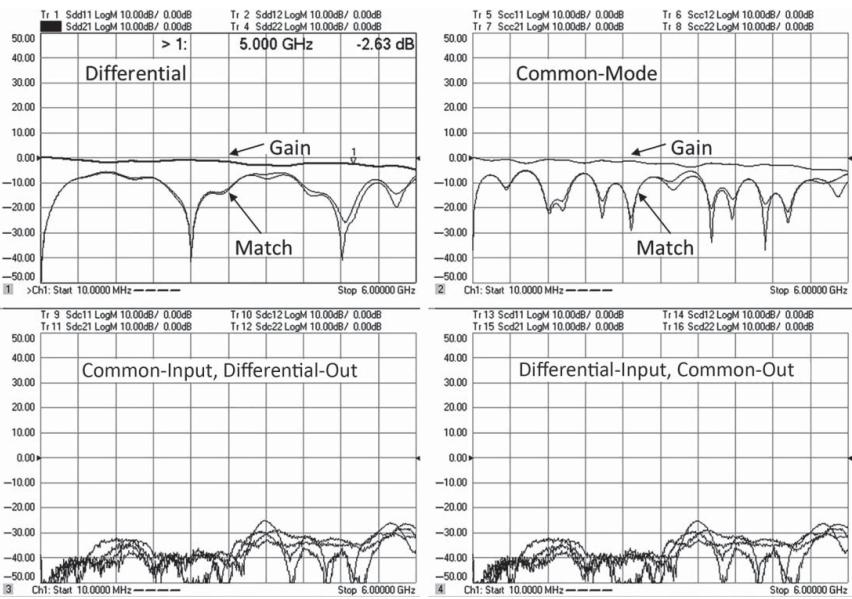


Figure 10.6 Display of all 16 mixed-mode parameters, each mode is displayed in its own window.

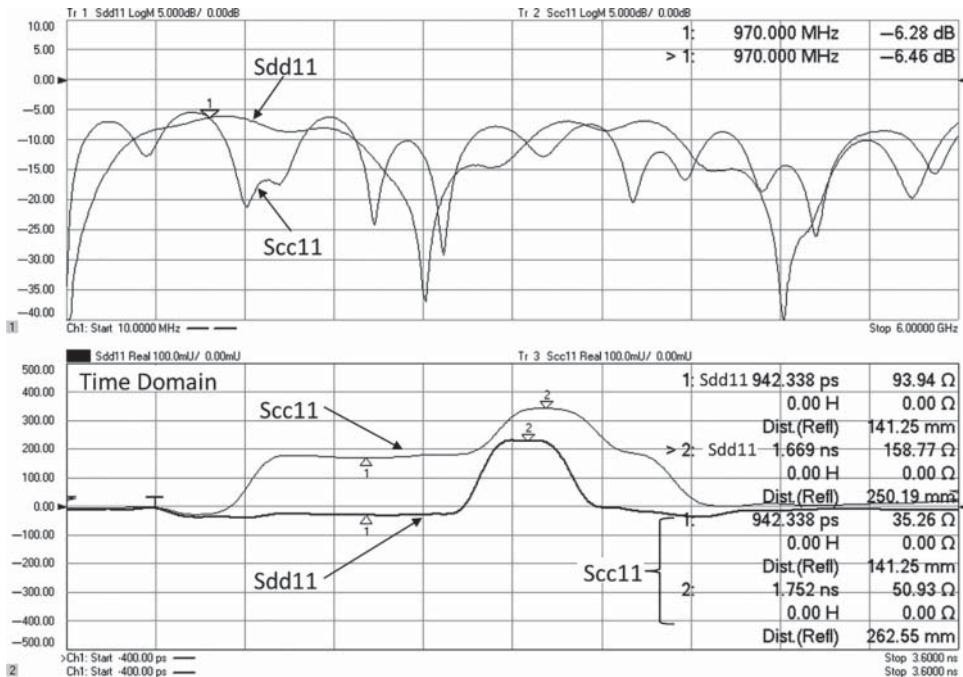


Figure 10.7 Frequency domain (upper); time domain of Sdd11 and Scc11 (lower).

10.3.2 Differential Amplifier Measurements

Differential amplifiers are common in low-frequency electronics; known as *operational-amplifiers* or *op-amps*, they have extremely high differential gain to a single-ended output, and they utilize feedback from the output to the input to perform many useful functions. However, their performance characteristics and definitions are almost entirely *unrelated* to the RF and microwave usage of the term *differential amplifier*. For most RF and microwave work, a differential amplifier refers to one of moderate gain (~ 20 dB) with balanced input and output ports, just as shown in Figure 10.1.

As with the amplifier tests described in Chapter 6, all manner of measurements desired for a single-ended amplifier are also desired for a differential amplifier. First and foremost are its linear gain attributes.

Figure 10.8 shows the 4-port single-ended measurement of one kind of differential amplifier. The input and output matches show that it is tuned to about 1 GHz. The forward gain parameters are interesting in that one of them, S_{21} (called out in the upper-right window) has substantially less gain than the others at only about 1 dB. S_{41} also has somewhat lower gain, at just 5 dB, where S_{23} and S_{43} have gain at about 8 dB. All the reverse isolation paths have reasonable isolation. This amplifier is a prototype low-noise limiting amplifier, and its behavior is typical of what one might find in the early stages of a design project. Because it is not completely symmetrical, many assumptions on differential behavior may not hold, and full differential analysis (also known as true-mode measurements) may be required. There is substantial coupling between ports 1 and 3, indicated by the high level of S_{31} and S_{13} .

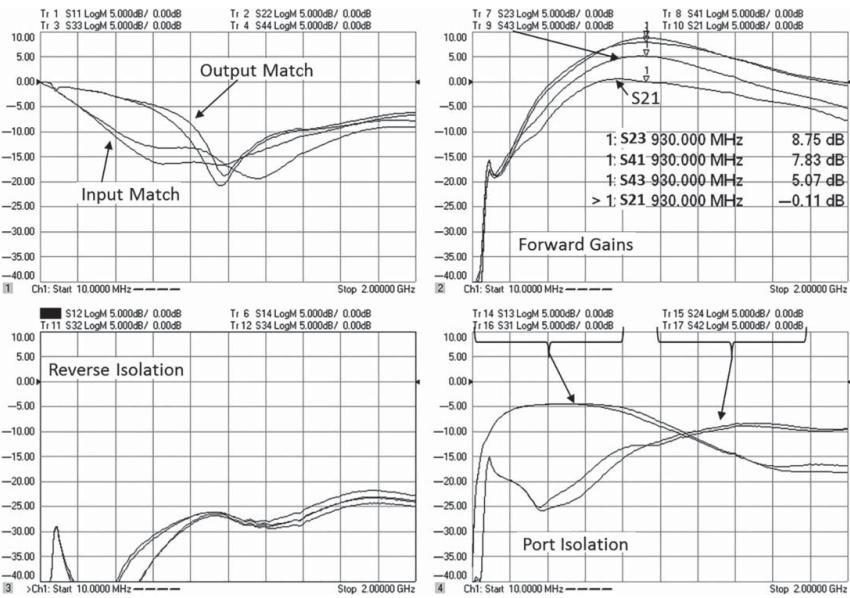


Figure 10.8 4-port single-ended S-parameters of a differential device.

In contrast, the output ports have more isolation at low frequencies, but have some coupling at high frequencies. Coupling between the ports is an indication that the common-mode match may not be very good. For this example, the amplifier does not follow the expected behavior of a true-differential amplifier as each of the forward gains are not identical. Later in this chapter, an example of a normal differential amplifier will also be shown, which is considerably less interesting.

The same amplifier is evaluated for its mixed-mode parameters, in Figure 10.9. Note that the differential gain is quite smooth and broad, even though the S_{21} gain does not match the S_{23} gain and peaks at a different frequency. However, likely due to the low S_{21} gain, the common-mode gain is greater than one might expect for a differential amplifier, and this amplifier has relatively low common-mode rejection. Common-mode rejection for a differential amplifier is often quoted as

$$CMRR = \frac{S_{dd21}}{S_{cc21}} \quad (10.23)$$

This definition came from the definition for low-frequency op-amps where, which were essentially as single-ended output. But this definition is of little use in radio frequency (RF) balanced systems, because the outputs are also differential. In an op-amp, if a large common-mode input-signal produced some output signal, then it would flow to the next section and be detected as the output of an op-amp is typically single-ended. However, in a full differential amplifier, a common-mode signal at the output does *not* propagate through the system, because the common-mode to differential-mode gain of the following sections (S_{dc21}) are usually very low.

In fact, when referring to the common-mode rejection of an RF differential-amplifier, one would be most concerned with S_{dc21} , as it is a measure of the effect on the desired output signal (the differential output voltage) due to a large common-mode input signal. Thus, for full differential amplifiers, common mode rejection ratio (CMRR) is of very little concern, and S_{dc21} should be the proper measure of common-mode isolation. In the example of Figure 10.9, the CMRR is about 9.5 dB, but if equal common-mode and differential-mode signals were at the input, the effect on the differential output due to the common-mode input would be down by 15 dB, the value of S_{dc21} .

Other interesting attributes is that the differential reverse isolation is much higher than any of the single-ended isolation terms, indicating very good balance in the reverse isolation. For these measurements, the amplifier is driven with quite low power and is operating in the linear region, so one expects that using Eq. (10.19) and computing the mixed-model parameters from the data taken with a 4-port VNA using single-ended S-parameters would work well.

10.3.3 Differential Amplifiers and Non-linear Operation

One common assertion is that while the mixed-mode parameters of an amplifier may be computed from single-ended 4-port measurements when the amplifier is operating in a low-power linear mode, the results will not be valid when the amplifier is driven with a large-signal and operated in a non-linear mode. Rather, for non-linear characterization, such as 1 dB compression point, it is presumed that the amplifier must be driven with a true-differential signal, usually called a *true-mode drive*.

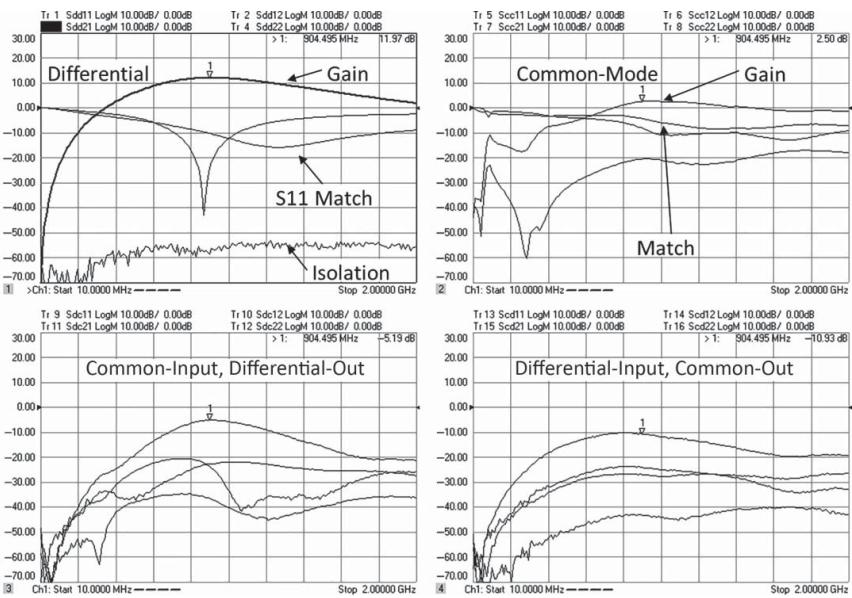


Figure 10.9 Mixed-mode parameters for a differential amplifier.

As it turns out, the truth of this assertion depends upon the configuration of the amplifier, and in most cases, single-end mixed-mode drive *does* provide a valid measurement of gain compression of a differential amplifier, provided it has normal differential behavior (Dunsmore 2003). Here, normal behavior for a differential amplifier has two main aspects, listed here:

1. A normal amplifier, differential or not, usually compresses because of limitations at the output of the amplifier. This is obviously because the RF voltages and currents are much greater at the output than at the input. Some amplifiers may have limiting elements at the input (as does the example amplifier from Figure 10.9) that clip the input signal before the output signal is limited by internal mechanisms in the amplifier, but this would be an unusual case.
2. A differential amplifier should have more differential gain than common-mode gain. In some discussions of balanced applications, two single-ended amplifiers are combined to create a so-called balanced amplifier, which will amplify a differential signal; but it will also amplify a common-mode signal with the same gain. This is not a differential amplifier in the normally understood definition.

To understand the effects non-linear response of differential amplifiers, consider what happens in when a differential amplifier is driven with a single-ended signal, as illustrated in Figure 10.10. Here, the single-ended signal can be decomposed into a differential portion and a common-mode portion. For an amplifier with substantially higher differential-mode gain than common-mode gain (as one would expect for a normal differential amplifier), the differential portion of the input signal is amplified, while the common mode is suppressed.

In this case, the single-ended signal has an input of 1 V on the positive input, and 0 V on the negative input, for a differential voltage of 1 V. The common-mode voltage in this case is $\frac{1}{2}$ volt. The output is 2 V differential, and only 0.1 V common-mode. The common mode has been suppressed by the behavior of the amplifier.

Many differential amplifiers are multistage so that there are one or more stages of differential gain at the input, often followed by single-ended output buffer stages at the output to obtain higher-power performance. An example schematic, in Figure 10.11, shows a design with two stages of differential gain (Input Diff-Pair and 2nd Diff-Pair) and one stage of emitter-follower

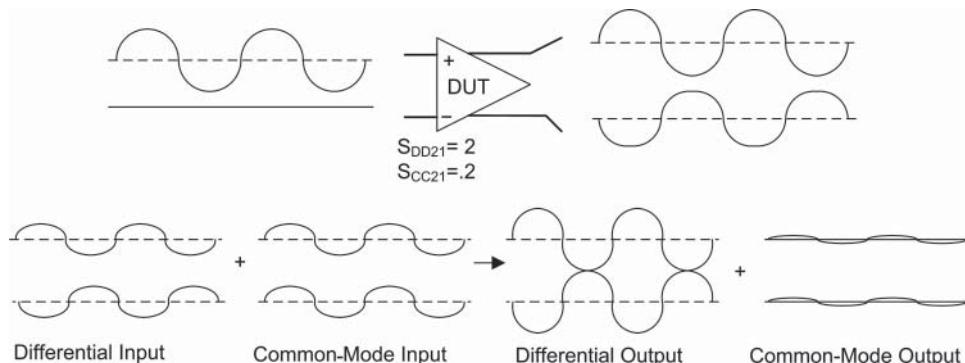


Figure 10.10 A differential amplifier driven with a single-ended signal.

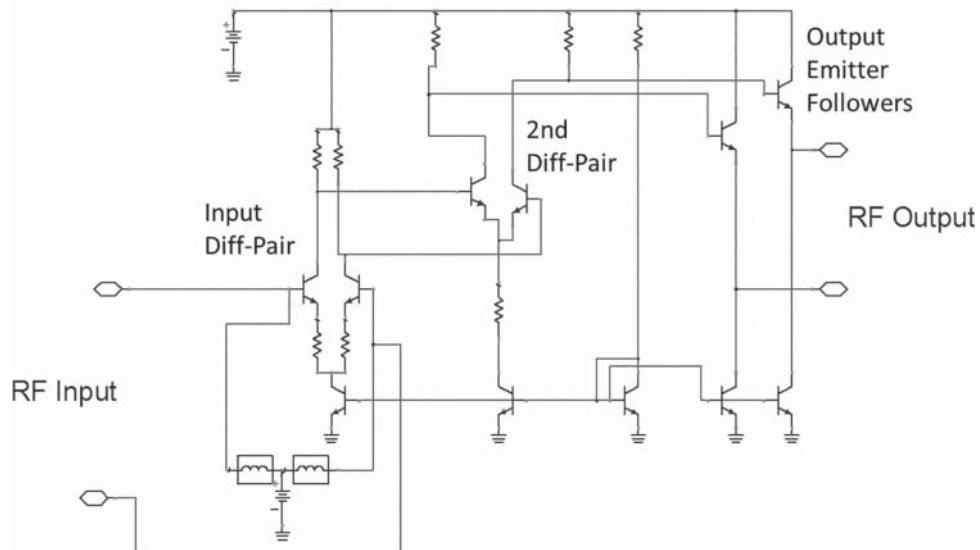


Figure 10.11 Example schematic of a differential amplifier.

output stage. It is quite typical to have at least the input stage be a differential amplifier using emitter-coupled or source-coupled pairs.

To understand the effect of non-linear operation of a differential amplifier, consider the two cases of the amplifier being driven with a differential-mode signal or a single-ended signal. The results of driving the amplifier of Figure 10.11 with these two drives, shown in Figure 10.12 (left), are shown in the plots of Figure 10.12 (right), where the output from the first stage is detected.

In this case, the differential input voltage is identical at 0.3 V. The non-linear response of the output of the first stage is nearly identical regardless of the input drive mode, as long as the differential input voltage is the same. This signal then drives the output stage, and since the drive to the next stage is the nearly identical, the output waveform should be identical as well.

Consider next a different design that does not have a differential input stage or, rather, that has a non-linear stage before the differential input stage, as illustrated in Figure 10.13. Here, the input stage is modeled as a balanced amplifier with equal common-mode and differential-mode gains, followed by a differential amplifier, and the input amplifier compresses at about 1.5 V on the positive half-cycle.

From this illustration, one can see that compression in the first stage will clip the upper portion of the waveform. This clipped waveform will then be amplified by the differential stage. The common-mode portion will be removed by common-mode suppression, leaving only a differential signal, but it is now distorted (compressed) on one half-cycle. On the other hand, if true-differential signal of the same differential amplitude were applied to this same amplifier pair, no compression will occur in the first stage (presuming the compression is due to the output signal rising above 1.5 V, in this example), so no output compression will be seen in the final signal, as shown in Figure 10.14. This is an example case where a true-mode

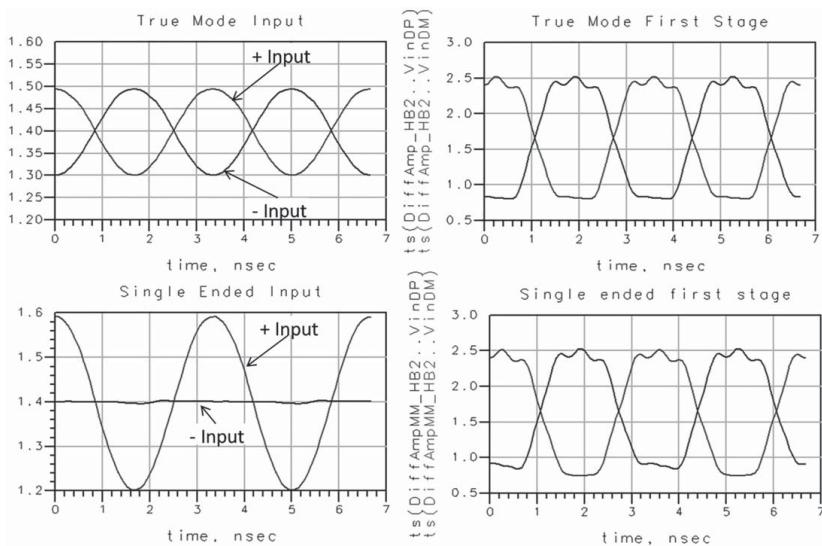


Figure 10.12 Driving an amplifier with true-mode (upper) and single-ended (lower) signals.

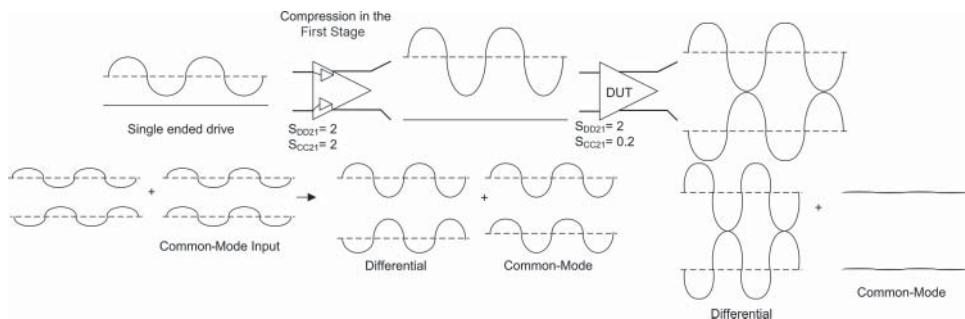


Figure 10.13 An example case where the input is non-linear and not-differential.

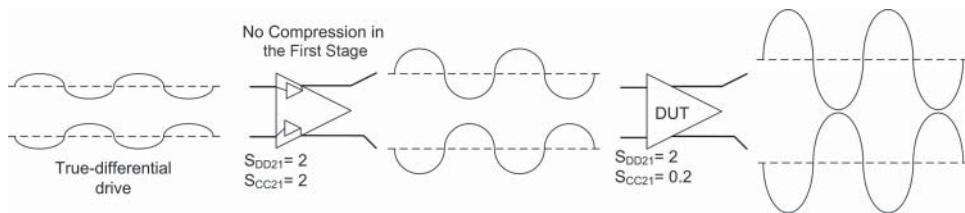


Figure 10.14 Non-linear non-differential input driven with a true-differential signal shows less compression.

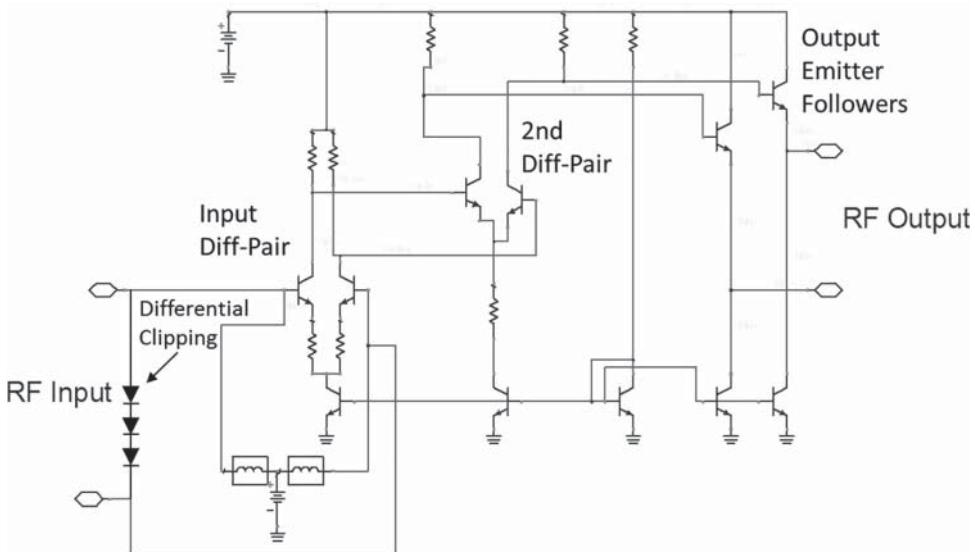


Figure 10.15 Amplifier with input clipping.

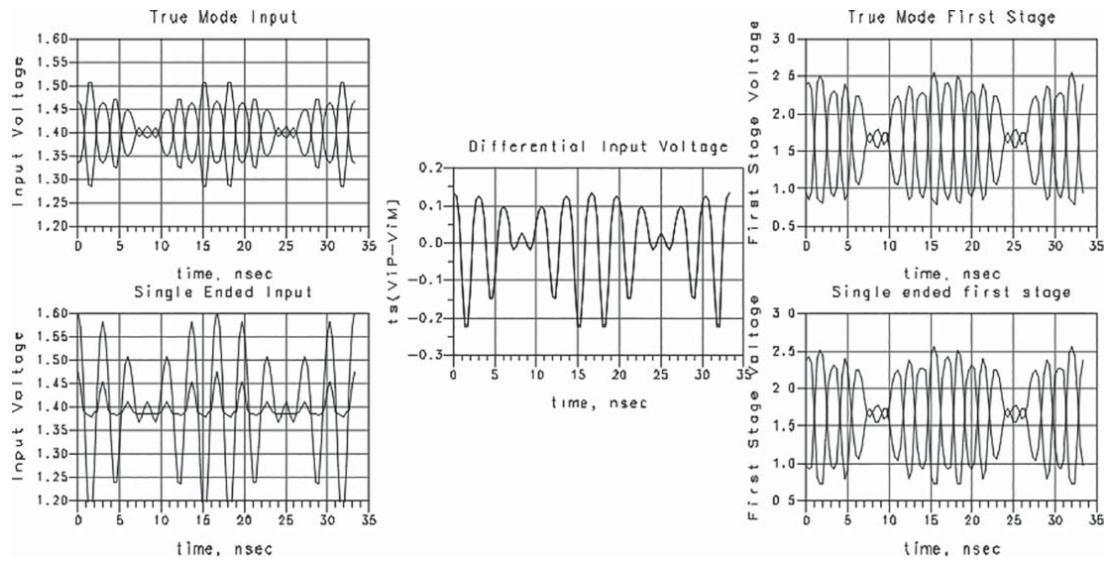


Figure 10.16 Two-tone response of a non-linear differential amplifier.

differential signal will provide a different answer for compression than will a single-ended signal that has the same differential content (Dunsmore 2004).

Finally, while this reasoning clearly applies to sinusoidal inputs used for normal gain and compression testing of amplifiers, a common question is whether the same arguments apply when the input signal has some modulation content. In fact, they do. Consider the amplifier from Figure 10.11 driven with a high-power two-tone signal. To further test the conditions on non-linearity, a RF clipping circuit (in the form of a diode string) is added across the input of the amplifier, as shown in Figure 10.15. This now has the condition of being non-linear on the input, but the non-linear mechanism is still differential.

The result of driving this amplifier with a two-tone signal is shown in Figure 10.16. The graphs on the left show the input signals, with the upper graph representing the true-differential drive and the lower graph representing the single ended drive. Interestingly, the voltage on the negative input of the SE-drive is non-zero because the clipping action of the input diodes drives some current into the negative input impedance and causes some small voltage to appear. However, the differential voltage across the amplifier input is identical, as shown in the middle graph where the differential portion of the input for each of the drive cases are super-imposed to see clearly that they are identical. Finally, the output from the first stages shows significant two-tone distortion, but it is identical for both drive cases; thus, the amplifier's second and final stage response to both signals will be the same.

This demonstrates that when an amplifier has a differential input stage before clipping or if the clipping itself is differential, then the measurement of an amplifier with a single-ended drive or a differential drive will yield the same result. The only time that true-differential-mode drive is required is when the input stage is not differential, and there is an input non-linearity that is also non-differential.

10.4 True-Mode VNA for Non-linear Testing

10.4.1 *True-Mode Instruments*

While it is true that many differential devices respond similarly to single-ended and differential signals, not all do, and one may not know if the device to be tested will behave as a normal differential device. In the past, testing a device for non-linear behavior, particularly compression, required that a balun or hybrid be placed before and after the device-under-test (DUT), so that normal, 2-port single-ended test-equipment could be used to drive a DUT with a true-differential drive.

In 2007, the first fully capable true-mode VNAs were developed based on 4-port network analyzers with dual sources (Dunsmore et al. 2008). Previous systems relied on baluns or hybrids to create the true-mode signals and relied upon matching of cables and connectors within the system to create a balanced signal. Some modern VNAs offer a means to provide independent amplitude and phase control of the two sources. In Chapter 6, these independent sources were utilized to create an active load to present to the output of a DUT. Here, the independent sources allow the system to create either a pure differential signal or a pure common-mode signal, at either the input or the output of the DUT. A representative block diagram is shown in Figure 10.17. Key to this system are two independent sources, each driving two of the ports, to allow independent signals at ports 1 and 3 and at ports 2 and 4.

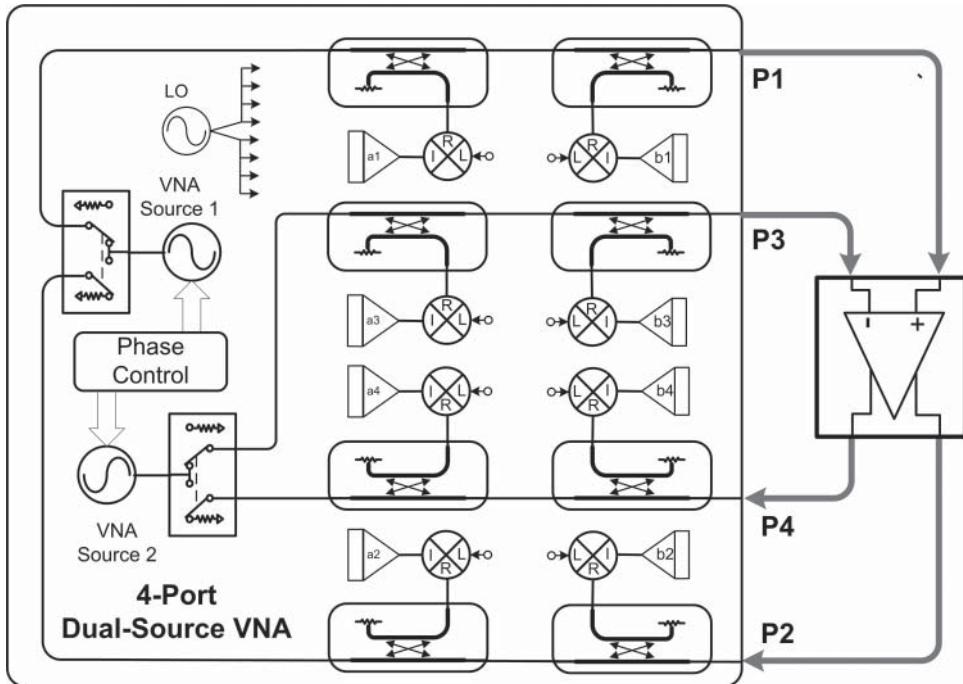


Figure 10.17 Block diagram of a 4-port dual source VNA.

As the magnitude and phase of the sources can be precisely controlled electronically, any deviation in the cables and connectors of the test system can be completely accounted for. Interestingly, the mismatch of the DUT may provide the largest source of imbalance in the drive signals.

In single-ended 2- or 4-port VNA usage, the phase of the source is of no concern because all the parameters are relative to the single-source phase. In a true-mode VNA, the relative phase of the source at port 1 and port 3 is very important, as it sets the absolute differential-amplitude. Chapter 3 described a means for characterizing a correcting the VNA receivers for absolute power. Now, these receivers can be used to characterize the power and relative phase of the sources in a VNA.

If one presumes that a full 4-port calibration can be performed on the VNA, then the corrected ratio of a_1 to a_3 should give exactly the relationship of the signals for the reference plane at the input of the DUT. For a true-mode drive, two input states are required. In the first state, differential drive, the ratio must be

$$\frac{a_1}{a_3} = 1 \cdot e^{j\pi} \quad (10.24)$$

That is, equal amplitude and 180° out of phase.

For the second state, common-mode drive or zero-phase, the ratio must be

$$\frac{a_1}{a_3} = 1 \cdot e^{j0} \quad (10.25)$$

The ratio of the $a1$ and $a3$ waves at the DUT interface is not the same as the raw-measured waves at the $a1$ and $a3$ receivers. However, the corrected waves can be computed directly as

$$\frac{a_1}{a_3} = \frac{(a_{1M}ERF_1 + b_{1M}ESF_1 - a_{1M}ESF_1 \cdot EDF_1) ETF_{31}}{(a_{3M}ERF_3 + b_{3M}ESF_3 - a_{3M}ESF_3 \cdot EDF_3) ERF_1} \quad (10.26)$$

where the subscript indicates the port for the reflection tracking (ERF), source-match (ESF), directivity (EDF), and transmission tracking (ETF). The subscript M identifies that as a measured wave. For the reverse direction, a similar formulation can be used to compute the ratio of $a2/a4$ as

$$\frac{a_2}{a_4} = \frac{(a_{2M}ERF_2 + b_{2M}ESF_2 - a_{2M}ESF_2 \cdot EDF_2) ETF_{24}}{(a_{4M}ERF_4 + b_{4M}ESF_4 - a_{4M}ESF_4 \cdot EDF_4) ERF_2} \quad (10.27)$$

This computation of the ratio of the incident waves at the DUT interface provides the proper correction including considerations of the mismatch of the DUT. Surprisingly, no direct computation of the input match is required, but one should recognize that the measured b waves contain information about the reflections from the DUT ports.

In practice, one of the sources is turned on, and the other source is adjusted until the resulting ratio of $a1/a3$ meets the requirements of Eqs. (10.24) or (10.25) for the differential or common-mode drive, respectively. Typically, the sources must be iterated to achieve the desired drive as the changing drive on the DUT can cause the match of the DUT to change, thus affecting the ratio. The mismatch correction of the $a1$ and $a3$ waves is important if the true-mode drive is to be maintained. Figure 10.18 shows the error in the phase of the drive

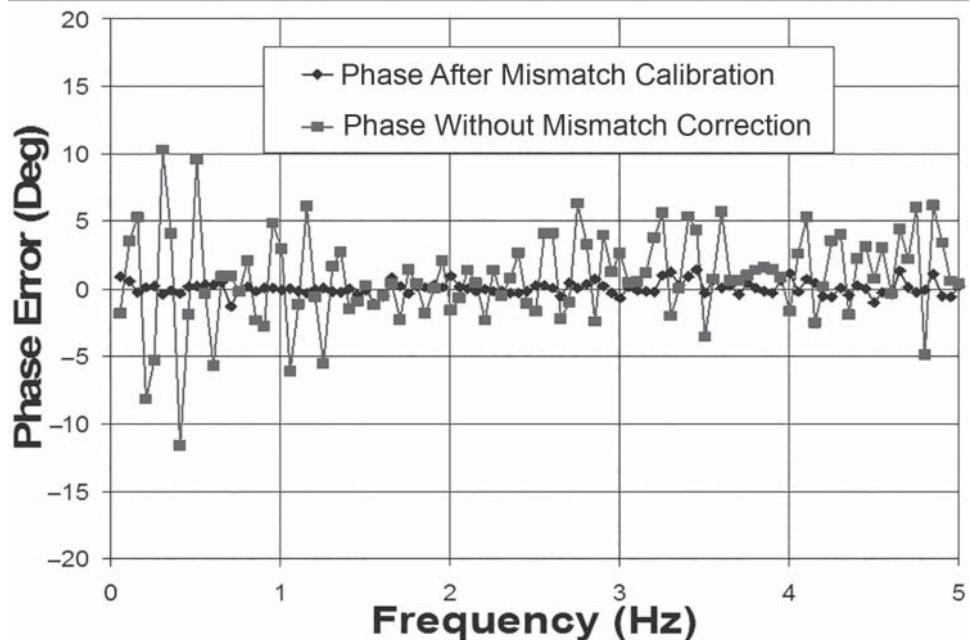


Figure 10.18 Error in phase due to DUT mismatch.

signal if the mismatch of a DUT is ignored. In this case, the DUT had an input match of about -15 dB.

The full measurement requires four total source stimulus settings: differential drive on port 1, common-mode drive on port 1, differential drive on port 2, and common-mode drive on port 2. For each drive, the two reference and all four test receivers must be measured. From these measurements, the values either single-ended S-parameters or mixed-mode S-parameters may be computed from Eqs. (10.1) or (10.10), respectively. Of course, each of the raw measurements must be corrected appropriately. Equation (10.1) provides four equations that describe the relationship between a and b waves. Taking 4 stimulus conditions provides 16 simultaneous equations that can be solved to find the 16 raw S-parameters, since

$$[\mathbf{b}] = [\mathbf{S}] \cdot [\mathbf{a}] \quad (10.28)$$

where $[\mathbf{b}]$ and $[\mathbf{a}]$ are 4×4 matrixes representing the various stimulus conditions, then

$$[\mathbf{S}] = [\mathbf{b}] \cdot [\mathbf{a}]^{-1} \quad (10.29)$$

Once the raw S-parameters are known, normal error correction methods can be applied to compute the corrected S-parameters, and from these the mixed-mode parameters are easily computed.

However, following this path, there are a few modifications to the normal S-parameter correction that must be made. First, since port 3 is active when port 1 is active, the port 3 load-match term ELF_{31} is not used; rather, the source-match ESF_3 is used. A similar substitution must occur for port 1, with ELF_{13} being replaced with ESF_1 . And of course, the same substitution is needed for ports 2 and 4, when they are driven. Finally, since the load match of the paired port is not the same as when the single-ended error correction terms are derived, the transmission tracking between paired ports must be modified, according to

$$ETF_{ji_TrueMode} = ETF_{ji} \cdot \left(\frac{ERF_j}{ERF_j + EDF_j \cdot ELF_{ji} - EDF_j \cdot ESF_j} \right) \quad (10.30)$$

One should recognize that similar substitutions must occur for the reverse tracking terms.

10.4.2 True-Mode Measurements

10.4.2.1 Measuring a Limiting Amplifier

With this understanding, the true-mode response of the amplifier from Figure 10.9 is compared with the single-ended measured mixed-mode parameters in Figure 10.19, for the differential parameters and the case of -25 dBm input and -5 dBm input. At -25 dBm input, there is no difference in the SE or true-mode derived S-parameters. At -5 dBm input, the differential gain is compressed 1.7 dB for the true-mode stimulus, but 2.7 dB for the SE stimulus.

The same two conditions are measured in Figure 10.20 for common-mode parameters.

From the previous figures, it is clear that this amplifier does respond differently to true-mode stimulus; in fact, this is a limiting amplifier designed to clip the input signal on the each input port. However, only the forward parameters change, and this makes sense as there is no gain in the reverse parameters, S_{dd12} and S_{dd22} , so there is not sufficient signal to create a non-linear change in the amplifier behavior.

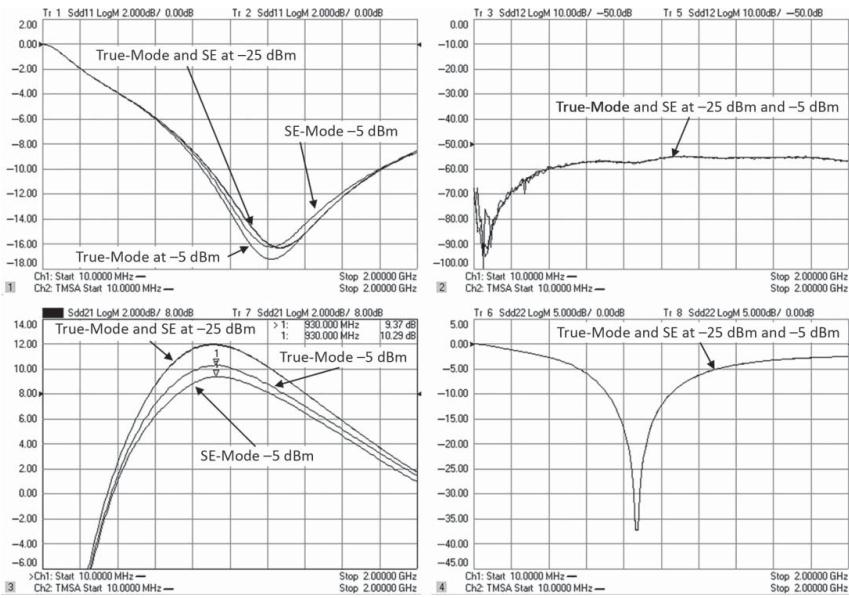


Figure 10.19 True mode vs. single-ended measurements for differential S-parameter.

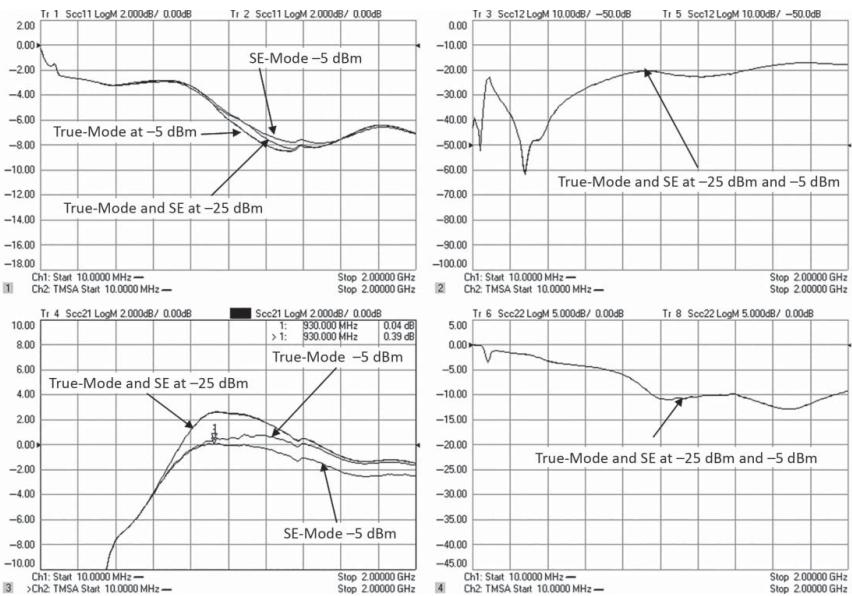


Figure 10.20 True-mode vs. single-ended measurements for common-mode S-parameters.

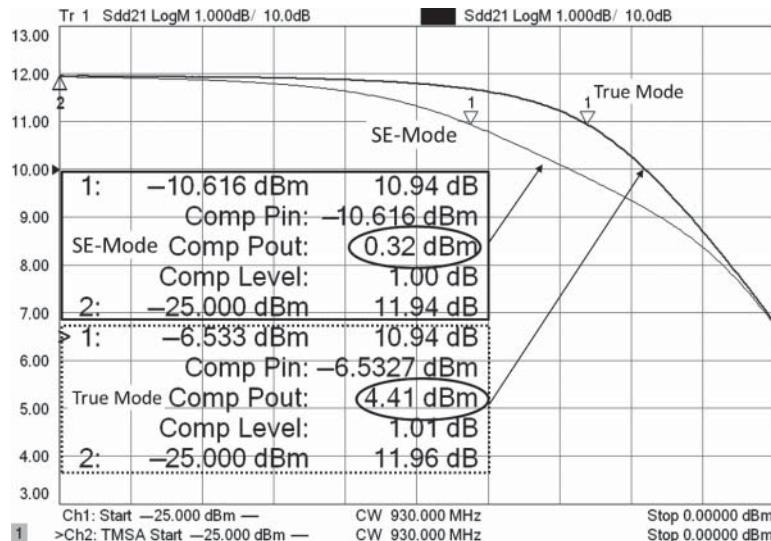


Figure 10.21 Fixed-frequency swept-power measurement showing differential gain compression.

Just as for normal measurements, the gain compression as a function of RF drive level, at a fixed frequency, can be measured and displayed. This measurement is shown in Figure 10.21 for just the differential gain, S_{dd21} , measured in a single-ended and true-mode measurement. This figure makes it very clear that the true-mode stimulus gives a different, higher, and more correct measurement of the 1 dB compression point than the single-ended measurement. This limiting amplifier is the perfect example of a device that must be measured with true-mode stimulus in order to accurately ascertain its non-linear behavior. However, and this is surprising to some engineers, at the low-power portion of this figure, as well as the -25 dBm measurements of Figures 10.19 and 10.20, the single-ended mixed mode and true-mode measurements give identical results, even though this is an active device. Measurements in the linear operating point of a device do not require true-mode drive, regardless of the nature of the device.

The common-mode gain compression measurement is not shown. While it may be performed, for the most part the common-mode signals normally present at the input of amplifiers are so small in actual use that they seldom create any non-linear behavior.

10.4.2.2 Measuring a “Normal” Differential Amplifier

The results on the limiting amplifier may be compared with the results of a more normal differential amplifier, the design of which more closely follows the design of Figure 10.11. In this case, the amplifier has a true differential input state. The differential gain, S_{dd21} , is shown in Figure 10.22, measured in both single-ended and true-mode, at both low power (light trace) and high power (dark trace). It is clear from these measurements that the non-linear differential response of this amplifier is the same regardless if true-mode is used or not, with only a slight difference at very high frequency, where the gain is low. Establishing the nature

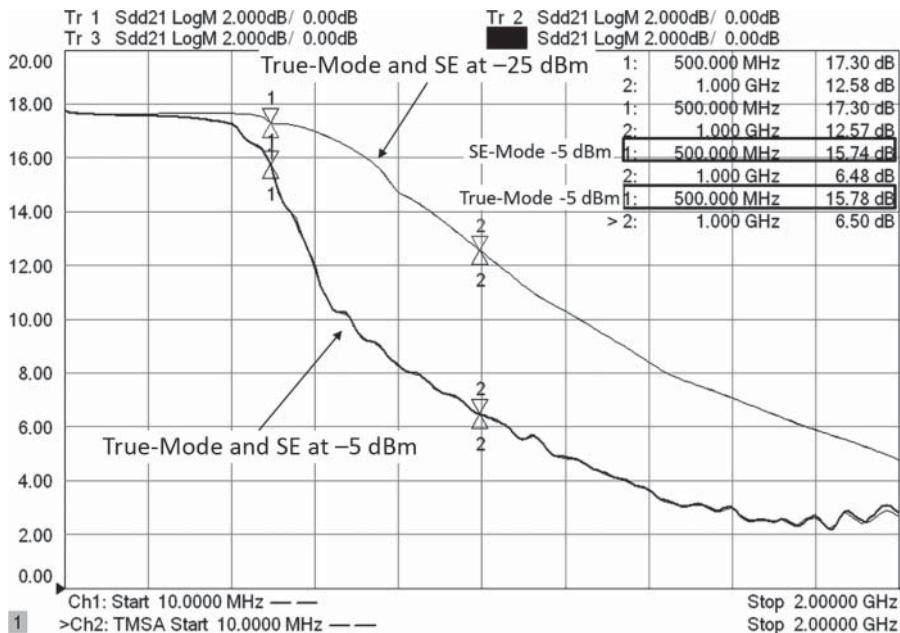


Figure 10.22 Differential gain for a normal differential amplifier, measured in single-ended and true-mode.

of the differential amplifier's response to single-ended drive is critical if one wished to make other, more advanced non-linear measurements such as two-tone intermodulation distortion (IMD) measurements.

For completeness, the four modes of forward gain, S_{dd21} , S_{cc21} , S_{dc21} , and S_{cd21} , are measured and compared in a power-sweep measurement, in Figure 10.23. As expected, the differential gain remains the same, when the drive level is of the SE drive is set to achieve the same differential input drive voltage. The common-mode gain does compress differently between true-mode drive and single-ended drive, as the common-mode signal is likely clipping at the input, since the output signal of the common-mode drive is very low, due to high suppression of common-mode gain. Similarly, the cross-mode terms see similar effects, where cross-mode terms with differential drive (S_{cd21}) shows the same compression as single-ended drives, and cross-mode drives with common-mode inputs show a change in compression between single-ended drive and true-mode drive. In fact, since the amplifier has very little common-mode gain, there is almost no compression in the common-mode drive parameters, since there is not sufficient output power to generate non-linear behavior. In the SE drive, when evaluating common-mode non-linear performance, the input signal has significant differential content, and so some non-linear behavior is evident.

The non-linear behavior of a differential amplifier may affect the phase response as well as the amplitude response. For this case of a normal differential amplifier, the phase and amplitude of S_{dd21} were measured in both SE mode and true-mode and are displayed in Figure 10.24. Here one can see that the phase response is also identical, regardless of drive mode, up to the point where the compression is more than 10 dB.

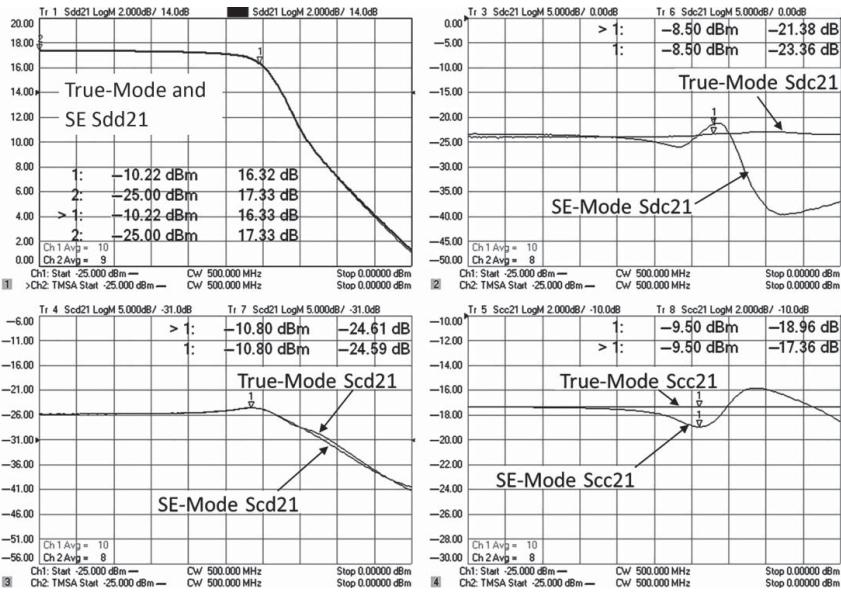


Figure 10.23 Swept power measurements of the mixed-mode transmission parameters.

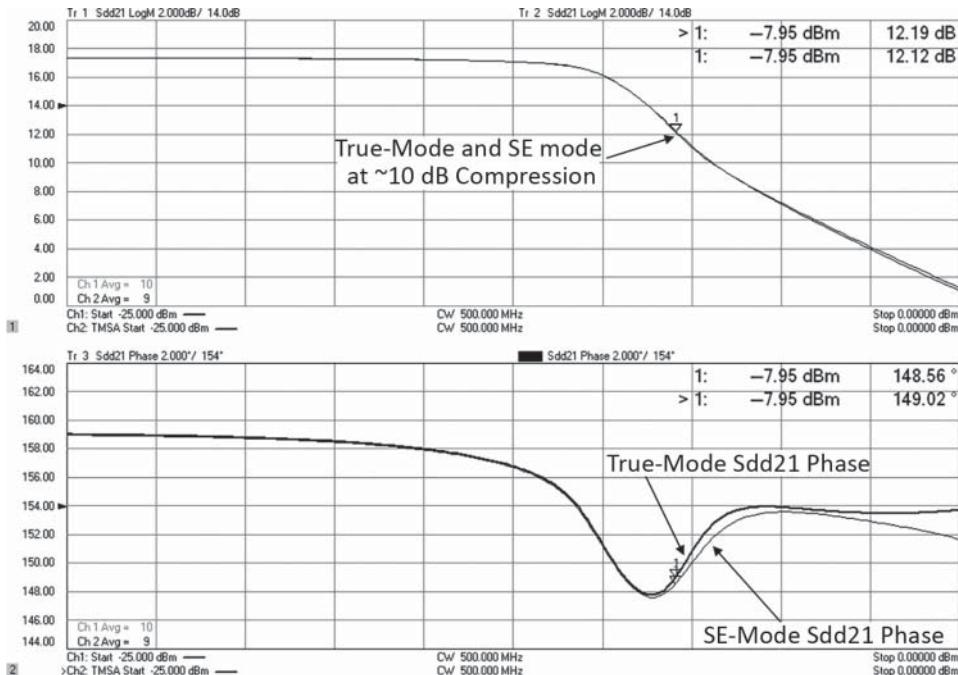


Figure 10.24 Non-linear magnitude and phase response of S_{dd21} for SE and true-mode power sweeps.

In all the measurements of the single-ended drive, the drive power must be increased by 3 dB over the differential-mode drive to ensure that the same differential voltage is applied in the single-ended case as in the true-mode case. This is because when power is applied in true-mode, half the voltage is applied to the positive input, and opposite-sign voltage is applied to the negative input of the amplifier. For single-ended to have the same total voltage across the input port pair, it must have 3 dB higher power. While it might seem that 3 dB higher power, which is twice the power, generates four times more voltage, in the single-ended case the load resistance is 50Ω , versus 100Ω in the differential case, so that 3 dB more single-ended power gives two times the voltage across the equivalent 100Ω resistance. In the previous graphs, the VNA had a source offset applied to account for this. If one does not use the built-in offset feature of a VNA, the SE drive can simply be set with 3 dB higher start and stop powers.

10.4.3 Determining the Phase Skew of a Differential Device

Differential devices are designed to have gain for differential inputs, and loss for common-mode inputs, but layout and other errors in the design can cause some phase skew in the design such that the gain is not maximized for a 180° differential input signal.

Some VNAs with true-mode drives facilitate investigating the phase-skew effects by allowing the phase of the drive sources to sweep over user-defined range of phases, while monitoring the differential gain of the DUT. In fact, in the normal sense, the gain of the

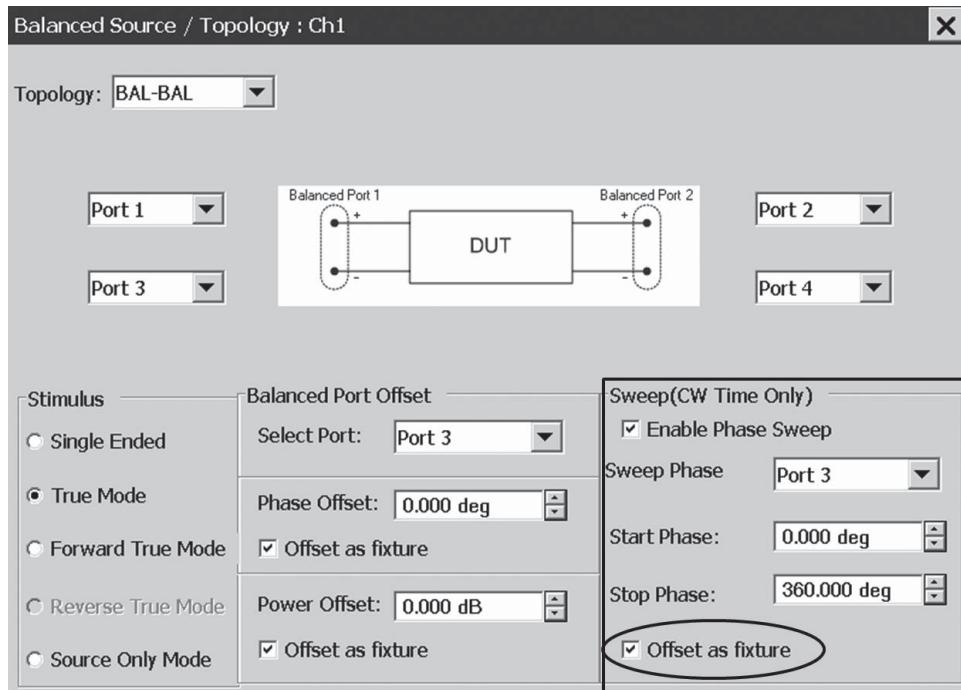


Figure 10.25 User interface for setting phase sweep.

DUT does not change with the drive signals; that is, without special consideration, any offsets in the drive-phase will be removed through the normal measurement and correction process, and the differential gain of the DUT, presuming it is in the linear mode of operation, will return a constant value. However, for the purpose of measuring phase skew, the offset in phase of the source is also de-embedded from the correction array to give the effect of adding a phase shifter in front of one port of the DUT. This virtual phase shifter allows the investigation of phase skew simply and conveniently.

Figure 10.25 shows one example of phase control for a true-mode VNA. In this example, the phase sweep settings are highlighted. Special note is made of the selection “Offset as fixture.” It is this selection that allows the phase shift to be applied as a virtual phase shifter, which allows one to directly measure phase skew.

If one suspects the phase skew may be due to external influences, such as uncompensated phase offsets in the test fixture, it is possible to take the value found in the phase-skew test and apply it as a fixed offset under the settings for Balanced Port Offset. In this case, the offsets should also be treated like a fixture, as they are compensating for externally induced effects. For completeness, offsets are available for both amplitude and phase in this example user interface.

A phase-skew test was performed on the amplifier used in Figure 10.22. Here the phase was swept from 0° to 360° . One might expect that the maximum differential gain should occur at 180° , but the offset here is offset from the expected phase; thus, 0° offset for a differential drive means the signals are exactly 180° out of phase. When the offset for differential is 180° ,

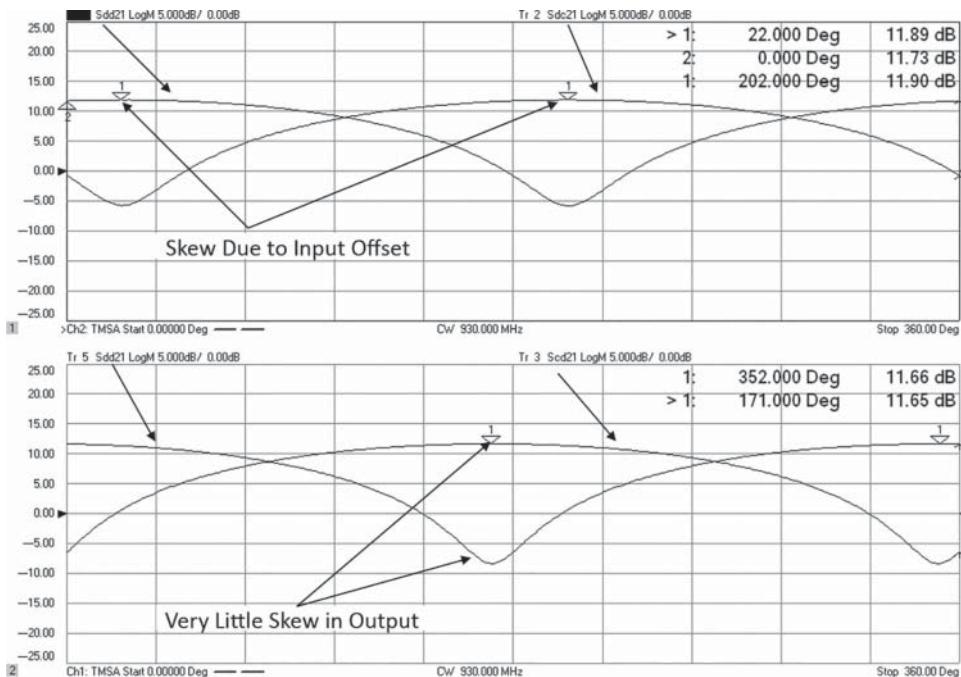


Figure 10.26 Phase-skew test on a differential amplifier.

the drive signals are both in-phase; that is, they are common mode. A marker search is used to find the maximum S_{dd21} gain; the marker's stimulus value represents the phase skew of the DUT, as shown in Figure 10.26.

In the upper window, two traces are shown while port3 is swept from 0° to 360° . S_{dd21} and S_{dc21} are sensitive to the skew in the input signal. The peak of S_{dd21} shows an offset of 22° , and the peak of S_{dc21} shows an offset of $180^\circ + 22^\circ$. One would expect to see the S_{dc21} take on a large value when the common-mode drive becomes offset by 180° , thereby stimulating the differential mode fully. The offset in phase here is a confirming indication that phase skew is at the input. For this test example, a small delay adapter was added to one of the inputs to generate some skew; its delay is about 60 psec, so at 930 MHz one would expect about 20° of skew.

In the lower window, port 4 is phase swept from 0° to 360° for the same configuration. Since the output has no additional delay, no skew would be expected. In fact, there appears to be a few degrees of skew in the out portion of this device, about 8° or 9° .

Skew in the input reduces the effective input differential signal, and thus the resulting output power is lower and so is the gain. Skew in the output reduces the differential portion of the output power and so also reduces the gain. Sweeping both the input and the output independently will show the skew values of each.

10.4.4 Differential Harmonic Measurements

Differential amplifiers are often used for driver amplifiers or receiving trans-impedance amplifiers for opto-electronic applications. These are typically very wideband applications,

some covering from nearly DC to above 67 GHz. One of the key figures of merit for these electro-optic amplifiers is total harmonic distortion (THD), usually expressed as a percentage. Measuring harmonics for these amplifiers is particularly difficult using traditional methods of baluns (discussed in the next section) because while the source drive may be limited in frequency, the received signal must be over a very wide band as the harmonics are measured. Since the figure of merit is a ratio of the harmonic power to the main output power, it is often not possible to find wideband baluns that can cover the frequency range desired; thus, it is desirable to find a means to measure the differential gain, power, and harmonics without the use of baluns.

The true-mode drive method described in Section 10.4.1 is available to measure the fully corrected linear differential S-parameters of an amplifier but requires the input and output frequencies to be the same. Recently, a new method of controlling the hardware advances the measurement capability to allow any source on any port to be controlled in frequency and power, and phase relative to any other port, and to make receiver measurements at any number of other frequencies. This application, called the “Differential and IQ” application (Keysight option S93089), was developed (by this author) to solve two different problems. The first, measurement of IQ mixers, is discussed in Chapter 7. The second use case, measurement of harmonics of differential amplifiers, was driven by a leading manufacturer of differential amplifiers used for optical applications, and there was no good way to measure their harmonics over a broad band.

The DIQ application allows one to set a series of related frequencies and to assign sources to some frequencies and measurements at all of the related frequencies. Figure 10.27 shows the setup diagram for differential harmonic measurements. The input frequency range, F1,

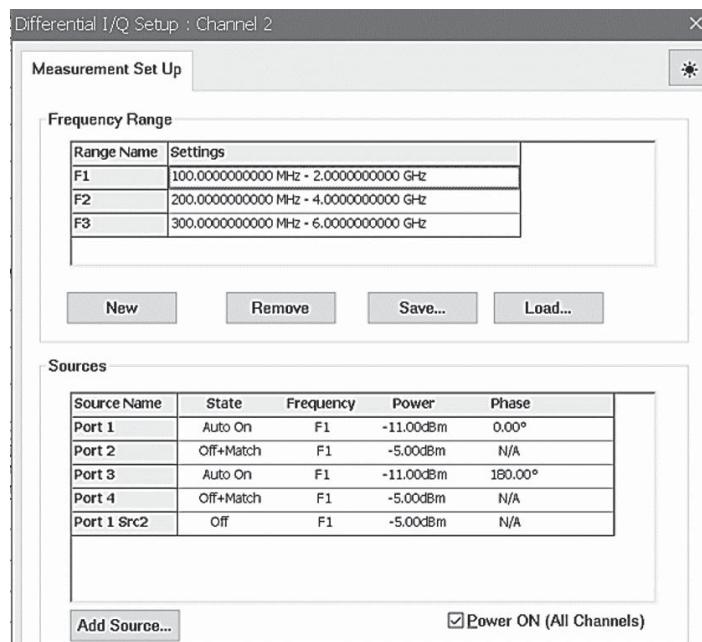


Figure 10.27 Setup for differential harmonics.

is assigned to ports 1 and 3 (from Figure 10.17). The other frequencies ranges are derived from two times the input (F2) and three times the input (F3), and these define the additional receiver measurement frequencies (any number of frequency ranges can be specified). The source sweep is defined by the F1 range and the number of measurement points. As the source steps through each of its measurement points, the source is first set to the F1 point and measured. The mismatch at each port may also be measured; the true, applied incident wave is computed (accounting for mismatch); and the phase of the pair of sources, port 1 and port 3, is determined. If the phase is not within the user-specified tolerance, the source phase is reset and remeasured; iteration continues until the phase is properly set, and the power is also adjusted to meet the specified power tolerance. After the sources achieve the proper power and phase (180° for differential drive), all the VNA receivers are moved to each of the other frequency ranges (F2, F3, ...), and the complex power is measured at each port. In this way, a complete capture of the fundamental and harmonic power at each source step is found.

In an additional step, if mismatch correction is desired, the return loss of each port of the DUT, at each of the specified frequencies, is performed (in a pre-sweep) so that the proper match correction can be determined. This is done before the main sweep so that the actual power applied to the port can be compensated for any mismatch as measured. To avoid any crosstalk from other sources, all the sources are turned off except the one making a port match measurement in the pre-sweep.

Displaying the results is a little more complex in the DIQ channel than in normal application channels as the measurement definition can change dramatically depending upon the device being measured. Because of this, the user can specify precisely the measurement math to be applied for any defined parameter; some predefined templates have been created to make this easier. The following is an example of a parameter template for differential harmonics (they are saved in a tagged XML format):

```

<ParameterList>
<Param1 ="IPwr_F1" Val="(a1_F1-a3_F1)/SQRT(2)"/>
<Param2 ="OPwr_F1" Val="(b2_F1-b4_F1)/SQRT(2)"/>
<Param3 ="OPwr_F2" Val="(b2_F2-b4_F2)/SQRT(2)"/>
<Param4 ="OPwr_F3" Val="(b2_F3-b4_F3)/SQRT(2)"/>
<Param5 ="Gain" Val="(b2_F1-b4_F1)/(a1_F1-a3_F1)"/>
<Param6 ="2nd_dBc" Val="(b2_F2-b4_F2)/(b2_F1-b4_F1)"/>
<Param7 ="3rd_dBc" Val="(b2_F3-b4_F3)/(b2_F1-b4_F1)"/>
<Param8 ="THD%" Val="100*sqrt(pow(mag(b2_F2-b4_F2),2)+pow(mag(b2_F3-b4_F3),2))/mag(b2_F1-b4_F1)"/>
<Param9 ="Sdd11_F1" Val="(b1_F1-b3_F1)/(a1_F1-a3_F1)"/>
<Param10 ="Sdd22_F1" Val="(b2_F1-b4_F1)/(a2_F1-a4_F1)"/>
<Param11 ="IPwr" Val="_F1"/>
<Param12 ="Sdd22_F2" Val="(b2_F2-b4_F2)/(a2_F2-a4_F2)"/>
<Param13 ="Sdd22_F3" Val="(b2_F3-b4_F3)/(a2_F3-a4_F3)"/>
<Param14 ="S22_F1" Val="b2_F1/a2_F1"/>
<Param15 ="S22_F2" Val="b2_F2/a2_F2"/>
<Param16 ="S22_F3" Val="b2_F3/a2_F3"/>
<Param17 ="S44_F1" Val="b4_F1/a4_F1"/>
<Param18 ="DiffBal" Val="(a1_F1*1)/(a3_F1*1)"/>
<Param19 ="H2outP" Val="(b2_F2*1)/(1*1)"/>
<Param20 ="H2outM" Val="(b4_F2*1)/(1*1)"/>
<Param21 ="H3outP" Val="(b2_F3*1)/(1*1)"/>
<Param22 ="H3outM" Val="(b4_F3*1)/(1*1)"/>
<Param23 ="S44_F2" Val="b4_F2/a4_F2"/>
<Param24 ="S44_F3" Val="b4_F3/a4_F3"/>
</ParameterList>
```

Let's examine a few specific examples.

The definition of the differential input wave in the context of the DIQ application, derived from Eq. (10.11), is

$$a_{d1} = \frac{1}{\sqrt{2}}(a_1-F1 - a_3-F1) \triangleq "IPwr_F1" \quad (10.31)$$

The differential output wave is

$$b_{d2} = \frac{1}{\sqrt{2}}(b_2-F1 - b_4-F1) \triangleq "OPwr_F1" \quad (10.32)$$

The differential harmonic waves are

$$\begin{aligned} b_{d2_H2} &= \frac{1}{\sqrt{2}}(b_2-F2 - b_4-F2) \triangleq "OPwr_F2", \\ b_{d2_H3} &= \frac{1}{\sqrt{2}}(b_2-F3 - b_4-F3) \triangleq "OPwr_F3" \end{aligned} \quad (10.33)$$

And the key figure of merit, Total Harmonic Distortion, is defined as the ratio of harmonic power relative to the main power expressed as a percent, or

$$THD\% = 100 \cdot \left(\frac{\sum_n |b_{d2_Hn}|^2}{|b_{d2}|^2} \right) \quad (10.34)$$

where n can be over any number of desired harmonics, but typically only the second and third.

Figure 10.28 shows the measured response of a differential amplifier (intended for use as part of a GSM mobile phone system). Unlike the optical drive amplifier, this amplifier does not have a flat, wideband response but is peaked at the frequency of interest. The upper trace shows the differential gain. The markers are shown at the fundamental (900 MHz), second harmonic (1800 MHz), and third harmonic (2700 MHz).

In the middle plot we see the swept frequency output power (from 100 MHz to 2 GHz), labeled OPwr_F1. Also shown in the dark trace is the second harmonic power, labeled OPwr_F2, and the third-harmonic power (light trace), labeled OPwr_F3. It is notable that the second-harmonic power is nearly the same as the third-harmonic power, as it is common for lower-order harmonics to have higher power than the higher-order harmonics.

Finally, in the lower plot we see the swept frequency THD, in percent. This amplifier, at this power, had less than 1% THD.

The curious case of the lower second harmonic power leads us to investigate this situation further.

We can do interesting experiments in the DIQ application; for example, we can view the harmonic power as a function of the phase difference of the input.

It is interesting to note the effect of gain variation on the THD. At low frequencies, near 300 MHz (the first grid line in Figure 10.28), the THD reaches a peak. We can expect this as THD increases when the fundamental power decreases. At 300 MHz the gain has dropped nearly 20 dB so the fundamental power drops by 20 dB, but at 300 MHz the third harmonic is in the gain peak region, and the second harmonic also sees reasonably high gain, so even

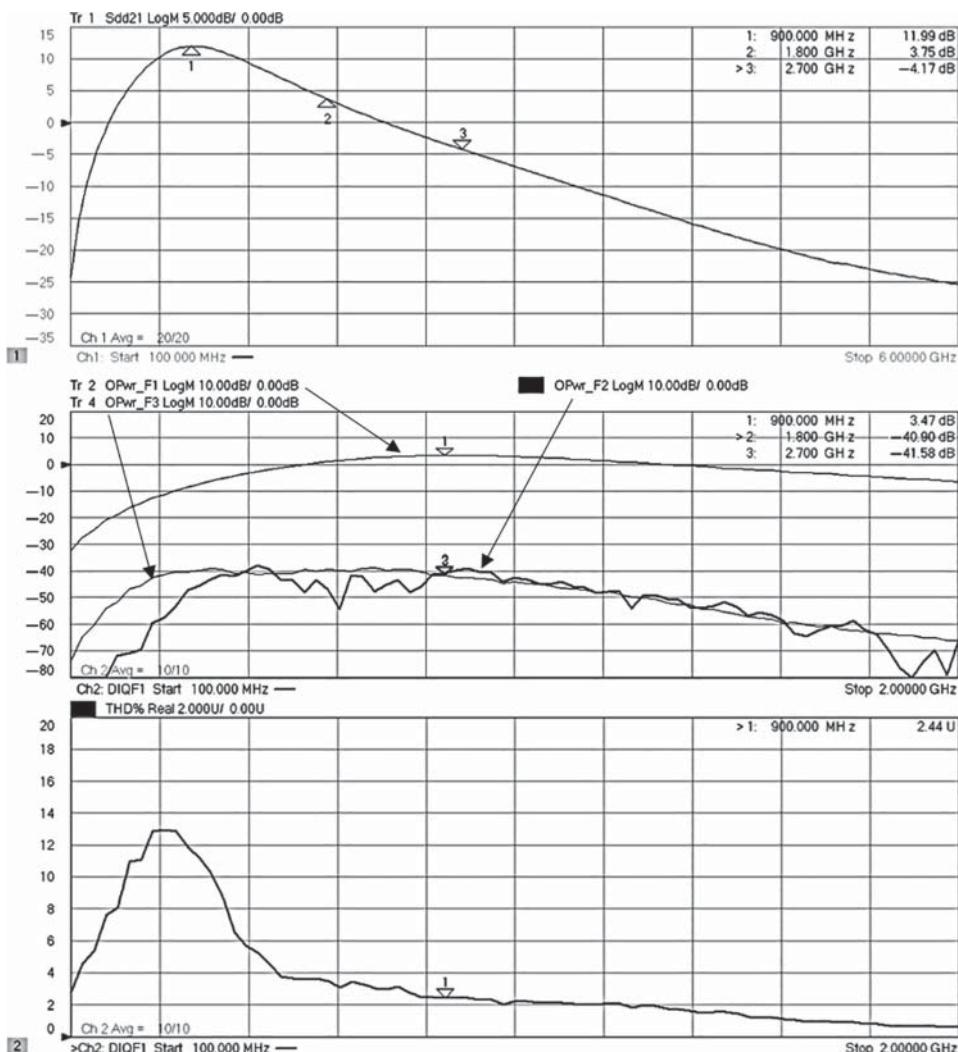


Figure 10.28 Gain, harmonic power, and THD of a differential amplifier.

though the harmonic power is relatively constant, the dropping fundamental power causes a large increase in THD.

Figure 10.29 shows the result of sweeping the phase of the differential input from 0° (common mode), through 180° (fully differential), and up to 360° (back to common mode). As expected, the output power is maximum when the input signal is fully differential. But the harmonics show some unusual behavior. One might intuitively expect that the higher the output power, the higher the harmonics, but we see that the second harmonic is minimum when the output power is maximum; that is, the second harmonic is minimum when the input is fully differential. The third harmonic also shows variation but has minimums near 126° and 230° . The powers shown are the *differential* power of the harmonics. Let's take a look at the

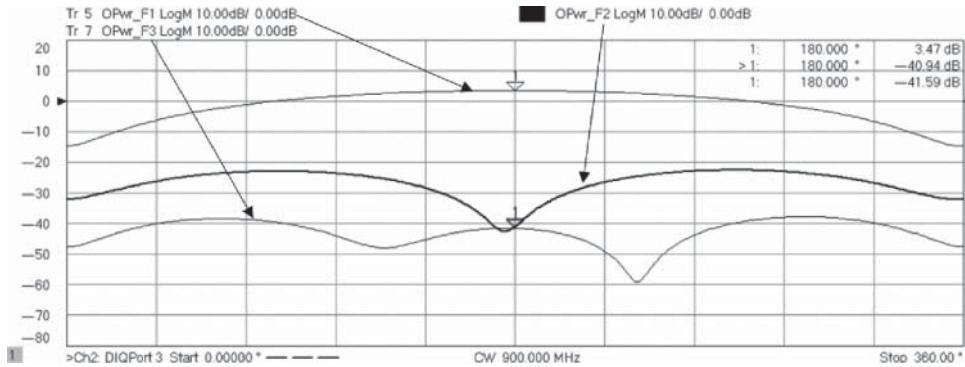


Figure 10.29 Output power, second and third harmonic versus input phase.

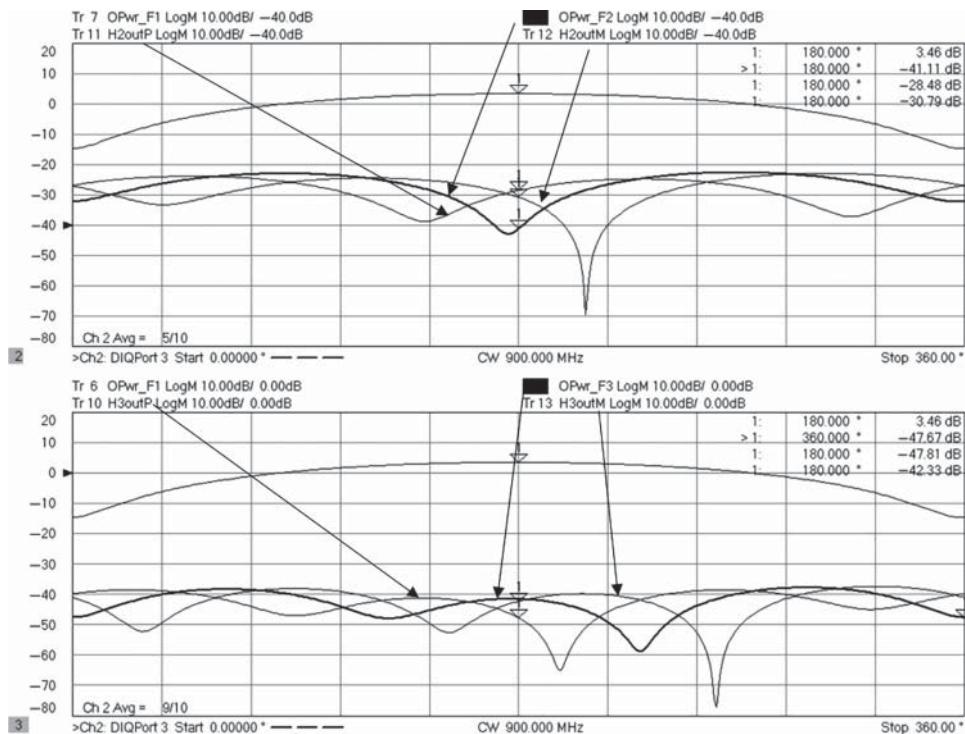


Figure 10.30 Second harmonic response (upper), third harmonic response (lower); showing differential and single ended power.

single-ended output of the harmonics, that is, the output power at the plus and minus port independently.

Figure 10.30 shows the same result as Figure 10.29, but here the individual positive and negative port single-ended power is shown. In the upper plot, the output differential power (OPwr_F1) is shown along with the differential second harmonic (OPwr_F2, dark trace).

Also shown is the output power of the positive (port 2) terminal ($H2outP$) and the negative (port 4) terminal ($H2outM$), both shown as light traces. At the 180° phase point, the two powers are essentially equal. Since the input signals are exactly out of phase, if we consider the case where the harmonics are generated with the same phase as the input, then at the second harmonic frequency they will have a doubling of the phase and be 360° out of phase, or back to being in-phase. Since the powers are equal and the definition of the differential power is the $b2$ power minus the $b4$ power, the result will be a cancelation at the second harmonic frequency.

Using a similar rational, we might expect cancelation of the third harmonic at 120° and 240° , and in fact we see on the lower plot that the $H3outP$ and $H3outM$ are nearly the same amplitude at these phase offsets, so a tripling of the phase offset would yield 360° and 720° : both in-phase conditions. Thus, the differential third harmonic would be minimized here. Note at 180° input offset, at the third harmonic we would have 540° : fully differential again, so the third harmonic is differential power and maximum at the 180° input shift.

This explains why we see the second harmonic power being much smaller than one might expect: each output is larger, but the differential combination is smaller or nearly the same as the third harmonic.

We can further explore the harmonic behavior by holding the frequency constant, setting the phase to fully differential, and sweeping the power. In Figure 10.31 we can see the swept

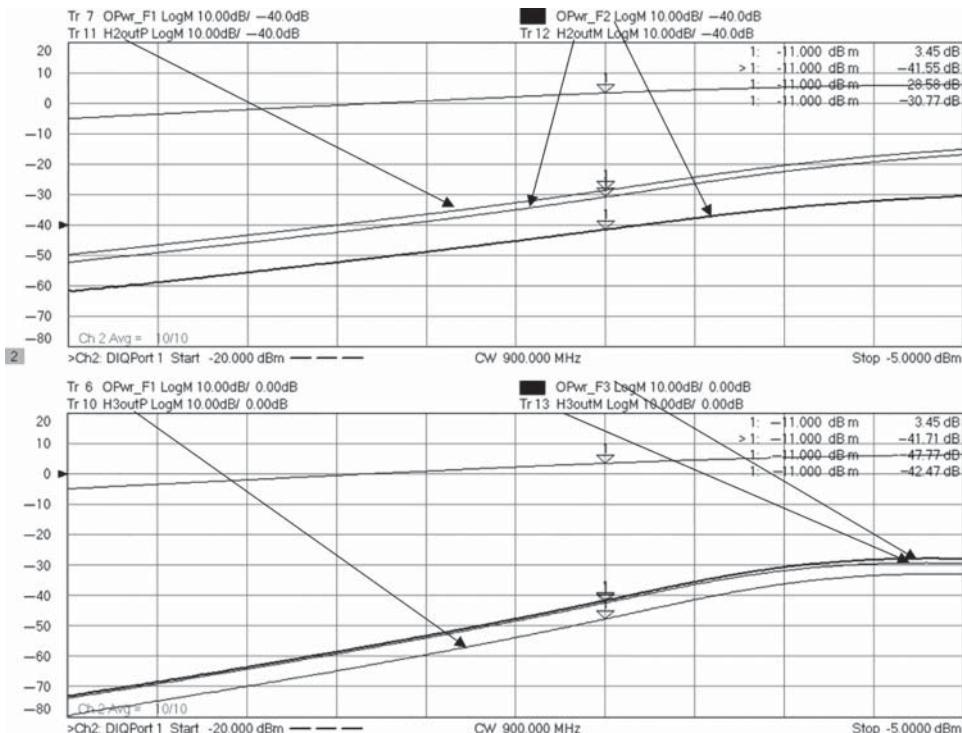


Figure 10.31 Swept power harmonics, with single-ended and differential results.

power response for second (upper plot) and third (lower plot) differential harmonics (dark traces in each plot), along with the differential output power and traces of the single-ended positive and negative outputs.

Here it is clear in the upper plot that while there is significantly more second harmonic single-ended power out of each of the b_2 and b_4 ports, compared with the third harmonic single-ended power (-28 dBm vs -42 dBm), the differential second harmonic is essentially the same as differential third harmonic. Conversely, the single-ended third harmonic at each of the positive (b_2) and negative (b_4) ports is actually lower than the differential third harmonic.

From this we can say that the harmonic power is mostly common-mode for the second harmonic and mostly differential for the third.

A final figure, Figure 10.32, shows the power sweep with the second and third differential harmonic power plotted on the same grid, and it is clear to see the 2:1 versus 3:1 slopes on the second harmonic with respect to the third harmonic. Also shown in the lower plot is the total harmonic distortion, in percent, as a function of the drive power.

From these plots, it is clear that the problem of making harmonic measurements of differential devices is essentially solved. The new application takes what was an almost intractable problem and makes it close to trivial. While the harmonic measurement problem is solved, there still exists difficulties in making more complicated measurements on differential devices, such as IMD and noise figure, and in these cases, one must often resort again to baluns, as discussed in the next section.

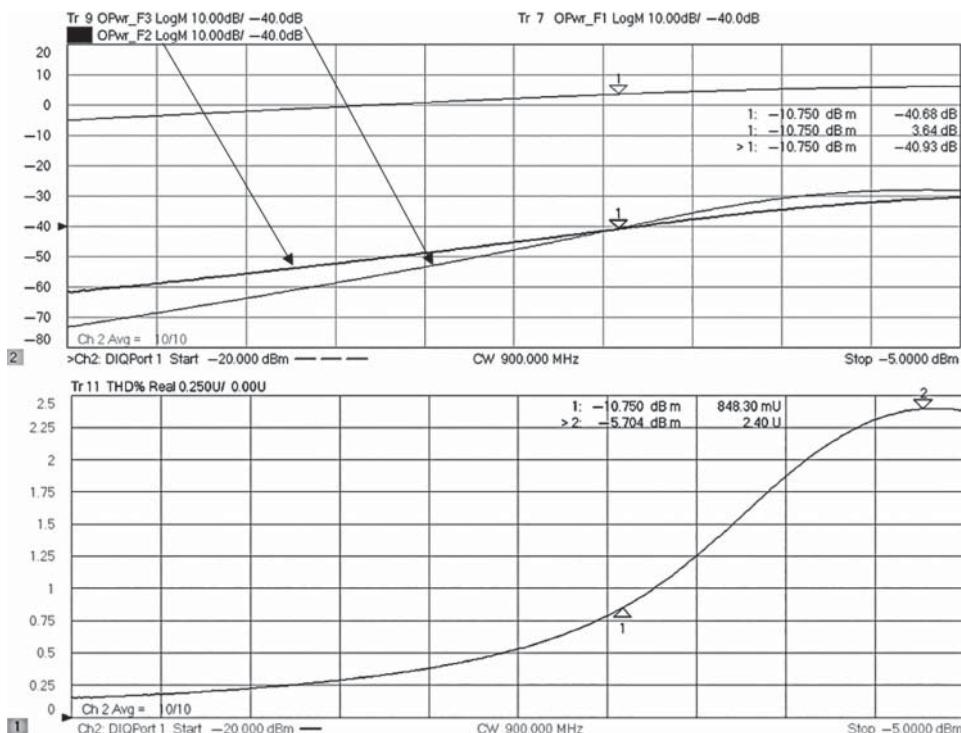


Figure 10.32 Swept-power harmonics (upper) and THD (lower).

10.5 Differential Testing Using Baluns, Hybrids, and Transformers

In some cases, such as when a true-mode 4-port VNA is not available, one must resort to the use of baluns to characterize differential devices. While they have been used for many years, it is only just recently that advances in understanding and advanced de-embedding capability has simplified and improved the measurements made with baluns. Any device that converts a single-ended signal into a balanced signal is considered a balun, and they come in two main forms: hybrids and balun-transformers. Most commonly, even though hybrids provide a balanced signal, they are not usually referred to as baluns.

10.5.1 Transformers vs. Baluns

Transformers and baluns are terms that are often used interchangeably and usually refer to 3-port devices that change a single-ended signal into a balanced signal. In many cases, these are actual small transformers, but occasionally transmission-line structures are used. Baluns have at least one single-ended input and a differential output pair. Several examples are shown in Figure 10.33. The common-mode impedance of a balanced port of a balun is often not defined but is generally either very high (open) for an ungrounded transformer (circuit a) or very low (short) for a center-tapped transformer (circuit c). Some baluns are created with series structures, and the common-mode impedance is set by the parasitic inductance of the ground leg of the balun (circuit b).

Some baluns are designed as 4-port devices and actually separate the mixed-mode on the right side to two discrete ports on the left side: a delta port that measures the differential mode and a sum port that measures the common mode. This structure (circuit d) is commonly used in simulation to create two separate ports for analysis of the two modes independently. Each transformer is 1:1, and the upper one is center-tapped to provide the common-mode connection. Normal mixed-mode S-parameter definitions have the differential port impedance

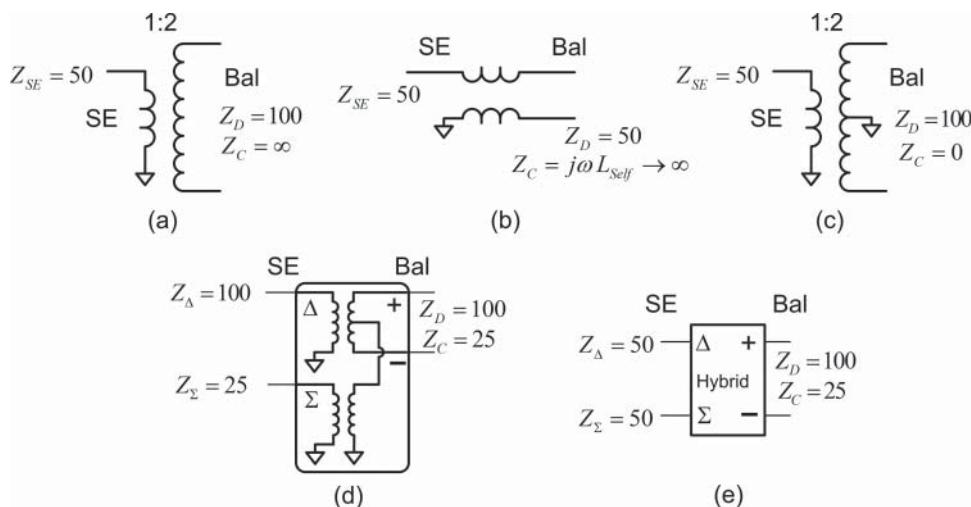


Figure 10.33 Examples of several RF transformers, baluns, and a hybrid.

as 100Ω , and the common-mode port impedance as 25Ω . This definition ensures that the single-ended view of the balanced port is 50Ω each to ground. Depending upon the nature of the balun, it can either supply impedance transformation or not. If the balun does not supply impedance transformation, one must decide the appropriate definition for the differential port impedance. For example, the balanced port of transformer (b) shown presents a balanced 50Ω impedance to a differential device. This structure is commonly fashioned by a coaxial cable with an RF choke around the outer conductor. In such a case, the common-mode impedance is set by the self-inductance of the core. Many VNAs supply fixturing software that allows one to redefine the port impedance mathematically, transforming the measured S-parameters into any desired impedance.

Hybrids are 4-port devices that have an in-phase port and difference port input (often called the sum or Σ port, and the delta or Δ port) and a pair of balanced ports as outputs. When the sum port is connected to an RF source, the output from the balanced pair is common mode. When the difference port is connected to the source, the output from the balanced pair is differential mode. Most hybrids require that the unused port be terminated in 50Ω . Hybrids have the nice attribute that the balanced ports are also isolated, and hybrids provide a 50Ω match on each one of the balanced pair ports, when the sum and difference port is terminated in 50Ω . This makes the balanced-port differential impedance 100Ω , and the common-mode impedance 25Ω . This implies that the hybrid is not a 1:1 transformer as in circuit (d) shown, since the impedance on the sum and delta ports are not Z_d and Z_c , but 50Ω . Using hybrids as baluns give results similar to the true-mode VNA, without requiring any additional impedance transformations.

The drive signal from the source is divided by the balun so that half the power (0.707 times the voltage) goes to each port. So when driving from a single-ended port, the differential power will be the same as the single-ended input-power, but the power is into both loads of the balanced device. Since each load sees half the power, the total power is the same. In terms of voltage, the plus port will see 0.707 times voltage, and the negative port will see -0.707 times the voltage, so the differential voltage will be 1.4 times the single-ended voltage. But the differential impedance is two times greater, so the differential power is the same as

$$P_{Diff} = \frac{V_{Diff}^2}{Z_{Diff}} = \frac{(\sqrt{2} \cdot V_S)^2}{100} = \frac{2V_S^2}{100} = \frac{V_S^2}{50} = P_S \quad (10.35)$$

Of course, all real hybrids have some non-zero additional loss as well as phase skew so that their outputs are not perfectly balanced. Proper calibration can account for some of these effects if the hybrid is measured as a 3-port, its single-ended to differential S-parameters computed, and de-embedded from the overall measurement. Some modern VNAs allow saving the mixed mode S-parameters directly as an S2P file, which greatly simplifies the task of creating the de-embedding file. Simply create a 3- or 4-port calibration, measure the single-ended to balanced parameters, and save the desired 2×2 matrix (SE-Differential or SE-Common) to an S2P file. An example user interface for this save function is shown in Figure 10.34. In this case, a hybrid was measured with two different SE inputs, one that gives differential signals at the output and one that gives common-mode signals at the output. Here the matrix for single-ended to differential de-embedding was chosen.

Examples of the measurements of an RF hybrid are shown in Figures 10.35 and 10.36. The first figure shows the measurement results from the delta-input to the balanced port. In the

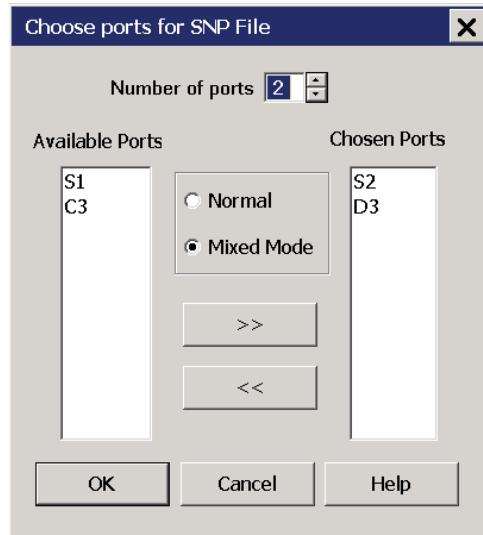


Figure 10.34 User interface to save mixed-mode parameters to an S2P file.

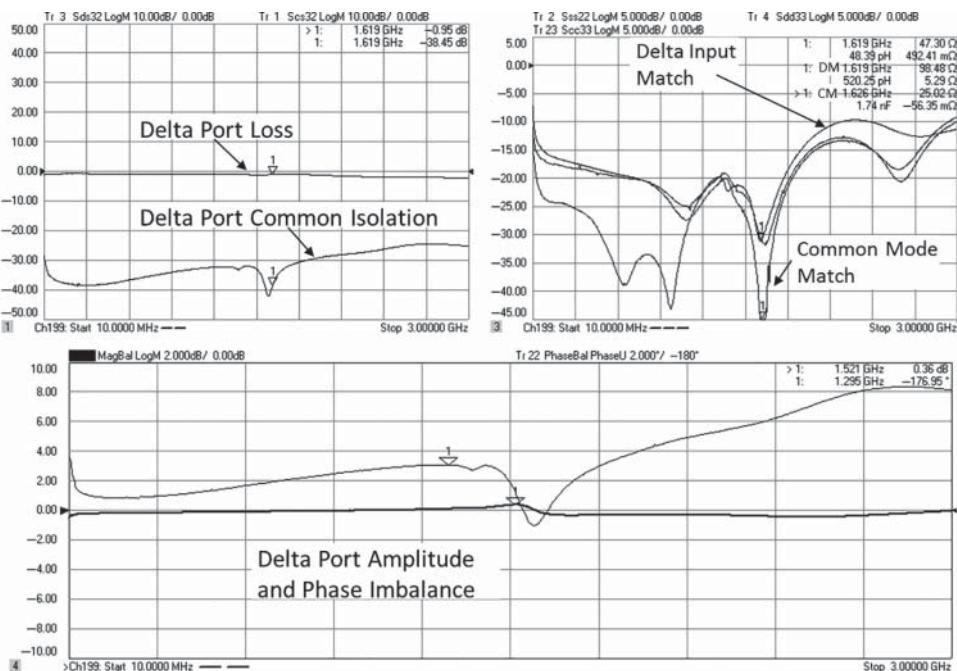


Figure 10.35 Measurement of SE to differential response from the difference port of a hybrid.

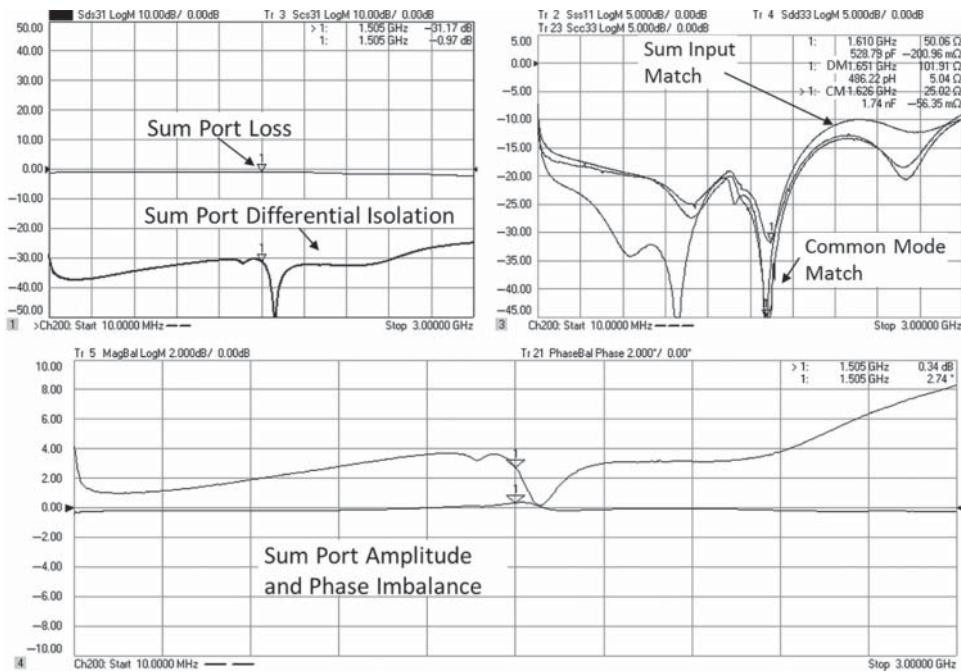


Figure 10.36 Measurement of SE to common-mode response from the sum port of a hybrid.

upper-left window is the loss from the delta-input to the balanced port differential mode, and the isolation of the common mode. In the upper right is the match of the delta-input and balanced differential mode and common mode. Notice that the differential impedance is near $100\ \Omega$, and the common mode is near $25\ \Omega$. In the lower window is a direct measurement from the input port to the plus and minus outputs, displayed as amplitude or magnitude balance and phase balance. Since this is driven from the delta port, the phase balance is relative to 180° .

Similar measurements were made from the sum or common input port, and the loss and isolation are nearly the same. Any imbalance in the hybrid usually shows up similarly in both the sum and delta ports. Note that in the sum-port input case, the amplitude balance is nearly the same, but of course the phase balance is relative to a 0° phase offset.

10.5.2 Using Hybrids and Baluns with a 2-Port VNA

If one has measured a hybrid with a 3-port or 4-port analyzer and the single-ended to balanced parameters are saved as an 2-port equivalent S2P file, then in many modern VNAs it is a simple task to using the built-in fixturing functions to de-embed the S2P file of a pair of hybrids from the input and output ports of a 2-port VNA. This has the effect of changing the 2-port VNA into a 2-port differential (or common if the VNA is connected to the sum port) network analyzer. Only the differential mode (to the limit of the common-mode isolation, about $-30\ dB$ in the case of the hybrid from Figure 10.35) drives the DUT, and only

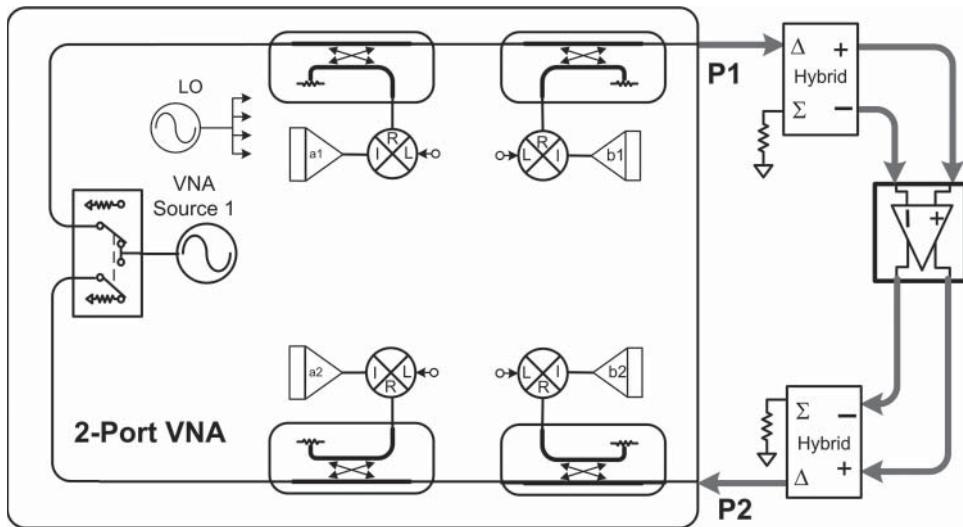


Figure 10.37 Test setup for using hybrids to test differential parameters.

differential-mode output signals are measured from the DUT. An example setup is shown in Figure 10.37.

This allows a very good measurement to be made on differential devices. A hybrid is ideally suited for this case as the terminating impedance is 50Ω for both inputs. Properly de-embedding these hybrids gives good results, even in the case where the device is in compression and the device is one that requires true-mode drive. As an example, the device from Figure 10.19, which was shown to require true-mode drive, was measured with both the full 4-port true-mode system, and with the system of Figure 10.37, for the case where the DUT was driven to the 1 dB compression point at 930 MHz. The measurement with the de-embedded hybrids is nearly identical to the true-mode measurement, with only about 0.1–0.2 dB variation, illustrated in Figure 10.38. This variation is consistent with the -30 dB isolation of the common-mode signal.

The same two conditions, true-mode drive and de-embedded hybrids, were used to measure a power sweep, as shown in Figure 10.39. This figure shows almost perfect response when compared with a true-mode measurement. At the same time, a single-ended mixed mode measurement shows a substantial error when measuring this limiting amplifier. Note that the gain in the linear region of the power sweep agrees within 0.04 dB comparing the de-embedding and the SE-method to the true-mode result. In the compression region, near 1 dB compression, the SE mode gives much lower compressed power, as noted earlier, but in the case of the de-embedded hybrids, the 1 dB compression level is within 0.01 dB of the true-mode measurement. This is remarkably good agreement and indicates that the de-embedding computations do a very good job of accounting for all the effects of the hybrid. Details on de-embedding are discussed in Chapter 11.

With this confirmation that de-embedding hybrids creates a measurement system that provides good differential measurement capability, one can use the same setup to make distortion and noise measurements on differential devices, as discussed in the next two sections.

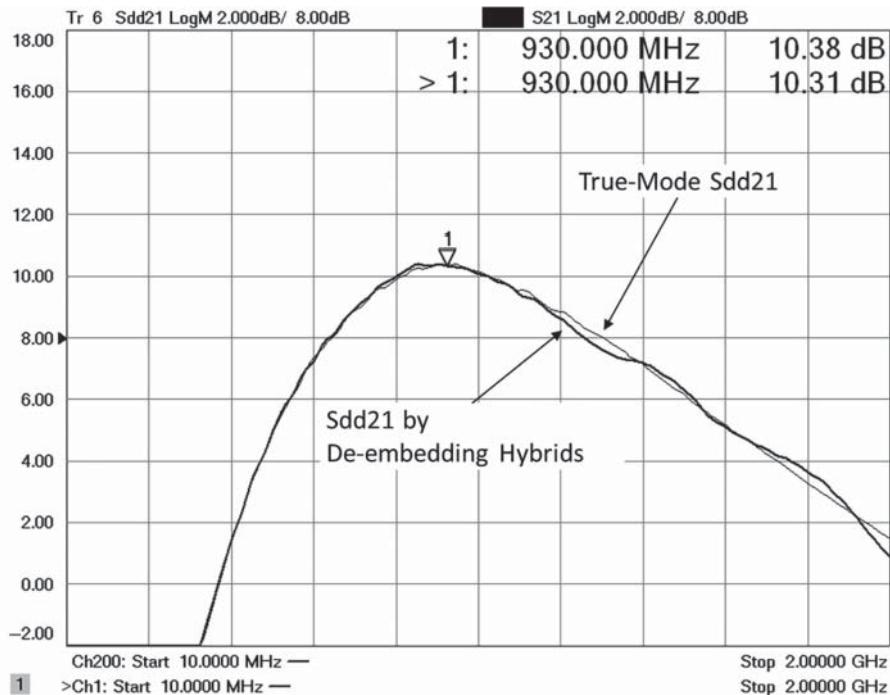


Figure 10.38 Frequency response for differential parameters using de-embedding for -5 dBm drive (1 dB compression).

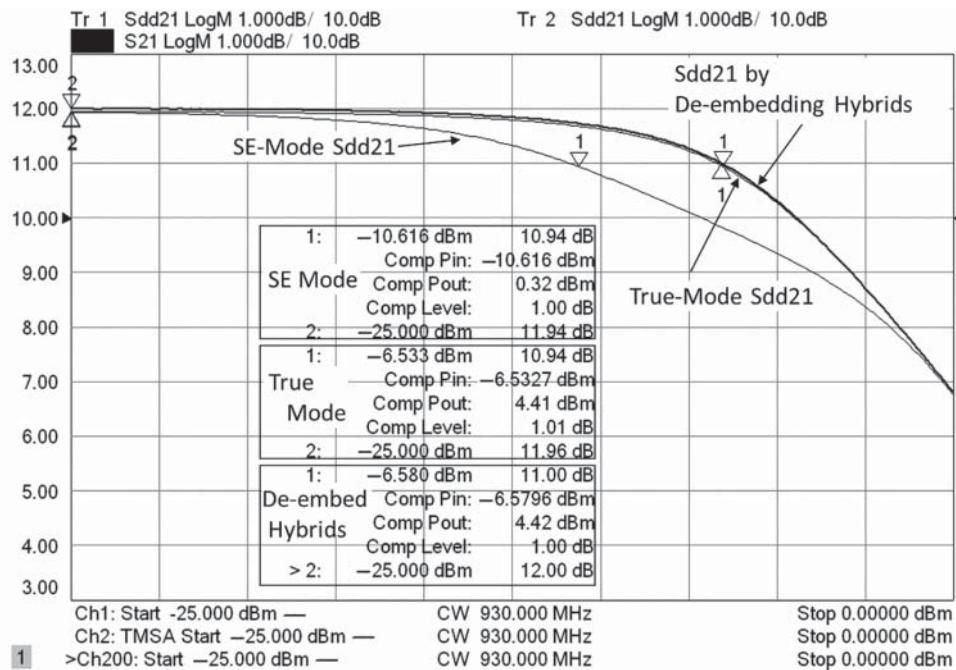


Figure 10.39 Non-linear response to a power sweep using hybrids to test S_{dd21} .

10.6 Distortion Measurements of Differential Devices

Distortion measurements on differential devices can be quite difficult to perform, especially if true-mode drive is required, as is the case for the limiting amplifier. But newer VNAs, with built-in applications and calibration for IMD measurements, allow the same de-embedding computations as the S-parameter gain and power. Thus, one may calibrate a 2-port VNA for IMD measurements, apply the hybrids as shown in Figure 10.37, and proceed to read the IMD results simply and directly from the VNA. In Figure 10.40, IMD measurements of the limiting amplifier are made across frequency, with an input power near the 1 dB compression point. A typical level for 1 dB compression of about -26 dBc IM3 holds true for this differential amplifier as well. The accuracy is ensured here as the input and output hybrids are de-embedded from the measurement result, including the appropriate source power offset to account for the loss in the port 1 hybrid.

The power spectrum of the IMD output can be shown for this power level, at a single-output frequency of 930 MHz, with the result matching the measurements shown for the marker position of this sweep, as shown in Figure 10.41.

Here, the markers show a slight skew in the output tones, likely due to a difference in gain or match in the amplifier. Marker 1 shows the low tone to be about 0.4 dB lower than the high-tone, for an average (in dB) power of -1.1 dBm, and matches closely with the power from the swept frequency plot. Also, the IM3 is within about 0.5 dB, and the IM5 is within about 0.5 dB as well, which is within the amplitude accuracy of the SA mode in this VNA.

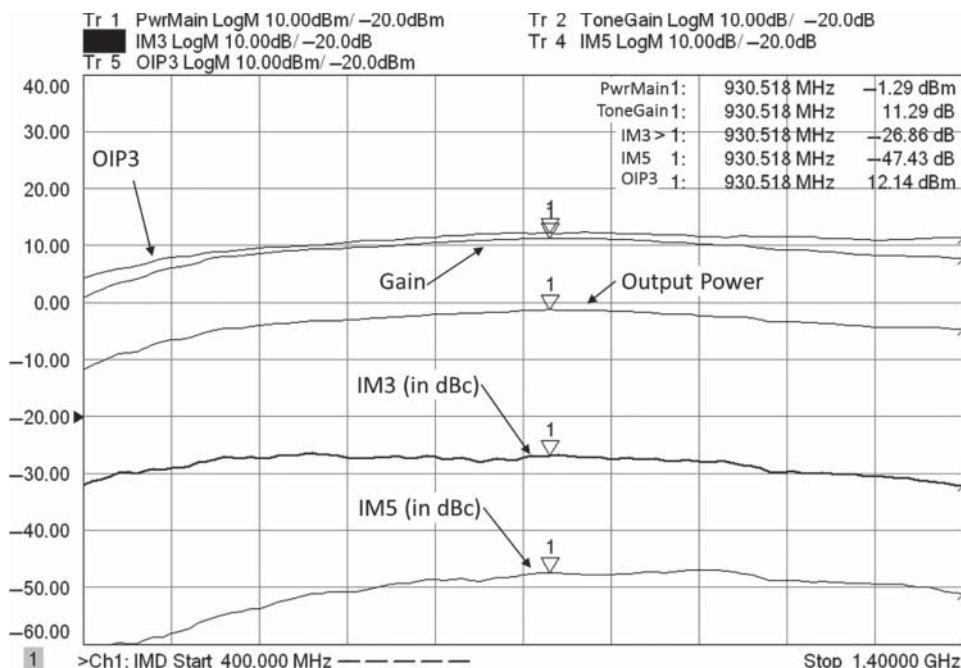


Figure 10.40 Swept-frequency IMD measurements.

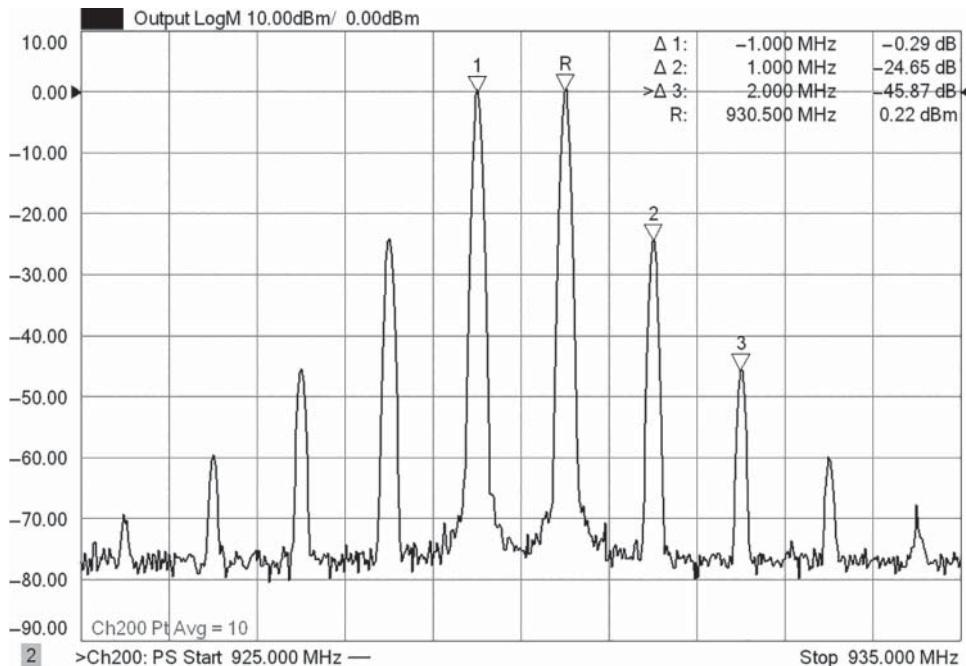


Figure 10.41 Spectrum plot of IMD for a differential amplifier measured with hybrids.

Finally, the IMD behavior of the amplifier with respect to drive power is also easily shown using a swept-power IMD measurement, illustrated in Figure 10.42.

From this plot, it is clear that the OIP3 level starts to degrade for powers above about -21 dBm input, and near the 1 dB compression point, the IM3 has increased substantially over the value one would project using the normal 2:1 degradation for IM3 dBc level versus RF drive. The value at -11 dBm is only -22.8 dBc, but the normal expectation would be -28.8 dBc, based on the level at -21 dBm. The fact that the IM5 tone goes down at high power might indicate that some additional IM3 is created by conversion of the IM5 power.

10.6.1 Comparing Single-Ended IMD Measurement to True-Mode Measurements

For a normal differential amplifier, one would expect that driving the single-ended input and measuring a single-ended output should give a very similar result for IMD as measurements with a balun, if the SE drive level is adjusted to provide for the same differential voltage as the true-mode measurement. This turns out to be nearly the case. Figure 10.43 shows an IMD measurement of a normal differential amplifier using hybrids de-embedded, as shown in Figure 10.37. Figure 10.44 shows the four single-ended measurements without a hybrid, from each input to each output. The SE measurement must be performed with 3 dB higher input tone powers to achieve the same differential voltage at the amplifier, and the output power

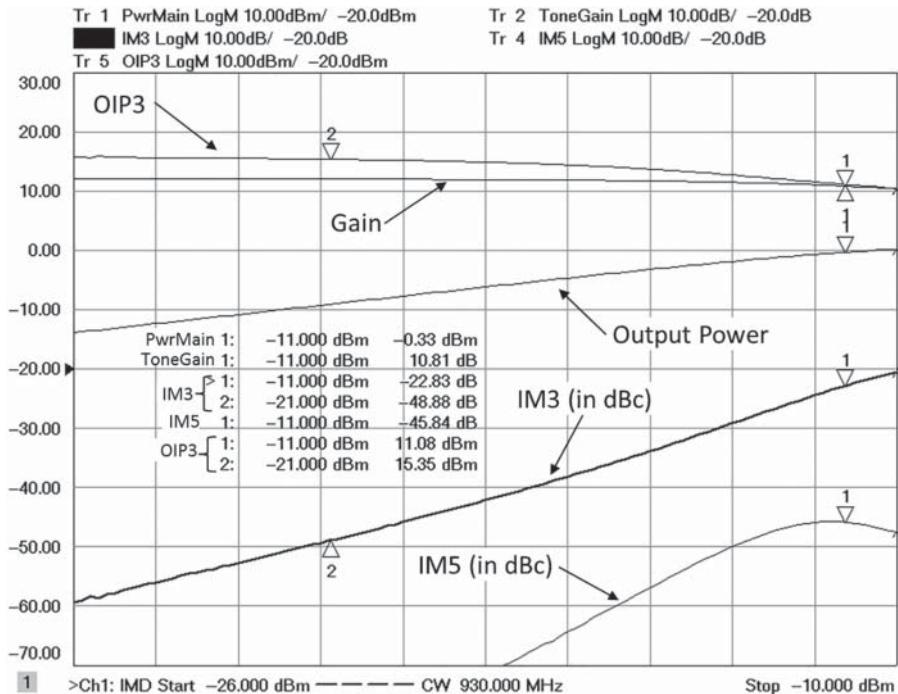


Figure 10.42 Swept-power IMD measurement of a limiting differential amplifier.

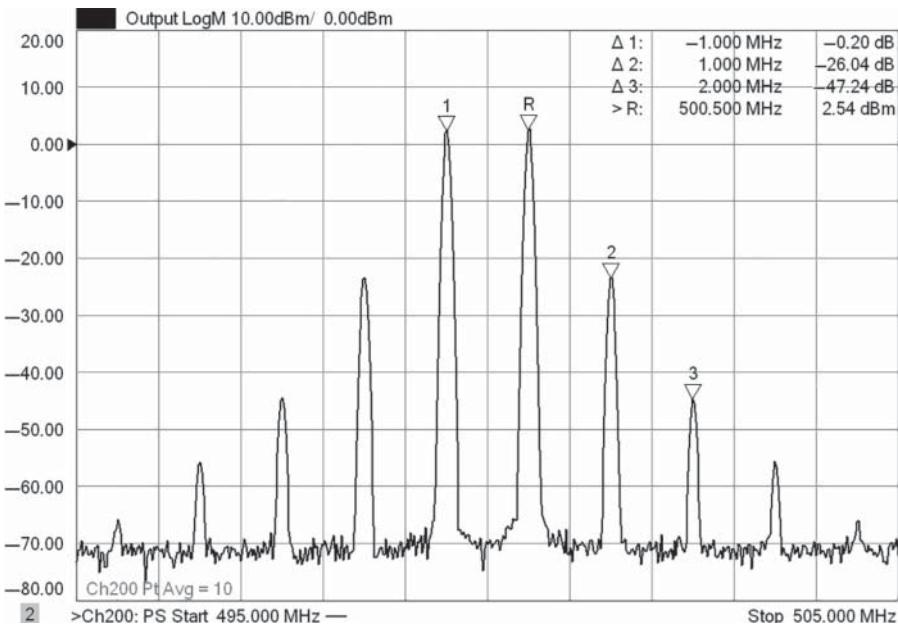


Figure 10.43 Measuring IMD on a normal differential amplifier, using hybrids, at the 1 dB compression power.

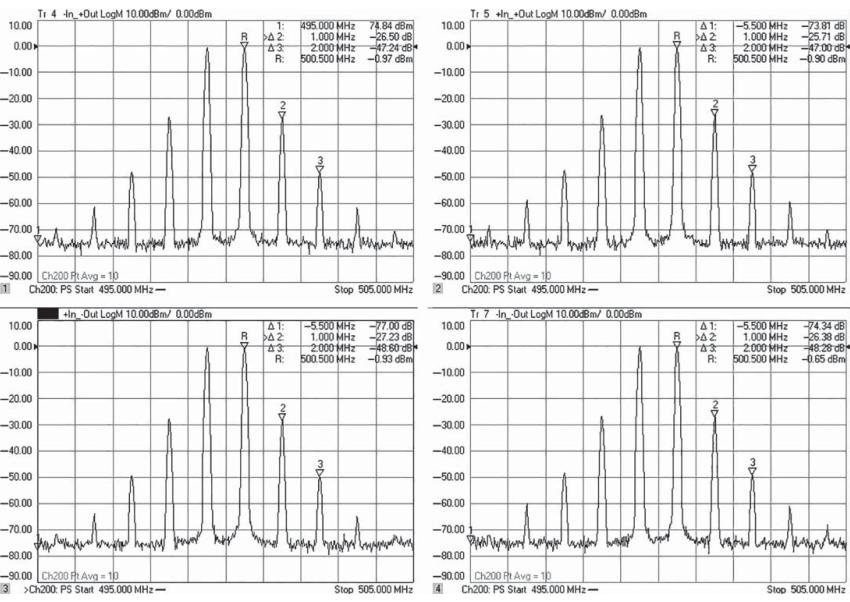


Figure 10.44 Measurement of IMD for a normal differential amplifier with SE drive and SE measurement.

will also read 3 dB lower since the SE output measures only one-half the output power, that is, the power in one SE load.

Here, each of the input drives shows similar output power and distortion. The lower-right window shows driving -In port and measuring -Out port, with a slightly higher output power (the R marker) but a similar lower IMD result. Each of the other outputs shows slightly lower output signal (one would expect -0.5 dBm from the result of Figure 10.43) and also very similar IMD levels. This measurement leads to the conclusion that for a normal differential-amplifier, with SE gain terms that are well matched, measuring the SE IMD gives a good prediction of the differential IMD, though at a power level 3 dB lower. Note that the unused ports were terminated in a matched load.

However, the same result is dramatically different in the case of the limiting amplifier, which has been seen to respond quite differently to SE drive versus true-mode drive (see Figure 10.39) for CW signals. Above, in the measurement of Figure 10.41, the drive power was -11.7 dBm for 1 dB compression in the DUT, which gives the resulting -27 dB IM3. Figure 10.45, below, shows the result of using an SE drive on the same amplifier; the SE drive is increased to -8.7 dBm, to produce the same differential voltage at the input. But for this amplifier, which is not a normal differential amplifier, the results are very unusual. The output power for each drive (one to the + port and one to the - port) is completely different, and it is also different depending upon which output port (+ or -) is measured. In fact, this is to be anticipated from the plot of Figure 10.8, considering that the gains are not equal. What is also remarkable is that there are vastly different values of IM3 for each path. For the path with the lowest gain, the IMD level is highest (in dBc terms), though lowest in absolute power terms. It is very hard to discern what the expected differential-mode IM behavior is from these SE measurements. This device is the same as the one used for Figures 10.40–10.42 and is an example of a case when only true-mode measurement will yield the correct differential behavior.

10.6.2 Differential IMD without Baluns

Recently, just as with harmonics, advanced capabilities in VNAs has made possible the ability to measure IMD without the use of baluns. The key requirement is to drive the input two-tone signal differentially, and to measure each of the output signals differentially as well. This is now possible with the new “Differential and IQ” application available on the Keysight PNA (option 089). While a more complicated setup than shown in Figure 10.37, as it requires additional sources, it does remove any need for band-limited baluns. The block diagram using a two-source VNA with two external sources is shown in Figure 10.46.

The two external sources are programmed to the two-tone frequencies, as well as the two internal sources; but the internal sources are phase-controlled to provide the true-differential two-tone signal at the input to the differential DUT. External source 1 (Ext1 in the figure) and VNA source 1 are both set to one of the two-tone frequencies, and external source 2 (Ext2 in the figure) and VNA source 2 are set to the other of the two-tone frequencies. Phase control of any source is possible, but it is much faster to control the internal sources of the VNA.

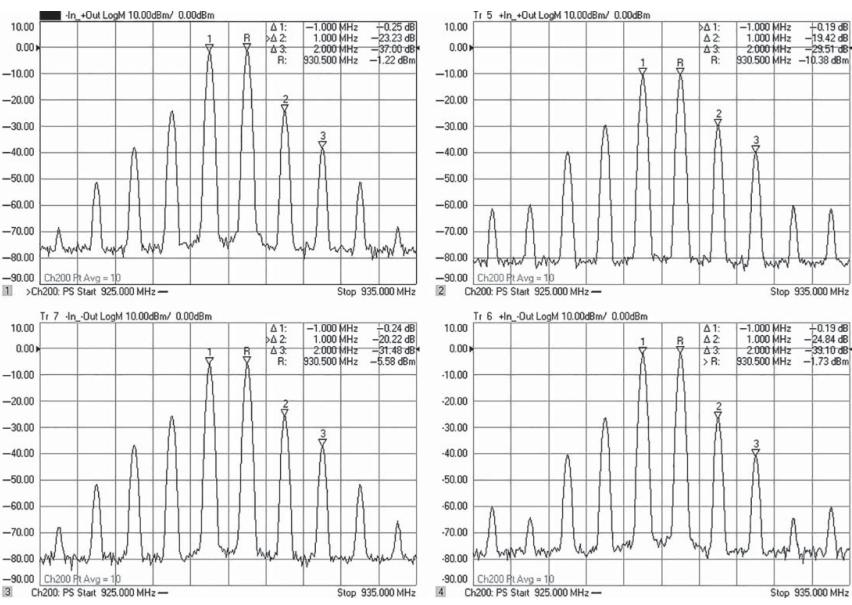


Figure 10.45 The SE measurements of a limiting amplifier shows poor prediction of differential behavior.

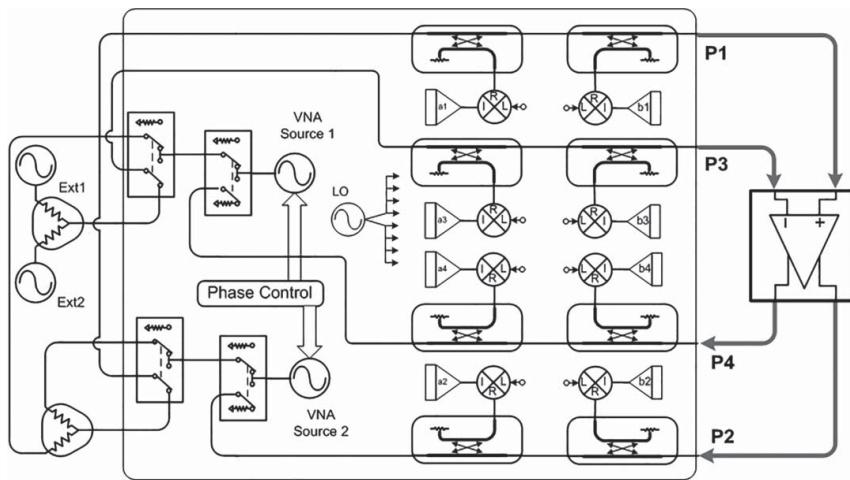


Figure 10.46 Block diagram for differential IMD with external sources.

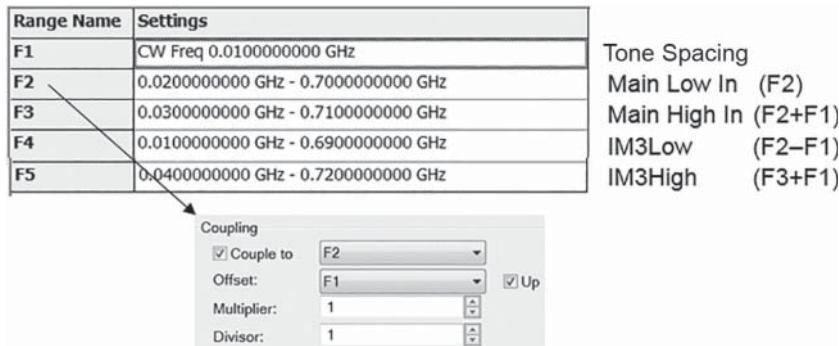


Figure 10.47 Frequency setup for differential IMD without baluns.

Source Name	State	Frequency	Power	Phase
Port 1	Auto On	F2	-15.00dBm	180.00°
Port 2	Off+Match	F1	-5.00dBm	N/A
Port 3	Auto On	F3	-15.00dBm	180.00°
Port 4	Off+Match	F1	-5.00dBm	N/A
Port 1 Src2	Off	F1	-5.00dBm	N/A
Ext1	Auto On	F2	-15.00dBm	0.00°
Ext2	Auto On	F3	-15.00dBm	0.00°

Figure 10.48 Source setup for differential IMD.

Figure 10.47 shows the frequency setup for differential IMD measurements. Setting the first frequency as the tone spacing, and the second frequency as the lower of the two tones, and deriving the other tones from these two allows one to easily change both the two-tone frequencies and the tone spacings.

In Figure 10.46 there are four sources that must be controlled. The frequency, power, and phase are set through the interface shown in Figure 10.48. In this setup, the second VNA source, normally set to port 3, is rerouted to combine with the port 1 source to form the two tones at port 1. They are each set to different frequencies, and they are set to be phase controlled relative to the two external sources, which are routed to the port 3 output of the VNA.

The measurement process proceeds like this: First the two external sources are set to their respective frequencies, and whatever phase they produce is defined as the reference phase, or zero degrees. The two internal sources are set to the same respective frequencies, and then the internal VNA receivers (which have been previously calibrated) are tuned in turn to F2 (lower of the two main tones) and F3 (higher of the two main tones). At frequency F2, an iterative process measures the signal at the port 1 (from internal source 1 on the a_1 receiver) and port 3 (from the external source 1 on the a_3 receiver), detects the amplitude and phase relationship, and resets the amplitude of the two sources, and rotates the phase of the internal source 1 to achieve a 180° phase difference. Then the same process is repeated for the internal

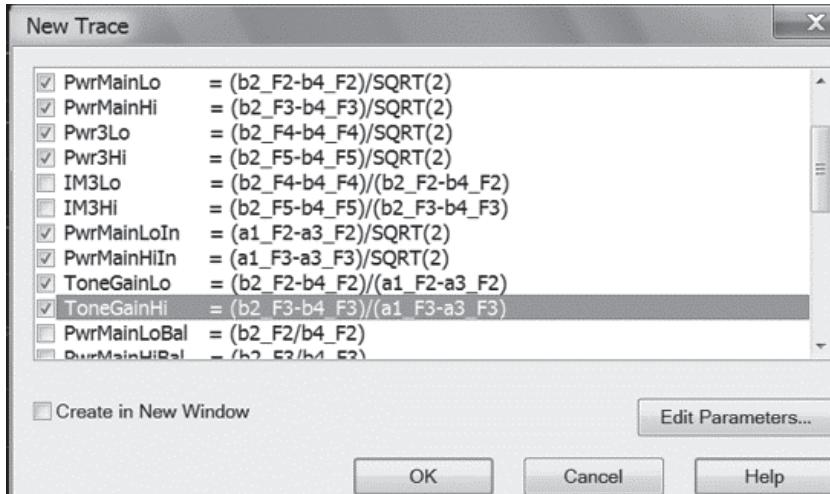


Figure 10.49 Differential IMD parameters.

source 2 and external source 2, at the F3 frequency. Since changing the amplitude of either source can change the phase relationship of the signals, due to AM-to-PM of the sources, an additional measurement is performed after each iteration. When the user-specified tolerance of both amplitude and phase is achieved, the iteration ends, and we have a measurement of each signal at each of the F2 and F3. Finally, the receiver is tuned to F4 and F5 (and any other frequencies defined in the frequency range list) to get a complete measurement of all the signals at all the frequencies at all the ports. From these receiver measurements, various parameters can be computed. Just like for harmonics, the IMD measurements must be computed from the difference of the signals at the receivers, either a_1 and a_3 for input signals, or b_2 and b_4 for output signals.

Figure 10.49 shows the setup for the IMD parameters of the differential amplifier measurement. The power measurements follow exactly the formulations of Eqs. (10.31) and (10.32). The Pow3Lo and Pow3Hi are defined as the absolute power of the lower and higher third-order intermodulation product, respectively. A pair of new parameters are defined here, IMD3Lo and IMD3Hi, and are defined as

$$\begin{aligned} IMD3Lo &= \left(\frac{b_2_F4 - b_4_F4}{b_2_F2 - b_4_F2} \right), \\ IMD3Hi &= \left(\frac{b_2_F5 - b_4_F5}{b_2_F3 - b_4_F3} \right) \end{aligned} \quad (10.36)$$

which is the ratio of the IMD power of the lower IMD product to the lower main tone, and the ratio of the IMD power of the higher IMD tone to the higher main tone. See Chapter 8 for more details on IMD definitions.

Finally, any of the traces can be displayed versus frequency sweep, or power sweep, or even input phase sweep. Figure 10.50 shows a swept frequency measurement where the gain of the amplifier at each tone, along with input power, output power, and IMD power are displayed.

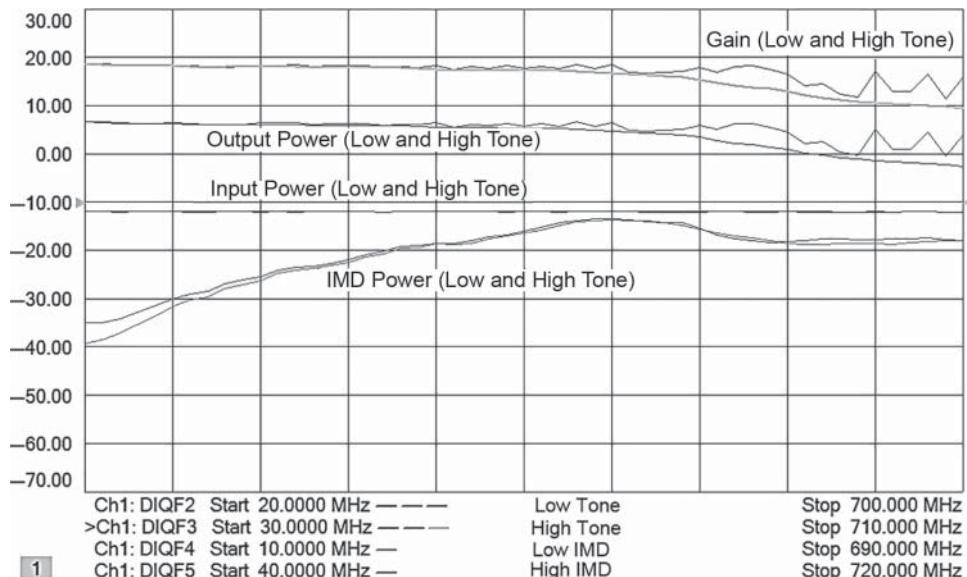


Figure 10.50 Swept-frequency IMD power of a differential amplifier, measured without baluns.

With this recent capability of creating multiple phase-controlled signals, calibrated at a reference plane, and measuring any set of other frequencies at other ports, the ability to make a wide range of differential measurements, even at mm-wave frequencies or in an on-wafer application (where it is difficult to use baluns), is greatly enhanced.

10.7 Noise Figure Measurements on Differential Devices

In previous sections it was demonstrated for low-power measurements, the linear behavior of a differential amplifier, measured as SE mixed-mode was identical to that measured in true-mode drive conditions. And for normal, differential amplifiers, the distortion and compression characteristics were also the same. For amplifiers that were not normal differential amplifiers, for example amplifiers that had limiting on the input or unequal gains on the different forward paths, the measurement at lower power matched between SE mixed-mode and true-mode, but the gain compression versus power, and the IMD response did not match at all. For gain compression, the 4-port true-mode method could easily be used, and it was demonstrated that it matched exactly with the non-linear response of properly de-embedding a hybrid. This hybrid was then used to characterize the IMD response.

Noise figure is generally considered a small signal measurement, and the two inputs, gain, and noise power measurements, are always performed when the amplifier is operating in a linear mode. However, an analysis of the definition of noise figure makes it very clear that at the least a balun or hybrid must be used at the output of a differential amplifier or the noise figure results may not be valid. This is true for any type of differential amplifier, whether it is normal or limiting or anything in between.

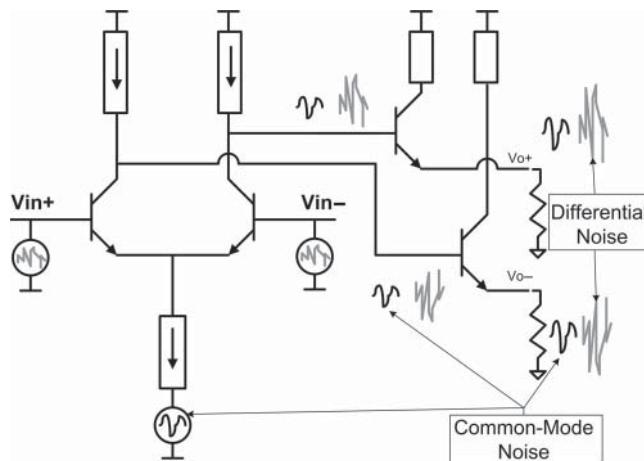


Figure 10.51 Amplifier with internal noise sources.

Consider the schematic diagram of an amplifier with internal noise sources shown in Figure 10.51. This amplifier has both common-mode sources of noise (such as noise in the common-leg of the current source) and differential sources of noise (noise at the input of the differential pair). In this system, the common-mode noise may not be amplified as much as the differential noise, but common-mode noise can be injected in any of the stages, which makes estimating the noise figure very difficult.

Also adding to the difficulties is that just as there are four gains in a differential amplifier, so too one might expect four different noise figures, depending upon the usage of the amplifier. The noise power at the output may be partially correlated (differential) or not, depending upon the internal characteristics of the DUT. In general, one cannot rely on single-ended measurements of the noise figure of an amplifier because there is no way to distinguish the differential noise (which will affect other stages) and the common-mode noise (which will not). Thus, some method is needed to provide true-mode noise-figure measurements.

There have been other methods proposed to make differential noise-figure measurements, based on the Y-factor technique, of multiport devices. In particular, a paper by Randa defines differential noise figure but does so only for the case where the detection, or measurement port, is purely differential, which does not apply to real-world measurements (Randa 2001). That is, in the theoretical configuration, the differential port is considered in isolation from the single-ended ports that comprise it. This is convenient for mathematical purposes but is not realizable in practice, unless one uses a hybrid or balun to separate the differential signal from the common-mode or SE signal. There have been proposals for an N-port representation of noise correlation matrix applied to a distributed amplifier, but not a differential amplifier (Moura et al. 2005). Other authors have proposed a de-embedding technique, based also on the Y-factor technique, which does not address the mixed-mode nature of the amplifier, and in fact, the amplifier that was used for proof of concept was not a differential amplifier in the commonly understood sense, in that the common-mode gain and differential-mode gain were identical, for a CMRR of 0 dB (Abidi and Leete 1999). This also does not represent an interesting case for differential amplifiers. Further, Y-factor provides only a noise-figure measurement and does not give any indication of the noise parameters.

10.7.1.1 Mixed-Mode Noise Figure

It is desirable to have a more complete measurement method for differential amplifiers that follows the ideas of a mixed-mode approach to noise figure of differential devices. Further, it is desirable that these noise characterizations be expressed in terms of noise parameters, thus illuminating the aspects of minimum noise figure, and optimum match for noise performance, in terms of mixed-mode parameters.

The previous discussion of differential signal waves, a_{d1} , a_{c1} , b_{d1} , b_{c1} , naturally leads to the concept of differential noise waves and further to a differential noise parameter matrix. From this groundwork, mixed-mode noise parameters can now be defined as

$$NF_{mn_{xy}} = \frac{S_{mn_y}/N_{mn_y}}{S_{mn_x}/N_{mn_x}} \quad (10.37)$$

where NF is the noise figure, m and n are the modes at the output and input of the device, respectively, and x and y are the output and input of the device, respectively. This can be re-configured, for example, to represent the differential noise figure as

$$NF_{dd21} = \frac{DUTRNPI_{d2}}{S_{dd21}} \quad (10.38)$$

where $DUTRNPI_{d2}$ is the differential relative noise power incident to a differential load connected balanced port 2 (as in other discussions of noise power, RNP is relative noise power, relative to kT_0B noise power at the input), and S_{dd21} is the mixed-mode differential gain. This configuration implies that to measure a mixed-mode noise figure, all which is necessary is a measurement of noise at the output, and another measurement of the device gain, each in the proper mode. This formulation represents the so-called cold-source method of noise-figure measurement, as discussed in Chapter 9, extended to differential amplifiers. This methodology removes the need for any noise sources to be used during the measurement process, which is used to advantage in developing the mixed-mode noise-figure measurement.

The method described next provides for full vector error correction of source and load effects, as well as vector noise receiver correction, because it employs the equivalent of a noise tuner (in the form of an Ecal module) during the noise calibration and measurement process. In fact, this system measures the noise parameters of the DUT and from this computes the noise figure in the system impedance.

One key attribute of the cold-source method is the ability to move the noise-calibration planes (which utilize coaxial noise-sources or power meters during calibration) to any reference plane for which an S-parameter calibration can be provided. Just as with compression and IMD, the 2-port noise calibration, combined with a hybrid de-embedding, provides a system that can characterize the mixed-mode noise-parameters and noise figure of the amplifier under test.

10.7.2 Measurement Setup

The setup from Figure 10.37 can be used for the differential noise measurement system, using scalar cold-source techniques, as described in Chapter 9. A tuner such as an Ecal module may be added behind the test port coupler for full vector noise measurements. The baluns

are hybrid circuits that have both sum and difference inputs. Note that the noise figure and noise-parameter calibration plane will be before the hybrid and thus uses only standard calibration procedures to provide a calibrated noise figure at this 2-port coaxial plane.

In this system, the noise parameters of the VNA noise receiver are fully corrected by using the tuner during the calibration stage to “pull” the noise receiver and establish noise parameters or noise correlation matrix, for it. Likewise, an S-parameter calibration is performed at the same reference plane, which provides a way to fully characterize the tuner for its impedances.

Then for differential measurements, one can use the 2-port equivalent of the hybrid input and output networks, based on the 3-port mixed-mode parameters of the hybrid, as shown in Figure 10.35, and de-embed them from the noise calibration. When a noise measurement is performed, only the modes (common or differential) of the input and output hybrid are exhibited in the results. Example S-parameter and noise-figure measurements are shown in Figure 10.52.

Also shown is the device under test relative-noise-power incident (DUTRNPI). This is the noise power delivered to a non-radiating, non-reflecting load, relative to the kT_0B noise floor. Highlighted in the upper-right window, two versions of DUTRNPI are shown. The first is the normal measurements, made with the tuner set to its nominal match condition; unfortunately, the source-match is rather high in this setup and so the DUTRNPI has some high variation at higher frequencies. Also shown is a measurement of the DUTRNPI with the differential input of the hybrid loaded with a well-matched load. Here the response is more in-line with the expected response, given the noise figure and gain. In practice, DUTRNPI should be equal to the S_{21} gain plus the input noise figure of the device (in dB), but DUTRNPI is not vector corrected for input mismatch. This parameter is useful in comparing noise powers out of a DUT in various conditions and can be thought of as the excess noise (above kT_0B) at the output of the DUT.

The noise figure is approximately 4.2 dB at 930 MHz and qualifies this as a relatively low-noise amplifier. This represents the case where the differential-mode input/output of the amplifier is measured.

In Figure 10.53, these same parameters are measured for the common-mode case of the amplifier. This is accomplished by using the same method, but in this case using the common mode (or summing) input of the hybrids. What is interesting here is that even though the common-mode noise figure is considerably higher (8.41 dB), the common-mode noise power, DUTRNPI (the excess noise), is in fact about 5.5 dB lower than the noise power measured the differential mode.

Still, the common-mode noise is much higher than one might expect, as the CMRR (differential gain/common-mode gain) is about 10 dB. If the noise were generated only at the input and were equal in each mode, we would expect the DUTRNPI to be 10 dB lower in the common-mode case: only 6 dB above the kT_0B noise floor. This result implies significant uncorrelated (in the differential sense) noise in the output stage or, rather, a significant amount of noise correlated in the common mode.

Figure 10.54 shows the noise figure of the amplifier for a single-ended case, where the input is at port 3, and the output is from port 2 of the amplifier, which is the highest gain path from Figure 10.8, with the other ports terminated in 50Ω .

The gain in this noise-figure measurement is 3 dB lower than the differential gain, but surprisingly, the noise figure is higher by about 1.7 dB. A common misconception is that measuring the single-ended noise figure will be a good estimate of the differential noise figure

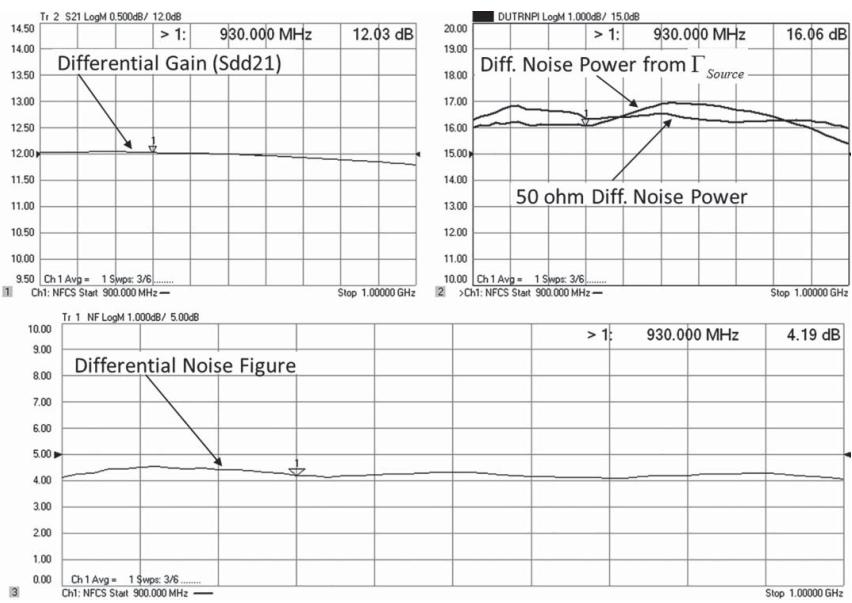


Figure 10.52 Noise figure and gain, and DUTRNPI measurement for a differential amplifier.

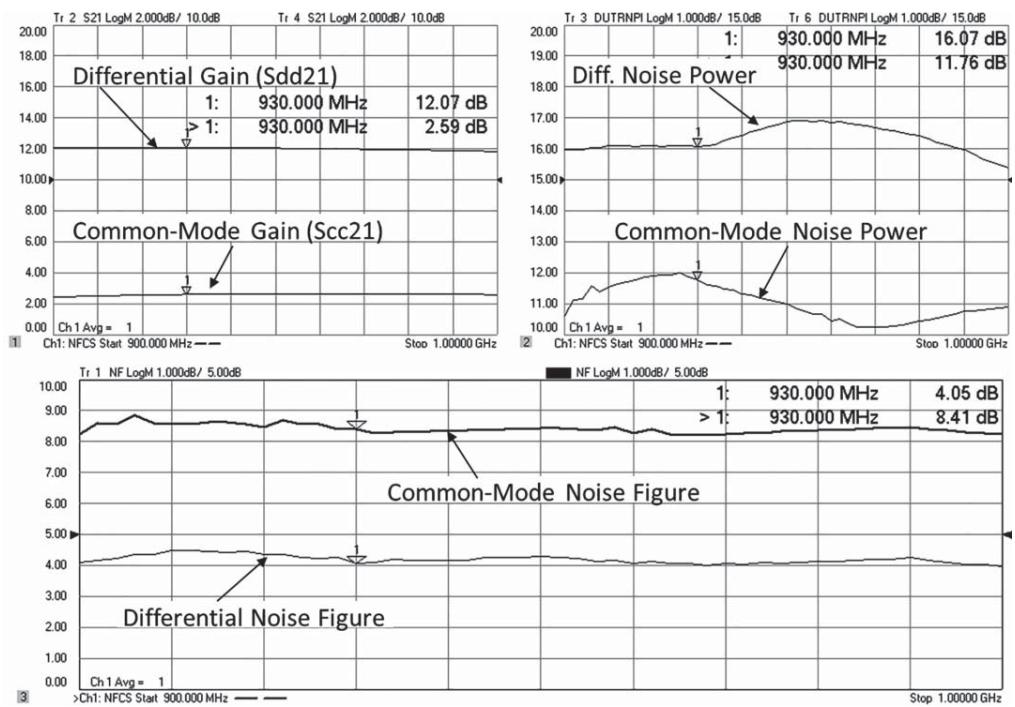


Figure 10.53 Common mode gain and noise figure.

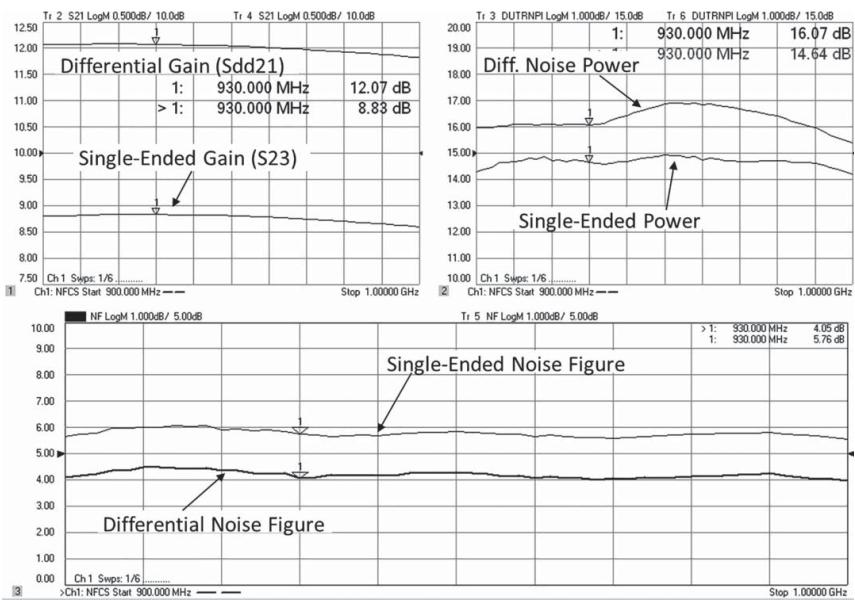


Figure 10.54 Noise figure measured on single-ended inputs.

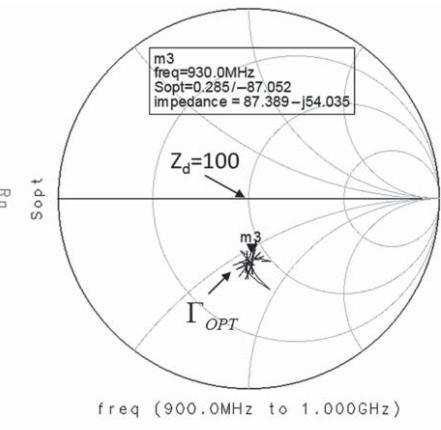
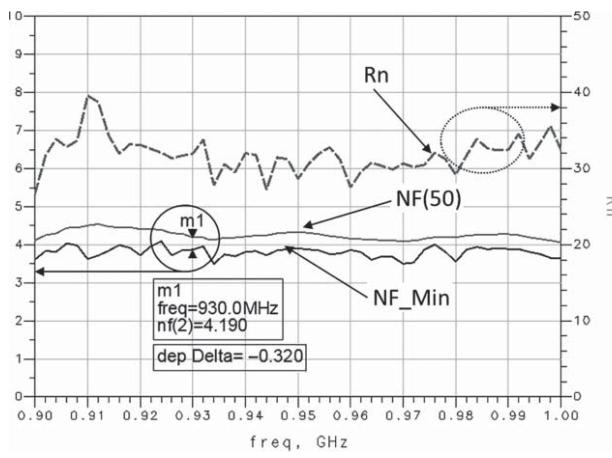


Figure 10.55 Differential noise parameters.

of an amplifier, but this experiment shows that such is certainly not the case. Even though the gain is 3 dB lower, the noise power is only 1.5 dB lower. One can estimate the contribution of the common-mode noise to the SE noise by taking half the differential noise, and adding it to half the common-mode noise (adding in the linear sense), since the noise power is the single-ended case is divided between two load resistors, then converting it back to dB, as in

$$DUTRNPI_{SE_estimate} = 10\log_{10} \left(\frac{10^{\frac{16.07}{10}}}{2} + \frac{10^{\frac{11.8}{10}}}{2} \right) = 14.44 \text{ dB} \quad (10.39)$$

This estimate gives a slightly lower than measured estimate of the SE excess noise, perhaps because the gain for this amplifier is quite uneven between the port pairs. Another reasonable source of difference may be because the DUTRNPI has some ripple error that changes as a function of source-match, as illustrated by the upper-right plot in Figure 10.52.

Finally, the noise parameters of a differential amplifier are shown in Figure 10.55. These can be extracted from the noise-correlation matrix of the vector noise figure correction. From this measurement one can see that the amplifier is very nearly matched for minimum noise figure over the gain region but could be improved by 0.3–0.4 dB. Because the tuner is applied to the differential input of the hybrid, only the differential impedance is varied during the measurement, yielding truly differential noise parameters, without depending on correlated noise sources. In this case, the noise calibration method uses a power meter for characterizing the receiver, so no noise source is needed at all.

From these noise parameters, it appears that some small series capacitance could optimize the noise figure for the matched $Z_d = 100 \Omega$ case, on the order of

$$C_{Match} = \frac{1}{2\pi f \cdot Z_{Match}} = \frac{1}{2\pi \cdot 930 \cdot 10^6 \cdot 54} = 3.17 \text{ pF} \quad (10.40)$$

Similar experiments on normal differential-amplifiers also yield unexpected excess common-mode noise in the output. From these experiments, it is clear that one must detect only the differentially correlated noise at the output to have a good estimate of the noise figure for a differential amplifier, as only the differential noise will be detected by the following stage, provided the following stage has common-mode rejection as well.

10.8 Conclusions on Differential Device Measurement

In this chapter, measurement methods were discussed for the differential device characterization, from which a few conclusions can be drawn.

For normal differential devices, ones that have common-mode rejection and do not have non-linear behavior in the input sections, the results from measuring single-ended-mixed mode measurements and from fully differential true-mode measurements are nearly the same. But this is true only for measurement methods (like the VNA), which can provide a mathematical difference (in the vector sense) for the output measurements.

For measurements that do not provide a magnitude and phase result, such as IMD or noise power and noise figure, one may sense the differential output using a hybrid or balun. Using modern techniques, the non-ideal hybrid effects can be almost completely de-embedded from the measurement. If the DUT has significant common-mode or cross-mode terms and does

not have a good common-mode termination, then there may be some resulting error due to common-mode reflection and mode conversion.

For devices that are *not* normal differential-amplifiers, for example an amplifier with limiting on the input stage or before a differential stage, they must be measured with true-mode methods, which include true-mode stimulus VNAs or incorporating the use of hybrids or baluns to generate and detect the true-mode signals. With these techniques, a complete characterization of differential devices is possible.

However, recent advances in the VNA capability now allow many non-linear measurements, such as harmonics or even IMD, to proceed without the use of baluns (as in the Keysight “Differential and IQ” measurement application) and thus provide a wideband method for performing these difficult characterizations.

References

- Abidi, A.A. and Leete, J.C. (1999). De-embedding the noise figure of differential amplifiers. *IEEE Journal of Solid-State Circuits* **34** (6): 882–885.
- Bockelman, E. and Eisenstadt, W.R. (1995). Combined differential and common-mode scattering parameters: theory and simulation. *IEEE Transactions on Microwave Theory and Techniques* **43** (7): 1530–1539.
- Dunsmore, J. (2003). New methods & non-Linear measurements for active differential devices. In: *Microwave Sym. Digest, 2003 IEEE MTT-S*, vol. **3**, 1655–1658. Piscataway, New Jersey: IEEE.
- Dunsmore, J. (2004). New measurement results and models for non-linear differential amplifier characterization. In: *Conference Proc. 34th European Microwave Conference*, vol. **2**, pt. 2, 689–692.
- Dunsmore, J., Anderson, K., Blackham, D., “Complete True-Mode Measurement System”, Symp Digest, 2008 IEEE MTT-S 2008.
- Moura, L., Monteiro, P. P., Darwazeh, I., “Generalized Noise Analysis Technique for Four-Port Linear Networks,” vol. 52, no. 3, pp. 631–640, Mar. 2005.
- Randa, J. (2001). Noise characterization of multiport amplifiers. *IEEE Transactions on Microwave Theory and Techniques* **49** (10): 1757–1763.

11

Advanced Measurement Techniques

For all the measurement techniques discussed, stable and reliable calibrations are the key to accuracy. There are many advanced situations where the basic coaxial calibration techniques are inadequate, usually because of the addition of external components or fixtures. A compendium of advanced techniques is discussed in this chapter.

11.1 Creating Your Own Cal-Kits

Radio frequency (RF) and microwave test engineers are often called upon to test a device that does not have connectors for which standard calibration kits exist. The most common case is for parts that are mounted on PC boards, followed by parts with specialized connectors such as blind-mate or push-on connectors. While there are many ways to deal with this situation, the most straightforward is to create a calibration kit for the particular interface of the device-under-test (DUT).

The two main types of calibration methods, short-open-load-thru (SOLT) and thru-reflect-line (TRL), each require different calibration standards. TRL is very straightforward requiring only an unknown reflect, and two transmission lines of known impedance, and is often applied to PC board test situations during R&D development where designing test boards with different length lines is possible. For low-frequency measurements, TRL is replaced with thru-reflect-match (TRM), and an ideal load element is needed as well, since very long line lengths become impractical. In manufacturing situations, where the test fixture may be embedded into a part handler, there may not be a way to create sufficiently long lines, so TRL cannot be applied.

SOLT or SOLR (unknown thru) require known standards for reflection, typically an open, short, and load, although sets of offset shorts can be used as well. For in-fixture calibration, where the fixture is short, the method of TRL is not practical, and so fixed standards become required. If a thru with good impedance can be created, TRM is preferred; if not, then SOLR using unknown thru is preferred. In this section, a complete PC board calibration kit is defined, and various methods are used to determine the quality of the calibration that results.

11.1.1 PC Board Example

An example PC board used for SMT part characterization is shown in Figure 11.1. This board has a short, two loads, and an open. It is not wise to have an open line near the short standard, as coupling between equal-length lines can cause resonances to appear in the standards. For this example, two different loads were created. The upper load was created with a two surface-mount technology (SMT) 100Ω resistors (1206 package) to ground, making a 50Ω impedance. The lower load was created using one 50Ω SMT resistor to ground. The SMA connectors for each device are soldered to top and bottom grounds. For this example, the single-resistor load will be used.

The open standard typically sets the reference plane. The thru standard is designed to be exactly twice the length of the open, but sometimes the modeling of the open adds delay to the open so that the values for fringing capacitance are reasonable. In this case, the thru might be longer than the open. To ensure a gap between the load and the open, the board dimension is increased for the portion of the board containing open and load, relative to the thru. For this board, the load was set to a shorter length on the assumption that the load is zero reflection and so its delay doesn't matter. But the evaluation will show that the load has significant reflection; by matching this to a model for the load, even a poor load can be used as a precision calibration standard. In such a case, the modeling method to incorporate the load inductance requires that the load position sets the reference plane. Also shown (far right in the figure) is a shunt device connected to a line that matches the thru line. The short standard is made by placing printed circuit board (PCB) vias at the same relative reference as the open.

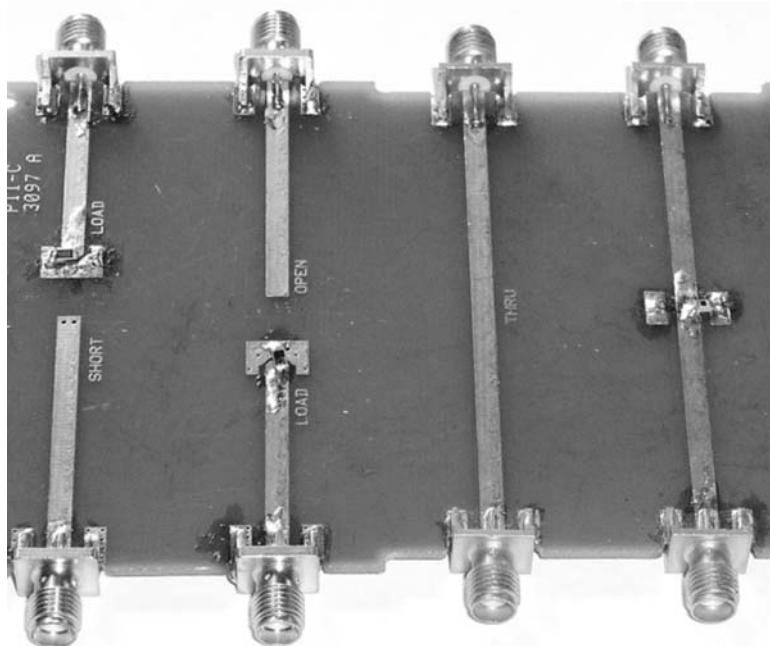


Figure 11.1 PC board designed for SMT part characterization.

11.1.2 Evaluating PC Board Fixtures

The investigation of PC board fixtures and standards really must start at the PC board fixture itself, more specifically, the PC board connector. A very common method of creating PC board fixtures is to create a board as in Figure 11.1 with sets of standards, lines, and even PCB RF features like embedding filters, and then measure each and try to determine their characteristics. But a critical and often ignored element in this work is the quality and especially repeatability of the SMA-to-PCB connector. Here, quality refers primarily to its return loss, and repeatability refers to the how one SMA connector compares to each of its neighbors. In all the work on PCB fixtures, it is always assumed that once they are calibrated out, the connector can be ignored, but this is only the case if each is identical. So, the first step in evaluating PCB fixtures is to look carefully at the connector repeatability. In this example, there are six total connectors for the standards. One can readily compare them by using a time-domain approach.

The general approach is to use time-domain gating to isolate the return loss of each connector. Using one connector as a reference, the gated response of each other connector will be compared to the reference connector to determine its vector difference. This difference displayed in dB will represent the connector repeatability. This fixture is intended to be used up to 6 GHz, but for proper time-domain analysis, a much broader frequency sweep is used; in this case 26.5 GHz. Using a frequency sweep of four times the intended operating frequency gives reasonable resolution of important features.

11.1.2.1 Characterizing the Thru Standard

The first step is determining the length of the fixture, so that the proper gates can be set, and for this the thru standard is used. Figure 11.2 shows several measurements of the thru: the upper-left window shows the time-domain response of transmission (S_{21}) with a marker at the peak (this is sometimes called T21). This is the total length of the fixture. The lower-left window shows the time-domain response of S_{11} and S_{22} . Note that marker 1 is set to the same time value (half the physical distance) as the S_{21} trace, this time represents exactly halfway through the fixture; remember that in reflection, signals travel the twice the distance to S_{11} and S_{22} as they must go forth and back from the reflection.

For reasons described in Chapter 4, it is best that the time gate be as wide as practical and centered on the reflection to be gated around. A bandpass gate is used, with the gate center-time set to 100 ps and the gate span set to 1040 ps, so the gate stop is near the center of the fixture. This gate will show the response of just one connector, the input connector for the S_{11} trace and the output connector for the S_{22} trace.

The normal S_{11} and S_{22} frequency response data for the thru is shown in the upper-right window, and the gated response for S_{11} and S_{22} is shown in the lower-right window, along with a trace that is the difference between the gated responses.

This shows us that each connector is similar, with about 28 dB return loss at 3 GHz, and about 20 dB return loss at 6 GHz. For the purposes of this exercise, the calibration kit will be determined for best performance up to 6 GHz. The difference shows about 33 dB return loss at 6 GHz; this difference represents the connector repeatability, and one cannot get better calibrations than this value. The same reference trace will be used for each of the other standards to determine their repeatability to the thru standard.

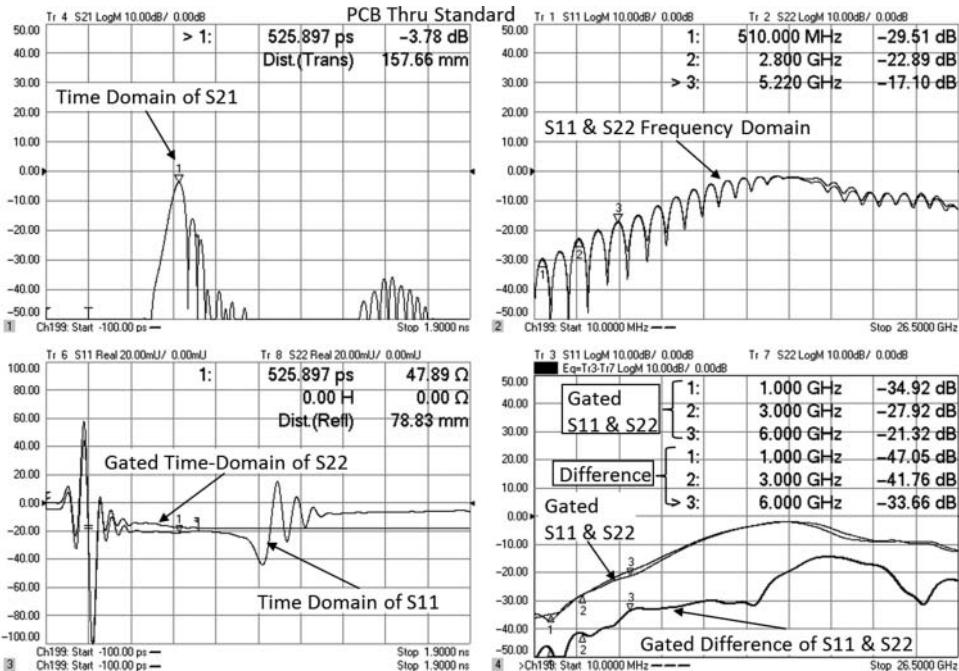


Figure 11.2 Analyzing the thru standard.

11.1.2.2 Characterizing the 1-Port Standards

The investigation of the 1-port PC board standards starts with the load, which sets the quality of the calibration for both the SOLR and the TRM calibrations. Figure 11.3 shows a similar measurement as earlier for the connector on one of the load standards. The upper left is S_{11} of the connector and load element. The lower left is the time-domain response, which shows the load response occurs just slightly before the halfway point of the thru (70.9 vs. 78.8 mm). The time-domain response peak gives a good first estimate of the delay to the load, but it is not necessarily the best way to establish the reference plane, as will be shown next.

The upper right shows the gated S_{11} response for the connector on the load standard. Here, the response is slightly worse than the gated response for the thru connector, shown in the lower-right window, along with the vector difference. In this case, the difference appears to be worse than between the two ends of the thru connector.

Further investigation is needed to understand the difference. From Figure 11.2 (lower left), the time-domain response shows a peak and dip near the connector. The response from Figure 11.3 shows only a dip. The peak-and-dip response of the thru connector may be improving the return loss in the low frequency region, at the expense of much worse return loss near 14 GHz, where it approaches 0 dB. A common mistake is to leave a gap in the backside ground between the connector and the ground plane of the PC board; this gap is typically required for normal PC mounting but will often cause inconsistency between standards, and this shows up as larger residual directivity and residual match. Inspection

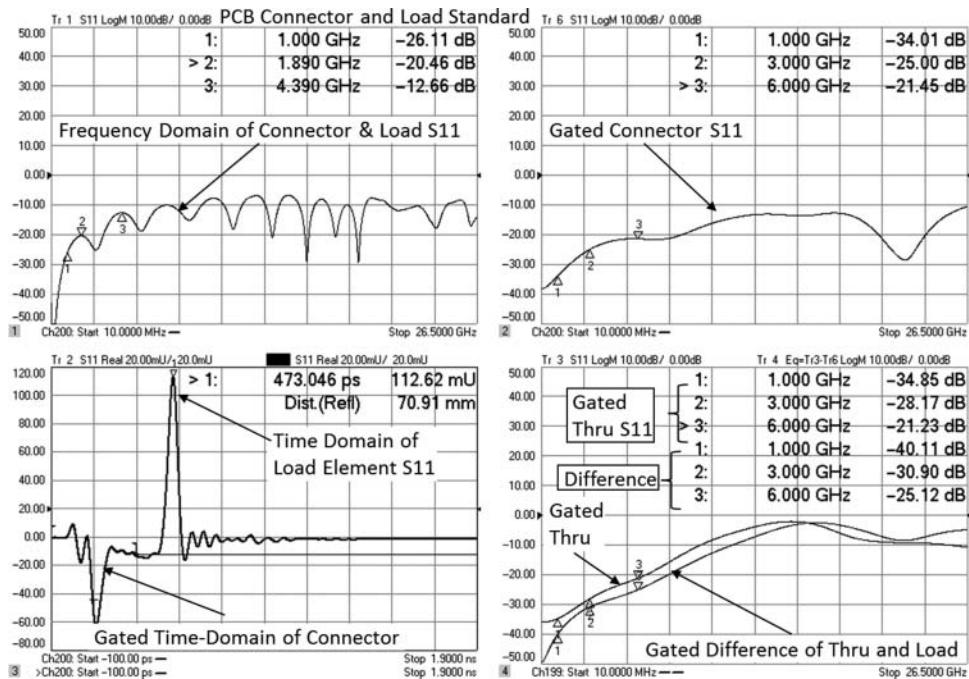


Figure 11.3 Comparing the load connector to the thru connector.

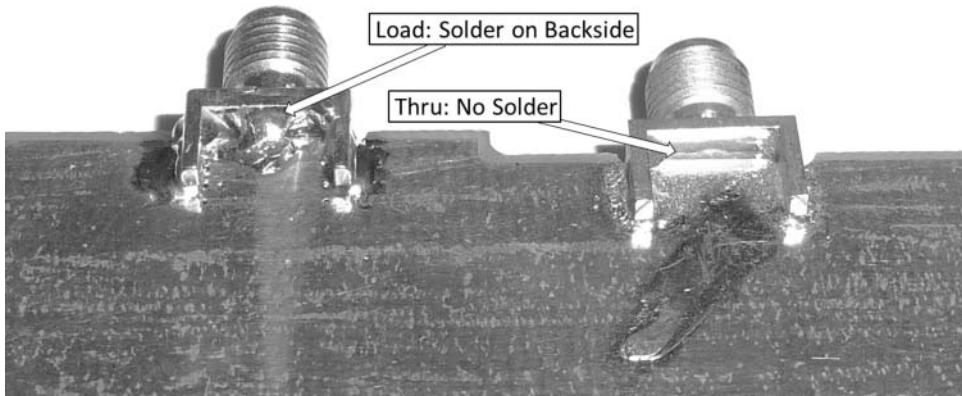


Figure 11.4 Comparing load connector soldering to thru connector.

of the PC board, as shown in Figure 11.4, reveals that the ground plane on the back of the connector for the thru is not soldered (as is normal), but the load is. The load connector has better high frequency performance and a lower deviation in the time domain.

The difference between the load connector and the thru connector is quite large, even though the S_{11} traces are not that much difference, because of the phasing of the reflections. Here, the best-case residual connector repeatability is only -25 dB, rather than -40 dB above.

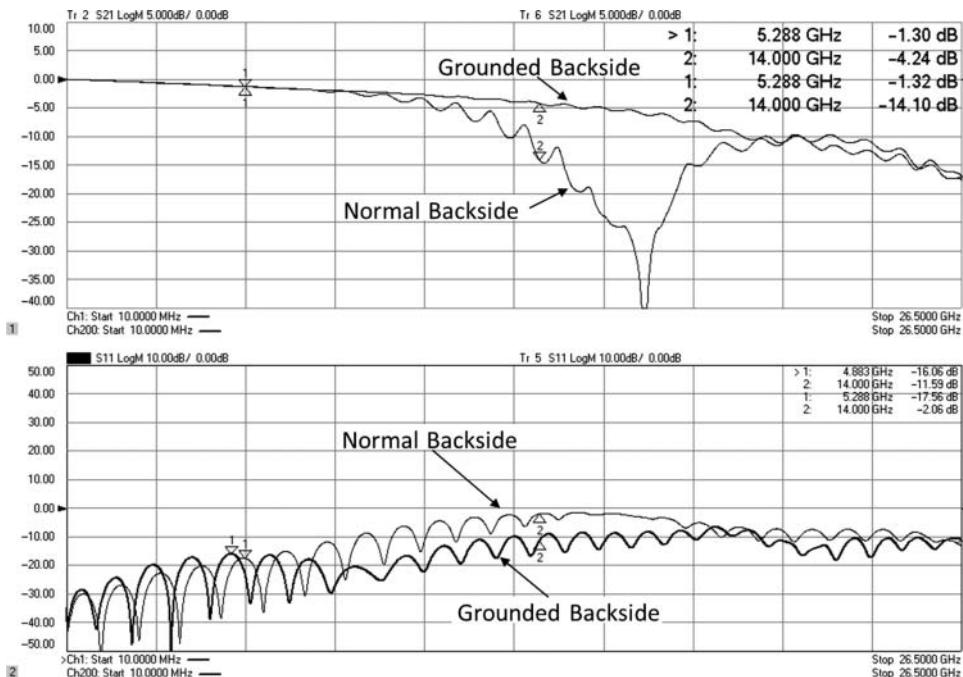


Figure 11.5 Thru measurements with and without ground solder.

The thru element should be modified to match the load element. In general, great care must be used to investigate and repair any differences between PC board fixture connectors to obtain good calibrations with them. The thru standard was measured before modifying the ground, and again after the ground of the SMA was soldered; Figure 11.5 shows the comparison.

The normal backside grounding actually shows better low-frequency response, up to 6 GHz, but has much worse response up at high frequencies, where it has a very large resonant loss in S_{21} . It is likely the case that the large inductive peak in the lower-left window of Figure 11.2 compensates for excess capacitive response, canceling somewhat the reflection at lower frequencies. With the thru connector re-soldered, it can now be compared again with the load connector, as shown in Figure 11.6. Here one can see that at 3 GHz, the difference is much improved to a worst case of -38 dB repeatability; however, the 6 GHz performance is nearly the same with one connector achieving -30 dB but the other still at around -25 dB. It appears that this level may be the limit for these connectors using normal care in assembly.

11.1.2.3 Investigating the Load Standard

From Figure 11.3, further investigation into the load is needed, as its time-domain response is very high relative to the connector. Figure 11.7 shows the result of applying a notch gate to the connector to remove its effect on the load. The upper right shows the resulting S_{11} trace with rather poor load response at 6 GHz. One might normally think this load would be

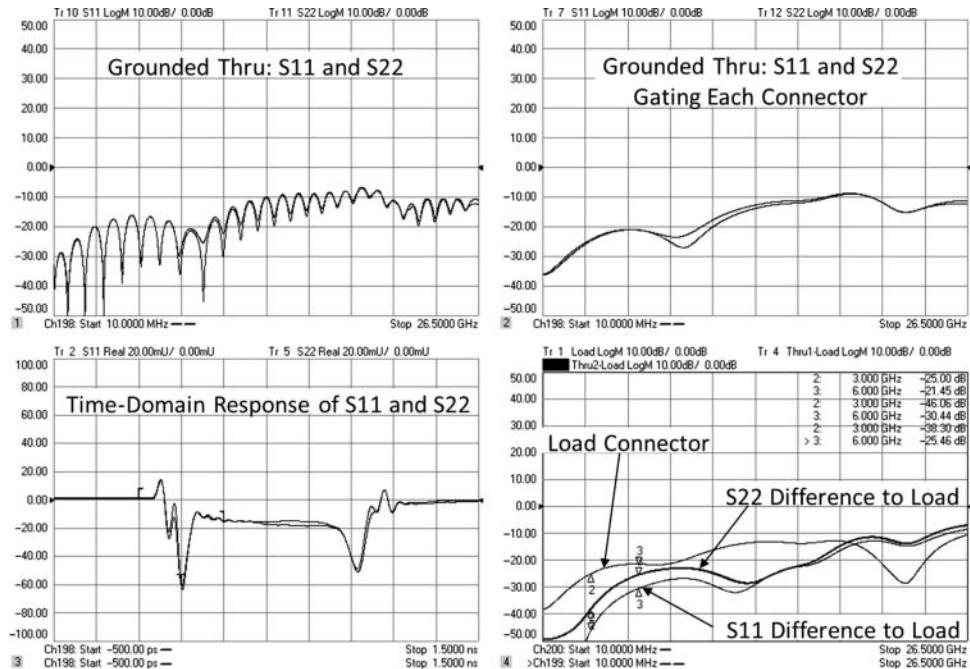


Figure 11.6 Difference between each end of the thru and the load connector.

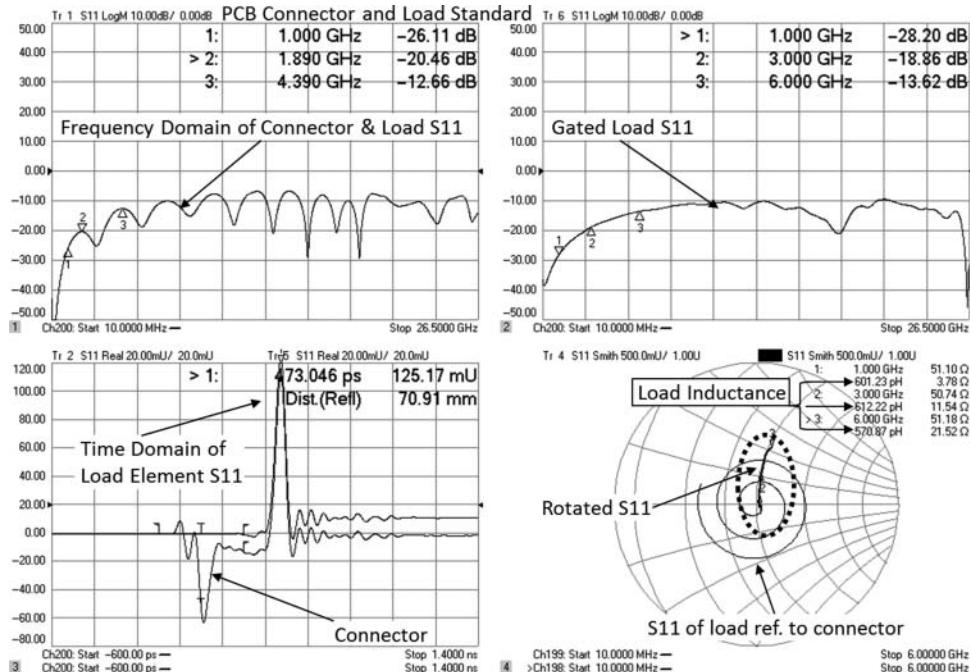


Figure 11.7 Measuring just the load by notch gating the connector.

unacceptable as calibration load, but in fact it may be satisfactory if it can be characterized. This is the single-SMT resistor, and previous publications indicate that such a configuration is typically too inductive.

The lower-right window of the figure shows a Smith chart plot of the gated S_{11} of the load, referenced to the input connector. For this plot, data is displayed only up to 6 GHz. Electrical delay of 462 ps was added to rotate the S_{11} plot until the values of the real part at several frequencies align as closely as possible (to nearly 50Ω in this case), so the trace follows a trajectory of constant resistances in series with an inductor. The equivalent value of the inductance is highlighted, and it too is reasonably constant, at about $590\text{ pH} \pm 3\%$. This is a very good estimate of the series inductance of the load. With this information, one can make an equivalent model of the load that works in almost all VNAs.

The model of a load was given in Chapter 3 as an offset-impedance transmission line in series with a resistor. The impedance of the transmission line and its length can be modified to account for the series inductance of the load found in the previous figure. The impedance and velocity factor of a transmission line are given as

$$Z_{Line} = \sqrt{\frac{l}{c}}, \quad v = \frac{1}{\sqrt{l \cdot c}} \quad (11.1)$$

where l and c are the inductance and capacitance per unit length. The delay of a transmission line is given as

$$\tau = \frac{d}{v} \quad (11.2)$$

where d is the distance or length of the transmission line. From these, one can compute an equivalent inductance to a line by

$$\begin{aligned} L_{Equivalent} &= l \cdot d \\ &= \sqrt{l} \sqrt{l} \frac{\sqrt{c}}{\sqrt{c}} d = \sqrt{\frac{l}{c}} \cdot \left(\frac{\sqrt{l \cdot c}}{1} \right) d = Z \frac{d}{v} \end{aligned} \quad (11.3)$$

Thus, one can simply define a line length and impedance to compute an equivalent inductance. The approximation is most accurate for high impedances; most VNAs allow up to 500Ω for the offset impedance of a load, so the delay for a given inductance is computed as

$$\tau_{Load_L} = \frac{L_{Equivalent}}{Z_{Offset_Line}} = \frac{L_{Equivalent}}{500} \quad (11.4)$$

Similarly, for a condition where the load has some shunt capacitance, the offset delay can be set to a very low impedance as

$$\begin{aligned} \frac{1}{Zv} &= \sqrt{\frac{c}{l}} \cdot \frac{\sqrt{l \cdot c}}{1} = c \quad F/m \\ C_{Equivalent} &= c \cdot d = \frac{1}{Z} \frac{d}{v} = \frac{\tau}{Z} \end{aligned} \quad (11.5)$$

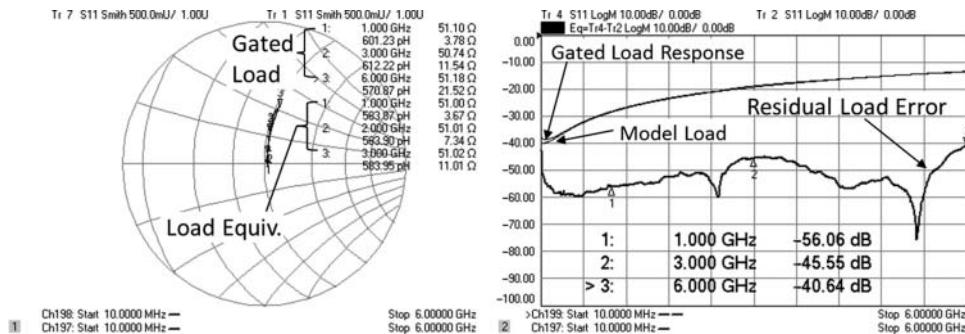


Figure 11.8 Residual load error from equivalent load.

For this example load, the equivalent inductance is 590 pH, giving an offset length of 1.18 ps for an offset impedance of 500 Ω . The key to the quality of calibration is match between the actual load and the model of the load. A simulation of such an offset load with a 500 Ω line terminated in 51 Ω was created and the data loaded into a trace on the VNA. Figure 11.8 shows the measured response of the time-gated load and the model of the calibration load on a Smith chart in the left window, and the return loss (in dB) of each, as well as the return loss of the vector difference, in the right window.

The characterization of this load is better than 40 dB up to 6 GHz, thus it is a fine load for use in calibration, especially considering that the connector repeatability is no better than -25 dB, from Figure 11.6.

11.1.2.4 Open/Short Characterization and Modeling

The final elements to be analyzed are the open and the short. The connector repeatability for each of these is compared to the load, with the results shown in Figure 11.9, with the worst case being on the order of -28 dBc at 6 GHz, and around -33 at 3 GHz. These values determine the overall quality of the PC board cal. This uses a time gate around the input connector. Using time gating when the line is terminated in a very large reflection, such as an open or short, is somewhat unreliable as the assumption that the response returns to zero at some far time is not really valid. The reason for this is the periodic nature of the VNA transform. Since the response must be periodic, signals from outside the time range of $1/\Delta f$ can fold back in an aliased way and distort the response of interest.

The final task is to determine the offset delay for the short and the open, as well as any excess fringing for the open. The same data is used as for the previous figure, but in this case a notch gate is applied to remove the effects of the connector, and the responses are shown in the Smith chart. Each response is rotated until the markers read a consistent value. For the short, we expect nearly zero inductance. For the open, the capacitance should be small, typically less than 100 fF. To find a model for the open and the short, first the approximate delay to the each must be determined. A simple way to do this is to use a marker search to find the point in the time-domain trace where the magnitude is 0.5 (-0.5 for the short). This is in the middle of the rise time of the standard, as illustrated in the lower window of Figure 11.10. A time-domain gate is placed around this region, to remove the effects of the

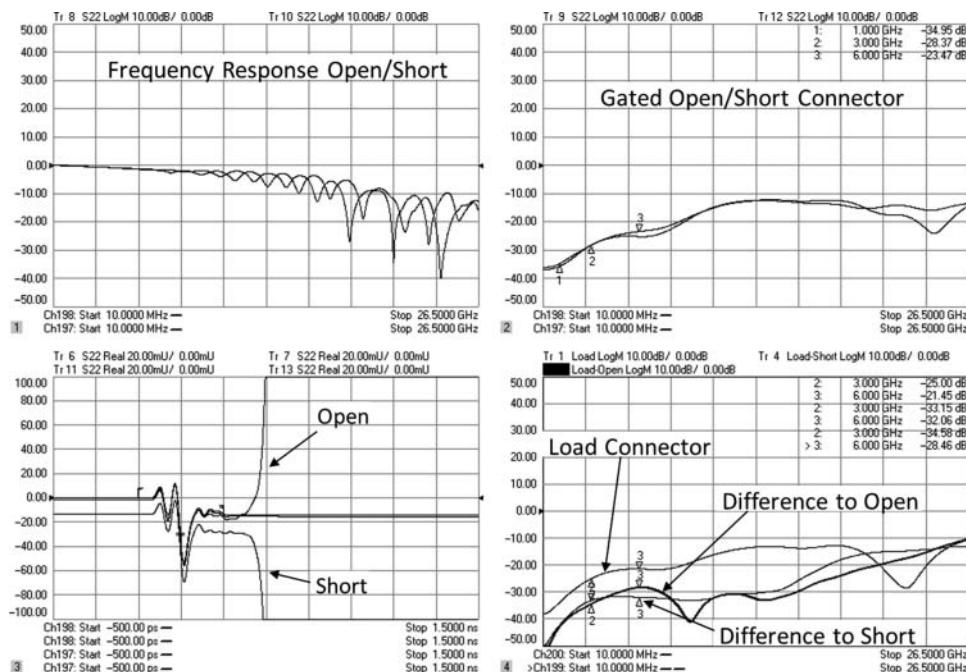


Figure 11.9 Open and short connector repeatability.

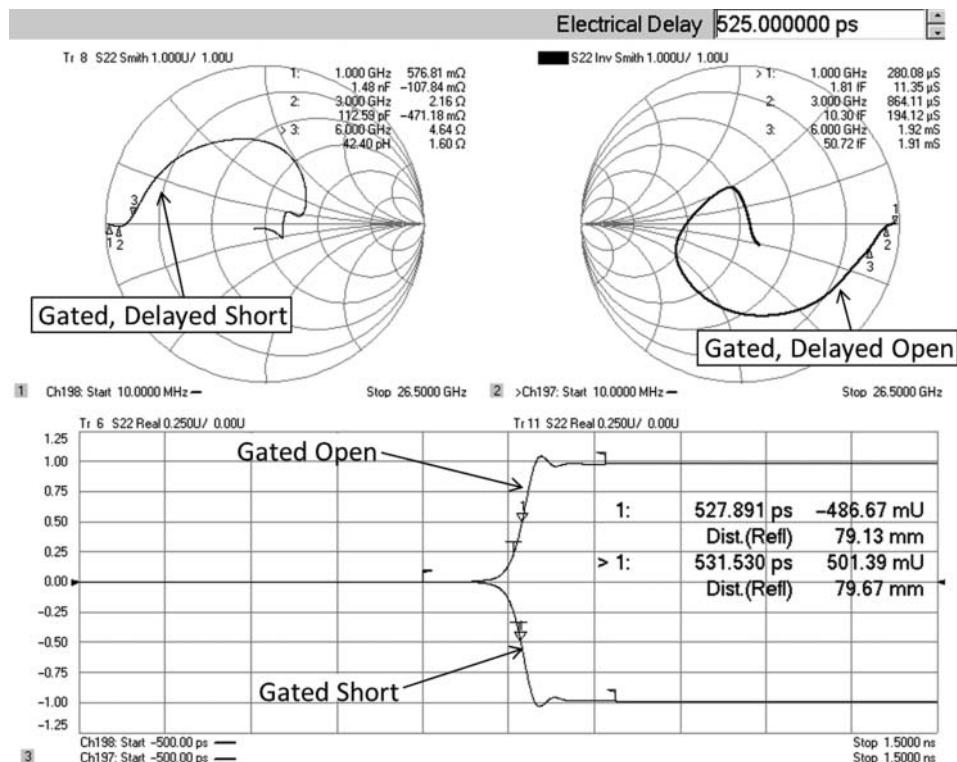


Figure 11.10 Finding a model for the open and the short.

connector, and any re-reflection further in time. In this case, the gate center is at 525 ps, with 450 ps of gate span.

The short is shown in the upper-left window, and the open in the upper right. The gated responses have been rotated by 525 ps of electrical delay. This value was chosen so that the marker values seem reasonable; this is slightly less than the time found in the time-domain trace marker, because there is some extra fringing capacitance for the open and some excess inductance for the short. The exact setting is up to the engineer's judgment, but it is good practice to try to achieve a small but constant reactance. In this case, it was not possible and the open and short were set to similar values that kept the marker phase close to zero or just slightly positive for the short, and just slightly negative for the open, without causing loops or clockwise rotations of the Smith chart. More delay would cause the traces to reverse; less delay results in excessive capacitance value. The loss of the open and short don't play a major role in the determination of the fringing elements, but using some form of loss compensation can help to generate a more consistent fringing element. Loss compensation is discussed in Section 11.4.

Now the models for the cal-kit can be determined.

11.1.2.5 Creating Cal-Kit Models

Since it is not ideal, the load element will set the reference plane. The open and short can be offset by an arbitrary amount using a Z_0 line, but the load must use the offset line to accommodate its series inductance, and only a single offset is offered in most cal-kits, so the reference of the load will be the reference of the kit. The load element values from earlier are $R = 51 \Omega$, and $Z_{Offset_delay} = 500 \Omega$, and the offset delay is $\tau = 1.18$ ps. The offset also provides for a loss, but at these levels, using the default loss is fine. These values, along with the open, short, and thru, are tabulated in Table 11.1.

In rotating the load to find the proper reactive element, 462 ps of electrical delay was added. For the open and short, 525 ps of delay is added to achieve the plot in Figure 11.10. Values for the open and short were approximated and matched to the equivalent models as described in Section 3.3.2. In this case, only three of the four coefficients were used, and in an example such as this, using even that many elements may be pushing the validity of the data. Since the measurements of the standards are lossy, the true values are also somewhat different than those shown in the figure.

Table 11.1 Cal kit values for example PCB

Standard	Element	Offset Z0	Offset Delay (ps)	Equivalent
Load	$R = 51$	500	1.18	585 pH
Open	$C0 = 0, C1 = 200,$ $C2 = 1000, C3 = 0$	50	31.5	
Short	$L1 = 0, L2 = 0,$ $L3 = 1000, L4 = 0$	50	31.5	
Thru		50	63	

The most straightforward method is to take three or four values and use them to solve directly for the coefficients, but this can sometimes lead to ill-conditioned values especially in the case of the inductor previously shown, where the first two frequencies are actually slightly negative in reactance. In this example, some matching to a 3-element model yielded the curves found in Figure 11.11 for the values given in Table 11.1; these values were adjusted from the directly computed values to avoid the situation of have two very large coefficients with opposite signs, and so the fit is not exact, but the curve's response is well controlled. Values computed directly from measurements using matrix techniques will create a curve fit that exactly matches the measured values but can have very large coefficients some positive

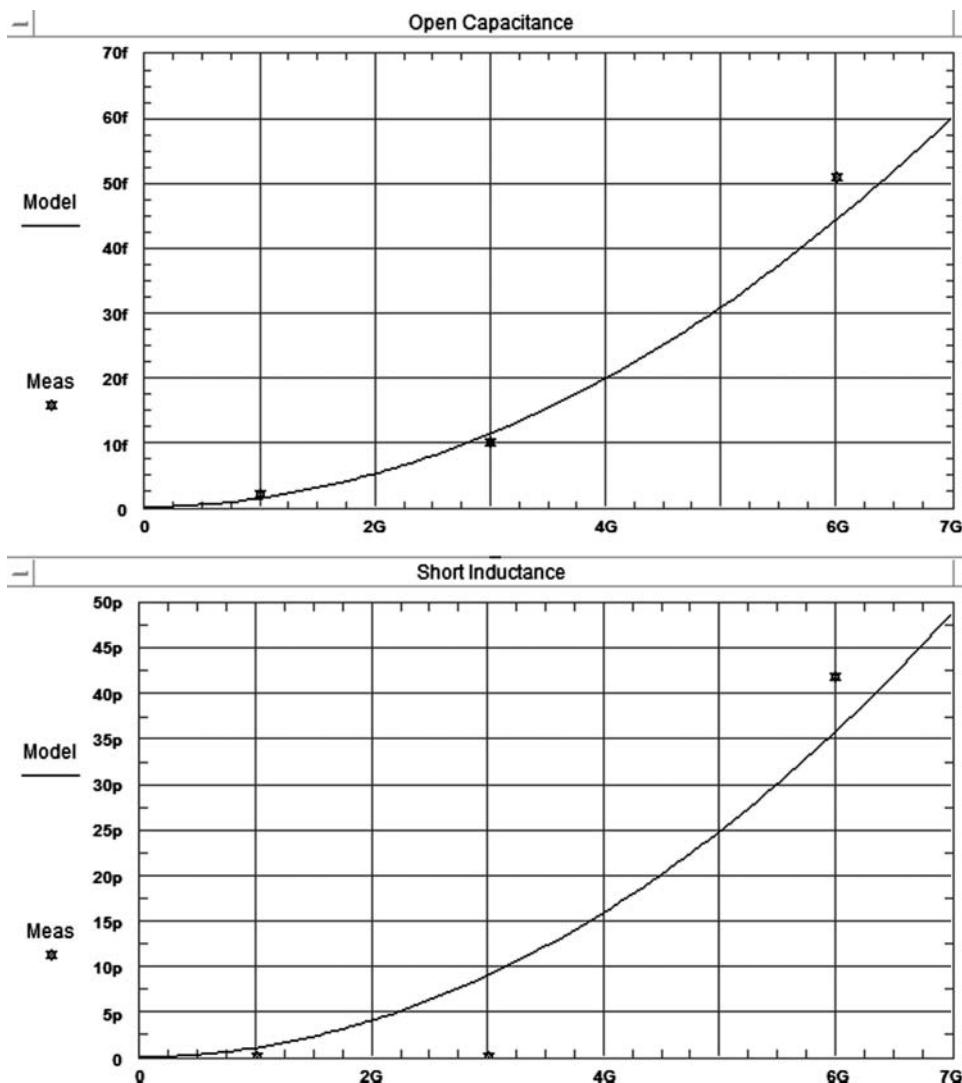


Figure 11.11 Models versus measured for capacitance and inductance.

and some negative, which almost completely cancel to give an appropriate curve between data points, but any extrapolation is very poor. Thus, unless one uses measured points at the extremes of the frequency range, one must be careful that the polynomial model does not “blow up” outside the exact measured ranges.

The offset delay for the open and short is one-half the difference between the electrical delay of the load and of the open or short. One must use one-half the delay, as electrical delay is set as the total round-trip delay, but the standards use only the one-way transmission delay. If port extensions is used to rotate the phase (port extensions operate on all traces, electrical delay operates independently on each trace), then the 2-way response is implied in the extension delay and the value read from the port extension is used directly. The offset delay values are given in Table 11.1 as well.

Finally, since the physical thru standard is twice as long as the open or short and longer than twice the load reference, a defined delay must be added, and in this case it is simply twice the open or short delay.

Using this calibration kit for the PCB calibration will give the reference plane at the load port, and measuring the thru, one will see 63 ps of delay. If it is desired to have the reference plane at the center of the thru line, one may simply add 31.5 ps of port extension to each port.

11.1.2.6 Database Standards

All the work in the previous section to match the measured response of a calibration standard to a model for the calibration standard is required only when using older model VNAs. Modern VNAs have the capability of using database models instead of the traditional polynomial model. A database model utilizes the frequency response information, either from a simulated model (e.g. 3-D EM model) or from direct measurements. The input for 1-port database models are either CITI files or S1P files. The input for the thru-standards is either 2-port CITI file or an S2P file.

For example, to create a database model of the load, one would use the result of the rotated, time-gated load shown in Figure 11.7. Save this result as an S1P and utilize the data directly. If this is our best estimate of the load response, then using it as the model for the load is the best choice. Similarly, using databased models for the open and short, as developed from the data in Figure 11.10, will give good results for these somewhat complex curves.

In the case of the thru, it depends upon how the thru is developed. In the case of the PC board fixture of Figure 11.1, the thru is essentially ideal, so it is not necessary to use a database standard for the thru. But in some cases, the thru cannot be modeled ideally; in this case, it should be characterized as well as possible and a databased model used, with a small exception. If the thru is easily modeled in a simulator, for example an on-wafer Thru standard, then the error in the simulated model might be less than the measurement errors trying to characterize the thru; here a database model based on the simulated data would be better. But if we have good database models for the open/short/load standards, we don't need any model of the thru as the unknown thru calibration method can be used.

11.1.2.7 Measurement Results with a PC Board Cal-Kit

The PC board and cal-kit definitions can now be used to calibrate the VNA to the PC board connectors, and measurement on components at that reference plane can be performed. It

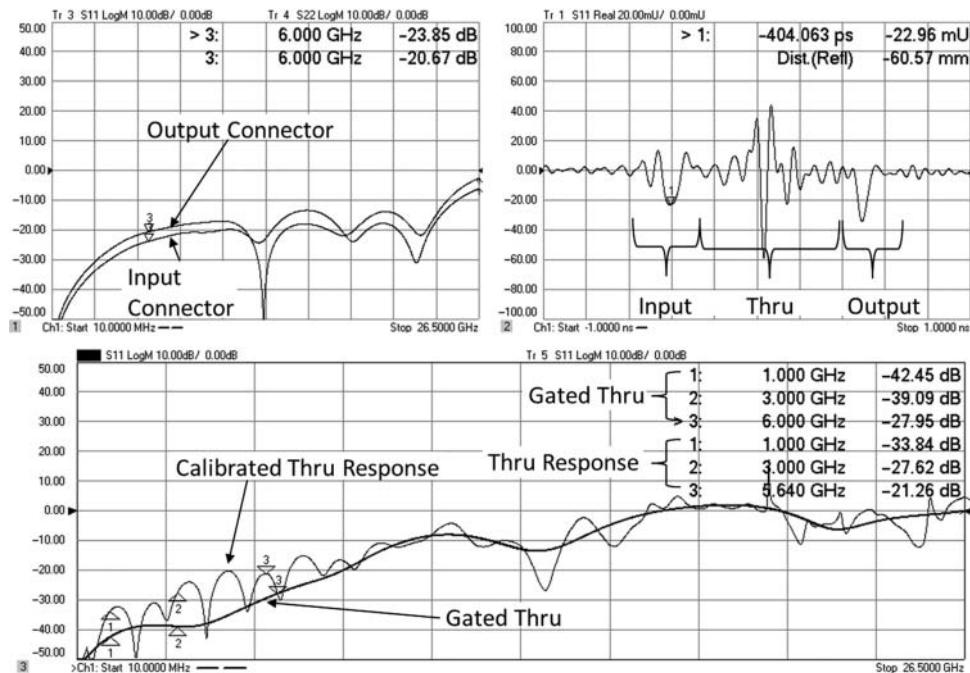


Figure 11.12 Measurement of the thru after an unknown thru calibration.

is highly recommended to calibrate over a very broad range at first to see if there are any issues with the cal-kit or standards. A common problem is a cal-kit is defined over a narrow range, for measuring a particular component in a particular application. Later, the same cal-kit is used over a different range without the user realizing the values had been optimized over only a narrow range. In particular, the fringing polynomials may extrapolate poorly if they were defined over only a narrow range. The results of using this calibration kit on a 26.5 GHz sweep is shown in Figure 11.12, after performing an unknown-thru (SOLR) calibration. In this case, there are large spikes in the results above 20 GHz, likely due to the open and short standards not maintaining sufficient separation at high frequencies. In such a case, the reflection from each might overlap, and the calibration becomes undefined at those points. This wideband sweep shows no issues well past twice the expected frequency of operation, 6 GHz.

Since the open, short, and load are all defined in the cal-kit, the unknown thru calibration can be used. After calibration, remeasuring the response of the thru can show a lot about the quality of the calibration, since it is essentially an independent standard. For these types of evaluations, it is still important to use wide frequency sweeps even if the cal-kit is intended only for use of a lower-frequency range; for example, this one is optimized for 6 GHz, but 26.5 GHz sweeps allow very good resolution in the time domain for further analysis. Figure 11.12 shows several views of the thru, with a key one, the time-domain response of S_{11} , shown in the upper-right window. In the lower plot, the calibrated thru response is shown for S_{11} of the thru. The definition of the thru should be 0 dB insertion loss and no return loss.

But here one sees a return loss of 21 dB near 6 GHz. Also shown in the plot is the time-gated response of the thru. Here now one sees a much better response, when the effects of the input and output connector are removed.

These connectors should have been removed by the calibration process, but because of connector-to-connector repeatability, residual connector match remains. The upper-left plot shows the results of time gating the input connector from the S_{11} trace and the output connector from the S_{22} trace. These are similar, though slightly worse, to the some of the individual estimates of connector repeatability in the previous figures. These represent not only the connector repeatability between connectors but also any errors in the estimation of the open/short/load, which will cause the characterization of the port to be in error. The primary error in the thru only (time gate thru response) is error in the load element, plus connector repeatability between the load and the thru (source match, which comes from the open/short measurements, does not greatly affect the return-loss of a well matched device like the thru). Thus, the measurement of the gated-thru is a good estimate of the residual error in the load. Residual error is the difference between the model of the load and the actual load performance. This measurement shows a very good load (nearly 40 dB) up past 3 GHz, with some degradation up to 6 GHz.

Measuring a device with a poor match will help to show the quality of the source math error term. For this example, a 100Ω resistor is added in shunt to ground at the center of the PC board. The response for this element is shown in Figure 11.13. Here the residual return loss for the input and output connector is better than in the thru case, indicating this set of

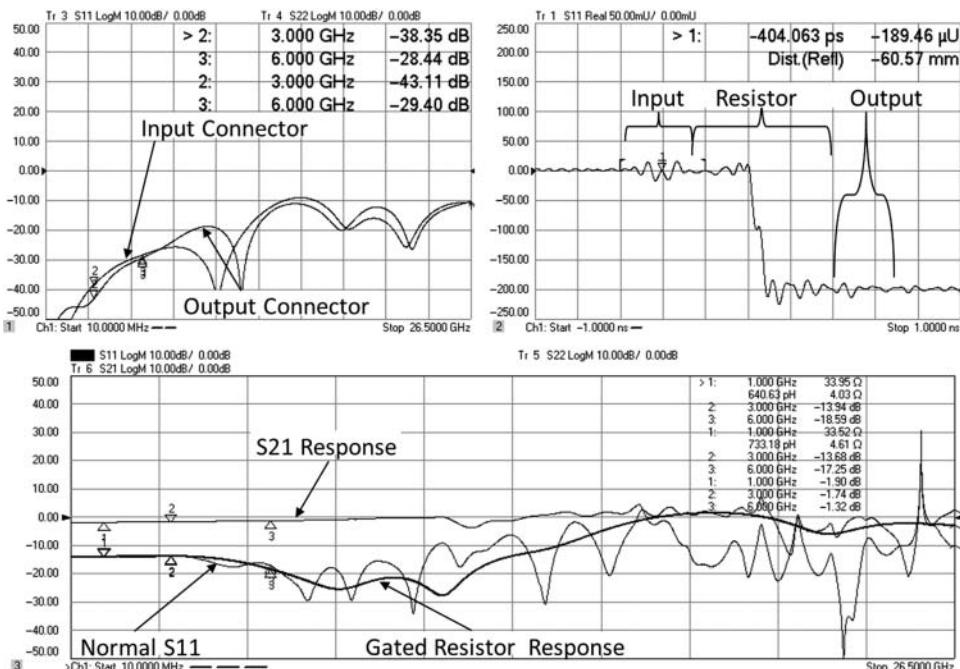


Figure 11.13 Measurement of a 100Ω resistor connected shunt to ground.

SMA connectors (far right in Figure 11.1) are much better matched to the SMA connector for the load, with the residual error around 40 dB for 3 GHz, and 30 dB for 6 GHz.

The normal S_{11} response, and an S_{11} response with the connectors gated out, are shown in the lower window, along with the S_{21} response. The ripple on the S_{11} response shows the evidence of the combination of the input and output connector mismatch, for a total error on the order of 20 dB below the signal being measured (as shown by the gated response). This would give a peak-to-peak ripple on the order of about 1.6 dB, which is quite close to what is measured between 3 and 6 GHz. Thus, the ripple in this measurement may be entirely attributed to the non-repeatable SMA connectors.

As a final evaluation, the same component was measured to just 6 GHz, using the PC board calibration kit, and a simulation model was created comprised of a shunt resistor with series parasitic inductance and with a short length of transmission line to represent the offset from the center of the thru to the reference plane, as shown in Figure 11.14. This simple model gives an indication of the expected S_{21} curve and S_{11} response.

The model's inductance was matched to the DUT measurement using the S_{21} trace. The transmission line delay was matched using the S_{11} trace on the Smith chart. Here again it is clear that source match is causing the ripples that produce the deviation from the model. This ripple can be evaluated as shown in Eq. (3.85) to estimate the sum of the directivity and source match error. In this case, there is 0.45 dB ripple at 3 GHz and some offset from the model. The ripple is very likely measurement error as the model shows $S_{11} = S_{22}$ for all frequencies. Of course, the

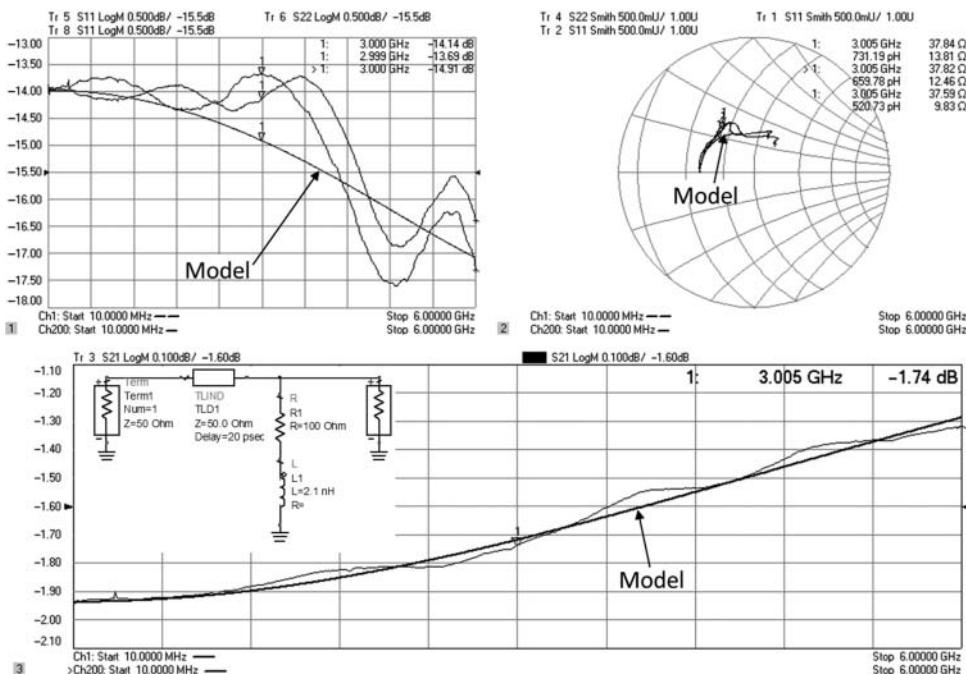


Figure 11.14 Model and measurements of shunt resistor.

simple model may not be an exactly correct estimate of the match of the shunt resistor, so the additional offset from the model may be due to the very simple nature of the model.

From the ripple of 0.45 dB, one can estimate the residual errors as about -40 dB. Some of this is from the connectors, but some might be because the definition of the open and short is not correct in the cal-kit, causing an error in the source match. Since the directivity error is estimated from the gated time-domain response in Figure 11.12 to be about -39 dB, this might entirely account for the offset. The connector residual match is about -38 dB at the input and -43 at the output, at 3 GHz. These can be added to the S_{11} value (estimated as the center of the ripple, about -14 dB at 3 GHz) to see the total error effect. If the sum of the ripples are evaluated from the residual directivity and each connector is smaller than the total ripple observed, the remainder must be source match; if there is no remainder, then the source match error (due to improperly defined opens and shorts) must be small. These errors as a contribution to the total ripple are computed as

$$EDF_{Linear} = 10^{-39/20} = 0.011, \Delta_{Conn1} = 10^{-38/20} = 0.013, \Delta_{Conn1} = 10^{-43/20} = 0.007$$

$$S11_{Linear} = 0.2, S21_{Linear} = 0.82$$

$$Error_{EDF} = \frac{20\log_{10}}{2} \left(\frac{1 + \frac{\Delta_{EDF}}{S11}}{1 - \frac{\Delta_{EDF}}{S11}} \right) = 20\log_{10} \left(\frac{1 + 0.055}{1 - 0.055} \right) = 0.48dB$$

$$Error_{Conn1} = \frac{20\log_{10}}{2} \left(\frac{1 + \frac{\Delta_{Conn1}}{S11}}{1 - \frac{\Delta_{Conn1}}{S11}} \right) = 20\log_{10} \left(\frac{1 + 0.013}{1 - 0.013} \right) = 0.56dB$$

$$Error_{Conn2} = \frac{20\log_{10}}{2} \left(\frac{1 + \frac{S21^2 \Delta_{Conn2}}{S11}}{1 - \frac{S21^2 \Delta_{Conn2}}{S11}} \right) = 20\log_{10} \left(\frac{1 + 0.005}{1 - 0.0005} \right) = 0.2dB \quad (11.6)$$

From this computation, it shows that the first connector can account for all of the ripple at 3 GHz, and the load response may also be a large contributor. However, because the load standard and the test DUT (100 Ω resistor) are located in the same physical plane, the error in the directivity term due to the load model error would show up as a very slowly moving error versus frequency, and so probably explains the offset of the trace. The fine-grain ripple most likely represents the connector errors as they have the electrical-length separation from the DUT to cause plus-and-minus phasing of the error.

This measurement points out that for moderate return loss (-14 dB), even small residual errors can give substantial errors in the overall measurement. Since each of these match terms are reasonably good, the effect on S_{21} is quite small and shows as a small ripple deviation from the model. The S_{21} error can be computed as

$$ESF_{Raw} \approx ELF_{Raw} \approx -12.66 dB, \text{ or } 0.23 \text{ linear, (worst case 3-6 GHz)}$$

$$ESF_{Residual} \approx ELF_{Residual} \approx -40 dB, \text{ or } 0.01 \text{ linear}$$

$$\begin{aligned} \Delta S_{21-p-p} &\approx 2 \cdot 20\log_{10}(1 + ESF_{Raw}ELF_{Residual} + ESF_{Residual}ELF_{Raw}) \\ &= 2 \cdot 20\log_{10}[1 + (0.23) \cdot (0.01) + (0.01) \cdot (0.23)] = 0.08dB \end{aligned} \quad (11.7)$$

which falls very well in line with the ripple seen in Figure 11.13.

11.1.2.8 Conclusions on PC Board Fixtures

Creating fixture calibration standards and kits, whether for PC boards or other fixtures, is a reasonable approach to improving component measurements. With some care, custom cal-kits can be created and characterized to provide a custom in-fixture calibration. This section presented a variety of approaches to characterize and verify the performance of an example calibration kit, including measurements of components and comparison to estimated model responses. Using SOLT with the unknown thru method allows one to use the thru as a verification standard, and with time-domain gating, one can verify the connector repeatability and the load model response. In this example, additional verification was attempted using a model of a well-controlled shunt element (100Ω resistor) to demonstrate the quality of the calibration.

In many cases, in-fixture calibration is not practical or possible, so other fixture compensation methods must be used, as discussed next.

11.2 Fixturing and De-embedding

Another approach to dealing with component fixtures is to calibrate in a common coaxial connector, such as SMA, and then attach the fixture and measure the combination of the fixture and the DUT. If some information about the fixture can be obtained, through independent measurements or modeling, then the fixture effects can be removed, and the DUT response can be de-embedded from the fixtured measurements. If in-fixture standards are available, it is relatively straightforward to characterize the fixture and determine its S-parameters (see Section 11.3.1).

But, it is often difficult to create and model in-fixture standards and so techniques have been developed to measure the combination of the fixture and some limited cal standards and determine at least some of the fixture effects. This then provides a means so that the S-parameters of the fixture are known, or at least estimated. Once the S-parameters are determined, de-embedding of the fixture can be accomplished by modifying the calibration coefficients such that the error-corrected results remove the fixture effects as well as the effects of match, directivity, and tracking of the VNA. Techniques for determining the fixture characteristics will be discussed in later sections.

Another use for de-embedding comes from the need for multiport testing. In many cases, the DUT has a $2 \times N$ path structure, such that the measurements are desired on any two ports of an N -port DUT. While normal calibration processes would require $(N-1)(N/2)$ calibration paths (e.g. a 10 port DUT requires 45 two-port calibrations), using de-embedding techniques, the calibration can be reduced to one 2-port calibration and $(N-2)$ 1-port calibrations.

De-embedding techniques can also be used to modify the effective port impedance of the VNA, allowing measurements made on a DUT in the system Z_0 to be translated to any other impedance. This is particularly useful in balanced devices where the input impedances are often not matched to the VNA system Z_0 .

Related to impedance transformation, port-matching is a de-embedding technique that allows the measured S-parameters to reflect the condition of additional virtual components added to the input or output of the DUT; this is called *port matching* or *embedding*. It is common in cases where devices such as integrated circuits (ICs) are not well matched with the intention that off-chip components will perform the matching function; it is desired to know

the response of the IC after the matching element is added. In the case of a DUT being used in a cascade of other devices, the S-parameters of which are already known, the response of the overall system can be modeled by measuring only the DUT and using port-matching to embed the characteristics of the other elements of the system. Often attributes of the DUT such as bias are adjusted to obtain the desired overall system performance. An example of this is using port matching to add the effect of a CATV cable loss to the measurement of a CATV amplifier. The specifications on the amplifier state a particular frequency-flatness when driven into a long length of cable (up to a km or more). In legacy measurement systems, actual rolls of CATV cables were used, but each cable had slightly different loss and results were not reliable from one test station to another. Emulating the loss of a cable, using embedding, removes one source of non-repeatable error from the results.

11.2.1 De-embedding Mathematics

The most common method for de-embedding fixture effects from a DUT is to modify the calibration coefficients to account for the effects of the fixture. The signal flow graph for the forward error terms is shown in Figure 11.15.

Computing the fixture effects on the systematic error terms can be accomplished through a variety of methods applied to the entire flow graph, but a convenient method that is not often discussed is through the use of intermediate variables in the flow graph and through the use of a select case for the S-parameters of the DUT. Consider the first case where the DUT is well matched and unilateral, that is, $S_{11} = S_{12} = S_{22} = 0$ and $S_{21} = 1$, which results in the flow graph of Figure 11.16. The choice of DUT has no effect on the error terms, so a convenient choice can be made to simplify the computation of the error terms.

For the fixturing computations, the error terms are converted to the equivalent simplified flow graph of Figure 3.23, the fixtured terms replacing the nominal terms. Note that here the source loss and the reference tracking are included in the signal flow diagram. During the embedding process, all of the error terms are modified such that the new error model includes the effects of the fixture. While this technique has been used for many years (Inter-Continental Microwave 2006), the application to source and receiver power calibrations was first presented in commercial VNAs in early 2011.

In the case of Figure 11.16, $b_{1A} = 0$. With this diagram, the only loop terms are from ESF and S11A on the input side, and ELF and S22A on the output side. By inspection one can

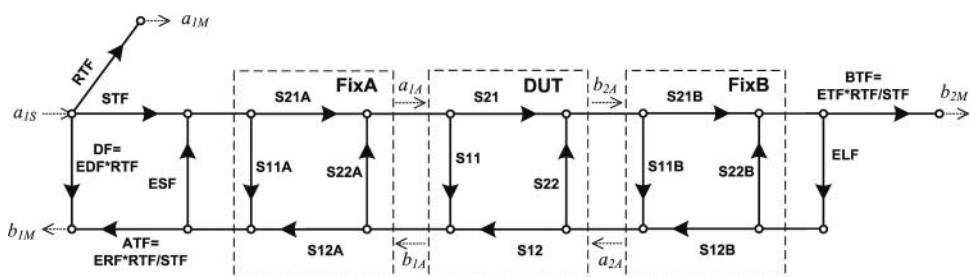


Figure 11.15 Signal flow diagram for a DUT in a fixture.

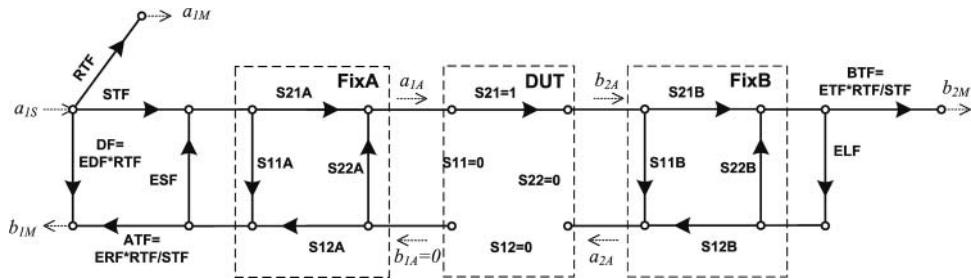


Figure 11.16 Fixture flow graph for a unilateral DUT.

see that

$$DF_{Fix} = \frac{b_{1M}}{a_{1S}} \Big|_{S11=S12=0} = DF + \frac{STF \cdot S11A \cdot ATF}{1 - ESF \cdot S11A} \quad (11.8)$$

$$STF_{Fix} = \frac{STF \cdot S21A}{1 - ESF \cdot S11A} \quad (11.9)$$

$$ATF_{Fix} = \frac{ATF \cdot S12A}{1 - ESF \cdot S11A} \quad (11.10)$$

$$ESF_{Fix} = S22A + \frac{S12A \cdot ESF \cdot S21A}{1 - ESF \cdot S11A} \quad (11.11)$$

Because the fixture has no effect on the signal to the reference channel, the value for RTF is unchanged,

$$RTF_{Fix} = RTF \quad (11.12)$$

From the output side, the error terms modified by the output fixture are computed as

$$BTF_{Fix} = \frac{S21B \cdot BTF}{1 - S22B \cdot ELF} \quad (11.13)$$

$$ELF_{Fix} = S11B + \frac{S21B \cdot ELF \cdot S12B}{1 - S22B \cdot ELF} \quad (11.14)$$

The traditional 12-term error model values are easily derived from Eqs. (11.8), (11.9), (11.10), and (11.13). The error term DF differs from the traditional EDF error term by the loss to the reference coupler,

$$EDF = \frac{DF}{RTF} \quad (11.15)$$

The fixture version of EDF can be related to the 12-term model as

$$\begin{aligned} EDF_{Fix} &= \frac{DF_{Fix}}{RTF} = \frac{DF}{RTF} + \frac{STF \cdot S11A \cdot ATF}{RTF(1 - ESF \cdot S11A)} \\ &= EDF + \frac{ERF \cdot S11A}{(1 - ESF \cdot S11A)} \end{aligned} \quad (11.16)$$

The reflection tracking term can get defined in terms of the 12-term model as

$$\begin{aligned}
 ERF_{Fix} &= \frac{ATF_{Fix} \cdot STF_{Fix}}{RTF_{Fix}} = \frac{1}{RTF} \cdot \frac{ATF \cdot S12A}{1 - ESF \cdot S11A} \cdot \frac{STF \cdot S21A}{1 - ESF \cdot S11A} \\
 &= \frac{ATF \cdot STF}{RTF} \cdot \frac{S21A \cdot S12A}{(1 - ESF \cdot S11A)^2} \\
 &= \frac{ERF \cdot S21A \cdot S12A}{(1 - ESF \cdot S11A)^2}
 \end{aligned} \tag{11.17}$$

And ETF of the fixtured calibration as

$$\begin{aligned}
 ETF_{Fix} &= \frac{BTF_{Fix} \cdot STF_{Fix}}{RTF_{Fix}} = \frac{1}{RTF} \cdot \frac{BTF \cdot S21B}{1 - S22B \cdot ELF} \cdot \frac{STF \cdot S21A}{1 - ESF \cdot S11A} \\
 &= \frac{BTF \cdot STF}{RTF} \cdot \frac{S21A \cdot S21B}{(1 - ESF \cdot S11A)(1 - S22B \cdot ELF)} \\
 &= \frac{ETF \cdot S21A \cdot S21B}{(1 - ESF \cdot S11A)(1 - S22B \cdot ELF)}
 \end{aligned} \tag{11.18}$$

For completeness, the final term is the crosstalk term, which like the RTF term, is not affected by the fixture.

$$EXF_{Fix} = EXF \tag{11.19}$$

The error terms for the reverse direction are computed in a similar manner.

11.3 Determining S-Parameters for Fixtures

The best possible way to de-embed the fixture is to determine the exact S-parameters of the fixture. But determining the exact S-parameters for fixtures can be quite difficult if a calibration kit for each fixture port is not available. One approach is to model the fixture in a linear simulator or 3-D EM structure simulator, and this approach is often used to remove effects of connection pads to IC devices from on-wafer probe tests. Another approach is to characterize the loss and delay of the fixture using port extension techniques. Still other techniques use time-domain gating applied to a pair of fixtures connected by a thru, with special mathematical functions use to split the fixture effects apart. These methods are discussed in detail next.

11.3.1 Fixture Characterization Using 1-Port Calibrations

The simplest technique for characterizing a fixture or adapter is to directly measure its characteristics using a 2-port calibration. Often, however, this is not possible as the full set of calibrations standards may not be available, or the physical limitations may make it impossible to do a full 2-port calibrated measurement on the adapter or fixture. However, the exact

S-parameters of the fixture can be determined if a 1-port calibration can be applied before and after the fixture. Formulations from Eqs. (11.11), (11.16), and (11.17) provide the basis for computing the fixture's S-parameters.

These error term equations, while derived from the premise of adding a fixture to an already calibrated test port, could equally well be used to relate the S-parameters of an unknown fixture to a 1-port calibration before and after the fixture. However, while there are four unknown S-parameters, a 1-port calibration presents only the three error term equations. The order of the problem is reduced by presuming that the fixture or adapter is passive and bilateral, such that $S_{21} = S_{12}$. With this choice there are only three equations and three unknowns. If the first-tier calibration (at the input of the fixture or adapter) has error terms EDF , ERF , and ESF , and the second tier calibration (at the output of the fixture or adapter) has error terms EDF_{Fix} , ERF_{Fix} , and ESF_{Fix} , then the S-parameters of the fixture or adapter are computed as

$$\begin{aligned} S_{11_Fix} &= \frac{(EDF_{Fix} - EDF)}{[ERF + ESF \cdot (EDF_{Fix} - EDF)]} \\ S_{21_Fix} = S_{12_Fix} &= \frac{\sqrt{ERF \cdot ERF_{Fix}}}{[ERF + ESF \cdot (EDF_{Fix} - EDF)]} \\ S_{22_Fix} &= ESF_{Fix} + \frac{ESF \cdot ERF_{Fix}}{[ERF + ESF(EDF_{Fix} - EDF)]} \end{aligned} \quad (11.20)$$

These are identical formulations to the adapter removal calibration as described in Chapter 3.

11.3.1.1 Computing the Square Root of a Complex Frequency Response

There is a subtle point to defining the phase of S_{21} : the square-root function can have a positive or negative result; in complex form, this manifests itself into a 180° uncertainty about the phase of the S_{21} term. For an isolated point, there is no way to determine the proper root, but if only a little information is known, particularly its nominal delay, then the proper root is chosen so that the phase response lies nearest to the phase predicted by the nominal delay. An alternative method for defining the phase response is to plot the unwrapped phase response of S_{21}^2 as a function of frequency and divide it by two. This gives the correct phase and delay provided that the data used has less than 180° phase change per frequency point. The only remaining uncertainty is the phase of the first point. This is found by determining the slope of the phase and projecting it back to zero frequency (DC) or, to put another way, using the delay of the adapter to predict the approximate phase offset at the first point of the data trace. Mathematically, computing the phase versus frequency with the proper offset from the S_{21}^2 phase data is derived as

$$\phi_{(f)Offset} = \frac{\phi_{(f)Unwrap}}{2} + \left(Int \left[\frac{(\phi_N - \phi_0)}{(f_N - f_0)} \cdot \frac{f_0}{360} \right] \right) \cdot \frac{360}{2} \quad (11.21)$$

where N is some point chosen in the trace to compute the group delay. Using the entire trace produces the lowest noise in the group delay estimation but may not be practical if the fixture has non-constant delay at higher frequencies. This works well for normal fixtures or adapters that have a nearly constant group delay but can have some difficulties when the adapter is band limited, as in the case of waveguide adapters. In particular, the delay of the waveguide

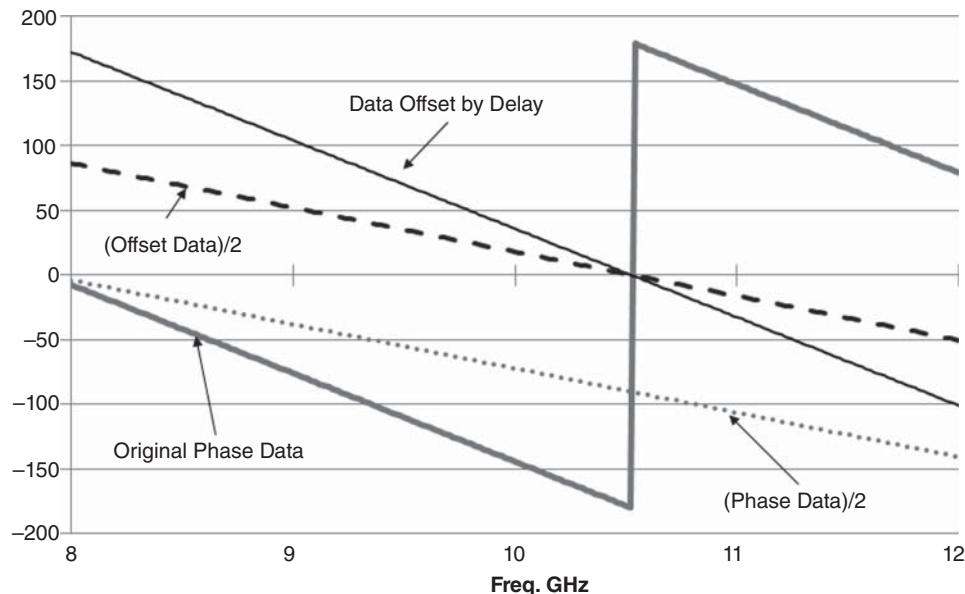


Figure 11.17 Determining the phase of S_{21} .

portion of the adapter is not constant with frequency, but the delay of the coax portion is constant with frequency.

An example of the phase response computation is shown in Figure 11.17, where an adapter is characterized over a narrow frequency span, starting at 8 GHz. The trace shows the normal (wrapped) phase response of the adapter, “Original Phase Data,” and the square-root phase based on dividing the unwrapped original phase data by 2, “(Phase Data)/2.” The projection of the 8 GHz point back to DC indicates that a 180° offset is needed if the original phase is to cross the DC value at 0 phase, and the original unwrapped phase, offset by 180°, is shown as “Data Offset by Delay.” Finally, the correct square-root phase is computed by dividing the phase response by two to obtain the final result labeled “(Offset Data)/2.” This result can be validated by measuring the S_{21} response of the adapter, if possible, or by doing the characterization over a broader frequency range extending down to DC, so that no phase wraps are neglected. In fact, whenever using relatively broadband devices, it is always good practice to measure them over an extended frequency range. Going lower in frequency avoids phase-wrapping issues as discussed here, and going higher in frequency can illuminate other effects, such as intermittent connections, more easily.

One final note for characterizing adapters for the purpose of de-embedding is that while the characterization results in the four S-parameters, the numbering of the ports is sometimes a point of confusion. In common practice, port 1 of the de-embedding network faces the VNA test port. This makes perfect sense for de-embedding from port 1, as that is the natural choice based on how the measurement of the adapter is performed. However, when the same device is measured in the same way, with the intention of de-embedding it from port 2, then the natural choice is port 2 of the adapter facing the test port. But this logic fails in the case of more than

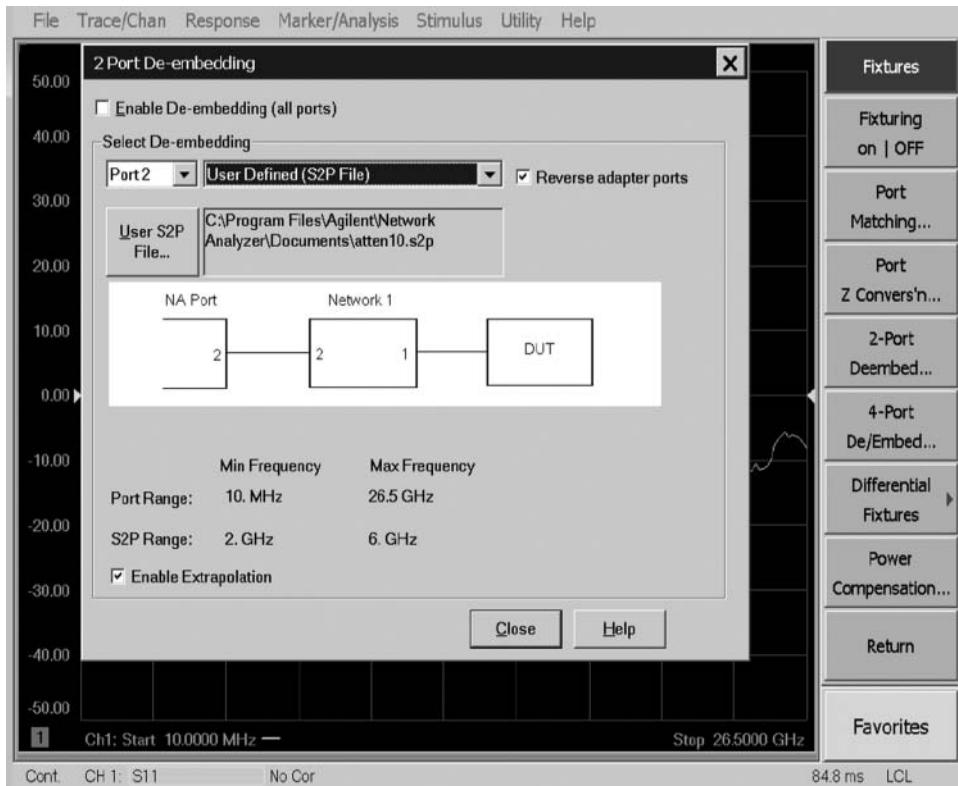


Figure 11.18 De-embedding setup dialog allowing arbitrary port selection.

two ports for the VNA. Thus, the convention of de-embedding networks always having port 1 facing the test port ensures a common practice for 1, 2, or N port VNAs. However, this requires the data for an adapter to sometimes be reversed, exchanging port 1 for port 2, so S_{11} becomes S_{22} , with no effect on passive adapters as $S_{21} = S_{12}$. In the most modern VNAs, provisions are made in the de-embedding settings to reverse the de-embedding networks ports, if desired, an example of which is shown in Figure 11.18.

Another important consideration when de-embedding is to ensure the de-embedding network matches the frequency range of the measurement to which it is applied. In this case, the S2P file range only covers a portion of the measurement range. Newer VNAs provide the ability to extrapolate S2P data to cover the entire frequency range. For some devices, such as attenuators, this is not too unreasonable, but one should carefully consider whether extrapolation is valid for the particular adapter or fixture used.

11.3.1.2 Port Extensions

When measuring microwave components, it is common to calibrate in a coaxial reference plane first, then add some small adapter or fixture to connect to the DUT. Often, these adapters

are well matched, but measurements of the DUT's S-parameters are in error by the phase shift or electrical delay of the adapter. The first VNAs used a mechanical line stretcher to compensate for delay in test fixtures or adapters, sometimes called *trombone lines* as the structure's center-conductor looked like a trombone when disassembled. These line stretchers were used in one path of the VNA ratio measurement to add delay to the reference channel to compensate for electrical delay in the test port due to some adapter or fixture.

Starting with the HP8510A, the function of the line stretcher was replaced with a mathematical function that added or subtracted delay to the phase response of the DUT; this function was called *port extension*. Port extensions worked by adding a phase shift, equivalent to an amount of electrical delay specified in seconds, to the measurement's phase data. The phase shift for a given delay is computed as

$$\phi = \begin{cases} 360 \cdot Freq \cdot Delay : S_{21}, S_{12} \\ 2 \cdot 360 \cdot Freq \cdot Delay : S_{11}, S_{22} \end{cases} \quad (11.22)$$

That is, the phase shift is applied twice per port for reflection measurements at that port, accounting properly for a signal's two-way transit through the adapter for reflection measurements. The phase shift is added once per port for transmission measurements so that the phase response of S_{21} or S_{12} is modified by the sum of port extensions applied at each port. For multiport measurements, any transmission parameter's phase change is similarly dependent on the port extension for each of the ports associated with that parameter.

One common, but improper, use for port extensions is to remove the effective delay of DUT, such as a filter, so that the phase deviation from linear is more easily perceived. The "electrical delay" function is a more appropriate choice as it applies only to the selected parameter rather than affecting all measurements on that port. Another common case is to remove small delay mismatch effects between drive ports in a balanced measurement, where the effective drive must represent exactly 180° for the balanced gain to be correctly measured. Details on these cases are presented in Chapters 5 and 10.

More recently, modern VNAs have added the concept of loss to the port extension function, allowing the small loss associated with an adapter to be removed, along with its phase response. The loss factors are entered as a known loss at one or two frequencies. If a single known loss is used, the formulation for computing loss at all frequencies follows the classic square root loss function as described in Chapter 1, which is

$$Loss(f) = A \cdot \sqrt{f} \quad (11.23)$$

Once a single loss factor is known, the loss at any frequency can be computed. However, this loss curve does not represent the loss of non-ideal coaxial, microstrip, or many other types of transmission lines. In the case of transmission lines other than airline, a more flexible function is used

$$Loss(f) = A \cdot f^b \quad (11.24)$$

where A and b are two factors computed by solving for each from the known loss at two select frequencies. Both of the previous equations presume no loss at DC, but occasionally a fixture will have some small resistive loss, which is accommodated by a DC loss term in the port extension setup; an example is shown for the Port Extension dialog of the Keysight PNA series of VNAs, as shown in Figure 11.19.

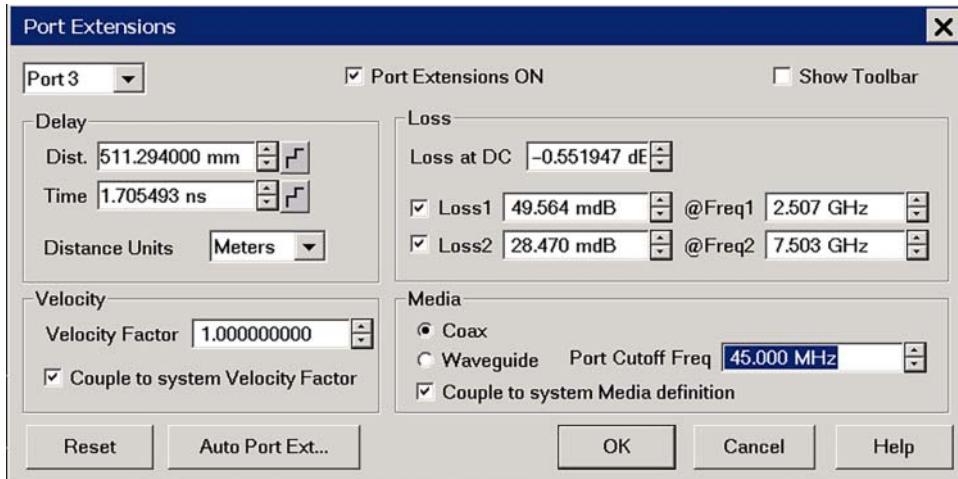


Figure 11.19 Port extension dialog including loss and waveguide compensation.

For convenience, newer analyzers provide either a time delay or a distance delay as the port extension entry, where the relationship between distance and time delay is

$$Dist_{Delay} = VF \cdot c \cdot Time_{Delay} \quad (11.25)$$

where c is the free-space velocity of light, VF is the velocity factor of the media, and 1 is the VF for a vacuum (and very nearly air).

Because waveguide transmission lines have strong dispersion, it is necessary to account properly for different phase shifts at different frequencies for a line that has a single port-extension delay time. If the Media choice is waveguide, the delay is based on the physical length of the guide, with the phase computed according to Eq. (3.36) as

$$\phi_{(f)} = \left(360f \sqrt{1 - \left(\frac{f_c}{f} \right)^2} \right) \cdot Time_{delay} \quad (11.26)$$

The losses and delay of the port extension are sometimes implemented as a de-embedding network that is applied to the calibration factor. While older analyzers applied only the phase offset term, and only as a scaling function in the display as the last step in data processing, modern analyzers have other fixturing attributes, and the proper use of port extension sometimes require that the port extension math come before other de-embedding functions such as port matching or impedance transformation. If the other fixturing and de-embedding functions are applied to the calibration set, the port extension must also be applied to the calibration set. This leads to a functional difference in how measurements are made and port extensions are applied between older legacy VNAs and more modern versions.

As a scaling function, the port extension simply adds phase shift to the final parameter result as described by Eq. (11.22) and can be applied to raw or uncalibrated results. But when applied as a calibration de-embedding, a full two-port measurement must be made to accommodate the necessary terms to apply port extensions. In this case, even if a raw measurement

is displayed, a background unity calset (one with idealized error terms of 1 for all the transmission terms and 0 for all the reflection terms) is created and the port extension attributes are de-embedded from that calset. Thus, an S_{21} trace with port extension will engender two sweeps in the VNA using the fixturing function to apply port extensions. In general, it is more robust to apply port extensions in this way, and it ensures proper computation in the case that additional fixturing is applied. Normally, turning on any fixturing function induces the de-embedding method for port extension.

11.3.1.3 Determining Port Extensions Values

For most circumstances, it is best to perform a full correction (1-, 2-, or N-port) at the test port, before applying the fixture or adapter and using port extensions. The 1-port calibration on each port removes all the mismatch and directivity effects up to the fixture. If the fixture is well designed, it will have small mismatches. A well-designed fixture should have less than 26 dB mismatch for RF frequencies and less than 20 dB mismatch for microwave frequencies.

One of the difficulties in using port extensions to correct for adapters or fixtures is determining the correct value for the port extension delay. In older analyzers, the loss could not be compensated and must simply be accepted. The delay value was often determined by looking at the phase trace of a fixture with nothing connected to the DUT plane. This open reflection should have approximately 0° of phase across the band. In actuality, this yielded a delay that was slightly too long, as the expected phase shift for an open circuit is a few degrees at low frequency due to fringing capacitance and can be much larger at microwave frequencies. One can account for the open circuit fringing by estimating the fixture's fringing capacitance based on its similarity to other connectors (in the case of a coax adapter) or by using some modeling to estimate the fringing capacitance of the open fixture (as in the case of a PC board microstrip fixture).

Many analyzers have an “electrical delay” scaling factor, which add or subtracts phase according the entered delay value. This delay offset is done as part of the formatting and does not include a two-way response as port extensions does for reflection measurements. But many VNAs include a “Marker to Delay” function that computes the group delay within a span around a marker on the phase trace (typically $\pm 10\%$) and then uses that delay value as the electrical delay. If using this in a reflection measurement, one-half of this value should be used as the port extension. If the S_{11} of the open fixture is displayed on the Smith chart, the electrical delay or port extension value can be adjusted until the marker reads 0° or, better still, reads same value as the expected fringing capacitance. This process must be repeated for each of the ports for which port extension compensation is desired.

11.4 Automatic Port Extensions (APE)

The process for determining the proper value for port extensions on fixtures can be tedious, and the values obtained can vary due to noise or small perturbations of the S_{11} trace from mismatch effects. Recently, the process for determining port extension values has been substantially automated so that it has become essentially as simple as a second-tier response calibration.

When the fixture is connected to the test port, a trace with ripple in magnitude and a rapidly changing phase response is the common result. The ripple is almost always due to

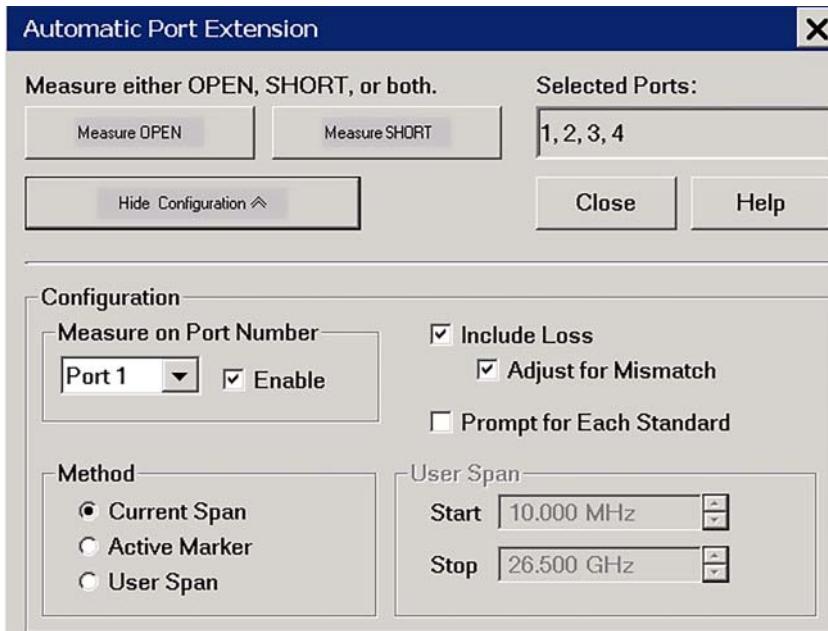


Figure 11.20 Automatic port extension setup.

the mismatch at the coax-to-PC board transition, re-reflecting with the open of the fixture. In cases where the DUT has ground connections in the same plane as the signal connections, it is a simple thing to create a “shorting block” that is essential a piece of conductor (metal) the same size as the DUT package. Sometimes a short provides a more reliable reference for port extensions than an open, due to radiation of an unshielded open.

When using automatic port extensions (APEs), the VNA measures either an open or a short, or both, and uses a least-squares method to fit the delay values to the phase response. An example setup dialog for APE is shown in Figure 11.20. If an open is used, the target phase response is 0° ; if a short is used, the target phase response is 180° . If both an open and short are used, the measurements are averaged (accounting for the 180° phase shift of the short) and the least-squares fit is applied to the averaged phase response.

The dialog provides for doing the least-squares fitting over the current span of the measurement, or over a user specified span within the current measurements. Sometimes, the fixture used for the DUT has some wildly varying phase response at the upper band edges, but the interest for the measurement is only at a particular portion of the band, usually the center or lower-frequency portion. This is the case for fixtures used to measure small ceramic or SAW filters used in cellular phone handsets. For these cases, it is convenient to use a narrow band around the center of the passband to set the port extension value. A third choice uses just the active marker and computes the delay value at just one point in the trace.

A similar process is used for the magnitude response of the fixture to determine the loss. For magnitude, the least-squares fitting attempts to find the best fit to Eq. (11.24). Once the best fit is found, the values for the curve are determined three frequency points: one-quarter and three-quarters the span above the start frequency and a DC point.

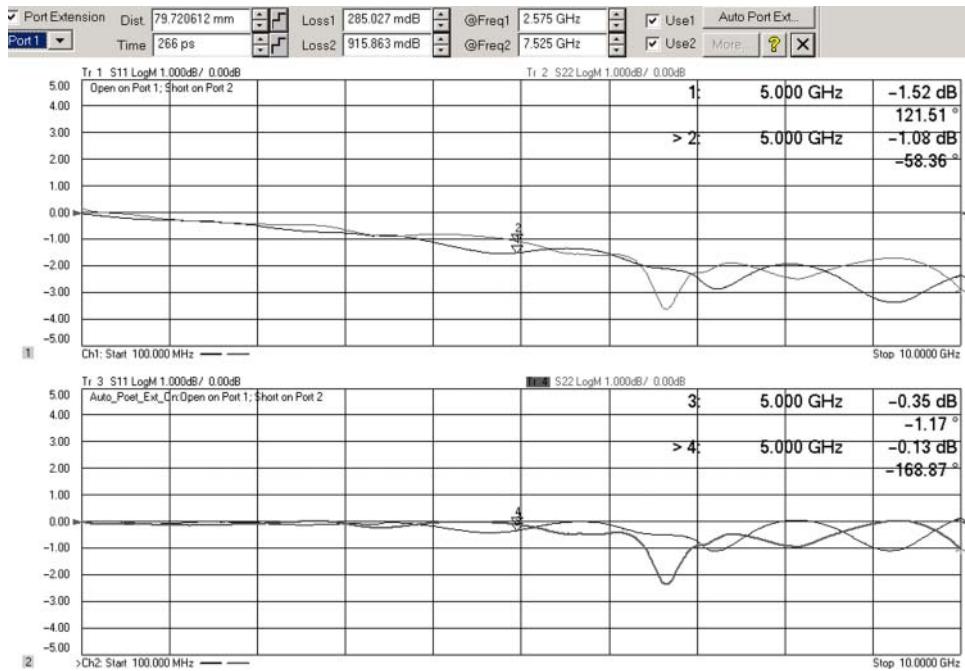


Figure 11.21 Measurement of an open and shorted fixture (upper); after automatic port extensions (lower).

Figure 11.21 (upper window) shows the measurement of an open and a short in a fixture. The ripple in the magnitude trace is apparent, and the open ripple is nearly the inverse of the short ripple. For wideband measurements, using the average of the open and the short or using the least square fit through either the open or the short yields nearly the same result. For narrow-band measurements, using both an open and a short can avoid an offset error due to the ripple. In this way, the effect of the mismatch at the input of the fixture on the port extension loss is minimized. A similar process is used for the phase response. The lower window shows the result of the APE, with the resulting delay and loss computations shown in the port extension toolbar above the figures.

Unfortunately, this straightforward approach has a drawback that if the loss is compensated for directly, the ripple in the open or short response will cause the S_{11} to go above 0 dB; that is, the magnitude of S_{11} will be greater than 1. While minimizing the maximum error, the result can cause a great deal of difficulty if the data for the DUT also contains high reflections and is used later in circuit simulators. For example, amplifiers often have pre-matching networks that can have high reflections out of the band of interest. If the S_{11} goes above 1, many computations “blow up” such as stability factor, and optimizations can become difficult. That is because all passives and most active devices have S_{11} less than 1.

A special feature of APE allows the user to choose between ignoring the mismatch of the port extension and offsetting the loss computed by the mismatch ripple to ensure that loss compensation does not result in S_{11} greater than 1. Figure 11.22 shows the result of simple APE compensation, using direct least-squares fit (lower window), and a result where

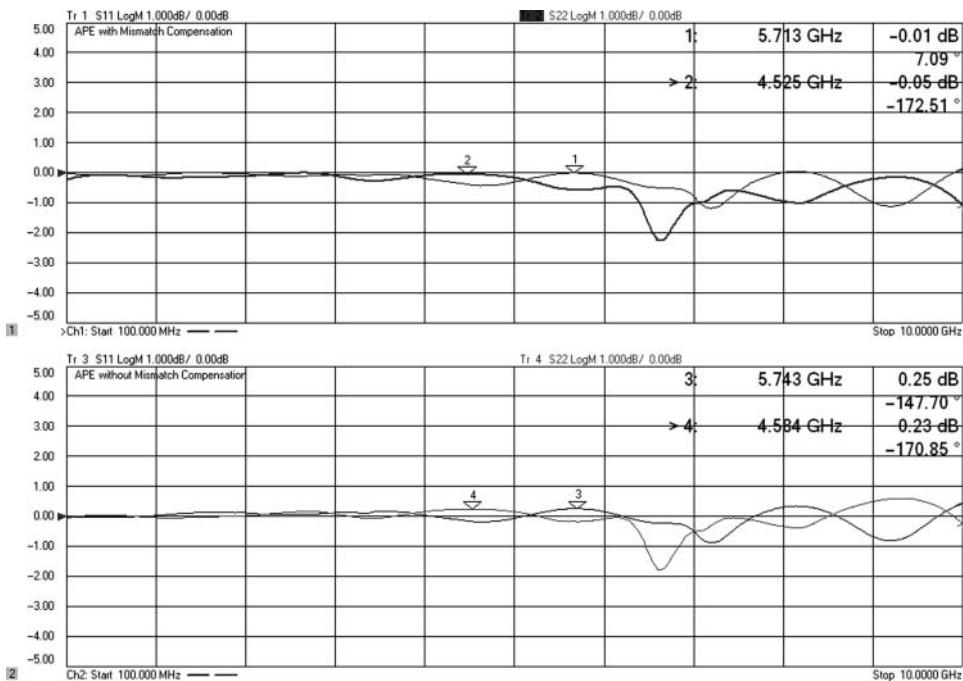


Figure 11.22 S_{11} response after APE with and without mismatch compensation.

APE compensates for mismatch (upper window). The algorithm for mismatch compensation essentially modifies the values in the loss table so that the peaks of the measured response (open or short) are below the $S_{11} = 1$ (or 0 dB) reference line. The theoretical max error is greater with mismatch compensation, with the positive error tending toward zero and the negative error being twice the ripple; however, the S_{11} trace is always well behaved, and this usually provides a better case for simulations.

APE is especially convenient for fixtured parts in a multiport or balanced fixture. It is even possible to use APE in a PC board fixture that has a DUT part soldered in place. Figure 11.23 shows the S_{11} measurement of a shunt 10 pF bypass capacitor, mounted from the center of the fixture thru trace to ground. The light gray trace shows the uncompensated S_{11} , from which it is impossible to infer any attributes of the DUT. The dark trace shows the result with APE applied. From the low-frequency response, near 300 MHz where the trace crosses the “ $j50 \Omega$ ” susceptance line (that is, where the susceptance is 0.02 S), marker 1 shows the correct value for the shunt capacitance. Marker 2 is positioned by searching for the maximum reflection and shows the self-resonant frequency of approximately 1.35 GHz. From this, we can compute the effective series inductance of this capacitance as

$$L_{SRF} = \frac{1}{(2\pi f_{SRF})^2 \cdot C} = \frac{1}{(2\pi(1.35 \cdot 10^9))^2 \cdot 10 \cdot 10^{-12}} = 1.4nH \quad (11.27)$$

This is quite in-line with other estimates made for the series inductance of an SMT capacitor. Using APE allowed a simple approach to removing the fixture effects for this device.

The complete S-parameters in LogMag format are shown Figure 11.24. As a side note for designers: From the S_{21} plot (dark trace, lower window), it is clear that this 10 pF capacitor

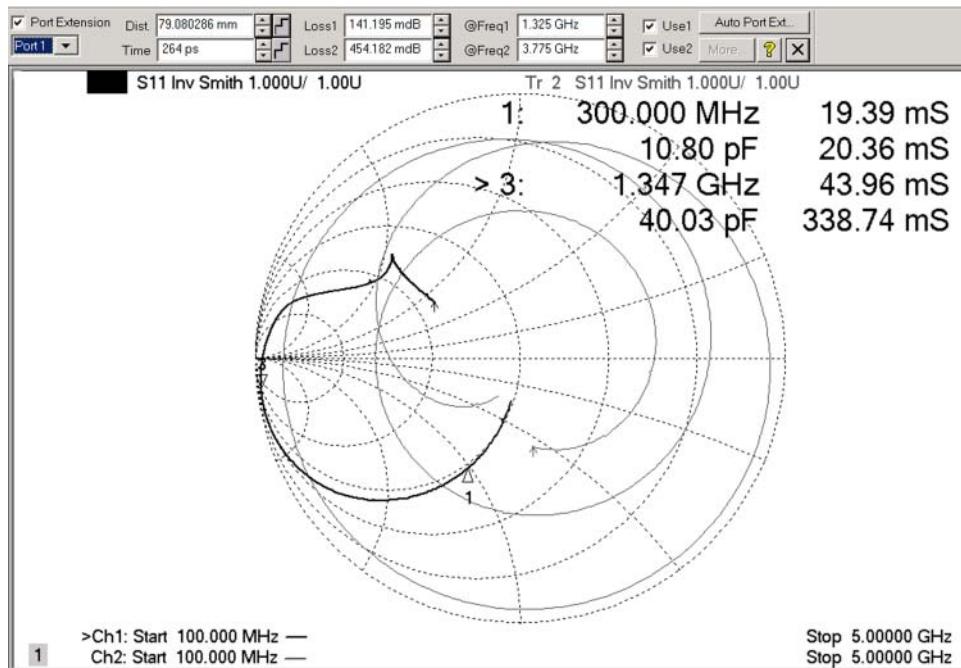


Figure 11.23 Measurement of a shunt 10 pF capacitor; as measured in the fixture (light gray), with APE applied (dark).

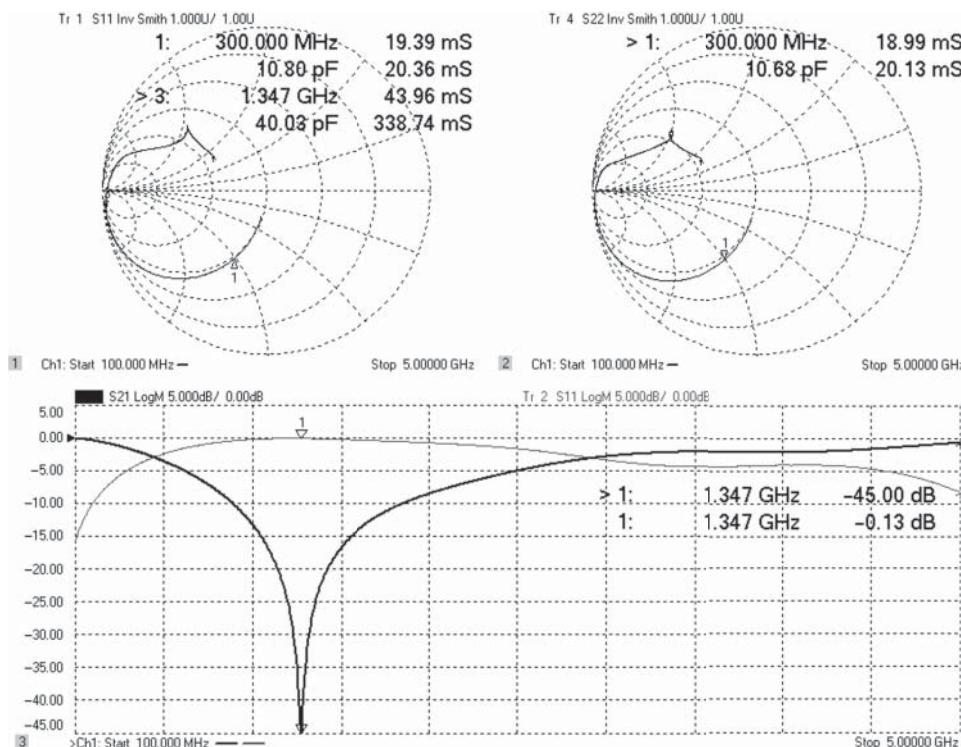


Figure 11.24 Complete measurement of a shunt 10 pF capacitance with APE.

does not provide a wide range of low impedance bypass, because at even low RF frequencies, the series inductance dominates the response. And when used as a filter element, the effects of the series inductance must be included in the design or the filter response is likely to be shifted much lower in frequency.

Most RF parts have high reflections when unpowered, so the APE process can be performed even on a fixture even if the DUT part is attached, provided the power is removed. The phase compensation will be valid if the part has a reasonably high reflection. In this case, the loss compensation should be disabled unless it is known that the part's impedance in the unpowered state provides a full reflection.

11.5 AFR: Fixture Removal Using Time Domain

APE is a simple way to compensate for a fixture's loss and delay and works on one-port fixtures. Another common method for compensating for loss in PC board or other fixtures is to create an identical test fixture as the DUT fixture but provide a thru connection. The simplest compensation using a thru fixture is to calibrate in a coax standard such as SMA, then measure the thru, and finally normalize the trace to the thru using data into memory and data over memory. While this provides a degree of normalization, mismatch in the input and the output of the test fixture can cause significant errors, up to ± 1 dB on a transmission measurement.

In recent years, advanced automatic fixture removal (AFR) techniques have been developed that make use of time-domain measurements (for more information on time domain, see Chapter 4) on PC board fixture to compensate for input and output mismatch, as well as loss, even if the input and output mismatch are not the same (Dunsmore et al. 2011).

11.5.1 2-Port AFR

The time-domain fixture removal starts by measuring the fixture thru response in the time domain. To achieve the best resolution, the widest possible span should be used, even if the fixture will be used only over a narrow span. The peak of the response represents the overall delay of the fixture, Figure 11.25. Alternatively, the average group delay response could be used. In many cases, the input and output of the fixture are designed to be the same length, with the DUT reference plane directly in the middle of the fixture.

Once the overall delay of the fixture is known, the input and output time-domain responses are measured. Figure 11.26 shows the time-domain response (T_{11}) of the thru fixture. The wide gray trace is the overall T_{11} , and the dark thin trace is the time-gated T_{11} . The time-domain response shows a capacitive discontinuity at the input of the fixture and an inductive discontinuity at the output of the fixture. It is best to have the time gate set symmetrically about the first reflection: the time difference from the first reflection (here about 46 ps) to the center (at 909 ps) is computed and is subtracted from the first reflection, to set the start time gate at -817 ps. In other words, make the time gate around the first fixture symmetric about the zero time point. The gated S_{11} response is shown in the figure as a narrow, dark trace. It is clear that after the stop gate, the trace has a constant value. The small offset in the base-line is a result of the DC loss of the transmission line in the fixture;

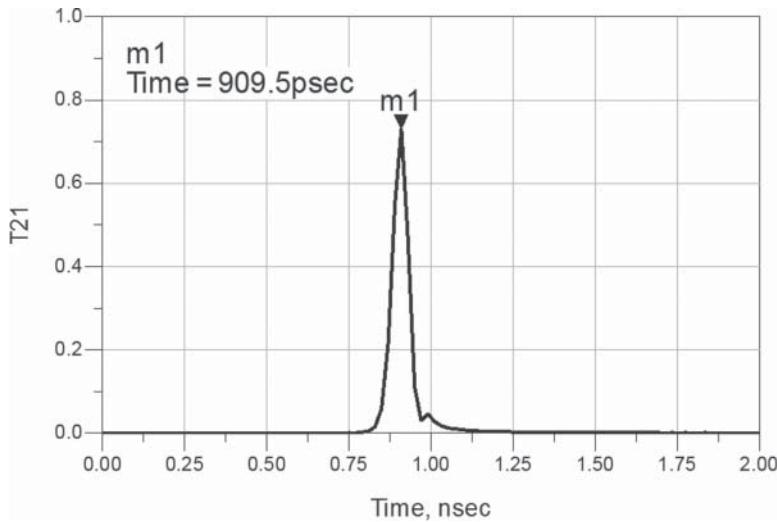


Figure 11.25 Time-domain transmission response of a fixture.

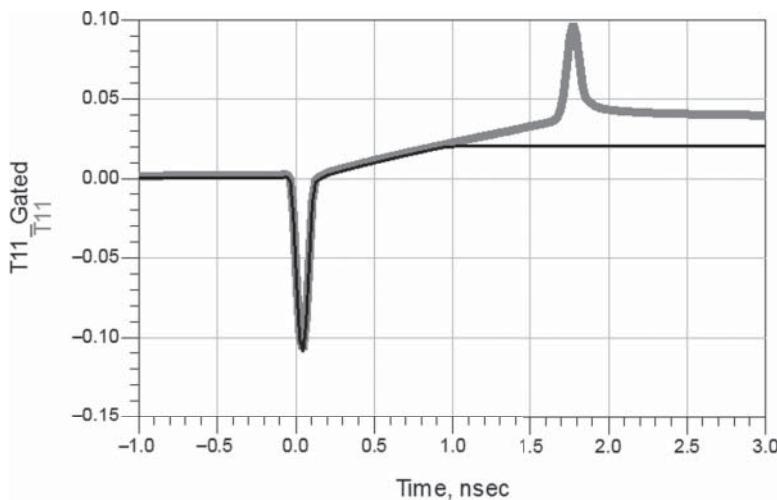


Figure 11.26 Time-domain response (gray, T_{11}) and gated response (black, T_{11_Gated}).

a review of the fixture shows about 1.5Ω of DC loss, which represents a reflection coefficient of 0.015, almost exactly matching the baseline offset of Figure 11.26.

This gated response represents the time-domain measurement of S_{11} of the left fixture, or S_{11A} of Fixture A from Figure 11.15.

In Figure 11.27, the overall frequency response of the thru fixture (a fixture with a thru in place of the DUT) is shown in a light, narrow trace, and shows substantial ripple. Also shown is the gated S_{11} response of the thru fixture, S_{11A} (dark trace) and the independently

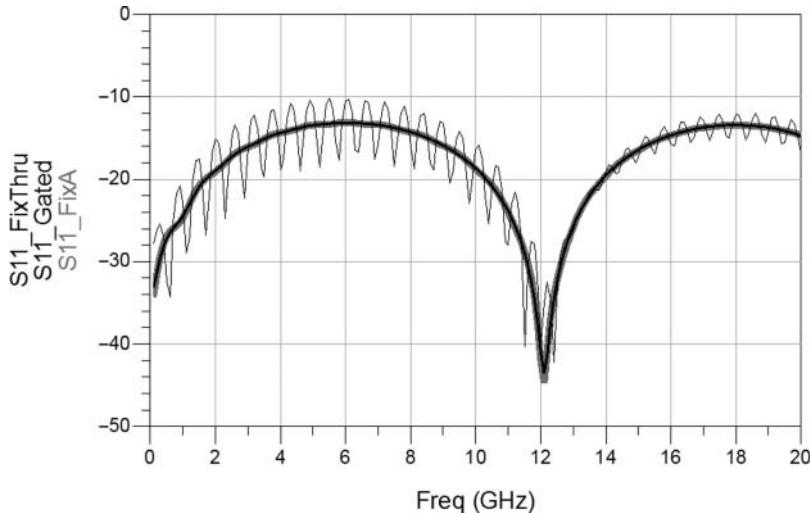


Figure 11.27 Frequency response of the thru ($S_{11_FixThru}$, narrow light trace), the gated S_{11} (S_{11_Gated} , wide black), and the actual fixture S_{11} (S_{11_FixA} , wide, gray trace).

obtained actual measurement of S_{11A} of fixture A (wide, light gray trace). It is clear that the gated response closely matches the actual value of the fixture.

The value of S_{22B} of Fixture B, the output portion of the thru fixture, is found by gating the S_{22} response of the thru fixture in a similar manner. In this way, six known values are obtained: S_{11A} of fixture A, S_{22B} of Fixture B, and the four S-parameters of the combined fixtures, which are designated as S_{11T} , S_{21T} , S_{12T} , S_{22T} of the thru fixture. This leaves three unknown S-parameters for each fixture.

The rest of the values of S-parameter for fixture A and fixture B can be obtained by presuming that the $S_{21A} = S_{12A}$, $S_{21B} = S_{12B}$, which leaves only four total unknowns: S_{21A} , S_{21B} , S_{22A} , S_{11B} . The four S-parameter measurements of the original thru give sufficient independent equations now to solve for these values.

Figure 11.28 shows the computed value for S_{22A} of the sample fixture as a dark trace, and the independently measured actual value as a wide gray trace. The results overlap almost completely with only slight difference at the edges of the bands.

Figure 11.29 shows the comparison of the value for S_{21A} (S_{21A_AFR} , narrow, dark trace) computed using the AFR technique, with that of the independently measured value for S_{21A} of fixture A (S_{21_FixA} , wide light gray trace), with almost perfect results.

Thus, even for a fixture with non-symmetric mismatch, it is possible to determine the individual values for the input fixture (Fixture A) and output fixture (Fixture B) using only a thru measurement. An example is shown in Figure 11.30 showing a measurement of filter in a fixture ($Filter_Fix11$, wide gray trace), and the same measurement but with AFR applied ($Filter_AFR$, dark narrow trace), as well as an independent measurement of the actual filter ($Filter_Actual$, narrow, light gray trace) for S_{11} on the left and S_{21} on the right. The filter response is greatly improved over the fixtured response, using nothing more than the measurements of the thru fixture and AFR techniques for compensation.

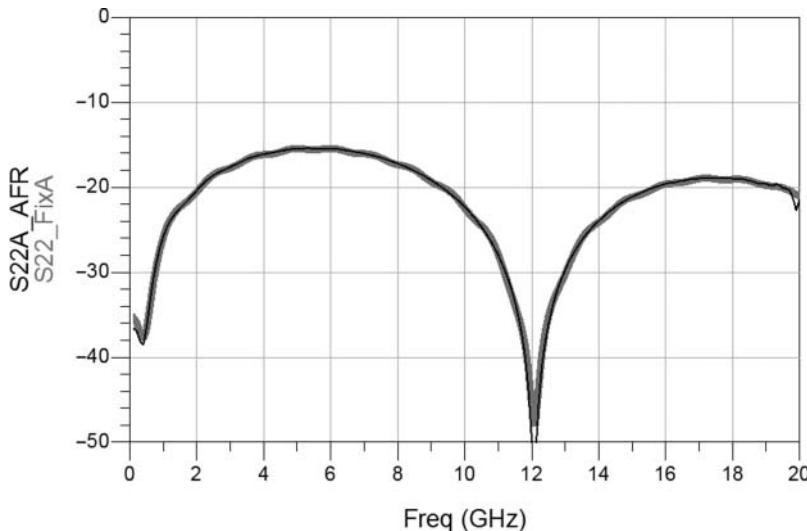


Figure 11.28 Computed value for S_{22} of the fixture (S22A) and actual fixture S_{22} (Fix_S22).

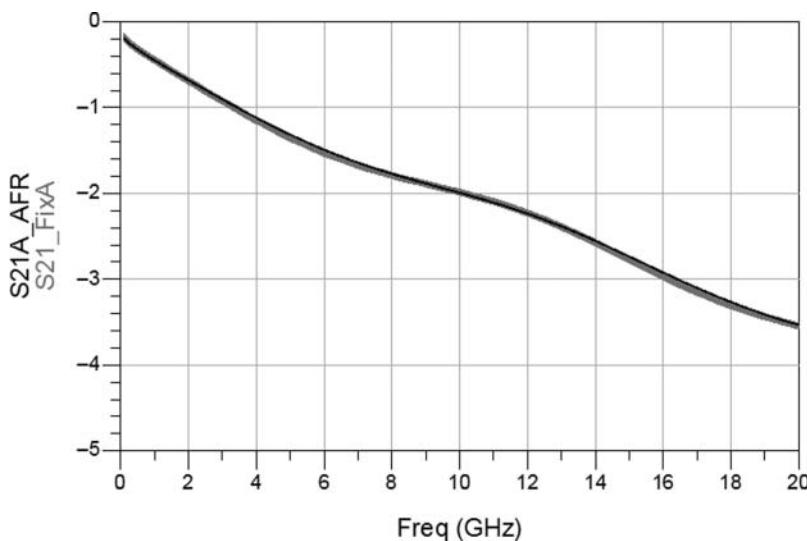


Figure 11.29 Computed value of S_{21} (S21A_AFR) and actual fixture S_{21} (S21_FixA).

In some cases, the DUT will not be centered in the fixture, so the loss and delay of fixture A and fixture B will not be identical. In these cases, the offset loss and delay can be easily accommodated by first performing AFR on a thru standard in the fixture and then adding an open measurement for the fixture, where the open fixture occurs the same physical point as the where the DUT will be inserted. Using the open response, an APE is applied as described in Section 11.4 to obtain a loss and delay offset from the fixtured results obtained using a thru

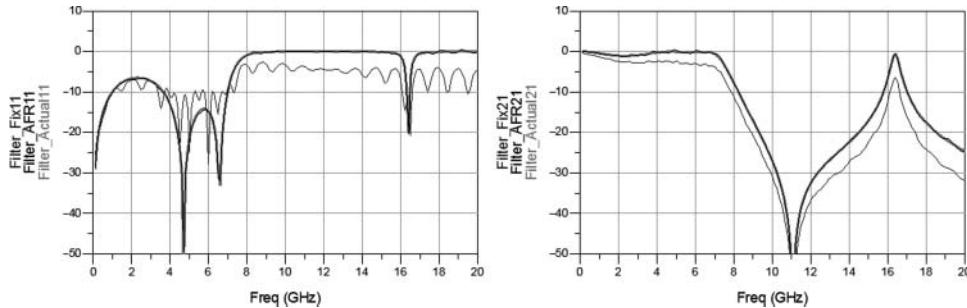


Figure 11.30 Filter measurement comparing the actual filter, filter in a fixture, and with AFR applied.

that was presumed to be centered. This will result in a small positive port extension for one port and an equal but opposite port extension for the other port.

These fixture removal techniques represent the state-of-the-art in dealing with PC board and similar fixtures. Further, the same techniques can be applied to balanced measurements, where the balanced parameters are substituted for the single-ended parameters.

11.5.1.1 AFR Measurement Example

A simple and effective example for evaluating the AFR technique in a real application is to use the same example PC board and DUT as in Section 11.1.1. In this case, a calibration is first performed over a broad frequency and the thru standard is measured. The previous AFR technique is applied, and the individual PC board input and output fixture is computed. The resulting S-parameters of these fixtures are shown in Figure 11.31.

The values for the input and output return loss for the fixture match quite closely with the gated estimates of the thru-standard input and output match from Figure 11.2.

The final step for AFR is to use the same calibration as when used to characterize the thru standard and de-embed the input and output fixtures from this calset. After the de-embedding step, the sample $100\ \Omega$ shunt resistor from the first example is remeasured, the comparison of the results shown in Figure 11.32.

In some ways, this is a remarkable result in that only one standard is used, the thru standard, but the S_{11} and S_{22} measurements, as well as the S_{21} measurements, follow very closely to the measurements performed with the PC board cal-kit.

The residual difference between the AFR measurements and the PCB cal-kit measurements is less than -40 dB below 3 GHz, and less than -30 dB up to 6 GHz. These residuals are quite small in the normal sense and on the same order as the absolute errors found for the PC board cal-kit. This measurement validates the AFR technique on a real-world device.

11.5.2 Fixture-Enhanced AFR

One key limitation of the AFR technique used in Figure 11.32 is that the connectors used for the thru standard, from which the AFR information is obtained, do not match exactly the connectors used for the DUT measurement. This slight mismatch, which was illustrated in

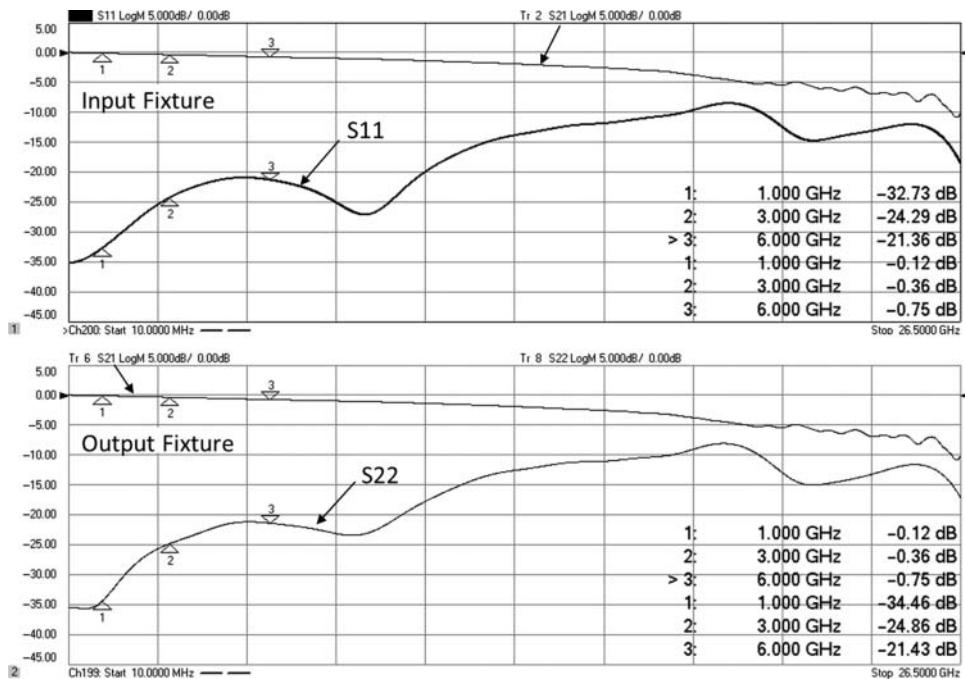


Figure 11.31 Using AFR to compute PC board input and output fixtures.

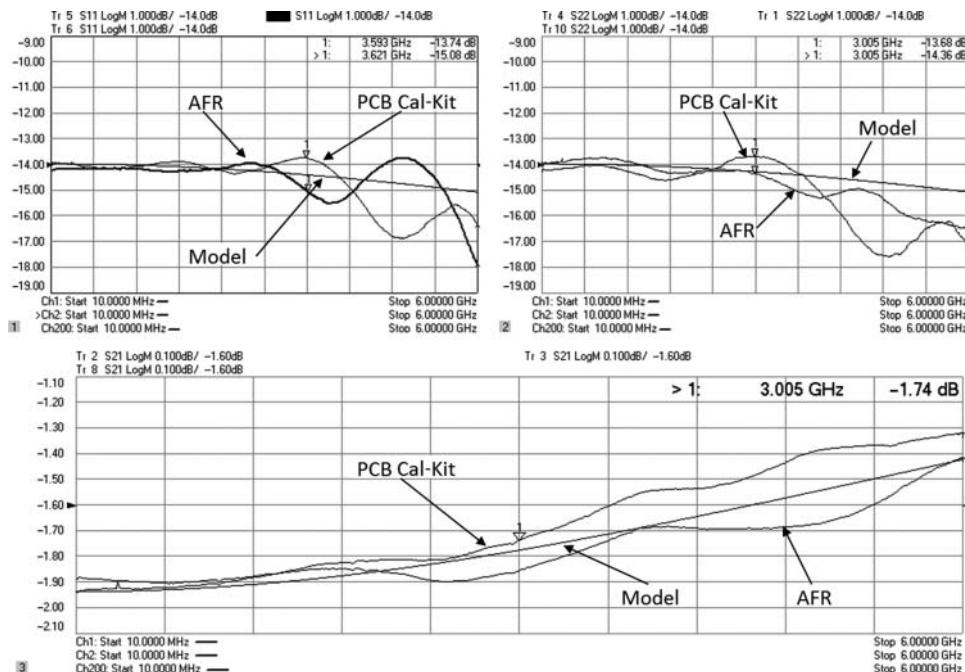


Figure 11.32 Comparing AFR measurement with in-fixture cal standards.

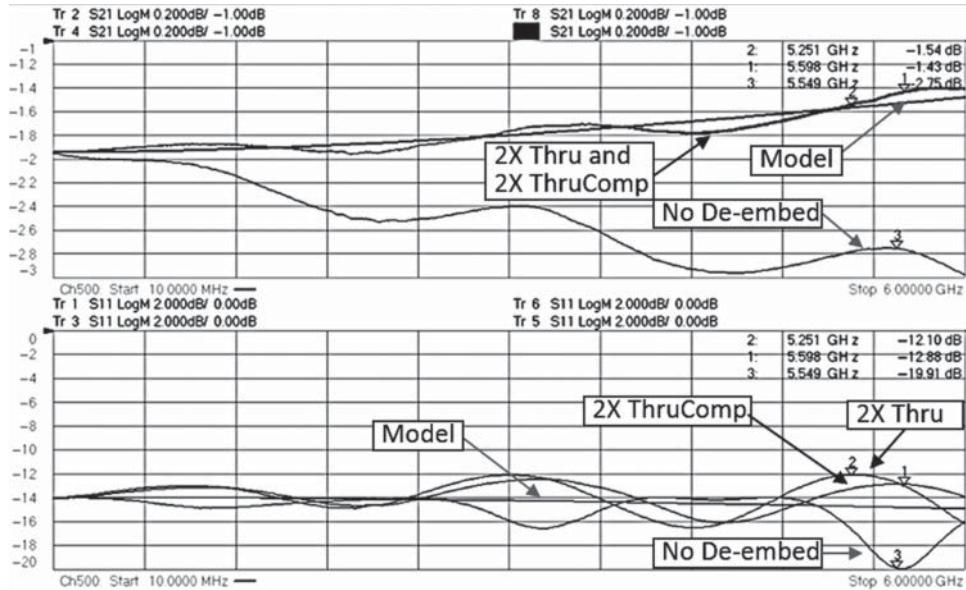


Figure 11.33 AFR with DUT fixture compensation.

Figure 11.3 and Figure 11.6, can cause errors in the estimation of the fixture used for the DUT. In the AFR computation of the fixture using the thru, the simultaneous equations were solved with the gated S_{11} and S_{22} of the thru fixture, but we know they are different from the DUT fixture. We can utilize the DUT fixture, with time gating around the S_{11} and S_{22} connectors, to recompute the estimated value of the Fixture A, S_{11A} , and Fixture B, S_{22B} , and apply them to the de-embedding parameters. In this way, the difference in the connectors can be removed as an error in the DUT measurement. Figure 11.33, while a little cluttered, shows the result of using the DUT fixture for enhancing the AFR results of S_{11A} and S_{22B} . The traces shown are the model of the resistor, the measurement of the shunt resistor with no de-embedding, and a comparison of 2X thru de-embedding and 2X thru with DUT fixture compensation (labeled 2X ThruComp). There is almost no difference in the S_{21} trace (mismatch is a second-order effect after correction), but there is a substantial difference in the S_{11} trace, there the peak ripple is reduced by about 0.6 dB in the 2X ThruComp (Marker 1 on the lower window) compared with the normal 2X thru (Marker 2 on the lower window).

11.5.3 1-Port AFR

While the 2-port AFR is a huge step forward in characterizing fixtures for packaged DUT applications, and has been widely utilized by industry, it has one major shortcoming: it requires a thru standard. In the case of a PC board fixture or a fixture used in an automatic part handler, it can be difficult to generate the thru standard. One must create a package that matches the DUT physically but has a thru connection instead of the DUT. Worse still, if the DUT is a multiport device, multiple thru standards must be created. In modern cell phones,

it is not unusual to have a front-end module (FEM) with an integrated antenna TR switch, integrated multiband low-noise amplifier (LNA), and integrated filters. These devices might have four or more TX inputs, two or more antenna outputs, and as many as eight receiver terminals (often the Rx port is differential). Thus, it becomes a 14-port device that is placed in the fixture. This requires so many thru standards to characterize the fixture that calibration becomes difficult to manage.

This problem of the high port-count fixture has been largely solved utilizing a new one-port AFR method, developed by this author. This patented (Dunsmore 2014) method utilizes a one-port standard for the AFR extraction, rather than a 2-port thru. The 1-port standard should supply a total reflection: either a short or an open. To characterize a fixture, the DUT is simply be removed, and the fixture is left open. Alternatively, a short can be developed and used. For many packaged devices, a short can be simply implemented by creating a solid metal standard matching the package dimensions. In most cases, the short is preferred as the open has a tendency to radiate if it is unshielded; the radiation is non-TEM and so is not properly calibrated for.

The example of the PC board test fixture similar to Section 11.1.1 is used to compare 1-port AFR using a short only or an open only, shown in Figure 11.34. Interestingly, the S_{11} measurement appears to match the model of the resistor better than the 2X Thru or 2X ThruComp in Figure 11.33, and match each other almost exactly (there is a very slight different near the marker).

The S_{21} of the one-port AFR de-embedded measurement shows a bit less loss than the 2X thru de-embedding. One reason might be the one-port AFR uses the reflection from the open to characterize the loss of the fixture, and with one-port unshielded measurements, the fixture standards can radiate some energy. This lost energy makes the fixture appear a bit more lossy

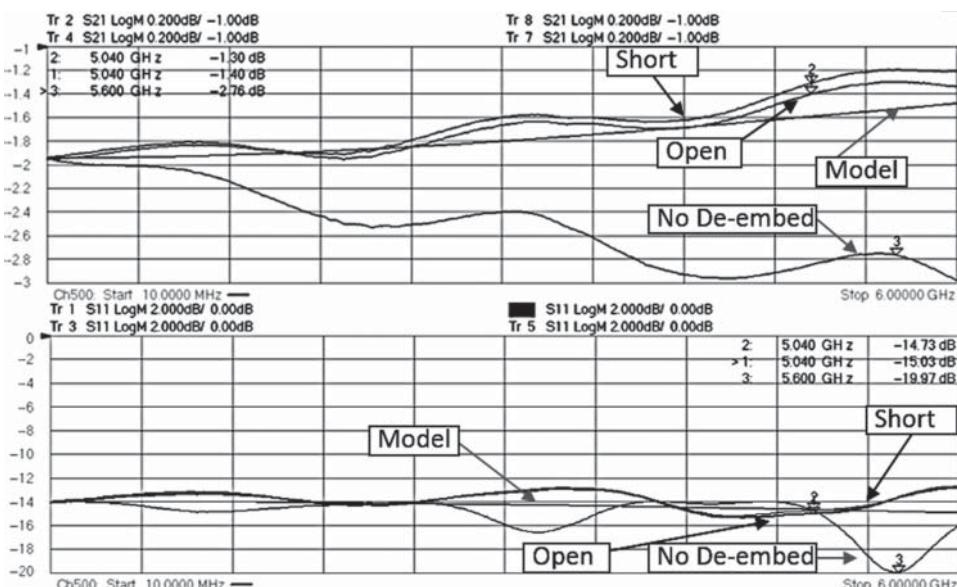


Figure 11.34 One-port AFR with a short and an open standard used.

than it really is (when the DUT is applied on the thru line, the radiation is much less), and thus, when the fixture is de-embedded, it makes the response a bit less lossy than it should be.

Still, it is remarkable that this response agrees very well with the result from using PC board standards to calibrate the fixtures, as shown in Figure 11.32. This also points to an interesting reason why both are showing a little high: the unknown thru method relies on the loss of the open and short to help characterize the loss of the test system, and if the open and short radiate a little bit, both the unknown thru and the 1-port AFR will show the effect in a similar way.

So, 1-port AFR does match very closely with a PC board calibration, but it is so much easier to do and simply requires leaving a fixture open (or shorting it) to provide full de-embedding. This is why, in many modern manufacturing situations, 1-port AFR has largely replaced in-fixture calibration kits for performing fixture compensation. Of course, a first-tier, coax-connector calibration is still required before the 1-port AFR can be used.

11.6 Embedding Port-Matching Elements

While de-embedding involves measuring a DUT in the presence of other elements such as a test-fixture, port matching and embedding involves measuring a DUT, directly at the VNA ports, but adding the effects of other virtual elements to the measured response of the DUT.

A common example of this application is an amplifier that may require external matching networks to bring the optimum power match condition from 50Ω down to some lower impedance. The amplifier designer may stipulate that the gain of the amplifier is to be measured when some exact value of external components, such as an inductor and/or capacitor, is placed in the input or output port path of the amplifier. In practice, providing a consistent value of external component can present a problem when multiple test systems are involved in manufacturing. An even greater difficulty occurs if the port matching should be done with an on-wafer device. Some probe manufacturers will create custom probes with port matching elements, but these can be quite expensive, have long lead times, and it is difficult to verify their correct values.

Instead, a virtual port matching can be performed mathematically using techniques quite similar to de-embedding. In practice, de-embedding techniques can be directly used by recognizing that embedding the effects of an S-parameter network is the same as de-embedding the parameters of the equivalent anti-network. The anti-network, S^A , is defined as that network, which when cascaded with the desired network S^N , forms a unity S-parameter matrix S^U , which has the characteristics of $S_{11} = S_{22} = 0$, and $S_{21} = S_{22} = 1$; the flow graph for this is shown in Figure 11.35. Note that as an alternative definition, an anti-network could be defined

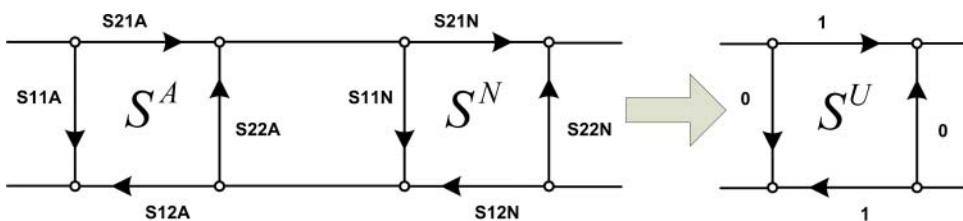


Figure 11.35 Signal flow for an anti-network.

such that the anti-network follows the desired network, rather than precedes it as below (Hong and Lancaster 2001), and will result in different values for the anti-network.

Using normal flow-graph analysis to correspond the left flow graph to the unity flow graph on the right, the values of the anti-network are computed. The output term of the anti-network is computed first as

$$S_{22}^A = \frac{S_{22}^N}{(S_{11}^N \cdot S_{22}^N - S_{21}^N \cdot S_{12}^N)} \quad (11.28)$$

and used to derive all the transmission terms for anti-network values, as

$$S_{12}^A = \frac{(1 - S_{11}^N \cdot S_{22}^A)}{S_{12}^N} \quad (11.29)$$

$$S_{21}^A = \frac{(1 - S_{11}^N \cdot S_{22}^A)}{S_{21}^N} \quad (11.30)$$

which are then in turn used to compute the input term of the anti-network as

$$S_{11}^A = \frac{S_{21}^A \cdot S_{12}^A \cdot S_{11}^N}{(S_{11}^N \cdot S_{22}^A - 1)} \quad (11.31)$$

In other publications, it is common to show the anti-network following the desired network, but such a configuration is not commensurate with the common understanding of de-embedding requiring that the network that is to be de-embedded has port 1 facing the port of the VNA.

Figure 11.36 shows the equivalent representation of a DUT with an anti-network/network pair preceding and following the DUT. Note that the S-parameters of the final cascade before

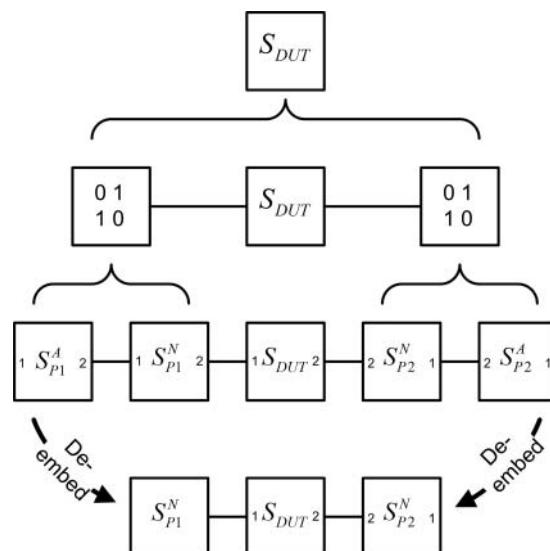


Figure 11.36 Representation of a DUT with input and output port matching, before de-embedding.

de-embedding are identical to those of the original DUT. From this diagram, it is clear that if the anti-network of the port 1 matching circuit is de-embedded from port 1 and that of port 2 is de-embedded from port 2, the S-parameter calculation will be exactly the DUT with the port 1 and port 2 networks' effects embedded in the result.

Using this technique allows port matching of networks to be accomplished using de-embedding functions. Modern VNAs typically include both de-embedding and embedding functions (often called *port matching*), so the step of computing the anti-networks is not required. One last note on computing anti-networks: because the computations include the difference between two values in the denominator (particularly the computation of S_{22A}), the values of the anti-networks can become undefined or infinite at some particular frequencies. Alternative computations, for example, using T-parameters or ABCD networks, while having their own numerical issues, may be computable where anti-networks are not.

11.7 Impedance Transformations

Most VNAs are available only with $50\ \Omega$ nominal impedance, although some RF models also provide a $75\ \Omega$ version. However, in many cases, it is desired to know the S-parameters of a DUT referenced to other than $50\ \Omega$. One method is to transform the S-parameters into some other parameter that does not have a port impedance dependency, such as Z-parameters (see Chapter 2). However, fixturing and de-embedding techniques can provide a simple alternative, by recognizing that the impedance transformation can be implemented by creating an S-parameter model of an ideal, lossless transformer and then embedding that into a measurement. To transform from one impedance to another, the S-parameters of an ideal transformer (Hong and Lancaster 2001), represented in Figure 11.37, are computed based on the ratio of impedances.

For a transformer, the impedance transformation effect goes as n^2 so that to transform from the reference impedance Z_0 to some load impedance Z_L , the turns ratio is computed as

$$n = \sqrt{\frac{Z_L}{Z_0}} \quad (11.32)$$

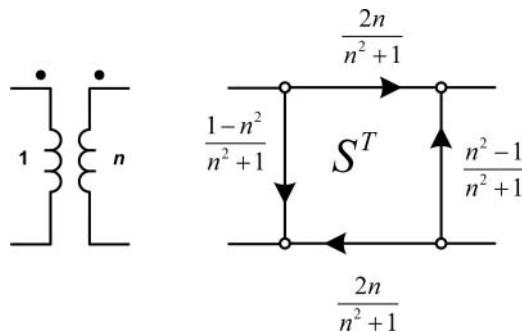


Figure 11.37 S-parameters of an ideal transformer.

and the S-parameters are computed as

$$\begin{aligned} S_{11}^T &= \frac{1-n^2}{n^2+1}, & S_{12}^T &= \frac{2n}{n^2+1} \\ S_{21}^T &= \frac{2n}{n^2+1}, & S_{22}^T &= \frac{n^2-1}{n^2+1} \end{aligned} \quad (11.33)$$

To change the effective value of a VNA test port impedance, one may calibrate in one impedance, compute the S-parameters for an ideal transformer to another impedance, and de-embed those S-parameters to allow the VNA readings to match those that would have been obtained if the VNA had been of the impedance. The S-parameters are correct, but to get the Smith chart to read the correct values of impedance, one must also change the system Z_0 , which sets the value for the center of the Smith chart.

But, of course, the actual impedance of the VNA port is unchanged. It is not likely to be exactly equal to either impedance, but rather its impedance will be the raw impedance of the test port. For RF VNAs, at 50Ω , the actual test port impedance typically ranges from 40 to 60Ω . For high-frequency VNAs, operating up to 50 or 70 GHz, mismatches in the VNA have a larger effect, and it is not uncommon for the VNA impedance to range from 25 to 100Ω . One should recognize that the raw S-parameters, while measured in whatever impedance the test port provides, are transformed to a 50Ω reference through the error-correction process. If this transformation is valid, it is also valid to transform to some other impedance.

11.8 De-embedding High-Loss Devices

It is common in the measurement of high-power devices to add a large attenuator at the output of the DUT to reduce the power level to the VNA to avoid damage. Because of the loss, the match measurements at the output port can be quite poor and normal 2-port calibration techniques fail to give good results. For example, Ecal modules cannot orient themselves because there is too little difference between impedance states apparent at the VNA test port. A good rule of thumb is that de-embedding devices or fixtures with up to 10 dB insertion loss (one way) can be easily performed. Up to 15 dB of loss and low reflections (an attenuator would qualify, but not a filter in its stop band) can be de-embedded with care, using averaging and low IF bandwidths. If the device to be de-embedded has more 20 insertion loss, de-embedding will likely yield poor results.

Consider the case of de-embedding a 20 dB attenuator placed at port 2, at the output of an amplifier. The S_{21} trace would have a large signal, due to the amplifier gain, even with the 20 dB of attenuation. But the S_{22} signal would be quite small and noisy. Likely, the reflection from the attenuator match itself would be larger than the reflection signal from the amplifier. After applying the de-embedding math, the value for ERF would be very small, and the error correction would have the effect of subtracting two very large numbers, one of them noisy, to generate the S_{21} trace; that is, the noise and error in S_{22} would be imposed on the S_{21} trace (this is discussed in detail in Chapter 6). The error correction for the port 2 match would be adding more error than it is removing, since the match of the attenuator presented to the DUT amplifier is typically quite good. Enhanced-response calibration, discussed in Chapter 3, can be used in such a case to good effect.

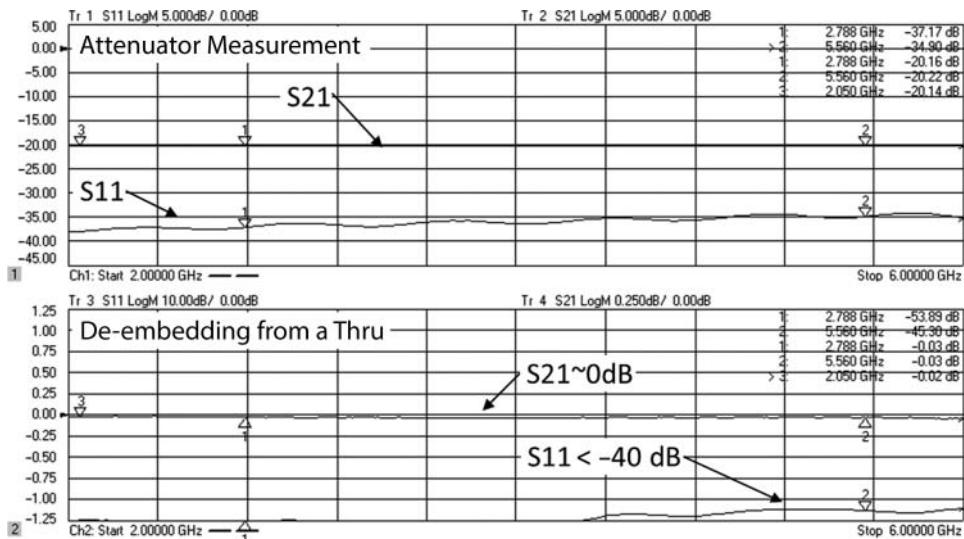


Figure 11.38 Attenuator S_{11} and S_{21} (upper); de-embedding the attenuator (middle); de-embedding while setting $S_{22} = 0$ (lower).

As an example of the limits of de-embedding, consider Figure 11.38. The upper window shows the S_{11} and S_{21} of a 20 dB attenuator. The return loss is quite good at -34 dB, and the insertion loss is pretty flat at -20 dB.

This adapter is added to port 2 of a VNA and is de-embedded; a thru-connection is measured and displayed in the lower window. In this case, de-embedding appears to work well, and the S_{21} measurement is very close to 0 dB. The S_{11} measurement is essentially the measurement of the S_{22} of the attenuator on port 2 (remembering that the common practice is to label it as port 1 of the port of the de-embedded device facing VNA port 2).

However, the measurement from port 2 of S_{22} of the DUT (a thru in this case) is very unstable. Thus, the correction for mismatch at port 2 between the attenuator and the DUT is also unstable. In many cases, the instability is worse than the errors being removed, and in these cases, changing the value of S_{22} of the de-embedded device to zero will actually improve the results.

The mismatch error effect would be even larger in the case of de-embedding a device with both large loss and high reflections; a filter in its stop band is a case in point. In such a case the raw S_{22} is quite large, due to the reflection back to port 2 of the filter to be de-embedded, and the filter's S_{22} (remember that port 2 of the de-embedded device faces the DUT) presents a very poor load match to the DUT. Any inaccuracy in measuring the error terms would yield large errors to the S_{21} measurements, because very large raw errors are being compensated.

One method to allow reasonable de-embedding of a transmission measurement but avoid difficulties in the match correction with very lossy networks is to set the reflection term of the de-embedding network that faces the DUT (normally S_{22}) to zero. When this is set to zero, and the loss is high, the mathematical effect of the load match for the fixture is nearly zero as well, and very little S_{22} match correction effect is applied. De-embedding of a filter can clearly demonstrate this effect.

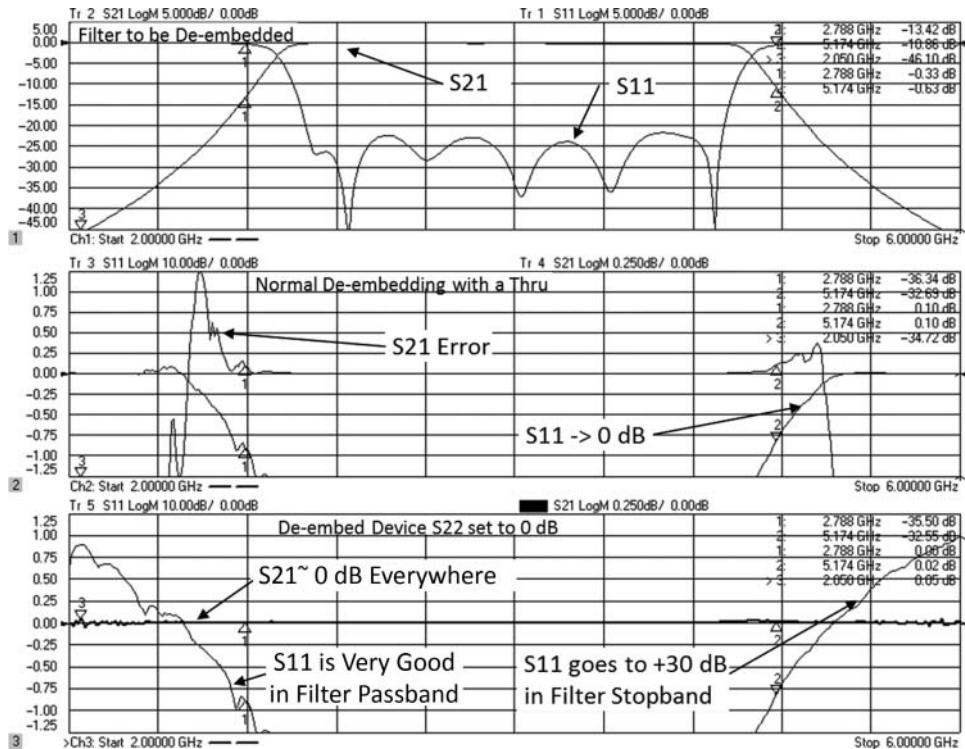


Figure 11.39 S-parameter of a filter (upper); after normal de-embedding (middle); with setting S_{22} to 0 (lower).

Figure 11.39 shows the S_{11} and S_{21} response of a filter to be de-embedded. Its measurement is saved as an S2P file. The filter is then de-embedded from port 1. The resulting middle S_{21} trace should be a flat line over all frequency, but noise, drift, and errors in measurement cause the S_{21} trace to deviate from a flat line when the loss of the filter is greater than about -10 dB, as shown by the markers at the lower and upper passband edges. The fact that the filter cannot be de-embedded successfully for even moderate loss is often a surprise to many users. The actual return loss of the filter is nearly 0 dB (total reflection) at these band edges, so this case is very sensitive to match correction.

Remarkably, a simple change to the filter S-parameters can greatly improve the measurement of S_{21} . The lowest trace is the same network de-embedding, but this time with S_{22} of the de-embedded network (the filter) set to zero. Here it is clear that in the lossy region of the filter, the quality of the S_{21} trace with this de-embedding is greatly improved. While normal de-embedding gives an error greater than 0.1 dB when the loss of the network is near 10 dB and greater than 1 dB for losses of 20 dB, the lower trace has less than 0.05 dB error even though the loss is greater than 45 dB. Of course, the de-embedded value for match is in substantial error outside the passband of the filter, but here the transmission measurement quality is the key figure of merit. Further, the normal de-embedding gives terrible results for the S_{11} as well; in this case the DUT is a thru line with -50 dB return loss, and both de-embedding

functions give S_{11} as 0 dB at the 20 dB of filter loss. It is merely academic that the new method of de-embedding is worse (providing positive S_{11}) at even higher losses.

Using the method of setting S_{22} of the de-embedded network to zero can remove much instability in the resulting measurement. Looking at the result from a mathematical aspect, it is quite similar to turning the 2-port cal with de-embedding into an enhanced response cal, since the load match correction is essentially removed with S_{22} of the de-embedding network set to zero. As discussed in Chapter 6, enhanced response calibration is a very good method for dealing with high loss at port 2.

11.9 Understanding System Stability

System stability describes how well the system maintains its calibrated measurement accuracy over time, temperature, and measurement connections. While the VNA system itself is often questioned as to its long-term stability, without question the largest source of instability is the test port cables. The next section describes methods to evaluate the quality and stability of test port cables.

11.9.1 Determining Cable Transmission Stability

The stability of a cable can be tested in three ways to determine its effect on the VNA test system. The first test is a simple test of transmission stability. A cable is connected between ports of a VNA, and two traces of S_{21} are displayed, one for magnitude and one for phase. The cable response is normalized (for example, using data into memory and data over memory). The cable is then flexed. If desired, it can be disconnected and reconnected. The worst-case magnitude and phase deviation are recorded. Some VNAs provide functions that allow tracking the minimum or maximum of a response automatically, for example in the Keysight PNA, the equation editor function Maxhold (mag (S_{21})) will give the worst-case deviation at each frequency. An example is shown in Figure 11.40 for magnitude stability of both a metrology-grade cable and a lower-cost flexible braided cable. Each cable was normalized, and each flexed 10 times then reconnected. The metrology cable has almost no change in transmission (less than 0.01 dB). The braided cable has approximately 0.1 dB deviation over much of the frequency response, with a large deviation (likely due to a cable mode) around 16 GHz. Even without the mode, the metrology cable is more than 20 times more stable than the braided cable.

11.9.2 Determining Cable Mismatch Stability

While transmission stability is typically specified for a cable, it is usually the match stability that causes most of the problems when doing fully corrected measurements, especially reflection measurements. The match stability of cables is often not specified, and measuring the match stability is not something that is well documented or widely agreed upon. A very good way of measuring match stability is to terminate the cable with various loads and then flex the cable while looking at the response.

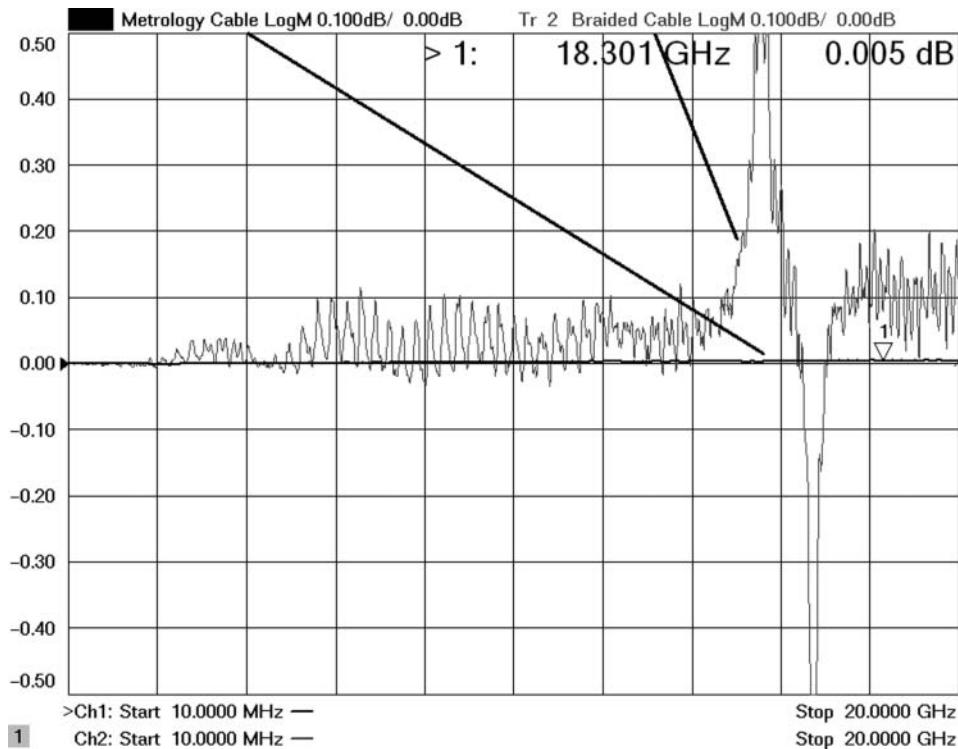


Figure 11.40 Stability in S_{21} of a metrology cable and a flexible braided cable.

The cable's input match gives a reflection that must be calibrated out. This reflection becomes part of the raw directivity, and stability of this match adds directly to the residual directivity. To test for match stability, the cable is terminated in a load, and then the response of the cable is put into memory, and the display math is set to data minus memory, or Data-Mem. Performing this step of subtracting (in a vector sense) the reflection of the cable match from itself is a critical step in assessing the cable performance. The display math occurs on the linear vector data, so taking the LogMag format shows the directivity stability directly in dB terms. This technique of using data minus memory for looking at mismatch effects can be useful in many other applications as well.

The cable is then flexed, and the worst-case return loss is recorded over several cable positions. One might also notice that the cable return loss stability may become quite poor when the cable is flexed but returns to a good value when the cable is returned to the same position as when the memory trace was stored. If this is the case, one can improve the calibration performance by positioning the cable similarly for calibration as for when it will be connected to the DUT, thus minimizing the flexure error. Examples of cable directivity stability with a metrology-grade cable and a braided flexible cable are shown in the plot of Figure 11.41.

Just as in Chapter 3, where a load was used to discern the residual directivity and the measurement of a short was used to discern the residual source match, so too here a short can

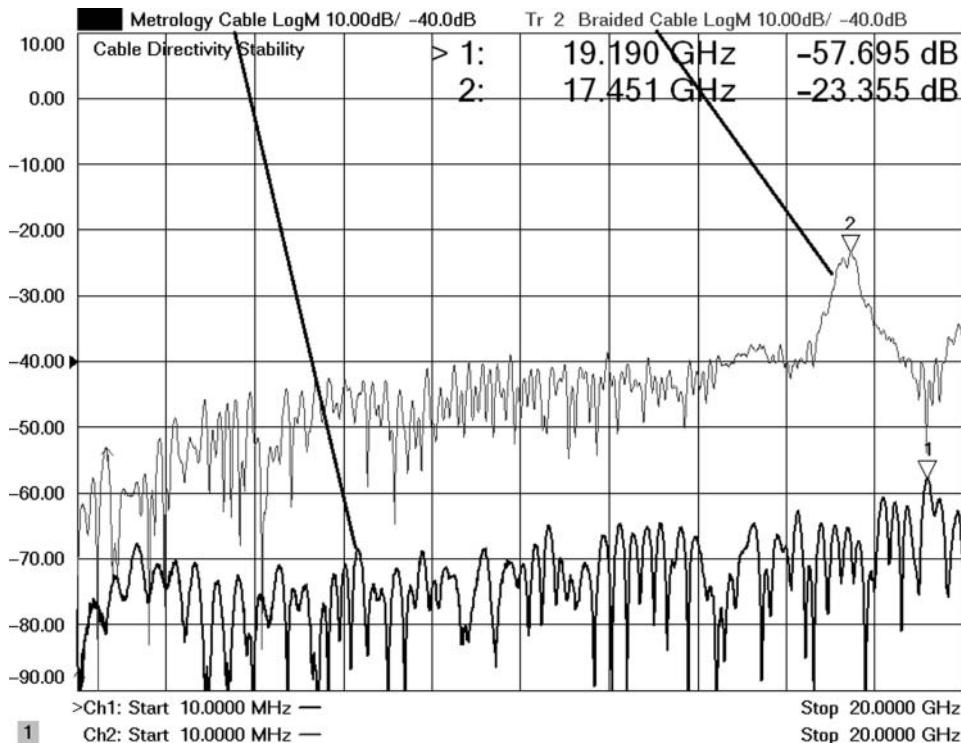


Figure 11.41 Directivity stability of a metrology cable with a load (dark); stability of a braided cable with a load (light).

be used to determine the source match stability of a cable. Just like the load, a short is added to the end of the cable, and the measurement result is stored in memory. Then the display math function data minus memory is used to display the residual source match. In the case of a terminating the cable with a load, mismatch at the far end of the cable will be absorbed by the load and not appear as a stability error. Using a short gives a combination of the directivity error from earlier, plus any mismatch error as well as any error in phase or delay of the cable. This is because the short produces a large vector reflection. If nothing but the phase changes, the vector difference between the memory and the flexed measurement will show up as a residual source match error, where the magnitude of the error is the arc tangent of the phase change.

Figure 11.42 shows the result of flexing both a metrology cable and a braided cable with short terminating the cable, and the display math set to data minus memory. It is clear here that the result is far worse than the directivity stability. This exactly illustrates that the source match stability is the sum of residual directivity plus errors in the open (or short) measurement. Source match stability is critical to the measurement of high-reflection devices, as is clear from the error analysis of Eq. (3.68) that shows the effect of residual source match, of which source match stability is a key component. From that equation, the error from the

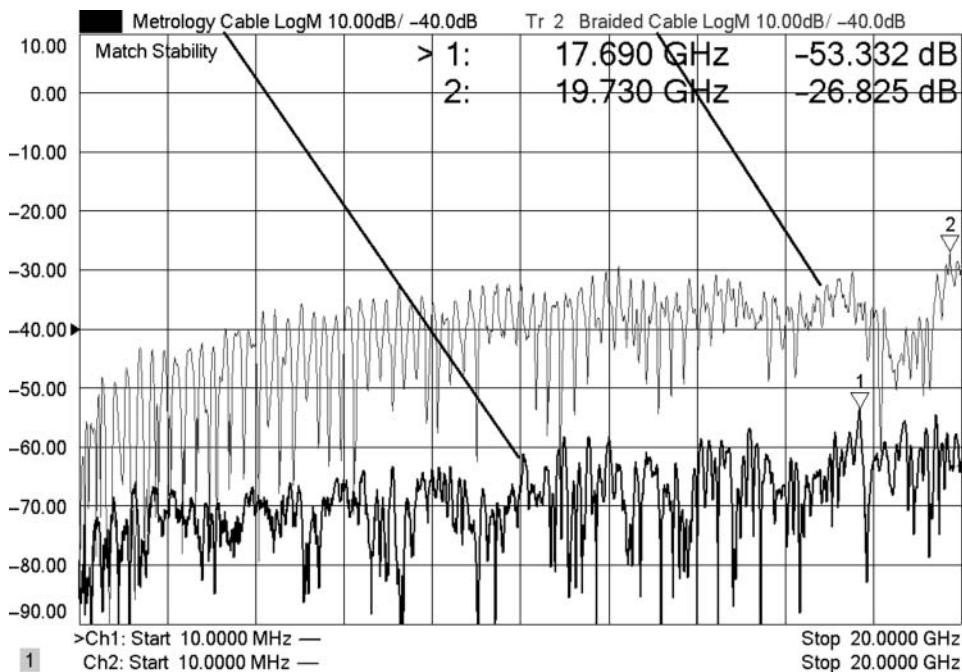


Figure 11.42 Source match stability of a metrology cable (dark), stability of a braided flexible cable (light).

source match goes as the square of the reflection coefficient. For the measurement of a load or other device with a very good match, the effect of source match stability is negligible.

11.9.3 Reflection Tracking Stability

Finally, with the same short attached, the display math can be changed to show data divided by memory, with an S_{11} trace of LogMag and one of phase displayed (both using data/mem). In this measurement, any change in the loss or the phase of the cable is directly shown by the 2-way path measurement after terminating in a short. The overall trend should compare well with the transmission stability measurements, and reflections measurements are sometimes easier to do as only one end of the cable needs to be connected to the VNA, so the other end is free to be flexed. However, if the cable has any mismatch, additional ripple will appear as the phasing of the mismatch to the short reflection will change. Figure 11.43 shows the reflection tracking phase stability of the metrology cable and the braided flexible cable.

Note that the ripple portion of the phase stability is in-line with the implied phase stability of the mismatch measurement. The phase stability should predict a mismatch stability on the order of the tangent of the phase stability, when measuring a short. In the figure, 3° of phase stability ripple translates to -26 dB of source match stability. Usually, the stability of the cable is the final limitation in the quality of the calibration.

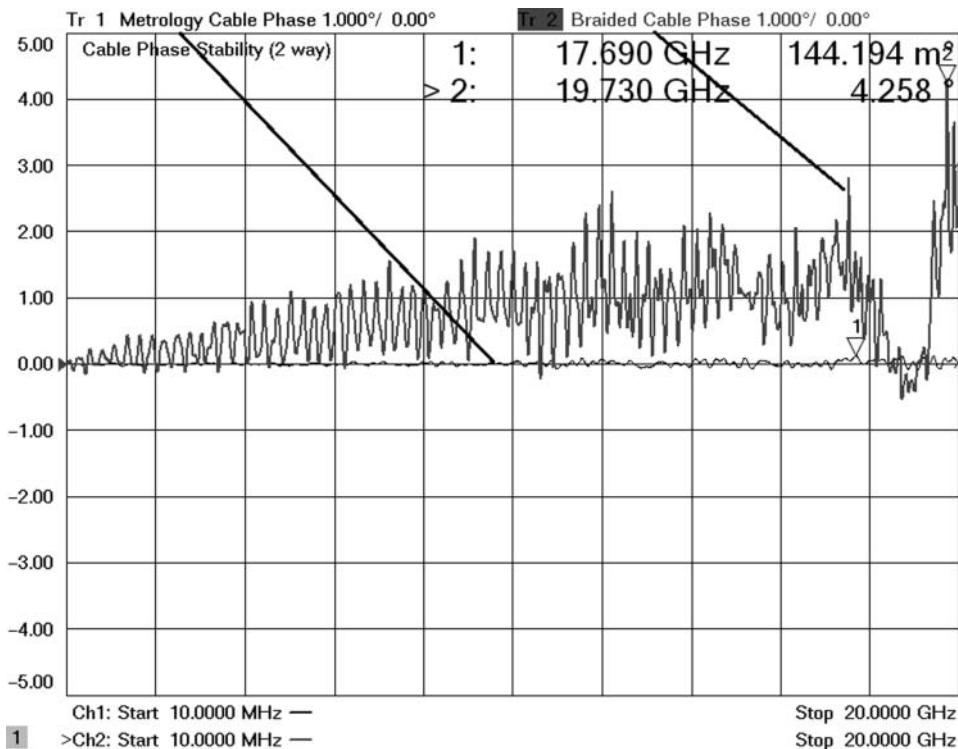


Figure 11.43 Reflection tracking phase stability of a metrology cable and a braided cable when terminated in a short.

11.10 Some Final Comments on Advanced Techniques and Measurements

In this final chapter, many advanced techniques were discussed to improve the microwave component measurement results for real-world situations where fixtures, cables, and adapters would otherwise distort the measurement results.

Users of older, legacy VNAs can use the mathematical functions described in this chapter to do offline processing of measured data, but for most users of modern VNAs, these functions are enabled directly in the user interface of the VNA.

With only a little care and the knowledge provided in this and previous chapters, the R&D or test engineer can create and optimize measurement results for almost any conceivable test condition.

Modern VNAs are rapidly replacing entire racks of test equipment, and using these advanced measurement methods, the test accuracy and test speeds are increasing at a fantastic rate, while the overall size and cost of the test systems are being reduced. Companies and test engineers who fail to utilize these new methods put themselves at a competitive disadvantage. The author hopes that the material presented in this book helps level the playing field.

References

- Dunsmore, J., Ning Cheng, Yong-xun Zhang. (2011) “Characterizations of asymmetric fixtures with a two-gate approach”, Microwave Measurement Conference (ARFTG), 2011 77th ARFTG, 10 June 2011, pp. 1–6.
- Dunsmore, J., Ning, C., and Zhang Y. P. (2014) Automatic fixture removal using one-port measurement. US Patent No. 9,086,376.
- Hong, J.-S. and Lancaster, M.J. (2001). *Microstrip Filters for RF/Microwave Applications*. New York: Wiley. Print.
- Inter-Continental Microwave. (2006) Agilent AN 1287-9 In-Fixture Measurements Using Vector Network Analyzers Application Note [Online], Available at: <http://www.icmicrowave.com/pdf/AN%201287-9.pdf>

Appendix A

Physical Constants

Symbol	Definition	Value
c	Speed of light in free space	$2.997\ 9 \times 10^8\ \text{m s}^{-1} \approx 300 \times 10^6\ \text{m s}^{-1}$
ϵ_0	Permittivity of free space	$8.854 \times 10^{-12}\ \text{F m}^{-1} \approx (1/36\pi) \times 10^{-9}\ \text{F m}^{-1}$
μ_0	Permeability of free space	$4\pi \times 10^{-7}\ \text{H m}^{-1}$
η_0	Impedance of free space	$\sqrt{\frac{\mu_0}{\epsilon_0}} \approx 120\pi\ \Omega \approx 376.73\ \Omega$
k	Boltzmann's constant	$1.380\ 65 \times 10^{-23}\ \text{J K}^{-1}$
kT_0B	Noise power at 290 K in a given bandwidth	$4 \times 10^{-21}\ \text{W Hz}^{-1} \approx -173.98\ \text{dBm in 1 Hz BW}$

Appendix B

Common RF and Microwave Connectors

Name/Notes	Outer conductor diameter (air dielectric only)	Rated frequency (GHz)	First mode (air dielectric only) (GHz)	Maximum usable frequency (GHz)
7/16	16 mm	7.5	8.1	7.5
Type N (50 Ω) precision	7 mm	18	18.6	26.5 ^a
Type N (50 Ω) commercial	7 mm	12	12.5	15
Type N (75 Ω) precision	7 mm	18	18.6	18
Type N (75 Ω) commercial	7 mm	12	12.5	15
7 mm (e.g. APC-7)	7 mm	18	18.6	18
BNC	Dielectric interface	4	N/A	11
TNC (Threaded BNC)	Dielectric interface	4	N/A	11
SMA	Dielectric interface	18	N/A	22
QMA (Snap-on SMA)	Dielectric interface	6	N/A	~18
SMB (Snap-on)	Dielectric interface	4	N/A	10
SMC (Threaded SMB)	Dielectric interface	4	N/A	10
3.5 mm	3.5 mm	26.5	28	33
SSMA	Dielectric interface	36	N/A	36
2.92 mm (“K”)	2.4 mm	40	44	44
2.4 mm	2.4 mm	50	52	55
1.85 mm (“V”)	1.85 mm	67	68.5	70
1 mm	1 mm	110	120	~125

^aSome instrument manufacturers place this connector on 26.5 GHz instruments because it is rugged; it has the same first modes as Type N and 7 mm.

Appendix C

Common Waveguides

Waveguide designation	Operating band	Normal frequency (GHz)	Lower cutoff frequency (GHz)	Next mode cutoff frequency (GHz)
WR-284	S (part)	2.60–3.95	2.08	4.16
WR-187	C (part)	3.95–5.85	3.15	6.31
WR-137	C (part)	5.85–8.20	4.30	8.6
WR-90	X	8.2–12.4	6.56	13.11
WR-62	Ku, P ^a	12.4–18.0	9.49	18.98
WR-42	K	18.0–26.5	14.05	28.10
WR-28	Ka, R ^a	26.5–40.0	21.08	42.15
WR-22	Q	33.0–50.0	26.35	52.69
WR-19	U	40.0–60.0	31.39	62.78
WR-15	V	50.0–75.0	39.88	79.75
WR-12	E	60.0–90.0	48.37	96.75
WR-10	W	75.0–110.0	59.01	118.03
WR-8	F	90.0–140.0	73.77	147.54
WR-6	D	110.0–170.0	90.79	181.58
WR-5	G	140.0–220.0	115.714	231.43
WR-4	Y	170.0–260.0	137.242	274.49
WR-3	J	220.0–325.0	173.571	347.14
WR-2		325–500	295.07	590.14
WR-1.5		500–750	393	786
WR-1		750–1100	590	1180

^aAlternative band designation

Appendix D

Some Definitions for Calibration Kit Opens and Shorts

Connector	Standard	Definition
7/16 mm (Agilent 85038A)	Open (male or female)	$\text{Delay} = 66.734 \text{ ps}$, $\text{Loss} = 0.63 \text{ G}\Omega \text{ s}^{-1}$ $C_0 = 32 \times 10^{-15} \text{ F}$ $C_1 = 100 \times 10^{-27} \text{ F Hz}^{-1}$ $C_2 = -50 \times 10^{-36} \text{ F Hz}^{-2}$ $C_3 = 100 \times 10^{-45} \text{ F Hz}^{-3}$
	Short (male or female)	$\text{Delay} = 66.734 \text{ ps}$, $\text{Loss} = 0.63 \text{ G}\Omega \text{ s}^{-1}$ $L_0 = 0 \times 10^{-12} \text{ H}$ $L_1 = 0 \times 10^{-24} \text{ H Hz}^{-1}$ $L_2 = 0 \times 10^{-33} \text{ H Hz}^{-2}$ $L_3 = 0 \times 10^{-42} \text{ H Hz}^{-3}$
7 mm (Agilent 85050D)	Open	$\text{Delay} = 0 \text{ ps}$, $\text{Loss} = 0.7 \text{ G}\Omega \text{ s}^{-1}$ $C_0 = 90.4799 \times 10^{-15} \text{ F}$ $C_1 = 763.303 \times 10^{-27} \text{ F Hz}^{-1}$ $C_2 = -63.8176 \times 10^{-36} \text{ F Hz}^{-2}$ $C_3 = 6.4337 \times 10^{-45} \text{ F Hz}^{-3}$
	Short	$\text{Delay} = 0 \text{ ps}$, $\text{Loss} = 0.7 \text{ G}\Omega \text{ s}^{-1}$ $L_0 = 0.3566 \times 10^{-12} \text{ H}$ $L_1 = -33.392 \times 10^{-24} \text{ H Hz}^{-1}$ $L_2 = 1.7542 \times 10^{-33} \text{ H Hz}^{-2}$ $L_3 = -0.0336 \times 10^{-42} \text{ H Hz}^{-3}$

(Continued)

Connector	Standard	Definition
Type N precision (Agilent 85054B)	Open (male)	Delay = 57.993 ps, Loss = 0.93 GΩ s ⁻¹ C0 = 89.939 × 10 ⁻¹⁵ F C1 = 2536.8 × 10 ⁻²⁷ F Hz ⁻¹ C2 = -264.99 × 10 ⁻³⁶ F Hz ⁻² C3 = 13.4 × 10 ⁻⁴⁵ F Hz ⁻³
	Open (female)	Delay = 22.905 ps, Loss = 0.93 GΩ s ⁻¹ C0 = 104.13 × 10 ⁻¹⁵ F C1 = -1943.4 × 10 ⁻²⁷ F Hz ⁻¹ C2 = 144.62 × 10 ⁻³⁶ F Hz ⁻² C3 = 2.2258 × 10 ⁻⁴⁵ F Hz ⁻³
	Short (male)	Delay = 63.078, Loss = 1.1273 GΩ s ⁻¹ L0 = 0.7563 × 10 ⁻¹² H L1 = 459.88 × 10 ⁻²⁴ H Hz ⁻¹ L2 = -52.429 × 10 ⁻³³ H Hz ⁻² L3 = 1.5846 × 10 ⁻⁴² H Hz ⁻³
3.5 mm (Agilent 85052D)	Short (female)	Delay = 27.99, Loss = 1.3651 GΩ s ⁻¹ L0 = -0.1315 × 10 ⁻¹² H L1 = 606.21 × 10 ⁻²⁴ H Hz ⁻¹ L2 = -68.405 × 10 ⁻³³ H Hz ⁻² L3 = 2.0206 × 10 ⁻⁴² H Hz ⁻³
	Open (male or female)	Delay = 29.243 ps, Loss = 2.2 GΩ s ⁻¹ C0 = 49.433 × 10 ⁻¹⁵ F C1 = -310.13 × 10 ⁻²⁷ F Hz ⁻¹ C2 = 23.168 × 10 ⁻³⁶ F Hz ⁻² C3 = -0.15966 × 10 ⁻⁴⁵ F Hz ⁻³
	Short (male or female)	Delay = 31.785 ps, Loss = 2.36 GΩ s ⁻¹ L0 = 2.0765 × 10 ⁻¹² H L1 = -108.54 × 10 ⁻²⁴ H Hz ⁻¹ L2 = 2.1705 × 10 ⁻³³ H Hz ⁻² L3 = -0.01 × 10 ⁻⁴² H Hz ⁻³
2.92 mm (Maury 8770D)	Open (male)	Delay = 14.982 ps, Loss = 1.8 GΩ s ⁻¹ C0 = 47.5 × 10 ⁻¹⁵ F C1 = 0 × 10 ⁻²⁷ F Hz ⁻¹ C2 = 3.8 × 10 ⁻³⁶ F Hz ⁻² C3 = 0.19 × 10 ⁻⁴⁵ F Hz ⁻³
	Open (female)	Delay = 14.883 ps, Loss = 1.8 GΩ s ⁻¹ C0 = 45.5 × 10 ⁻¹⁵ F C1 = 100 × 10 ⁻²⁷ F Hz ⁻¹ C2 = 0.3 × 10 ⁻³⁶ F Hz ⁻² C3 = 0.21 × 10 ⁻⁴⁵ F Hz ⁻³
	Short (male or female)	Delay = 16.83 ps, Loss = 1.1273 GΩ s ⁻¹ L0 = 0 × 10 ⁻¹² H L1 = 0 × 10 ⁻²⁴ H Hz ⁻¹ L2 = 0 × 10 ⁻³³ H Hz ⁻² L3 = 0 × 10 ⁻⁴² H Hz ⁻³

Connector	Standard	Definition
2.4 mm	Open (male or female)	Delay = 20.837 ps, Loss = $3.23 \text{ G}\Omega \text{ s}^{-1}$ $C_0 = 29.722 \times 10^{-15} \text{ F}$ $C_1 = 165.78 \times 10^{-27} \text{ F Hz}^{-1}$ $C_2 = -3.5386 \times 10^{-36} \text{ F Hz}^{-2}$ $C_3 = 0.071 \times 10^{-45} \text{ F Hz}^{-3}$
	Short (male or female)	Delay = 22.548 ps, Loss = $3.554 \text{ G}\Omega \text{ s}^{-1}$ $L_0 = 2.1636 \times 10^{-12} \text{ H}$ $L_1 = -146.35 \times 10^{-24} \text{ H Hz}^{-1}$ $L_2 = 4.0443 \times 10^{-33} \text{ H Hz}^{-2}$ $L_3 = -0.0363 \times 10^{-42} \text{ H Hz}^{-3}$
1.85 mm	All	Data-based standards ^a
1 mm	All	Data-based standards ^a
Type N 75 Ω (Agilent 85036B/E)	Open (male)	Delay = 17.544 ps, Loss = $1.13 \text{ G}\Omega \text{ s}^{-1}$ $C_0 = 41 \times 10^{-15} \text{ F}$ $C_1 = 40 \times 10^{-27} \text{ F Hz}^{-1}$ $C_2 = 5 \times 10^{-36} \text{ F Hz}^{-2}$ $C_3 = 0 \times 10^{-45} \text{ F Hz}^{-3}$
	Open (female)	Delay = 0 ps, Loss = $1.13 \text{ G}\Omega \text{ s}^{-1}$ $C_0 = 63.5 \times 10^{-15} \text{ F}$ $C_1 = 84 \times 10^{-27} \text{ F Hz}^{-1}$ $C_2 = 56 \times 10^{-36} \text{ F Hz}^{-2}$ $C_3 = 0 \times 10^{-45} \text{ F Hz}^{-3}$
	Short (male)	Delay = 17.544, Loss = $1.13 \text{ G}\Omega \text{ s}^{-1}$ $L_0 = 0 \times 10^{-12} \text{ H}$ $L_1 = 0 \times 10^{-24} \text{ H Hz}^{-1}$ $L_2 = 0 \times 10^{-33} \text{ H Hz}^{-2}$ $L_3 = 0 \times 10^{-42} \text{ H Hz}^{-3}$
	Short (female)	Delay = 0.93, Loss = $1.13 \text{ G}\Omega \text{ s}^{-1}$ $L_0 = 0 \times 10^{-12} \text{ H}$ $L_1 = 0 \times 10^{-24} \text{ H Hz}^{-1}$ $L_2 = 0 \times 10^{-33} \text{ H Hz}^{-2}$ $L_3 = 0 \times 10^{-42} \text{ H Hz}^{-3}$

^aData-based standards do not follow the normal polynomial fit but use a data file based on an explicit reflection versus frequency for both amplitude and phase.

Appendix E

Frequency, Wavelength, and Period

$$F = \frac{c}{\lambda}, \quad \lambda = \frac{c}{F} = c \cdot T, \quad T = \frac{1}{F}$$

Frequency	Common band names ^a	Wavelength (km, m, microns, nm)	Wavelength (miles, feet, inch, mils)	Period
1 Hz	Sub-acoustic	300 000 km	186 411 miles	1 s
10 Hz	ELF	30 000 km	18 641 miles	100 ms
100 Hz	Acoustic	3000 km	1864 miles	10 ms
1 kHz	Acoustic	300 km	186 miles	1 ms
10 kHz	VLF	30 km	18.6 miles	100 μ sec
100 kHz	LF	3 km	1.86 miles	10 μ sec
1 MHz	MF	300 m	984 feet	1 μ sec
10 MHz	HF	30 m	98 feet	100 ns
100 MHz	VHF	3 m	9.8 feet	10 ns
1 GHz	RF	30 cm	\sim 1 foot	1 ns
10 GHz	Microwave	3 cm	1.2 inch	100 ps
100 GHz	mm-wave	3 mm	0.12 inch	10 ps
1 THz	Sub mm-wave	300 micron	12 mils	1 ps
10 THz	Quasi-optical	30 micron	1.2 mils	100 fs
100 THz	Infrared	3 micro	0.12 mils	10 fs
1000 THz	UV (optical)	300 nm	0.012 mils	1 fs

^aThe common band names apply from approximately 0.3–3 times the listed frequency.

Index

- ABCD-parameters, 124
absolute mixer p, 473
active load, 418
actual waves, 181
adapter-removal, 152
additive-white-gaussian-noise (AWGN), 24, 546
adjacent-channel, 23
adjacent channel leakage ratio (ACLR), 23, 584
adjacent channel power (ACP), 23, 584
adjacent channel power ratio (ACPR), 23, 584
admittance, 113
airline, 29
amplifier, 58
 analysis, 357
 booster, 2, 105, 390, 394
 calibration, 347
 compression, 344
 DC power, 355
 differential, 680
 differential limiting, 692
 distortion, 571
 doherty, 410
 harmonics, 404
 high-gain, 354
 high-power, 390
 linear, 343
 low-noise, 58
 matching, 360
 measurement optimization, 346
 measuring, 343
 output power, 351
 power, 59, 389
 pretest, 346
 pulsed, 394
 stabilization, 359
 system, 58
AM-PM, 376
antenna, 52, 338
antenna noise figure, 638
arbitrary waveform generator (AWG), 65
A transmission in the forward flow graph (ATF), 132
attenuation, 83, 278
 long cables, 287
 measurement, 278
attenuator
 receiver, 104
 source, 83
 step, 223
automatic fixture removal (AFR), 764
 example, 768
 fixture, 768
 one-port, 770
automatic port extensions (APE), 759
available gain, 363
backside ground, 737
balanced measure, 667
balanced transmission line, 675
balun, 86
 differential, 708

- band pass response, 238
 band pass transform, 251
 band power, 589
 density, 591
 beta, 246
 block-diagram, 73–75
 spectrum analyzer, 539
 VNA, 70–73
 bolometer, 63
 B transmission forward tracking (BTF), 132

 cable
 airline loss, 29
 characteristic impedance, 27
 delay measurement, 305
 differential, 328
 electrically long, 287
 flexure, 220
 generalized loss, 29
 impedance measurement, 303
 length measurement, 277
 loss, 29
 measurement, 271
 one-way, 292
 SRL, 299
 stability, 778
 structural return loss, 31
 velocity of propagation, 30
 cables are like, 220
 Cal All, 182
 calibration
 adapter removal, 152
 DC, 355
 devolved, 194
 Ecal, 162
 enhanced-response, 196
 errors, residual
 directivity, 199
 load match, 206
 reflection tracking, 202
 source match, 202
 transmission tracking, 205
 flattening, 321
 in-situ, 290
 mixer, 476
 one-port, 136
 phase cal, 479
 phase-reference, 485
 QSOLT, 161
 reciever power, 48, 173
 reciprocal mixer, 484
 SOLR, 158
 source power, 167, 347
 split-cal, 478
 swap equal adapter, 151
 three-mixer, 479
 TRL, 153
 TRM, 157
 twelve-term, 134
 unknown-thru, 158
 using airlines, 199
 waveguide, 166
 calibration standards
 isolation, 148
 load, fixed, 145
 load, sliding, 146
 one-port, 139
 open, 141
 short, 144
 thru, 149
 TRL line, 155
 TRL reflect, 156
 TRL thru, 154
 two-port, 148
 cal-kit, 734
 characterizing, 735
 load, 738
 open, 741
 short, 741
 verification, 747
 calorimeter, 63
 CalPod, 191
 as Ecal, 194
 initialization, 192
 in-situ, 290
 re-correction, 192
 capacitor, 56
 CATV, 279
 attenuation, 287
 cable measurements, 280
 multiport, 327

- return loss, 295
- circulator, 51
 - measurement, 334
- CITI file, 367
- coherent
 - NPR, 594
 - receiver, 461
 - SA, 559
 - time averaging, 561
- cold source NF, 626
- common-mode, 671
 - impedance, 671
- CompGain21, 378
- CompIn21, 378
- complementary composite distribution function (CCDF), 547
 - AWGN, 548
 - QPSK, 548
- complex square root, 754
- CompOut21, 378
- compression, 344
 - amplifier, 377
 - from back-off, 374
 - 2-D, 380
 - from linear, 372
 - from max, 374
 - mixer, 497
 - from sat, 376
 - smart sweep, 378
 - swept frequency, 377
 - X-Y, 374
- concatenate S-parameter, 118
- connectors, 31
 - 2.4, 40
 - 2.92, 40
 - compensation, 279
 - gating, 283
 - in1line, 296
 - K, 40
 - measurement, 279
 - 1 mm, 42
 - 1.85 mm, 40
 - 3.5 mm, 37
 - 7-mm, 34
 - mode, 33
 - PC board launch, 43
- SMA, 37
- type N, 34
- V, 40
- conversion-loss, 444
- conversions
 - reflection, 120
 - transmission, 121
 - Z, 120
- converters *see* frequency converters
- coplanar waveguide (CPW), 46
- correction
 - IMD, 582
 - noise figure, 637
- couplers, 49
 - directional, 49
 - measurement, 328
- cross-mode parameter, 672
- crosstalk, 94
- cumulative distribution function (CDF), 547
- current, 4
- databased standards, 745
- dB_i, 53
- DC power, 382
- de-embedding, 751
 - high-loss, 775
- delay, 13, 273
 - deviation, 13
 - group, 13
 - IF, 274
 - measurements, 320
 - mixer, 461
- DeltaGain, 378
- delta-match, 155
- detectors, diode, 64
- deviation, 324
 - linear phase, 324
- device-under-test (DUT), 2
- devolved calibration, 194
- differential
 - amplifier, 680
 - compression, 688, 695
 - current, 671
 - distortion, 714
 - harmonics, 700

- differential (*contd.*)
 impedance, 671
 match correction, 691
 mixed mode, 672
 mode, 670
 modulated, 685
 noise figure, 723
 noise parameters, 731
 phase compression, 698
 S-parameters, 672
 THD, 701
 to-single, 674
 true-mode, 689
 voltage, 670
 waves, 671, 703
- differential and IQ mixer application
 (DIQ), 470
- diplexer, 47, 307
- DIR, 132
- directional bridge, 85
- directional coupler, 89
- directivity, 130
 Coupler, 329
- discrete Fourier transform (DFT), 238
- distortion
 differential, 714
 harmonic, 19, 404
 mixer, 501
 optimizing, 579
 swept, 576
 two-tone, 20, 571
- dividers, 61, 62
- Doherty amplifier, 410
- doubler, 61, 508
- double sideband, 644
- down/up, 457
- drift, 228
- diplexer, 47, 307
- DUT EVM, 603
- DUTRNP, 618
- DUTRNPI, 618
- dynamic range, 315
 extended, 317
- dynamic uncertainty, 215
- Ecal, 162
- user-characterized, 164
- efficiency
 contours, 421
 PAE, 382
- electrically long, 272
 attenuation, 287
 delay, 305
- embedded LO, 515
- embedding, 772
- enhanced-respons, 196
- equalization, 26
- error correction, interpolation, 229, 308
- error-model
 8-term, 136
 12-term, 134
 T-matrix, 137
- errors in phase, 218
- error-terms, 129, 130
- error-vector mag (EVM), 25, 376, 596
- even functions, 236
- excess-loss, 312
- excess noise ratio (ENR), 623
- excess noise temp, 624
- extrapolation, 756
- eye diagram, 266
- Fast Fourier Transform (FFT), 235, 238
 SA, 538
- filters, 47, 306
 diplexer, 307
 low-loss, 309
 return loss, 309
 SAW, 323
 transmission, 310
 tunable, 308
- fixed LO, 453
- fixtures
 characterization, 735
 one-port characterization, 573
 PCB, 734
 port matching, 750
 time-gating, 746
- fixturing, 750
- flattop filter, 535
- Fourier transform, 236, 240
 definition, 241

- inverse, 241
windowing, 246
- frequency converter, 59, 445
- frequency doubler *see* doubler
- frequency multiplier, 61
- Friis equation, 624
- fringing capacitance, 743
- gain
available, 15, 363
compression, 372
max available, 357
max stable, 359
power, 10
transducer, 365
- gain compression application (GCA) *see* compression
- gamma
1, 9
2, 9
In, 9
S, 14
1+gamma, 90
- gamma opt, 19, 628
noise, 19, 631
- gating, 252
compensation, 257
connector, 283
examples, 256
loss, 253
masking, 257
renormalization, 253
- Gaussian filter, 535
- Gaussian noise, 546
- GBrcr, 624
- GMS, 359
- Group delay *see* delay
- G/T, 638
- guided power cal, 174
- harmonics, 19, 404
enhancement circuit, 409
- hermitian function, 237
- high-gain amplifier, 384
calibration, 387
setup, 386
- high-power amplifier, 389
calibration, 393
receiver power, 391
- Hot S_{22} , 421
hot S-parameters, 384, 413, 421
- H-parameters, 125
- hybrid load-pull, 420
- hybrids, 331, 708
balanced, 332
- IF delay, 274
- IF response, 543
- IF-RF conversion, 448
- image enhancement, 450
- image rejection
coherent, 562
digital, 542
hardware, 539
software, 541
- IMD optimization, 579
receiver, 581
source, 579
- impedance
cable, 303
conversion, 126
definition, 9
transformation, 116
transformer, 774
- impulse response, low-pass, 236
- inductor, 57
- in-line connector, 296
- input referred third-order intercept point (IIP3), 22, 572
- insertion loss, 10
- in-situ calibration, 191
- intercept, 19, 572
second-order, 19
third-order, 22, 572
- interpolation, 229
- Inverse Fast Fourier Transform (IFFT), 238
- IP2, 19
- IP3, 22, 572
- IQ balance, 524
- IQ downconverter, 527
- IQ mixer, 519

- isolation, 94
- isolator, 51
 - measurement, 334
- isotropic radiation, 53
- jitter, 621
- Kaiser-Bessel, 246
- Kelvin, 1
- k-factor, 357
- leakage, 95, 130
- leveling, 354
 - pulse, 397
 - receiver, 354
- limiter, 61
- limit lines, 313
- linearity
 - receiver, 221
 - source, 171
- line-strecher, 80
- load
 - contours, 417
 - fixed, 145
 - model, 740
 - pull, 417
 - sliding, 146
 - standard, 738
- load-match, 130
- LO multiplier, 507
- loss, 10
 - insertion, 10
 - return, 10
 - structural, 31
- low gain, 647
- low pass transform
 - DC extrapolation, 249
 - impulse, 248
 - step, 249
- LRL, 153
- marker functions
 - marker-to-delay, 325
- match stability, 779
- max stable gain, 359
- measurement
 - analysis, 2
 - calibrating, 2
 - DC bias, 355
 - DC current, 356
 - harmonics, 404
 - measuring, 2
 - optimizing, 2
 - power, 101
 - pretest, 2
 - pulsed, 394
 - saving data, 3
- mechanical devices, 184
- microstrip, 44
- mismatch, power sensor, 169
- mixed-mode noise, 725
- mixed mode S-parameter, 670, 673
- mixer cold-source noise, 644
- mixer noise figure, 642
- mixers, 59, 94, 435
 - amplitude, 452
 - calibration, 476
 - conversion gain, 451
 - embedded LO, 515
 - excess LO noise, 467
 - gain compression, 500
 - higher order products, 445
 - high-gain, 517
 - image, 439
 - image-reject, 520
- IMD, 501
- LO Phase, 443
- measurement, 451
- model, 438
- noise figure, 642
- normal, 438
- phase-matching, 470
- phase response, 456
- power sweep, 493, 497
- reciprocity, 442
- re-conversion, 448
- single-balanced, 435
- spurs, 445
- SSB, 520
- TOI, 501
- mm-wave, 106
- modulated waveform, 546

- power detection, 562
modulation theorem, 237
mu factor, 361
multiplexer, 47
multipliers, 61
multiport, 97
calibration, 186

network
linear distortion, 11
lossless, 11
 $N_{F\min}$, 19, 628
noise
bandwidth, 619
correlation, 631
factor, 15
figure, 15
floor, 107
high-level, 109
jitter, 621
like signal, 549
parameters, 18, 628
power, 619
power-detector, 18
power-ratio, 24, 587
resistance, 19, 628
temperature, 16
trace, 108
voltage, 619
noise figure
analyzer, 67
cold-source, 626
correction, 637
definition, 618
verification, 655
Y-factor, 623
noise figure analyzer (NFA), 67
noise power
available, 15
density, 17
excess, 17
incident, 16
noise power ratio (NPR), 24, 587
non-insertable, 159
number of paths, 186
odd functions, 236
OIP3, 22, 572
on-wafer cal, 160
open
cal-std, 141
model, 741
shielded, 143
optimizing
amplifier, 346
IMD, 579
orthogonal frequency domain multiplexing (OFDM), 26
output power (Opwr), 353

passband measure, 311
phase
compression, 376
skew, 698
stitching, 465
unwrapped, 13
unwrapping, 114
phase-reference, 485
point-in-pulse, 397
1-port AFR, 770
1-port cal, 136
2-port cal, 135
port extensions, 756
automatic, 759
loss, 757
waveguide, 578
port-matching, 772
1-port standards, 736
power
added efficiency, 355
available, 14
calibration, 131
changes, 221
contours, 421
delivere, 11, 14
dividers, 333
enhancedcal, 348
excess-noise, 17
incident, 17
leveling, 354
linearity, 171
meter, 63

- power (*contd.*)
 - receiver, 173, 347
 - reflected, 13
 - source, 131, 347
 - splitters, 333
 - waves, 181
- power-added efficiency (PAE), 355, 382
- pre-selector, 540
- propagation, 30
 - velocity, 30
- pulsed measurement
 - averaging, 397
 - DC, 401
 - narrowband, 396
 - PAE, 402
 - pulse profile, 398
 - pulse-to-pulse, 401
 - receiver, 399
 - spectrum, 609
 - wideband, 395
- loaded, 336
- unloaded, 337
- quick-short-open-load-thru (QSOLT),
 - 161
 - multiport, 190
- receiver
 - calibration, 173
 - phase coherent, 461
 - tracking, 132
- reciprocal mixer, 482
- reciprocal thru, 158
- re-conversion, 449
- repeatability
 - connector, 225
 - drift, 228
 - noise effects, 227
- residual errors
 - directivity, 199
 - load-match, 206
 - reflection, 205
 - source-m, 202
 - transmission, 209
- resistor, 53
- resolution bandwidth (RBW), 534
- effects, 556
- speed, 537
- resonators, 336
- response correction
 - match corrected, 178
 - output power, 179
 - reference channel, 175
- return loss, 9
- reversed coupler, 317
- rho, 10
- ripple, 313
- RN, 19, 628
- samplers, 92
- sampling theorem, 242
- saving
 - CITI file, 367
 - CSV, 369
 - S2P, 369
 - touchstone, 369
- scalar mixer, 444
- scalar-network analyzer, 68
- S_{dd11} , 672
- S_{dd21} , 673
- segment sweep, 317
 - mixers, 509
- self-resonant frequency, 57
- sensitivity, receiver, 564
- shift theorem, 237
- short
 - cal-std, 144
 - model, 741
- short-open-load-reciprocal (SOLR), 158
- signal-source, 64
 - analog, 64
 - vector, 65
- skew, phase, 698
- skin effect, 29
- Smith chart, 112
- smoothing *see* group delay
- source
 - correction, 168
 - impedance, 82
 - match, 76, 130
 - power, 78
 - power calibration, 168

- ratio match, 76
residual, 202
source calibration factor (SCF), 168
source transmission forward (STF), 132
S-parameter
 definition, 4
 mixed mode, 672
specifications, 232
spectral-regrowth, 23
spectrum analysis
 fundamentals, 534
 RBW, 534
spectrum analyzer, 65
splitters, 333
S2P, SnP, 369
spurious, 605
stability
 circles, 360
 conditional, 358
 K-factor, 357
 reflection, 781
 transmission, 778
 unconditional, 358
standards, 141
 calibrat, 141
 load, 145
 open, 141
 plynomia, 143, 144
 short, 144
 thru, 149
statistics, 313
step attenuator, 223
stepped sweep, 273
stripline, 47
structural return loss (SRL), 31, 299
surface acoustic wave (SAW)
 filter, 323
 triple travel, 324
surface mount technology (SMT), 53
swap-equal-adapt, 151
sweep-time, 553
swept
 frequency, 49
 IMD, 576
 segments, 317
swept LO, 469, 493
group de, 513
swept-tuned SA, 537
switch terms, 154
switch tree, 99
symbol skipping, 604
system-directivity, 131
system-relative noise power incident
 (SYSRNPI), 618
tansducer gain, 365
TDT, 265
Te, 624, 627
telegraphers equation, 28
temperature
 effective, 17
 noise, 17
 relative, 17
8-term error, 136, 153
12-term error, 134
test set
 extension, 100
 switch tree, 99
thermistor, 63
thermocouple, 63
thru
 cal-std, 149
 defined, 149, 151
 flush, 149, 164
 non-insertable, 149
 TRL, 154
thru-reflect-line (TRL), 153
thru-reflect-match (TRM), 157
time domain
 band pass, 237
 discontinuity, 257
 gating, 252
 gating fixtures, 764
 impedance step, 256
 low pass impulse, 240
 low pass step, 249
 masking, 259
 renormalization, 253
 sampling effects, 239
 truncation, 244
 windowing effects, 246
time-domain reflectometer (TDR), 267

- TOI, 22, 571, 572
total harmonic distortion (THD), 701.703
touchstone, 369
T-parameters, 119
trace noise
 averaging, 273, 347
 vs. IFBW, 272
 vs. power, 272
trace-noise, 108
tracking, 130
transformation, impedance, 116
transmission lines
 coplaner waveguide, 46
 differential, 328
 microstrip, 44
tripler, 61
triple-travel, 324
true-mode drive, 682, 689
true-multiport, 103, 188
truncation, 244
two-tone delay, 467
two-tone IMD, 571
- uncertainty
 calculations, 210
 noise figure, 655
 phase, 218
 receiver, 212
 reflection, 210
 source power, 211
 transmission, 212
unknow-thru, 158
up/down, 457
- vector averaging, 561
vector mixer, 445
vector mixer characterization (VMC),
 459
vector mixer correction, 459
- vector network analyzer (VNA)
 block diagram, 72
 converter, 94
 crosstalk, 94
 digital IF, 95
 isolation, 94
 mm-wave, 106
 noise floor, 94
 receiver, 91
 reflectometer, 80
 source, 73
 test set, 82
vector-signal analyzer, 66
verification, 655
 noise fi, 655
Via, 57
video bandwidth (VBW), 540, 553
voltage, 4, 7
voltage standing wave ratio (VSWR), 9,
 80, 140
VSA, 66
- waveguide, shim, 166
waves
 actual, 181
 incident, 13
 power, 181
 reflected, 13
white noise, 546
windowing, 246
- X-parameters, 413
- Y-factor, 353, 623
Y-parameters, 123
- Z-parameters, 122
Z-transformation, 774